Long Memory Dynamics in a Discrete-Time Real Business Cycle DSGE Model and a Continuous-Time Macro-Financial Model

Von der Fakultät Wirtschafts- und Sozialwissenschaften
der Universität Stuttgart zur Erlangung der Würde
eines Doktors der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)
genehmigte Abhandlung

Vorgelegt von

Frank Calisse

aus Geislingen an der Steige

Hauptberichter:	Prof. Dr. Frank C. Englmann
Mitberichter:	Prof. Dr. Bernd Woeckener
Tag der mündlichen Prüfung:	24. Juli 2023

Institut für Volkswirtschaftslehre und Recht der Universität Stuttgart

2023

Danksagung

Die vorliegende Dissertationsschrift entstand während meiner Tätigkeit als akademischer Mitarbeiter am Lehrstuhl für Theoretische Volkswirtschaftslehre der Universität Stuttgart.

Besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Frank C. Englmann. Er hat sich auf das Experiment eingelassen, einem Mathematiker freie Hand bei der Themenwahl einer Dissertation im Kontext der theoretischen Volkswirtschaftslehre zu lassen. Diese Offenheit aber auch seine stete Beratung und unsere langen Gespräche haben es mir ermöglicht, vertiefte Erkenntnisse in einem für mich weitestgehend neuen Themengebiet zu erlangen. Herzlich bedanken möchte ich mich auch bei Herrn Prof. Dr. Bernd Woeckener für die Übernahme des Zweitgutachtens und für seinen kritischen Blick auf das Promotionsvorhaben während der gemeinsamen Doktorandenseminare.

Für die wertvollen Anmerkungen zu meinen Vorträgen in unseren gemeinsamen Doktorandenseminaren und für ihr Durchhaltevermögen während diesen möchte ich mich bei Benjamin Bätz, Yonas Ogbamicael, Dr. Jonathan Ulrich Baumann, Magdalena Auer und Dr. Markus Till Berkefeld herzlich bedanken. Mein Dank gilt auch meinen weiteren Kolleginnen und Kollegen am Institut. Mit Dr. Susanne Becker, Dr. Marion Aschmann, Dr. Klaas Macha, Claudia Feucht und Roman Wedel wurde der Arbeitsalltag am Lehrstuhl nie eintönig. Die gemeinsame Zeit werde ich vermissen.

Für den fachlichen Austausch und für ihre hilfreichen Kommentare zu meiner Forschungsarbeit möchte ich ebenfalls Dr. Alexandra Mitschke von Herzen danken.

Nicht zuletzt danke ich meiner Frau Simone und meiner Tochter Sophie. Worte vermögen kaum zu beschreiben, welche Unterstützung sie mir während meiner Promotionsphase gegeben haben. Für ihre Geduld mit mir bin ich unendlich dankbar. Auch meinen Eltern, Schwiegereltern und meinen Freunden danke ich für ihre bedingungslose Unterstützung.

Contents

List of Figures

List of Tables

List of Abbreviations

ACF	autocorrelation function
AIC	Akaike information criterion
ARFIMA	autoregressive fractionally integrated moving average
ARMA	autoregressive moving average
BIC	Bayesian information criterion
CD	Cobb-Douglas
CIR	cumulative impulse response
DSGE	dynamic stochastic general equilibrium
ELW	exact local Whittle estimator
fBm	fractional Brownian motion
FELW	fully extended local Whittle estimator
fGn	fractional Gaussian noise
GDP	gross domestic product
GNP	gross national product
GPH	Geweke-Porter-Hudak
i.i.d.	independent and identically distributed
IRF	impulse-response function
ML	maximum likelihood
p.c.	per capita
RBC	real business cycle
SS	steady state
TFP	total factor productivity
UK	United Kingdom
US	United States (of America)
VARMA	vector autoregressive moving average

Notation and Symbols

General Notation, Mathematical Operators and Functions

$\lvert\cdot\rvert$	absolute value
$\approx$	approximately
$\mathrm{AR}(p)$	autoregressive process of order p
$\mathrm{ARFIMA}(p,d,q)$	autoregressive fractionally integrated (of order d) moving average process involving p autoregressive and q moving average terms
$\arg\min_{x\in\Theta} f(x)$	subset of Θ containing the elements for which f attains its minimum value
$\mathrm{ARMA}(p,q)$	autoregressive moving average process involving p autoregressive and q moving average terms
B	backshift operator
$(B_t^H)_{t\geq 0}$	type I fractional Brownian motion
$:=$	the left-hand side is defined by the right-hand side
$\longrightarrow_{\mathcal{D}}$	convergence in distribution
$\cos(\cdot)$	cosine function
$\mathrm{cov}(\cdot,\cdot)$	covariance
d	order of integration/long memory parameter
$\hat{d}$	estimator of d
$\hat{d}_{GPH}$	GPH estimator of d
$\hat{d}_{LW}$	local Whittle estimator of d
$\det(A)$	determinant of the matrix A
$\mathbb{E}$	expected value
$e, \exp(\cdot)$	exponential function
$\mathbb{E}_t(\cdot)$	conditional expectation given the information set $\mathcal{F}_t$ in period t, i.e., shorthand notation for $\mathbb{E}_t(\cdot\vert\mathcal{F}_t)$

$a \in A$	a is an element of the set A
$\equiv$	is identically equal to
$k!$	factorial of the non-negative integer k
$f_X(\cdot)$	spectral density of the stochastic process X
$\bar{f}_X(\cdot)$	normalized spectral density of the stochastic process X
$\Gamma(\cdot)$	gamma function
$\gamma_X(\cdot,\cdot)$	covariance function of the stochastic process X
${}_2F_1(a,b;c;\cdot)$	Gaussian hypergeometric function
A'	Hermitian transpose of the matrix A, i.e., the transpose of the element-wise complex conjugated matrix
H	Hurst index
$\mathbf{I}_{n\times n}$	identity matrix of size $n \times n$
$[\cdot]$	integer part of a real number
$\mathcal{I}(d)$	class of processes that are integrated of order d
$I_{n,X}(\cdot)$	periodogram of the time series X up to period n
$\lim\limits_{t\to a} x_t$	limit of a function/process x_t for t converging to a
$\log(\cdot)$	natural logarithm
$\mathbb{N}$	natural numbers, i.e., $1, 2, 3, \ldots$
$\mathbb{N}_0$	non-negative integers, i.e., $0, 1, 2, 3, \ldots$
$\lVert\cdot\rVert$	Euclidean norm
$N(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$n \times m$	specifies the size of a matrix containing n rows and m columns
$\partial f/\partial x$	partial derivative of f with respect to x
$\partial^2 f/\partial x \partial y$	partial derivative of $\partial f/\partial x$ with respect to y
$\partial^k f/\partial^k x$	k-th-order partial derivative of f with respect to x
$\phi(\cdot)$	autoregressive polynomial
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space containing the sample space Ω, the sigma algebra $\mathcal{F}$ and the probability measure $\mathbb{P}$
$(\psi_j)_{j=0}^{\infty}$	series of coefficients of the infinite moving average representation/impulse-response function
$\langle\cdot\rangle_t$	quadratic variation of an Itô process
$R(n)$	adjusted range statistic
$\mathbb{R}$	real numbers
$\mathbb{R}^n$	n dimensional Euclidean space
$\rho_X(\cdot)$	autocorrelation function of the stochastic process X
$R/S(n)$	rescaled adjusted range statistic

σ_X^2	variance of the stationary stochastic process X
Σ_X	covariance matrix of the Gaussian random vector X
$\sin(\cdot)$	sine function
$\sqrt{\cdot}$	square root
$A \subset B$	the set A is a subset of the set B
$\theta(\cdot)$	moving-average polynomial
$f(x) \sim g(x), x \to a$	$\lim_{x\to a} f(x)/g(x) = 1$, i.e., both functions show the same asymptotic behavior as $x \to a$
$\mathbb{T}$	(time) index set
A^T	transpose of the matrix or vector A
$\mathrm{var}(\cdot)$	variance
$\mathrm{VAR}(k)$	k-th order VAR process
$\Rightarrow$	weak convergence
$(X_t^H)_{t\geq 0}$	fractional Gaussian noise
$\mathbb{Z}$	integers
$(Z_t^H)_{t\geq 0}$	type II fractional Brownian motion
$0_{n\times m}$	$n \times m$ dimensional zero matrix

Model Variables and Parameters of Chapter 4

$\bar{A}_t$	growth component in TFP in period t/ labor augmenting technological progress
$A_t; \tilde{A}_t; A_{ss}$	transitory TFP component in t; percentage deviation from steady state; steady state value
TFP_t	total factor productivity in period t (corresponds to $A_t\bar{A}_t^{1-\alpha}$)
$b_t; \tilde{b}_t; b_{ss}$	growth factor of $\bar{A}_t$, i.e., $b_t = (1 + g_t)$; percentage deviation from steady state; steady state value
$C_t; \tilde{C}_t; C_{ss}; c_t; c_{ss}$	consumption expenditures in period t; percentage deviation from steady state; steady state value; stationarized variable; steady state of stationarized variable
$g_t; g_{ss}$	growth rate of $\bar{A}_t$; steady state value
$H_t; \tilde{H}_t; H_{ss}$	hours worked in period t; percentage deviation from steady state; steady state value
$I_t; \tilde{I}_t; I_{ss}; i_t, i_{ss}$	investment expenditures in period t; percentage deviation from steady state; steady state value; stationarized variable; steady state of stationarized variable

$K_t; \tilde{K}_t; K_{ss}; k_t; k_{ss}$	capital stock available for production in t; percentage deviation from steady state; steady state value; stationarized variable; steady state of stationarized variable
$L_t; \tilde{L}_t; L_{ss}$	hours of leisure in period t; percentage deviation from steady state; steady state value
P_t	final goods price
Π_t	the representative firm's objective function/profits
$R_t; \tilde{R}_t; R_{ss}$	real rental rate on capital in period t; percentage deviation from steady state; steady state value
R_t^n	nominal rental rate on capital in period t
$U(\cdot, \cdot)$	utility function
$U_{C,t}; U_{H,t}$	shorthand notation for $\partial U(C_t, H_t)/\partial C_t$; $\partial U(C_t, H_t)/\partial H_t$
$W_t; \tilde{W}_t; W_{ss}; w_t; w_{ss}$	real wage in period t; percentage deviation from steady state; steady state value; stationarized variable; steady state of stationarized variable
W_t^n	nominal wage in period t
$Y_t; \tilde{Y}_t; Y_{ss}; y_t; y_{ss}$	output produced in period t; percentage deviation from steady state; steady state value; stationarized variable; steady state of stationarized variable
$Y_t^p; \tilde{Y}_t^p; Y_{ss}^p$	permanent income in period t; percentage deviation from steady state; steady state value
$\varepsilon^A = (\varepsilon_t^A)_{t\in\mathbb{Z}}$	transitory technology shock process
$\varepsilon^g = (\varepsilon_t^g)_{t\in\mathbb{Z}}$	growth shock process
$\nu^A = (\nu_t^A)_{t\in\mathbb{Z}}$	fractionally integrated process of ε^A, i.e., $\nu_t^A := (1-B)^{-d}\varepsilon_t^A$
α	capital share in production/output elasticity with respect to capital
β	time preference rate
δ	depreciation rate
$\bar{\gamma}$	composite parameter equal to $(1-\gamma)(1-\tau)-1$
γ	exponent of leisure (parameter of the Cobb-Douglas utility function)
$\varkappa$	relative distaste of supplying labor (parameter of the additive separable utility function)
φ	inverse of the Frisch elasticity of labor supply (parameter of the additive separable utility function)
ς	inverse of the intertemporal elasticity of substitution (parameter of the additive separable utility function)

τ	inverse of the intertemporal elasticity of substitution for the composite good $C_t^{1-\gamma}(1-H_t)^{\gamma}$ (parameter of the Cobb-Douglas utility function)
ϱ_A	short memory parameter of TFP's transitory component (A_t)
ϱ_g	short memory parameter of the growth rate g_t
d	long memory parameter of TFP's transitory component (A_t)

Additional symbols of the Appendix to Chapter 4

$\mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{G}, \mathbf{M}, \mathbf{N}, \mathbf{P}$ $\mathbf{Q}, \mathbf{R}, \mathbf{S}, \mathbf{T}, \mathbf{U}, \mathbf{V}, \mathbf{W}$	matrices of state transition equations or canonical forms possibly appearing with certain indexes
$b_j, j = 1, 2, \ldots$	coefficients of the infinite moving average representation in the long memory model
$e_1; e_2$	first and second two-dimensional standard unit vector
$\lambda_1, \ldots, \lambda_n$	generalized eigenvalues
Λ, P, D	matrices involved in diagonalization during Klein's solution method
$\hat{G}_1, \hat{G}_2$	matrices used during Klein's solution method
Q, Z, S, T	matrices of the generalized Schur decomposition possibly appearing with certain indexes
n	number of variables in the model
$n_b; n_f$	number of backward-looking variables; number of forward-looking variables
$n_s; n_u$	number of stable eigenvalues; number of unstable eigenvalues
n_z	number of elements in the vector z
s_t	auxiliary vector associated with the stable subsystem in Klein's solution method
u_t	auxiliary vector associated with the unstable subsystem in Klein's solution method
x_t	vector containing all model variables
$x_t^b; x_t^f$	part of the vector x containing the backward-looking variables; part of the vector x containing the forward-looking variables
z_t	vector containing all exogenous shock processes

Model Variables and Parameters of Chapter 5

B_t	value of the risk-free asset at time t
$c_t^i; \hat{c}_t^i$	consumption of expert i; corresponding optimal value

$\tilde{c}_t^j; \hat{\tilde{c}}_t^j$	consumption of household j; corresponding optimal value
$\eta_t; \eta_t^{\mathrm{BS}}$	experts' wealth share at time t; corresponding value in the benchmark model of Brunnermeier and Sannikov
$\mathbb{I}; \mathbb{J}$	index set of experts; index set of households
$\iota^i; \hat{\iota}_t^i; \iota$	expert i's re-investment rate; corresponding optimal value; corresponding aggregate value
$K_t; K_t^{\mathrm{BS}}$	the economy's aggregate capital stock at time t; corresponding value in the benchmark model of Brunnermeier and Sannikov
k_t^i	expert i's capital holdings at time t
$K(t,s)$	weighting Kernel
$\mu_\eta^{H,\varepsilon}\eta_t$	drift of η_t
μ_t^q	drift rate of the price of capital
N_t	experts' aggregate total wealth at time t
$n_t^i; \hat{n}_t^i$	expert i's wealth at time t; corresponding optimal value
$\tilde{n}_t^j; \hat{\tilde{n}}_t^j$	household j's wealth at time t; corresponding optimal value
$\pi_t^i; \tilde{\pi}_t^j$	amount of risk-free assets held by expert i; amount of risk-free assets held by household j
$q_t; q$	price of capital; corresponding equilibrium value
$R_{Y,n}; R_{Y,n}^{\mathrm{BS}}$	growth rate of the economy's aggregate output; corresponding value in the benchmark model of Brunnermeier and Sannikov
$r_t; r_t^{\mathrm{BS}}$	interest rate; corresponding value in the benchmark model of Brunnermeier and Sannikov
$\sigma_\eta^{H,\varepsilon}\eta_t$	volatility of η_t
σ_t^q	volatility rate of the price of capital
$\vartheta_t^i; \hat{\vartheta}_t^i; \vartheta$	expert i's market price of risk vector; corresponding optimal value; corresponding aggregate value
$(W_t)_{t\geq 0}$	standard Brownian motion/Wiener process
$x_t^i; \hat{x}_t^i$	share of expert i's wealth invested in capital; corresponding optimal value
$Y_t; Y_t^{\mathrm{BS}}$	the economy's aggregate output; corresponding value in the benchmark model of Brunnermeier and Sannikov
$(Z_t^{H,\varepsilon})_{t\geq 0}$	approximation of a type II fractional Brownian motion
a	productivity parameter of the AK production technology
δ	depreciation rate
ε	parameter measuring how close $Z_t^{H,\varepsilon}$ is to a type II fBm
H	Hurst index

$\bar{k}_0^i; \bar{\tilde{k}}_0^j$	initial value of expert i's capital holdings; initial value of household j's capital holdings
κ	parameter of the investment adjustment cost function $\Phi(\cdot)$
$\bar{\tilde{n}}_0^j$	initial value of households j's wealth
$\Phi(\cdot)$	investment adjustment cost function
$\varphi_t^{H,\varepsilon}$	drift component of $Z_t^{H,\varepsilon}$
$\Psi(\cdot)$	inverse function of $\Phi'(\cdot)$
ρ	agents' time preference rate
σ	volatility rate in the evolution of capital

Zusammenfassung

Long memory ist eine Eigenschaft stationärer stochastischer Prozesse oder Zeitreihen. Die Autokorrelationsfunktion eines long memory Prozesses konvergiert so langsam gegen null, dass die kumulierte Summe ihrer Absolutwerte divergiert. Long memory Prozesse weisen daher eine sehr starke Abhängigkeitsstruktur auf. Traditionelle Zeitreihenmodelle wie ARMA Prozesse sind jedoch sogenannte short memory Prozesse, weil sie eine sehr schnell abfallende Autokorrelationsfunktion besitzen und damit nur eine begrenzte Abhängigkeitsstruktur zulassen.

Die vorliegende Dissertationsschrift ist durch die folgende Beobachtung motiviert. Seit den Arbeiten von Mandelbrot in den 1960er Jahren, die die Forschung zu long memory Prozessen initiiert haben, und den Arbeiten von Hosking, Granger und Joyeux zu Beginn der 80er Jahre, die eine wichtige Klasse von long memory Prozessen entwickelt haben, gibt es einerseits zunehmend empirische Evidenz, dass viele makroökonomische Zeitreihen gut durch long memory Prozesse beschrieben werden können. Zudem gibt es einige theoretische Erklärungsansätze, die das Vorliegen von long memory in (makro)ökonomischen Zeitreihen begründen. Etwa kann die Aggregation von Mikro-Daten auf Makro-Daten long memory in den Makro-Daten hervorrufen.

Andererseits bilden stochastische Modelle in der modernen Makroökonomik wichtige Werkzeuge, um gesamtwirtschaftliche Zusammenhänge zu erklären, kontrafaktische Szenarien zu analysieren oder Prognosen zu erstellen. Zwei Repräsentanten von solchen stochastischen Modellen sind die dynamischen stochastischen allgemeinen Gleichgewichtsmodelle (DSGE Modelle), die überwiegend in diskreter Zeit formuliert sind, sowie die überwiegend in stetiger Zeit formulierten Makro-Finance Modelle. Beide Modellarten verwenden exogene stochastische Prozesse zur Beschreibung von Dynamiken der Modellvariablen. Allerdings werden zur Modellierung häufig exogene stochastische Prozesse unterstellt, die weit überwiegend short memory Prozesse sind. Evident wird dies am Beispiel von DSGE

Modellen, in denen Technologieschocks, Geldpolitikschocks, Präferenzschocks und ähnliche mit autoregressiven Prozessen erster Ordnung (ein sog. AR(1) Prozess) modelliert werden. Da DSGE Modelle jedoch typischerweise mit Makro-Daten geschätzt werden, kann es aber sinnvoll sein, zur Modellierung im Rahmen eines DSGE Modells einen long memory Prozess anstelle eines short memory Prozesses zu verwenden.

Diese Dissertationsschrift möchte daher einen Beitrag zur Zusammenführung dieser beiden Literaturstränge leisten, indem long memory Dynamiken in DSGE Makro-Finance Modellen abgebildet werden.

Bevor in den Kapiteln 4 und 5 die Einführung von long memory in die zwei genannten Modelltypen realisiert wird, werden in Kapitel 2 neben den mathematischen Rahmenbedingungen die diskreten und stetigen long memory Prozesse eingeführt, die für die Modellierung herangezogen werden. Kapitel 3 gibt einen Überblick über long memory in der ökonomischen und ökonometrischen Forschung und unterstreicht dabei die Relevanz von long memory.

In Kapitel 4 wird dann ein Real Business Cycle (RBC) Modell betrachtet, welches durch long memory im exogenen Technologieschock erweitert wird. Damit es sich um eine tatsächliche Verallgemeinerung des bestehenden Modells handelt, wird der seither angenommene exogene AR(1) Technologieschock durch einen exogenen long memory ARFIMA$(1, d, 0)$ Prozess ersetzt. Letzterer weist im Vergleich zum ersteren einen weiteren Parameter d auf, der die Zerfallsrate der Autokorrelation angibt und daher die Stärke des long memory im Prozess kontrolliert. Setzt man diesen auf null, erhält man das bekannte Standardmodell (AR(1) Prozess) als Spezialfall zurück. Jedoch ist die Herleitung der Lösung eines solchen Modells nicht trivial. Nimmt man anstelle eines AR(1) Prozesses einen ARMA Prozess höherer Ordnung, kann dies recht einfach im Modell durch Erweiterung des Zustandsvektors abgebildet werden. Bei ARFIMA Prozessen funktioniert dies nicht, da diese keine endlich dimensionale Zustandsraumdarstellung haben. Somit steht im Rahmen des Kapitels 4 zunächst die Frage der Lösbarkeit eines solchen DSGE Modells im Fokus. Es stellt sich heraus, dass die Lösungsmethode von Klein (2000) auf solche Prozesse erweitert werden kann. Dies eröffnet die Möglichkeit, die Reaktionen der Modellvariablen auf unterschiedliche Spezifikationen der exogenen Schocks anhand von Impuls-Reaktionsfunktionen (IRF) zu analysieren. Dabei werden neben reinen short und long memory Prozessen auch gemischte Prozesse betrachtet sowie sogenannte Trendschocks, die einen permanenten Charakter aufweisen.

Es stellt sich heraus, dass sich die Modellreaktionen auf reine long memory Schocks nicht stark von reinen short memory AR(1) Schocks unterscheiden. Aus der Modellperspektive erscheint dies zunächst überraschend, da man annehmen könnte, dass ein unendlich lang lebender repräsentativer Agent mit rationalen Erwartungen seine intertemporale Konsum- und Arbeitsangebotsentscheidung anders trifft, wenn er weiß, dass der Schock lange in die Zukunft wirkt. Dass dies nicht so ist, kann damit erklärt werden, dass der Haushalt im Modell seinen erwarteten Nutzen mit einer exponentiellen Rate diskontiert. Somit beeinflussen die Schock-Auswirkungen in späteren Perioden seine Konsum- und Arbeitsangebotsentscheidung unmittelbar nach Auftreten des Schocks nur in geringem Maße. Allerdings zeigt sich auch, dass sich die Modellreaktionen im gemischten Fall von short und long memory deutlich von den Reaktionen der jeweiligen reinen Fälle unterscheidet. Es wird gezeigt, dass der Effekt des Schocks in der Periode nach seinem Auftreten gerade der Summe der beiden memory Parameter entspricht. Somit wirkt sich long memory nicht nur in der langen Frist auf das Modell aus, sondern kann auch in der kurzen Frist die Modelldynamiken beeinflussen. Abschließend werden die Modellreaktionen noch mit jenen auf einen permanenten Wachstumsschocks verglichen. Dabei zeigt sich, dass sich die Modellreaktionen bei hohen short und long memory Parametern in der kurzen Frist ähnlich zu jenen eines Trendshocks verhalten. In der langen Frist erreicht die abgebildete Volkswirtschaft einen neuen gleichgewichtigen Wachstumspfad als Reaktion auf den Wachstumsschock wohingegen sie im gemischten short und long memory Fall langsam zu ihrem alten Gleichgewicht zurückkehrt.

In Kapitel 5 wird ein zeitstetiges Makro-Finance Modell ebenfalls so erweitert, dass es long memory in den Wachstumsraten des gesamtwirtschaftlichen Outputs zulässt. Zur Modellierung wird die im Referenzmodell unterstellte Brownsche Bewegung durch eine Approximation einer gebrochenen Brownschen Bewegung ersetzt. Dadurch kann der exogene Schock in einen Drift- und Volatilitätseffekt aufgespalten werden und das Modell mit bestehenden Lösungsmethoden gelöst werden. Es zeigt sich, dass die Entwicklung der Vermögensverteilung zwischen den zwei verschiedenen Agenten, die als Zustandsvariable in dem Modell fungiert, lediglich vom Volatilitätseffekt abhängt. Insbesondere verlangsamt die Präsenz von long memory die Konvergenz der Zustandsvariablen hin zu ihrem stationären Wert. Zudem kann die Entwicklung des gesamtwirtschaftlichen Wohlstands ein Stück weit von der Entwicklung der Vermögensverteilung und damit von der Entwicklung der Zustandsvariablen abgekoppelt werden.

Zwar können beide betrachteten Modelle mit den allgemeineren long memory Dynamiken gelöst werden, doch scheint der Preis dafür hoch. So gestaltet sich etwa die Schätzung eines long memory DSGE Modells schwierig, da das zugehörige DSGE Modell keine endlich dimensionale Zustandsraumdarstellung mehr besitzt, die typischerweise zur Schätzung von DSGE Modellen herangezogen wird. Im zeitstetigen Modell erlaubt zwar die skizzierte Abkopplung der gesamtwirtschaftlichen Entwicklung von der Vermögensverteilung eine differenziertere Modellierung, sie scheint aber der Modellstruktur dieser Modellklasse entgegenzustehen. So ist es Kern dieser Modelle, dass alle Modellvariablen als Funktionen der zugrundeliegenden Zustandsvariablen ausgedrückt werden können. Dies ist im betrachteten Modell nicht mehr der Fall, sodass eine Verallgemeinerung auf komplexere Modelle schwierig erscheint.

Abstract

Long memory refers to a property of a stationary stochastic process or a time series. More specifically, a stationary time series is called a long memory process if its autocorrelation function (ACF) decays very slowly to zero. Indeed, the convergence is so slow that the sum of the ACF's absolute values diverges. In contrast, traditional time series models such as ARMA processes are so-called short memory processes as their ACF decays rapidly, such that these processes permit only a limited dependency structure.

This dissertation is motivated by the following observation. In the 1960s, Mandelbrot initiated research on long memory processes. After the work of Hosking, Granger, and Joyeux in the early 1980s, who developed a class of long memory processes (the so-called ARFIMA processes), there is increasing empirical evidence that many macroeconomic time series can be well-described by long memory processes. Moreover, some theoretical explanations exist for the presence of long memory in (macro)economic time series. For instance, the aggregation of microdata can induce long memory in macro data.

On the other hand, stochastic models build a cornerstone in modern macroeconomics to explain macroeconomic relationships, analyze counterfactual scenarios, or make forecasts. Two representative types of stochastic models are the discrete-time dynamic stochastic general equilibrium (DSGE) models and the continuous-time macro-financial models. Both types of these models use exogenous stochastic processes to describe the dynamics of the model's variables. However, the exogenous stochastic processes often assumed for modeling are predominantly short memory processes. This becomes evident for DSGE models, in which technology shocks, monetary policy shocks, preference shocks, etc., are described by first-order autoregressive processes (AR(1) processes). However, since DSGE models are typically estimated with macro data, it may be appropriate to use a long memory process instead of a short memory process in a DSGE model.

This dissertation aims to contribute to the integration of these two strands of the literature by introducing long memory dynamics in a DSGE and a macro-financial model.

Before Chapter 4 and Chapter 5 introduce long memory into these two types of models, Chapter 2 introduces the mathematical framework and the discrete-time and continuous-time long memory processes that will later be used for modeling purposes. Chapter 3 gives an overview of long memory in economic and econometric research and underlines the relevance of long memory.

Chapter 4 considers a real business cycle (RBC) model extended by long memory in the exogenous technology shock. In order to ensure that this is a true generalization of the existing model, the class of so-called ARFIMA processes is used. More precisely, the assumption of an exogenous AR(1) technology shock is replaced with an exogenous long memory ARFIMA$(1, d, 0)$ process. Compared to the former, the latter has an additional parameter d that specifies the ACF's decay rate and controls the strength of the long memory in the process. Setting this parameter to zero returns the well-known standard model (the AR(1) process) as a special case. However, the derivation of the solution of such a model is not trivial. If one considers a higher-order ARMA process instead of an AR(1) process, this can be done quite easily by expanding the model's state space. For ARFIMA processes, this procedure does not work since they do not have such a finite-dimensional state space representation. Thus, Chapter 4 focuses on the solvability of such a long memory DSGE model. It turns out that the solution method of Klein (2000) can be extended to such models. This opens the possibility of analyzing the responses of the model's variables to different specifications of the exogenous shocks using impulse-response functions (IRF). In addition, besides pure short and long memory processes, mixed processes, as well as so-called trend shocks with a permanent character, are considered and contrasted with each other.

It turns out that the model's responses to pure long memory shocks do not differ qualitatively from pure short memory AR(1) shocks. At first glance, this seems surprising from a model perspective. One might have expected an infinitely-lived representative agent with rational expectations to account for the long-lasting shock effects in his intertemporal consumption and labor supply decision. That this is not the case can be explained by the fact that the household in the model discounts its expected utility at an exponential rate. Thus, the shock effects in later periods have little impact on his consumption and labor supply decisions immediately after the time of shock occurrence. However, it is also shown that the model's responses in the mixed short and long memory cases differ significantly from the responses in the corresponding pure cases. It is shown that the effect of the shock in the period after its occurrence is equal to the sum of the two memory parameters. Thus, long memory not only affects the model in the long term but can also affect model dynamics in the short term. Finally, the model responses are compared with those of a permanent growth shock. It is illustrated that the model's responses to shocks with high short and long memory parameters are similar to those of a trend shock in the short run. In the long run, the economy reaches a new balanced growth path in response to the growth shock, whereas, in the mixed short and long memory case, it slowly returns to its old steady state.

In Chapter 5, a continuous-time macro-financial model is extended to allow for long memory in the economy's aggregate output growth rates. For modeling purposes, the Brownian motion assumed in the reference model is replaced with an approximation of a fractional Brownian motion. This replacement allows the exogenous shock to be split into a drift and volatility effect and the model to be solved using existing solution methods. It turns out that the evolution of the wealth distribution between the two agents in the model, which serves as a state variable, depends only on the volatility effect. In particular, the presence of long memory slows down the convergence of the state variable toward its steady state value. Moreover, the evolution of the aggregate wealth can be decoupled to some extent from the evolution of the wealth distribution and, thus, from the evolution of the state variable.

While both models considered in this thesis can be solved given the more general long memory dynamics, the price for introducing long memory this way seems high. For example, estimating a long memory DSGE model is difficult because the associated DSGE model no longer has a finite-dimensional state space representation, which is typically used for estimating DSGE models. In the continuous-time model, the outlined decoupling of an economy's total wealth from wealth distribution allows for more sophisticated modeling. However, this feature seems to contradict the general model structure of this kind of models. Generally, in this model category, all variables can be expressed as functions of the underlying state variables. This no longer holds true in the model under consideration, so generalizations to more complex models appear difficult.

1

Introduction

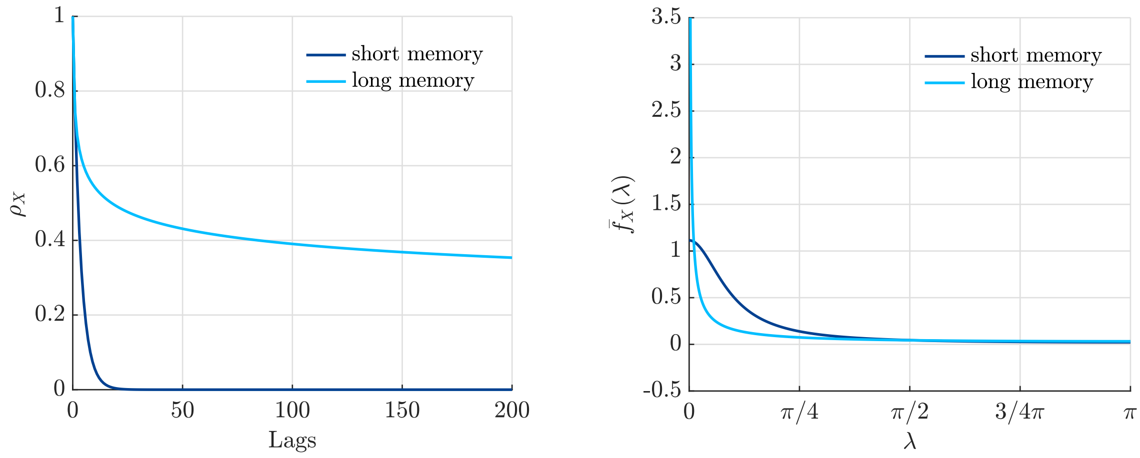
Long memory refers to a property of a stationary stochastic process or a time series. More specifically, a stationary time series is called long memory process if its autocorrelation function (ACF) decays very slowly to zero. Indeed, the convergence is so slow that the sum of the ACF's absolute values diverges.[1] Conversely, stationary stochastic processes whose ACF decays rapidly such that the ACF's absolute values converge are called short memory processes. An example of the ACF of a long and a short memory process can be seen in Panel a) of Figure 1.1.[2] The first-order correlation is the same for both processes. However, the higher-order autocorrelations decay much faster for the short memory process (dark-blue line) than for the long memory process (light-blue line).

That long memory processes can contribute to explaining empirical puzzles left unexplained by short memory processes was illustrated by Mandelbrot and Wallis (1968/2002). The empirical phenomenon Mandelbrot was interested in dates back to the work of Hurst in the early 1950s. Hurst was working on the optimal size of a water reservoir for the Nile River. A reservoir is optimal if it is the smallest reservoir that satisfies the following three properties: the reservoir's outflow is uniform, the reservoir's level at the end of a period is as high as at the beginning of the period, and the reservoir never floods.[3] Let $R(n)$ be the optimal capacity of the reservoir for n periods and $S(n)$ be the standard deviation of n subsequent reservoir outflows. Hurst observed that the rescaled adjusted

[1] More precisely, if ρ_X is the ACF of a stationary stochastic process X, then X is called a long memory process if $\sum_{k=0}^{n} |\rho_X(k)| \to \infty, n \to \infty$. A precise definition of the ACF is given in Chapter 2.

[2] Figure 1.1 was computed using Matlab code written by the author.

[3] See Mandelbrot and Wallis (1968/2002, p. 239).



a) Autocorrelation function b) Normalized spectral density

Figure 1.1: ACF and normalized spectral density of a short and a long memory process. Panel a) shows the ACF of a short memory process (AR(1) process with $\varrho = 0.75$) and a long memory process (ARFIMA$(0, d, 0)$ process with $d = 0.43$) and Panel b) the corresponding normalized spectral densities. In order to equalize the area under the spectral density, normalization is performed by dividing the spectral density by the process's variance. Note the different scaling of the axes.

range statistic $R(n)/S(n)$ is proportional to n^H for large n and $0.5 < H < 1$.[4] Strikingly, this behavior holds not only for river statistics or rainfalls but also for the size of tree rings, the number of sunspots, and wheat prices.[5] These observations, also known as the "Hurst phenomenon", are remarkable since independent normally distributed variables would have implied that $H = 1/2$.[6]

Inspired by the empirical works of Hurst, Mandelbrot and Wallis (1968/2002) introduced in 1968 the colorful name "Joseph effect" for stochastic processes whose corresponding $R(n)/S(n)$ statistic grows as Cn^H with a constant C and $H \in (0, 1)$ with $H \neq 1/2$.[7] His intention for naming the Joseph effect was derived from the biblical story of Joseph, son of Jacob. More precisely, Genesis 41, 29-30 states "Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them."[8] In the context of water resources management, where Mandelbrot introduced his concepts, the Joseph effect states "that a long period of unusually high or low precipitation can be extremely long."[9] In the end, the Joseph effect describes that periods of precipitation and droughts do not alternate regularly; instead, several wet years are followed by several dry

4 See Hurst (1951, Figure 4 on p. 787). Hurst uses the variable K instead of H.
5 See Hurst (1951, Figure 4 on p. 787).
6 See Hurst (1956, p. 14).
7 See Mandelbrot and Wallis (1968/2002, p. 246).
8 This passage was taken from the new international version of the Bible, see Biblica (2021), URL in list of references.
9 Mandelbrot and Wallis (1968/2002, p. 236).

years. Clearly, an optimal water reservoir must have a greater capacity if the rainfall data shows the Joseph effect.[10] Processes showing the Joseph effect are characterized by local trends and cycles, i.e., by considering a part of a time series, trends and periodic cycles appear. However, when looking at the whole time series, the periodicity and the trends disappear from the series, and the process shows non-periodic cycles of varying lengths. In addition, periods where the process is above (or below) the mean appear to be quite long, thereby resembling the mentioned seven good and seven bad years mentioned in the biblical story of Joseph.[11]

In 1968, Mandelbrot and van Ness (1968) introduced a continuous-time stochastic process called fractional Brownian motion (fBm) that depends on a parameter $h \in (0, 1)$. Fractional Brownian motion is a generalization to the ordinary Brownian motion, as the latter can be recovered from fBm by setting $h = 1/2$.[12] The discrete-time increment process of fBm (also depending on h) is called fractional Gaussian noise (fGn). It turned out that fGn was a stationary Gaussian stochastic process that could replicate Hurst's findings with $H = h$ if $h > 1/2$.[13] This was the first time a stationary Gaussian process was presented that was able to explain Hurst's phenomenon.[14] Indeed, it can be shown that fGn is a long memory process if $h > 1/2$.[15]

A possible reason why a stochastic process shows the Joseph effect is strong dependencies or autocorrelation in the process. A strong degree of autocorrelation implies a tendency in the process to keep its direction, i.e., if a process starts showing an upward movement, there is a tendency for keeping this movement for several periods (the periods of great abundance). Similarly, if the process shows a downward direction, there is a tendency to keep this movement for some time (the periods of famine). In contrast, a process's path appears to be rather jagged if there are negative autocorrelations since positive movements tend to alternate frequently with negative movements.[16] That the autocorrelations of a stochastic process must be strong to ensure that the process shows the Joseph effect was already mentioned by Mandelbrot and Wallis (1968/2002, p. 246). Later in 1976, it was shown by Siddiqui (1976) that all processes with a summable ACF have an R/S statistic that scales as $n^{1/2}$.[17] Consequently, all short memory processes fail to show Mandelbrot's Joseph effect. In order to replicate Hurst's empirical findings, one has to resort to long

[10] See Mandelbrot and Wallis (1968/2002, p. 246).
[11] See Mandelbrot (1972, pp. 260 and 275) and Beran (1994, p. 33).
[12] See Lemma 2.5.3 below.
[13] See Mandelbrot and Wallis (1968/2002, p. 250).
[14] See Graves et al. (2017, p. 9).
[15] See Lemma 2.5.3 below.
[16] See Hassler (2019, p. 25).
[17] See Siddiqui (1976, p. 1274).

memory processes. Therefore, long memory processes seemed to be an important class of stationary processes as they were able to replicate empirical phenomena that were left unexplained by short memory processes.[18]

Long memory may not only be accessible through the ACF of a stationary process but also through its spectral density. For stationary processes, the spectral density at the zero frequency corresponds to the sum of the ACF.[19] Consequently, since long memory is associated with a non-summable ACF, one would expect the spectral density of a long memory process to blow up as the frequency tends to zero. This can be seen in Panel b) of Figure 1.1, where one can see that the spectral density of a short memory process is bounded and finite at $\lambda = 0$ while it is unbounded for a long memory process.

It has already been noted by Adelman (1965) and Granger (1966) that such behavior is of great interest in economics. Adelman (1965) provided evidence that many economic variables have a spectral density blowing up at the zero frequency and Granger (1966) described such behavior as the "typical spectral shape of an economic variable".[20] More precisely, the typical spectral density is a smooth function monotonically decreasing with increasing frequency and blowing up for frequencies converging to zero. Following Granger, many economic variables show this typical spectral shape after removing a time trend and seasonal components.[21] Regarding business cycles, Granger (1966) restates the typical spectral shape as "events which affect the economy for a long period are more important than those which affect it only for a short time"[22]. It was these early works in economics that led to the recognition that long memory is also present in many economic time series.

Mandelbrot proposed his stationary fractional Gaussian noise process as a suitable model for replicating Hurst's phenomenon. However, a drawback of Mandelbrot's single parameter fGn process was that the parameter h can only control for the asymptotic properties of the ACF. More precisely, the parameter h controls how slowly the ACF converges to zero, whereas it is impossible to simultaneously match the ACF at small lags. This is in contrast to the rich-parameter autoregressive moving average (ARMA) models that allow, with their autoregressive and moving average parameters, for shaping the ACF at small lags.[23] In the

[18] Some authors do not distinguish between long memory or long-range dependence on the one hand and the Joseph effect on the other and use both terms synonymously, see, e.g., Cutland et al. (1995, p. 330) or Graves et al. (2017, p. 10). It should, however, be noted that a long memory process needs to be stationary. At the same time, no stationarity requirements have to be fulfilled by a process given the definition in terms of the R/S statistic, see Mandelbrot (1972, p. 267).

[19] See Section 2.1.4 for more details.

[20] Granger (1966, p. 150).

[21] See Granger (1966, p. 150).

[22] Granger (1966, Footnote 7 on p. 155).

[23] See Siddiqui (1976, p. 1275).

end, they remain short memory processes and, thus, fail to replicate Hurst's phenomenon. The limited flexibility of Mandelbrot's fBm and fGn might be why they had limited success in economics. However, a breakthrough for long memory in economics dates back to Granger and Joyeux (1980/2001) and Hosking (1981). Both propose a generalization of ARMA processes to make them long memory processes. This generalization refers to the class of autoregressive fractionally integrated moving average (ARFIMA) processes. ARFIMA processes are formulated in discrete time and have an additional parameter d governing the long memory properties of the process. More specifically, if $d = 0$, the corresponding ARFIMA process coincides with the standard ARMA process showing short memory, while for $0 < d < 1/2$, the ARFIMA process shows long memory.[24] Allowing for additional autoregressive and moving average parameters that control for the short-run dynamics is a major advantage of ARMA models over Mandelbrot's single-parameter fGn process. Thanks to this flexibility in reflecting both the long- and short-run properties, ARFIMA models became a standard model for describing long memory dynamics, thereby asserting against Mandelbrot's fGn and fBm.

However, things look different in finance, where models are often formulated in continuous time. In 1973, the seminal works of Black and Scholes (1973) and Merton (1973) propose a continuous-time stochastic framework, where the price of a stock is modeled as a geometric Brownian motion.[25] These works paved the way for modern finance, where stochastic differential equations are extensively used to model interest rates, stock prices (and their volatility), and other financial assets. Many of these finance models implicitly assume that markets are efficient, i.e., prices in such markets reflect all the available information. If new information is revealed, it is arbitraged away instantaneously.[26] In a Black-Scholes setting, this is equivalent to the existence of an "equivalent martingale measure" under which the discounted stock price becomes a martingale.[27] In the end, these financial markets are assumed to be free of arbitrage in order to establish a rational option pricing theory.[28] Although the modern finance theory went beyond the initial assumption of a geometric Brownian motion made by Black and Scholes (1973) and Merton (1973), it still makes use of martingale methods in their models.[29]

The original Black-Scholes model implies that stock returns are stochastically independent

[24] See Section 2.4 for further details.
[25] See Cutland et al. (1995, p. 328).
[26] See Cutland et al. (1995, p. 338) and Mandelbrot (1971, p. 225).
[27] This is the "Fundamental Theorem of Asset Pricing", see, e.g., Lamberton and Lapeyre (2008, p. 20).
[28] See Merton (1973, p. 143).
[29] An early critical voice was Maheswaran and Sims (1993, p. 306) who state that the focus on semi-martingales in finance is a substantive restriction, not just a regularity condition.

(and thus uncorrelated).[30] Since this contradicts empirical findings, some authors propose to replace the Brownian motion in the Black-Scholes setting with Mandelbrot and van Ness's fractional Brownian motion.[31] In such a setting, stock market returns have long memory and, consequently, are positively autocorrelated, a fact for which there appears to be empirical evidence. However, that long memory and the absence of arbitrage and the martingale methods seem incompatible was already suggested by Mandelbrot (1971, pp. 227f.) and later by Maheswaran and Sims (1993, p. 306) and Rogers (1997, p. 101). Nevertheless, the financial mathematics literature in the 2000s proposed fractional Brownian motion as a suitable process for modeling asset prices in a continuous-time stochastic framework. Since fBm is incompatible with the traditional martingale methods and the Itô calculus, these approaches came along with other difficulties that may have prevented them from becoming widely applied.[32] More recent approaches suggest the use of stochastic processes that combine long memory and the applicability of martingale methods in order to make long memory compatible with modern finance theory. One such suggestion followed in this thesis is based on Dung (2013), who uses an approximation of fractional Brownian motion as a source of uncertainty in his continuous-time model.

Not surprisingly, ARFIMA models and long memory have become popular in economics and econometrics since the work of Granger and Joyeux (1980/2001) and Hosking (1981). There is a large and still growing economic literature dealing with long memory. This economic literature on long memory can roughly be grouped into three strands, where the distinction between the first and second strands is not clear-cut.[33] The first strand deals with developing new or evaluating existing statistical or econometric methods for detecting, estimating, or testing long memory in time series. The second strand applies the techniques developed by the first strand in order to find evidence for or against long memory in various data sets. The third strand of the literature works on theoretical explanations for why long memory is likely to occur in an economic time series. The early work of Granger (1980) can be assigned to the third strand. Granger (1980) showed that long memory is likely to occur in aggregate time series, which frequently occur in macroeconomics. More precisely, he showed that the sum of a large number of individual short memory processes could, under specific circumstances, show long memory. His work made it plausible why the time series of an economy's output or inflation exhibit long memory. The former can be viewed as the sum of the output of many individual

[30] See Cutland et al. (1995).
[31] See Section 5.1 for a discussion of this literature.
[32] A detailed consideration of the drawbacks of these suggestions can be found at the beginning of Chapter 5.
[33] All of these strands are considered in detail in Chapter 3 of this thesis.

firms, while an economy's price level appears to be an aggregate of the prices of different goods. As illustrated so far, there is an extensive history of long memory in economics and vast and still growing literature dealing with long memory in economics and finance. Additionally, there seems to be evidence why long memory processes should primarily be used for modeling aggregate (macroeconomic) relationships.

On the other hand, in recent years, stochastic models have become a cornerstone in modern macroeconomics. Two prominent types of models are the discrete-time dynamic stochastic general equilibrium (DSGE) models and the stochastic continuous-time macro-financial models. DSGE models have become popular not only among researchers but also in policy institutions. Essentially, DSGE models are systems of stochastic difference equations involving exogenous stochastic processes to explain patterns and dependence structures in the data; they try to explain how shocks affect economies. Additionally, continuous-time macro-financial models try to merge the two strands of economics and finance. They may be seen as a continuation of traditional finance theory in a macroeconomic context.

Interestingly, the most common exogenous shock processes involved in DSGE models are assumed to be small-order ARMA processes having short memory. Additionally, to a large extent, the choice of the exogenous processes is arbitrary, and most models involve simple first-order autoregressive processes (AR(1) processes).[34] DSGE models require these exogenous processes. Otherwise, the model would not generate the persistence or degree of autocorrelation in the data.[35] On the other hand, most models assigned to the continuous-time macro-financial literature build on the Itô calculus and stochastic differential equations and the Brownian motion as their source of uncertainty in their models.[36]

Overall, there appear to be two opposite strands in the literature. The first (subdivided into three sub-strands, as already mentioned) deals with various aspects of long memory in an economic and econometric context. The second strand uses exogenous short memory stochastic processes in their models to gain insights into the economy's operating principles. So, why do economic modelers not fall back on long memory processes when they set up their stochastic models?

Clearly, this thesis cannot answer this question. Instead, it tries to contribute to integrating both strands by considering long memory processes in the context of stochastic macroeco-

[34] See, e.g., Schorfheide (2011, p. 22).
[35] See, e.g., Schorfheide (2011, p. 12).
[36] Of course, there are finance models involving other stochastic sources such as Poisson processes or Lévy processes. However, both types of processes share the property of stochastically independent increments.

nomic models. More specifically, this thesis focuses on a real business cycle (RBC) DSGE model and a simplified continuous-time macro-financial model. The following research questions are addressed.

The first question addresses the theoretical aspects of whether and how a long memory process can be involved in these two models. Does a DSGE model remain solvable given this richer exogenous stochastic dynamics? How can it be solved? If a DSGE model stays solvable, given a long memory process, is its solution stable? Unsurprisingly, the first question can be affirmed: DSGE models remain solvable. Consequently, the implicated research question addresses the effects on the model more deeply. More precisely, does the model economy respond differently to a shock exhibiting long memory than to a shock exhibiting short memory? If so, can these differences be quantified? and how can they be quantified? Given the infinitely-lived representative agent in the model, will he make a different consumption and labor supply decision when faced with a long memory shock than when faced with a short memory shock?

Similar questions arise in the context of the continuous-time macro-financial model. In contrast to DSGE models, these models try to capture the whole dynamics of a model's variable instead of focusing on the model's dynamics in a vicinity of a steady state. Consequently, the question to be addressed in the continuous-time model is whether and how the endogenous model dynamics are affected by long memory.

These questions regarding the DSGE model are tackled in Chapter 4, where a discrete-time RBC-DSGE model that builds the framework for the proposed long memory approach is set up. Although medium to large-scale DSGE models incorporating dozens of exogenous shocks and parameters have been built and analyzed in the literature nowadays, this thesis focuses on a simpler structure with a transitory technology shock and a technology growth shock as the only sources of uncertainty. Since most modern medium- and large-scale DSGE models are built on such a core RBC model, it seems reasonable to focus on this core as a first step. The focus on an RBC model is owed to the first research question; since it is apriori not clear whether and how a DSGE model can be solved, given this richer long memory dynamics, it seems reasonable to reduce the model-implied intricacy. To illustrate the different shock impacts on the model economy, the transitory technology shock considered in this thesis is an ARFIMA$(1, d, 0)$ process. This process has two parameters ϱ and d. If $d = 0$ one can recover the AR(1) setup frequently used in the literature. For $0 < d < 1/2$, the process shows long memory. This specification allows distinguishing three cases: The standard case of pure short memory ($\varrho > 0$, $d = 0$), the case of pure long memory ($\varrho = 0$, $0 < d < 1/2$), and the case of short and long memory

($\varrho > 0$, $0 < d < 1/2$). Additionally, the impacts of these shocks' specifications on the economy are compared to a growth shock in technology that pushes the economy on a higher balanced growth path. The growth shock has a permanent character, while both, the short and long memory shocks, have a transitory character as their impacts finally die out. Considering all these cases enables a direct comparison between the effects of these shock specifications. Moreover, since the original case of an AR(1) process is nested within these processes, the direct implications of long memory can be revealed.

The continuous-time model is treated in Chapter 5. Since fBm is inconsistent with the martingale framework of the modern finance theory from which the continuous-time macro-financial literature borrows extensively, the path for introducing long memory in such a model seems not as straightforward as in the discrete-time model. Consequently, the approach followed in this thesis is inspired by a stochastic process that can be regarded as an approximation of fBm while keeping the martingale methods applicable. The model into which this process is plugged in comes from Brunnermeier and Sannikov (2016, Section 2 on pp. 1504ff.). Similar to the discrete-time DSGE model, the structure of the model is relatively simple to abstract from model-implied complexities. However the results derived in Chapter 5 indicate that a generalization to more sophisticated models appears difficult.

By answering the research questions and following the sketched research agenda, this thesis contributes in various ways to the existing literature. A growing literature questions the somewhat restrictive assumptions regarding the exogenous stochastic processes made by the DSGE literature. This literature proposes more sophisticated exogenous processes (mostly higher-order ARMA processes or their vector-valued counterparts) in the context of DSGE models.[37] To the best of the author's knowledge, non of these deal with the possibility of long memory in the exogenous processes. Therefore, this thesis represents the first step towards fractionally integrated DSGE models that allow for richer persistence dynamics than the already discussed ones. More precisely, it is shown that such a long memory DSGE model is indeed solvable, and the model's solution is derived with the widely applied solution method of Klein (2000). Another contribution is given by the considerations of the continuous-time macro-financial model: This thesis aims to combine the above-mentioned developments in the financial mathematics literature, which allows for long memory in the exogenous dynamics, and the continuous-time macro-financial literature, which frequently deals with Brownian or Poisson uncertainty, both associated with stochastic independence.

The structure of the thesis closely follows the line of reasoning given by the research

[37] This literature is considered in more detail in Section 4.3.

questions mentioned above. Chapter 2 is an extended introduction to long memory. More precisely, Section 2.1 summarizes the mathematical preliminaries necessary for the following analysis. There, the traditional short memory ARMA processes are also treated briefly, focusing on the AR(1) process, which is used later in Chapter 4 as the reference technology shock in the DSGE model. The aim of Section 2.2 is to go deeper into the works of Mandelbrot and the Hurst phenomenon. This section illustrates that ARMA processes cannot replicate the Joseph effect. The link between Hurst's findings and the non-summability of a stationary process's ACF is also carved out in more detail. It will turn out that the R/S statistic mentioned at the beginning of this introduction has some disadvantageous properties. Therefore, Section 2.3 introduces the definition of long memory more rigorously. This section also distinguishes between the notions of "memory" and "persistence". At the beginning of Section 2.4, it is outlined that all ARMA processes are short memory processes, thereby calling for more sophisticated time series models to replicate long memory. Consequently, Section 2.4 introduces the class of ARFIMA processes. The last section of Chapter 2, Section 2.5, bridges Mandelbrot's continuous-time fractional Brownian motion and discrete-time ARFIMA processes. The considerations of this section will then be caught up again in Chapter 5.

Chapter 3 discusses relevant literature dealing with long memory in economics and econometrics. The structure of Chapter 3 corresponds to the three already mentioned strands of the economic long memory literature. Consequently, Section 3.1 (the first strand) provides a short review of existing estimation techniques. Additionally, some difficulties in accessing long memory from an empirical perspective are illustrated. The following section, Section 3.2, summarizes empirical evidence of long memory. The focus lies on the gross domestic product (GDP) and related time series (such as GDP growth rates or the gross national product etc.) as GDP appears to be the major variable in the RBC model of Chapter 4. Nevertheless, three additional examples of long memory in economic time series and their theoretical implications are given in this section. Section 3.3 finally reviews some theoretical approaches for generating long memory discussed in the literature. Overall, the aim of Chapter 3 is to illustrate that long memory is a relevant topic from an economic perspective and still an active research field to which this thesis aims to contribute.

Chapter 4 deals with the RBC-DSGE model in more detail. Section 4.1 outlines the structure of the model and specifies production functions, utility functions, and the exogenous technology shocks. Section 4.2 treats DSGE models from a theoretical perspective. This section shows that if the exogenous processes, plugged into a DSGE model, are short memory processes, so is the model's solution. This result justifies the approach followed in this thesis by introducing long memory via an exogenous stochastic process into a DSGE model. Additionally, this section outlines the concept of stability in the context of DSGE models. Section 4.3 provides additional justification for the approach followed in this thesis by summarizing the literature that also questions the AR(1) assumption of the DSGE literature. The method for solving the long memory DSGE model is given in Appendix B.5 and is not discussed in the main text to maintain readability. Section 4.4 analyzes the model responses to the different specifications of the exogenous processes.[38]

Chapter 5 considers the continuous-time model. Section 5.1 outlines the difficulties when introducing fBm into stochastic differential equations and briefly discusses some proposed approaches in the literature to overcome these difficulties. This section finally defines the shock-generating process for the model showing long memory. Section 5.2 outlines the model setup and Section 5.3 derives the model's solution. Section 5.4 discusses the results and the implications for the benchmark model in the presence of long memory.

Chapter 6 concludes and points to some topics for future research.

[38] The reader may wonder why the author decided to treat two rather theoretical sections (Section 4.2 and Section 4.3) between the introduction of the model and the presentation of its results. In the author's opinion, these two sections are better accessed by the reader if they have a concrete model structure in mind when reading them.

2

Long Memory, the Joseph Effect and the Hurst Phenomenon

As the thesis aims to investigate long memory processes in economic models, this chapter provides the necessary methodological foundations for the model of Chapter 4 and Chapter 5. Section 2.1 outlines briefly the mathematical concepts needed for the considerations of the following sections. After a brief introduction to stochastic processes and their properties such as stationarity and persistence, a closer look is taken at so-called autoregressive moving average (ARMA) processes that play a significant role in dynamic stochastic general equilibrium (DSGE) models such as those considered in Chapter 4.

Section 2.2 is a short historical review of Mandelbrot's Joseph effect, which is nowadays referred to as long memory or long-range dependence. It is also shown that traditional time series models, such as ARMA processes, fail to replicate the Joseph effect.

The aim of Section 2.3 is to introduce a formal definition of long memory. Additionally, Section 2.4 focuses on a class of discrete-time long memory models, so-called autoregressive fractionally integrated moving average (ARFIMA) processes, which generalize ARMA processes. Furthermore, a closer look is taken at the ARFIMA$(1, d, 0)$ process which is at the core of Chapter 4's DSGE model.

Section 2.5 introduces a continuous-time stochastic long memory process called fractional Brownian motion (fBm) and illustrates some relations to ARFIMA processes. The theoretical considerations of Section 2.5 are picked up again in the continuous-time macro-financial model of Chapter 5. The accompanying appendix, Appendix A, contains additional material such as the proofs of the lemmas given in the text (Appendix A.1), some details on the definition of the gamma and the Gaussian hypergeometric function (Appendix A.2) and some moments of the ARFIMA$(1, d, 0)$ process (Appendix A.3).

2.1 Some Mathematical Preliminaries

2.1.1 Stochastic Processes, Moments and Stationarity

As mentioned in the introduction, long memory refers to a property of time series or stochastic processes. Therefore, the author briefly reviews the key mathematical concepts necessary for the following. A formal definition of a random variable and a stochastic process is as follows.[39]

Definition 2.1.1 ———

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

i) Then, an $\mathcal{F}$-measurable function $Y : \Omega \to \mathbb{R}$ is called random variable.

ii) Let $\mathbb{T}$ be an index set. A real-valued stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is a family of random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assuming values in $\mathbb{R}$.

$\rtimes$

The index set $\mathbb{T}$ in the former definition is assumed to be a subset of $\mathbb{R}$ through the whole thesis and can be either a continuum, e.g., the non-negative real line $\mathbb{T} = [0, \infty)$ or discrete, e.g., $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{Z}$.[40]

The definition of a stochastic process offers two ways of interpretation. First, by fixing a certain instant of time, say t_0, $X_{t_0}(\cdot)$ defines a function mapping from Ω to $\mathbb{R}$, i.e., for each instant of time, the process is a random variable. Second, by fixing a specific random event ω. The function $t \to X_t(\omega)$ is called the path of the stochastic process, i.e., for different values of ω, one obtains different functions of time, yielding the interpretation of a stochastic process as a random function.[41]

Probabilistic determinants such as distributions, moments, or the degree of dependence are essential to describe the properties of a stochastic process. An important subgroup of

---

39 See, e.g., Øksendal (2013, p.8 and Definition 2.1.4 on p. 10).

40 Note that the previous definition is restricted to the univariate case that is assumed throughout the thesis. Generalizations to higher dimensions are mentioned in the text if they become necessary.

41 See Øksendal (2013, pp. 10f.) and Brockwell and Davis (1987, Definition 1.2.2 on p. 9).

stochastic processes is given by (strictly) stationary processes that are characterized by a time-invariant distribution, i.e., X_{t+h} has the same distribution as X_t for all $t, h \in \mathbb{T}$.[42] In the context of time series analysis, the consistency is not assumed to hold for the whole distribution but for different moments, leading to the notion of weak stationarity or covariance stationarity. Before this concept is outlined in more detail, some key moments of a random variable and stochastic processes are repeated briefly.

In the following, the (unconditional) expected value of a random variable X is denoted by $\mathbb{E}(X)$.[43]

Assume that $\mathbb{E}(|X|^2) < \infty$, then the variance of X is defined as $\text{var}(X) := \mathbb{E}((\mathbb{E}(X) - X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X)^2)$. If Y is another random variable defined on the same probability space as X with $\mathbb{E}(|Y|^2) < \infty$, then the covariance between X and Y is defined by $\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.[44]

Definition 2.1.2

Let $(X_t)_{t\in\mathbb{T}}$ be a real-valued stochastic process as defined in Definition 2.1.1. Then $(X_t)_{t\in\mathbb{T}}$ is called (covariance) stationary, if[45]

i) $\mathbb{E}(|X_t|^2) < \infty$ *for all* $t \in \mathbb{T}$.

ii) $\mathbb{E}(X_t) \equiv \mu \in \mathbb{R}$.

iii) the autocovariance function $\gamma_X(t, s) : \mathbb{T} \times \mathbb{T} \to \mathbb{R}, \gamma_X(t, s) = \text{cov}(X_t, X_s)$, *satisfies* $\gamma_X(s + h, t + h) = \gamma_X(s, t)$, *for all* $s, h, t \in \mathbb{T}$.

⋊

It follows immediately from the previous definition that covariance stationary processes have a constant expected value and variance at each instance of time. However, no inference can be made about higher moments. Strictly stationary processes have time-consistent distributions that pin down all moments, i.e., strict stationarity implies covariance stationarity if $\mathbb{E}(|X_t|^2)) < \infty$.[46] In the following, a stochastic process is said to be stationary if it is covariance stationary.[47]

To describe the dependence structure of a stochastic process, the autocorrelation function

[42] See, e.g., Klenke (2013, Definition 20.1 on p. 449).
[43] Here, it is implicitly assumed that the expected value of X exists. A necessary condition for this is $\mathbb{E}(|X|) < \infty$, see, e.g., Klenke (2013, Definition 4.7 on p. 90 and Definition 5.1 on pp. 103f.).
[44] See, e.g., Klenke (2013, Definition 5.1 on pp. 103f.).
[45] This definition is adapted from Brockwell and Davis (1987, Definition 1.3.2 on p. 12).
[46] See Brockwell and Davis (1987, p. 13).
[47] This is a usual convention in the time series literature, see, e.g., Brockwell and Davis (1987, p. 12) and Hamilton (1994, p. 46).

(ACF), which is now formally defined, becomes essential.

Definition 2.1.3 ———

Let $(X_t)_{t\in\mathbb{T}}$ be a real-valued stationary stochastic process. The autocorrelation function of X is defined as follows[48]

$$\rho_X : \mathbb{T} \to [-1, 1], \rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)}.$$

⋊

The ACF measures how strongly the process is correlated with a lagged version of itself.[49] If the process shows positive autocorrelations, subsequent observations tend to lie on the same side of the mean, i.e., a positive k-th order ACF indicates the observation at time t to lie above the mean if the observation at $t-k$ already lied above the mean.[50] If negative autocorrelations are present, the affected observations tend to lie on different sides of the mean, implying a more jagged path.[51]

2.1.2 Wold Decomposition and Impulse-Response Functions

This section focuses on stationary discrete-time (i.e., $\mathbb{T} = \mathbb{Z}$) stochastic processes. By the well-known Wold decomposition, such processes can be uniquely decomposed in a purely deterministic and a purely non-deterministic part. To be more precise, let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process, then it yields[52]

$$Y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k} + V_t \tag{2.1}$$

with a white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}, \psi_0 = 1$ and $\sum\limits_{k=0}^{\infty} \psi_k^2 < \infty$.[53]

The process $(V_t)_{t\in\mathbb{Z}}$ is called deterministic as it is "perfectly predictable"[54] from past values of Y.[55] It is important to note that deterministic in this sense does not mean that the

48 From Definition 2.1.2, it follows that $\text{cov}(X_s, X_t) = \gamma_X(s,t) = \gamma_X(s-t, 0)$. Thus, it is convenient to write $\gamma_X(k) = \text{cov}(X_t, X_{t+k})$, for $k \in \mathbb{T}$. See again Brockwell and Davis (1987, Remark 2 and 3 on p. 12).

49 See Hamilton (1994, p. 45).

50 See Brockwell and Davis (1987, Remark 7 on p. 29).

51 See Brockwell and Davis (1987, Remark 7 on p. 29).

52 See Brockwell and Davis (1987, Theorem 5.7.1 on pp. 180f.).

53 According to Definition 3.1.1 in Brockwell and Davis (1987, p. 78), $(\varepsilon_t)_{t\in\mathbb{Z}}$ is called white noise if $\mathbb{E}(\varepsilon_t) \equiv 0$, $\text{var}(\varepsilon_t) \equiv \sigma_\varepsilon^2$ and $\gamma_\varepsilon(k) = 0$ for $k > 0$. Additionally, it is often assumed that $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of Gaussian independent and identically distributed (i.i.d.) random variables. In such a case, the process is called Gaussian white noise.

54 Brockwell and Davis (1987, p. 180).

55 See also Hamilton (1994, Proposition 4.1 on p. 109).

process is not random. It can be shown that the processes ε and V are uncorrelated and that $\mathbb{E}(X_t) = \mathbb{E}(V_t)$ for all $t \in \mathbb{Z}$.[56] Within the scope of this thesis, however, processes with at most a constant value of $V_t \equiv V$ are relevant. Then the process Y can be transformed to have zero-mean by defining $X_t := Y_t - V$. Eventually, (2.1) becomes

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}. \tag{2.2}$$

Processes satisfying (2.2) are called infinite moving average processes.[57]As will turn out in the following sections, the discrete-time processes the reader is concerned with can be represented as an infinite moving average process. Thus, at this stage, it is justified to describe and analyze such processes in more detail.[58]

The infinite moving average representation reveals an important interpretation. By taking the derivative of (2.2) with respect to ε_t, one obtains

$$\frac{\partial X_{t+j}}{\partial \varepsilon_t} = \frac{\partial}{\partial \varepsilon_t} \sum_{k=0}^{\infty} \psi_k \varepsilon_{t+j-k} = \psi_j. \tag{2.3}$$

This expression illustrates that ψ_j is the impact of an innovation of size one at time t on X_{t+j}. This observation motivates the following definition.[59]

Definition 2.1.4

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process with an infinite moving average representation as in (2.2). The series of coefficients $(\psi_j)_{j=0}^{\infty}$ is called impulse-response function (IRF) of the process $(X_t)_{t\in\mathbb{Z}}$. ⋊

Note that the right-hand side of (2.3) does not depend on t.[60] Thus, the response to a shock is independent of the time of the shock occurrence; it just depends on the number of

[56] See Brockwell and Davis (1987, Theorem 5.7.1 on pp. 180f.).

[57] In addition, Hassler (2019, pp. 33ff.) motivates the removal of the deterministic component V_t due to ergodicity concerns.

[58] Note that the condition $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ is not very strong. It follows from Brockwell and Davis (1987, Theorem 3.2.1 on p. 90) that the autocovariance function of an infinite moving average process X is given by $\gamma_X(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$, where σ_ε^2 is the variance of the white noise process in the moving average representation. Hence, $\text{var}(X) = \gamma_X(0) = \sum_{k=0}^{\infty} \psi_k^2$ and the imposed condition is necessary to ensure a finite variance.

[59] Definition 2.1.4 adapted from Hassler (2019, p. 33) and Hamilton (1994, p. 2ff.).

[60] Clearly, the IRF $(\psi_j)_{j=0}^{\infty}$ depends on the stochastic process X, i.e., one could have written $(\psi_j^X)_{j=0}^{\infty}$ instead of $(\psi_j)_{j=0}^{\infty}$ to express this dependency. However, the superscript X is suppressed in the following to keep the notation simple. Additionally, the context of the formulas given in the text shall make clear to which process the coefficients ψ_j belong.

periods following the shock.[61,62] Furthermore, as can be deduced from the linear structure of (2.3), the response to a shock of magnitude other than one, say $\sigma \neq 1$, is given by $(\sigma\psi_j)_{j=0}^{\infty}$.

As will turn out in Chapter 4, the analysis of IRFs play a major role in the context of (discrete-time) dynamic stochastic general equilibrium (DSGE) models. There, IRFs show the economy's responses to exogenous shocks hitting the economy's equilibrium state. Thus, they allow investigations on how the treated economic variables are affected by the shock under consideration and how long it takes for the economy to reach (if at all) its equilibrium state again. In the end, the coefficients of the infinite moving average representation in such a DSGE context depend on the specified model parameters and the properties of the exogenous stochastic processes.[63]

2.1.3 Persistence

It follows from the square summability of the IRF, i.e., $\sum_{k=0}^{\infty} \psi_k^2 < \infty$, that $\psi_j \to 0$, as $j \to \infty$. Consequently, the response to a shock dies out over time for all stationary moving average processes of the form (2.2).

However, the square summability of the IRF does not indicate how fast or slow this decay will ultimately be. Therefore, the decay of the IRF is referred to as the persistence of the process, or more precisely "[...] the magnitude of the IRF across different time horizons indicates how much persistence is present in the series."[64,65]

The IRF is an infinitely long series of numbers, so it seems not quite tractable to describe or compare the persistence of different processes along their IRF.[66] For this reason, D. W. K. Andrews and H.-Y. Chen (1994) propose a scalar measure of persistence, namely the cumulative impulse response (CIR) for which the formal definition is given now.

Definition 2.1.5 ———

Let $(X_t)_{t\in\mathbb{Z}}$ be as in (2.2) with IRF $(\psi_k)_{k=0}^{\infty}$ and $\sum_{k=0}^{\infty} \psi_k^2 < \infty$, then the cumulative impulse

61 See Hamilton (1994, p. 3).
62 For these reasons Hamilton (1994, p. 2f. and 442f.) calls the series of coefficients $(\psi_j)_{j=0}^{\infty}$ "dynamic multiplier".
63 See Chapter 4 for more details.
64 D. W. K. Andrews and H.-Y. Chen (1994, p.187).
65 A similar definition is provided by Pivetta and Reis (2007, p. 1329).
66 See D. W. K. Andrews and H.-Y. Chen (1994, p. 189) and Pivetta and Reis (2007, p. 1329).

Response (CIR) of X is given by

$$\mathrm{CIR} = \sum_{k=0}^{\infty} \psi_k.$$

⋊

Thus, the CIR is simply the cumulative sum of the IRF and summarizes the properties of the IRF in a wield scalar value.[67] As the IRF shows the effects of a unit shock on the process in the periods after the shock occurrence, the CIR is the cumulative effect of this shock.[68]

On the other hand, consider that in each period following t_0, all values of ε_{t_0+j} for $j = 0, 1, 3, \ldots$ are increased by one unit, then the effect on X_{t_0+j} is given by

$$\sum_{i=0}^{j} \frac{\partial X_{t_0+j}}{\partial \varepsilon_{t_0+i}} = \sum_{i=0}^{j} \psi_{j-i} = \sum_{k=0}^{j} \psi_k \longrightarrow \mathrm{CIR}, \text{ as } j \to \infty.$$

Hence, the CIR may further be viewed as the effect of a permanent increase by one unit of the exogenous shock ε on the process X. This justifies the term "total multiplier"[69] or "long-run effect"[70] for the CIR. Thus, the CIR may be seen as the cumulative effect of a transitory shock or as the long-run effect (defined in the above sense) of a permanent shock.[71]

However, there are some drawbacks to summarizing persistence by CIR. Two processes may have the same CIR, but the IRF's "mass" is not equally distributed over time between the processes. Such an unevenly distributed "mass" would be the case, e.g., if one process's IRF is characterized by a strong increase followed by a rapid decrease compared to a process's IRF showing a slight decrease over all periods.[72] Information about the shape of the IRF is lost by looking only at the CIR. For this reason, the CIR is used to roughly classify stochastic processes into three main groups:[73]

Definition 2.1.6

Let X be as in Definition 2.1.5. If i) $\mathrm{CIR} = 0$*, X is called anti-persistent; ii)* $0 < \mathrm{CIR}^2 <$

[67] Similar to Footnote 60, an index expressing the dependency of the CIR on X is suppressed in the following.

[68] This can be seen clearly by plugging (2.3) into the definition of the CIR, i.e., $\mathrm{CIR} = \sum_{k=0}^{\infty} \frac{\partial X_{t+k}}{\partial \varepsilon_t}$.

[69] Hassler (2019, p. 37).

[70] Hamilton (1994, p. 6).

[71] See Hamilton (1994, p. 7).

[72] See D. W. K. Andrews and H.-Y. Chen (1994, p. 190).

[73] Definition 2.1.6 is inspired by Hassler (2019, Definition 2.5 on pp. 23f.). Note that Hassler (2019) motivates the persistence from the long-run variance of a process. As will turn out, both concepts lead to the same conclusions, see Hassler (2019, Proposition 3.4 on p. 38).

∞, X is called moderately persistent; iii) $\text{CIR}^2 = \infty$*, X is called strongly persistent.* ⋊

The three cases do not contradict the condition $\sum_{k=0}^{\infty} \psi_k^2 < \infty$. In the case of an anti-persistent process, the IRF has at least some negative values, i.e., the positive responses to the shock are balanced by negative ones such that the cumulative response disappears. Shocks to a moderately persistent process die out quickly to ensure the convergence of the CIR. In this case, the positive effects outweigh the negative or vice versa. Finally, a divergent CIR, i.e., CIR $= \pm\infty$, indicates a very slowly decaying IRF and, thus, justifies the notion of strong persistence.

Overall, it can be argued that processes with higher CIR are more persistent than those with lower CIR. It should be noted, however, that there is no single definition of the term "persistence" in the literature. In the context of ARMA processes, additional measures are discussed and briefly reviewed in Section 2.1.5.2. However, these measures are partly context-dependent and require more knowledge of the underlying process than the more general structure of an infinite moving average process assumed here.

2.1.4 Spectral Density

So far, some general properties of stochastic processes have been described in the so-called time domain. In this subsection, some additional properties of stationary discrete-time stochastic processes are described in the frequency domain. As shown in Section 3.1.2, the frequency domain provides opportunities for parameter estimation, so the basic foundations are presented here for convenience.

It is a classical theorem in time series analysis that a zero-mean stationary stochastic process can be viewed as a superposition of (possibly infinitely many) sinusoidal functions with different frequencies.[74] Thus, each time series may be seen as a sum of cyclical fluctuations at different frequencies with stochastic amplitudes. Further, there is an analogous representation for the autocovariance function of stationary stochastic processes:[75]

Definition 2.1.7 ——
A function f_X is the spectral density of a stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ with autocovariance function $\gamma_X(\cdot)$ if

i) $f_X(\lambda) \geq 0$ *for all* $\lambda \in (-\pi, \pi]$*, and*

[74] See Brockwell and Davis (2016, pp. 102f.) and Brockwell and Davis (1987, Theorem 4.8.2 on p. 140) for a proof of the "Spectral Representation Theorem".

[75] See Brockwell and Davis (2016, Definition 4.1.1 on p. 99).

ii) $\gamma_X(k) = \int_{-\pi}^{\pi} e^{ik\lambda} f_X(\lambda)\, d\lambda = \int_{-\pi}^{\pi} (\cos(\lambda k) + i \sin(\lambda k)) f_X(\lambda)\, d\lambda$ *for all* $k \in \mathbb{Z}$.

⋊

Although complex numbers occur in the preceding definition, the spectral density is a real-valued function, as one can deduce from symmetry properties.[76] Especially, it yields

$$\gamma_X(k) = 2 \int_0^{\pi} \cos(\lambda k) f_X(\lambda)\, d\lambda. \tag{2.4}$$

On the other hand, by assuming $\sum_{k=0}^{\infty} |\gamma_X(k)| < \infty$, one can recover the spectral density from $\gamma_X(\cdot)$ by[77]

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik\lambda} \gamma_X(k) = \frac{1}{2\pi} \left(\gamma_X(0) + 2 \sum_{k=1}^{\infty} \gamma_X(k) \cos(\lambda k) \right), \tag{2.5}$$

where the second equality uses some symmetric properties of the sinus and cosine functions. Both (2.4) and (2.5) emphasize that the spectral density and the autocovariance function of a stationary stochastic process contain the same information, which can be easily transformed into each other. The serial correlation represented by γ_X can be mapped into the frequency domain by (2.4), and the frequency domain information stored in f_X can be transformed into the time domain by (2.4).[78]

By setting $k = 0$ in (2.4), one further obtains that $\mathrm{var}(X_t) = \gamma_X(0) = \int_{-\pi}^{\pi} f_X(\lambda)\, d\lambda$. This expression states that the variance of a stationary stochastic process is given as the integral of the spectral density over the whole cycle frequencies. Thus, one can analyze how certain frequencies contribute to the variance. To be more precise, the integral

$$\frac{1}{\gamma_X(0)} \int_{-\bar{\lambda}}^{\bar{\lambda}} f_X(\lambda)\, d\lambda = \frac{2}{\gamma_X(0)} \int_0^{\bar{\lambda}} f_X(\lambda)\, d\lambda$$

represents the percentage contribution of the cycle frequencies less than $\bar{\lambda}$ (in absolute terms) to the process's variance.[79,80]

Note that the frequency of a general sinusoid $s(t) = a \sin\left(\bar{\lambda} t + b\right) + c$ with some real constants a, b, c is given by $\bar{\lambda} \in [0, \pi]$, i.e., the frequencies summarizes how many cycles

76 The spectral density is non-negative, i.e., $f_X(\lambda) \geq 0$ and symmetric, i.e., $f_X(\lambda) = f_X(-\lambda)$, see Brockwell and Davis (1987, Remark on p. 120).

77 See Brockwell and Davis (1987, Corollary 4.3.2 on p. 118).

78 See Hassler (2019, p. 66).

79 See Hamilton (1994, pp. 156-162).

80 For better comparability, it is convenient to consider the normalized spectral density $\bar{f}_X(\lambda) = f_X(\lambda)/\gamma_X(0)$ in order to unitize the area under the spectral density.

are completed during 2π periods.[81] The corresponding time needed for one complete cycle duration is inversely related to the frequency. It is defined as the period of the cycle, i.e., period $= \left|2\pi/\bar{\lambda}\right|$.[82] Thus, small (high) frequencies correspond to longer (shorter) cycle periods. If low frequencies contribute mainly to the process's variance, one would expect the process's path to be smoother than that of a process whose variance is mostly explained by high-frequency components.[83]

Hence, the spectral density reveals much information about which cycles the process is driven by. As stated above, a smooth path of a stochastic process is associated with a higher degree of autocorrelation and a large mass of the spectral density on low frequencies. Thus, high autocorrelations come along with long cycles contributing to the process's variance.

In the context of an infinite moving average process as the one given in (2.2), the spectral density offers insights into the persistence properties of the process. To be more precise, it follows from (2.5) that[84]

$$f_X(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) = \frac{1}{2\pi} \left(\gamma_X(0) + 2 \sum_{k=1}^{\infty} \gamma_X(k) \right). \tag{2.6}$$

Plugging the values of γ_X (see Footnote 58) into (2.6) yields

$$f_X(0) = \frac{\sigma_\varepsilon^2}{2\pi} \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|} = \frac{\sigma_\varepsilon^2}{2\pi} \left(\sum_{k=0}^{\infty} \psi_k \right)^2 = \frac{\sigma_\varepsilon^2}{2\pi} \mathrm{CIR}^2, \tag{2.7}$$

where the second equality is derived similarly to Brockwell and Davis (1987, p. 103), and the third equality uses Definition 2.1.5. Equation (2.7) highlights that the value of the spectral density at the zero frequency is proportional to the squared CIR. Thus it follows immediately from Definition 2.1.6 that an infinite moving average process is moderately persistent if $0 < f_X(0) < \infty$ and strongly persistent if $f_X(\lambda) \to \infty$, as $\lambda \to 0$. Correspondingly, the process is anti-persistent if $f_X(0) = 0$.[85,86]

Spectral analysis is of general interest in economic and econometric contexts as early studies by Adelman (1965) and Granger (1966) indicate some stylized facts about the

81 See Hamilton (1994, p. 708).
82 See Hamilton (1994, p. 708).
83 See Brockwell and Davis (2016, pp. 104f.).
84 This is due to $\cos(0) = 1$ and $\gamma_X(k) = \gamma_X(-k)$.
85 See further Hassler (2019, pp. 66f.).
86 Note that $\sum_{k=1}^{\infty} \gamma_X(k) \geq 0$ due to the non-negative definiteness of the autocovariance function, see Brockwell and Davis (1987, pp. 26f.) .

"typical spectral shape"[87] of economic variables. To be more precise, Adelman (1965) investigates the spectral density of a battery of economic variables in the U.S., such as output, consumption, labor productivity, and others (each in terms of deviations from a deterministic trend) and founds that all spectral densities look similar, showing a sharp increase as the frequency approaches zero.[88] It follows from (2.6) and (2.7) that a peak at (or near) the zero frequency is associated with a high value of accumulated autocovariances and a high level of persistence, respectively.

2.1.5 ARMA Processes

In this subsection, an important family of stationary stochastic processes, namely ARMA processes, is described briefly. Besides their statistical relevance, they frequently occur as the source of stochastic shocks in various economic models.[89] As to be shown later in the thesis, their properties differ mainly from them of long memory processes that build the focus of this thesis. Thus, ARMA processes will serve as a benchmark for further considerations. Additionally, fractionally integrated ARMA processes, as a vital representative of discrete-time long memory processes (considered in more detail in Section 2.4), are easily constructed from ARMA processes. In this regard, restating some necessary notations and properties of ARMA processes seems convenient. The following definition is adapted from Brockwell and Davis (1987).[90]

Definition 2.1.8 ——
Let p and q be non-negative integers. Then $(X_t)_{t\in\mathbb{Z}}$ is called ARMA(p,q) process if $(X_t)_{t\in\mathbb{Z}}$ is stationary and if for all $t \in \mathbb{Z}$,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \tag{2.8}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process with variance σ_ε^2. ⋊

Equation (2.8) may also be written in terms of the lag or backshift operator B defined as $BX_t = X_{t-1}$, i.e., (2.8) becomes

$$\phi(B)X_t = \theta(B)\varepsilon_t, \tag{2.9}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ denotes the autoregressive polynomial and $\theta(z) =$

[87] Granger (1966, p. 150).
[88] See Adelman (1965, Figure 4-11 on pp. 457f.).
[89] See, for example, the DSGE models of Chapter 4.
[90] See Brockwell and Davis (1987, Definition 3.1.2 on p. 78).

$1+\theta_1 z+\cdots+\theta_q z^q$ the moving average polynomial. Thus, ARMA processes are represented (at each instant of time) by linear combinations of their own past values and the present and past values of an exogenous white noise process. ARMA processes reveal a high empirical relevance as each given autocovariance function (e.g., from an observed time series) can be approximated arbitrarily well by the autocovariance function of a certain ARMA process.[91]

Assuming that $\phi(\cdot)$ and $\theta(\cdot)$ do not have common roots, and further, that all (possibly complex-valued) roots of $\phi(\cdot)$ lie outside the unit circle, i.e., $\phi(z) \neq 0, |z| \leq 1$, ensures the existence of an infinite moving average representation as in (2.2) where the coefficients $(\psi_k)_{k=0}^{\infty}$ are obtained from a time series expansion of the function $\psi(z) = \theta(z)\phi(z)^{-1}$.[92] In this case, the ARMA process is called causal, as it can be expressed as a linear combination of the past values of $(\varepsilon_t)_{t\in\mathbb{Z}}$.[93] On the contrary, if the roots of $\theta(\cdot)$ lie outside the unit circle, the white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ can be expressed as a linear combination of past values of $(X_t)_{t\in\mathbb{Z}}$. In such a case, the ARMA process is called invertible.[94] Commonly (and throughout this thesis), it is assumed that an ARMA process is causal and invertible.[95]

Without moving average terms, the corresponding ARMA$(p, 0)$ process is simply an autoregressive process of order p for which the shorthand notation AR(p) is used in the following.

2.1.5.1 ACF, Spectral Density and CIR

The derivation of closed-form expressions of the autocovariance function of a general ARMA process involves the solution of possibly high-order linear difference equations. The derivation of these equations is beyond the scope of this thesis.[96]

In order to gain some intuition for the autocovariance function of an ARMA process, the closed-form expressions are, however, not needed since the autocovariance function for all ARMA processes is geometrically bounded. To be more precise, let X be an ARMA(p, q) process with autocovariance function γ_X, then, there are constants $0 < C$ and $0 < \beta < 1$

[91] See Brockwell and Davis (1987, p. 77).
[92] See Brockwell and Davis (1987, Theorem 3.1.1 on p. 85).
[93] See Brockwell and Davis (1987, Definition 3.1.3 on p. 83).
[94] See Brockwell and Davis (1987, Definition 3.1.4 on p. 86 and Theorem 3.1.2 on pp. 86f.).
[95] It seems reasonable to make this assumption, as one can find a similar representation of (2.9) if $\phi(\cdot)$ and $\theta(\cdot)$ have common roots such that the corresponding process $(X_t)_{t\in\mathbb{Z}}$ is causal and invertible as long as the roots of $\theta(\cdot)$ and $\phi(\cdot)$ lie not on the unit disk, i.e., $\phi(z) \neq 0 \neq \theta(z)$ for $|z| = 1$, see Brockwell and Davis (1987, Remark 5 on p. 88). If $\phi(z) = 0$ for $|z| = 1$, then X is not stationary, see Brockwell and Davis (1987, Remark 3 on p. 86).
[96] See Brockwell and Davis (1987, Section 3.3 on pp. 91ff.) for details. The special case of an AR(1) process is treated separately in Section 2.1.5.3.

such that[97]

$$|\gamma_X(k)| \leq C\beta^k., \tag{2.10}$$

Consequently, the autocovariance function and, thus the ACF, decays exponentially fast in absolute terms. The values of C and β depend on the parameters of the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ in (2.9).[98]

The corresponding spectral density is given by

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{\left|\theta(e^{-i\lambda})\right|^2}{\left|\phi(e^{-i\lambda})\right|^2}, \tag{2.11}$$

provided that $\phi(z) \neq 0$ for $|z| = 1$.[99] Equations (2.7) and (2.11) imply that

$$\text{CIR} = \sqrt{\frac{2\pi}{\sigma_\varepsilon^2} f_X(0)} = \frac{|\theta(1)|}{|\phi(1)|} = \frac{\left|1 + \sum\limits_{k=1}^{q} \theta_k\right|}{\left|1 - \sum\limits_{k=1}^{p} \phi_k\right|}. \tag{2.12}$$

It follows from (2.12) that the CIR of all ARMA processes is finite since $\sum\limits_{k=1}^{p} \phi_k \neq 1$ due to stationarity reasons.[100] In general, an ARMA process can be anti-persistent, i.e., CIR $= 0$ if $\sum_{k=1}^{q} \theta_k = -1$. In this case, however, $\theta(1) = 0$, and the corresponding ARMA process is not invertible. Thus, it follows from (2.12) that all causal and invertible ARMA processes are moderately persistent. The magnitude of the CIR depends on the autoregressive and moving average parameters, but they cannot be chosen to meet the other persistence regimes mentioned in Definition 2.1.6.

2.1.5.2 A Note on Persistence

Note that there is not one uniform definition of the term “persistence” in the literature. Generally, the notion of persistence in terms of the CIR used in this thesis is rather general and comes along with some drawbacks already mentioned in Section 2.1.3. In the context of ARMA processes, which are all moderately persistent in the sense of this thesis, there are additional measures of persistence that allow for further persistence comparison. Some are related to the CIR, and some carve out the relationship between the autoregressive and moving average coefficients and the IRF in more detail. In order to fit the CIR in the

97 See Hassler (2019, Proposition 3.5 on pp. 49f.).

98 Note that (2.10) does not rule out a cyclical decay of the ACF. If the ACF decays cyclically, (2.10) states that the amplitudes decay exponentially.

99 See Brockwell and Davis (1987, Theorem 4.4.2 on p. 121).

100 If $\sum_{k=1}^{p} \phi_k = 1$, it follows that $\phi(1) = 0$ and thus the process is not stationary, see Footnote 95.

persistence measures discussed in the literature, a brief review of commonly used measures is provided. Again, the CIR applies to general processes of the form (2.2), while some of the other measures presented below assume less general representations as, e.g., (2.8).

In the context of purely autoregressive models, the sum of the autoregressive coefficients is used as a measure of persistence in the empirical literature.[101] Indicated by (2.12), the higher this sum is, the higher the CIR, and thus, the more persistent the process. Without moving average terms, both measures contain the same information. However, the sum of the autoregressive coefficients may underestimate the degree of persistence (in the sense of the CIR) in the presence of moving average terms, see (2.12).

Another measure of persistence in the context of purely autoregressive models is the largest root of the autoregressive polynomial $\varphi(\cdot)$.[102] The autoregressive polynomial can be written as $\varphi(z) = (1 - b_1 z) \cdots (1 - b_p z)$. Let $b = \max_{i=1,\dots,k} \{|b_i|\}$, then b is called the largest autoregressive root.[103,104] Since b determines the decay of the IRF of an AR(p) process, it is often used as a measure of persistence.[105] On the other hand, it may be criticized that the largest autoregressive root neglects (by definition) the effect of the other autoregressive roots on the IRF.[106]

Instead of referring to the properties of the autoregressive polynomial, the half-life is an additional indicator of persistence. It is equal to the number of periods needed for the IRF to reach a value of 1/2 after a shock of size one.[107] Some drawbacks of this measure arise if the IRF is characterized by up-and-down swings or a fast drop of the IRF with a slow decay afterward.[108] The latter can be observed in Section 2.4.2 (more specifically in Figure 2.7), where the IRFs of some strongly persistent processes, i.e., $\text{CIR}^2 = \infty$, show a smaller half-life than the one of an AR(1) process which is moderately persistent, i.e., $\text{CIR}^2 < \infty$.

A related measure of persistence is the number of periods needed for the IRF to reach

101 See Paya et al. (2007, p. 1523f.).
102 See Paya et al. (2007, p. 1524) and Pivetta and Reis (2007, p. 1329).
103 See Pivetta and Reis (2007, p. 1329).
104 Formally, $1/b$ is a root of $\varphi(\cdot)$ and not b, but this notation seems to be well-established in the literature, see, besides Pivetta and Reis (2007, p. 1329), DeJong and Whiteman (1991, p. 226) and Stock (1991, p. 437). However, the causality and invertibility conditions mentioned above refer to the “true” roots $1/b_1, \dots, 1/b_k$ instead.
105 See Pivetta and Reis (2007, p. 1329).
106 See Paya et al. (2007, p. 1524). D. W. K. Andrews and H.-Y. Chen (1994) show that the roots other than the largest have too much influence on the shape of the IRF. For this reason, they advocate the CIR as a better measure of persistence than b.
107 See Paya et al. (2007, p. 1524).
108 See Pivetta and Reis (2007, p. 1330).

50% of the CIR.[109] This measure is, however, only applicable as long as $0 < \text{CIR}^2 < \infty$.

An additional measure of persistence, completely independent from a certain model, is determined by how often a process crosses its mean value or, to be more precise, by the "unconditional probability of a stationary stochastic process [...] not crossing its mean in period t".[110] Assume that a shock pushes a process far from its mean value, then a more persistent process is expected to return to its mean value slower than a less persistent process. This low degree of mean reversion causes the process not to cross its mean value very often overall. Hence, the number of mean crossings can be seen as a measure of persistence.[111] As with the other measures, this one has some drawbacks. It can be shown that a purely white noise process is indicated to have the same persistence as an AR(2) process with parameters $\varphi_1 = 0$ and $\varphi_2 = 0.8$.[112] If the CIR is used instead, it follows from (2.12) that a white noise process has a CIR of one, while the AR(2) process has a CIR of five. Thus, according to the CIR, the AR(2) process would be more persistent than a white noise process, but they would be equally persistent if the number of mean crossings were used.

In summary, there are several measures of persistence, each with advantages and disadvantages. Sometimes the way one measure indicates persistence is contrary to what another measure suggests. Overall, the rather coarse distinction between anti, intermediate, and strong persistence made in Section 2.1.3 still seems appropriate since the scope of this thesis is to compare processes with infinite CIR and finite CIR and not between the nuances of two moderately persistent processes. There is, however, one process for which all of the mentioned persistence measures lead to equivalent results, namely the AR(1) process with parameter $\varphi_1 = \varrho$.[113] A higher value of ϱ is associated with a higher degree of persistence by all measures mentioned above. This property may be why the parameter ϱ is called the parameter of persistence in the context of DSGE models.[114]

It seems appealing to control the persistence of a process with a single parameter. However, stationary AR(1) processes are moderately persistent and, thus, unsuitable for modeling the other persistence scenarios mentioned in Definition 2.1.6. Additional properties of the AR(1) process are discussed in the next section.

109 See Paya et al. (2007, p. 1524).
110 Dias and Marques (2010, p. 264).
111 See Dias and Marques (2010, p. 264).
112 See Dias and Marques (2010, p. 266).
113 See Dias and Marques (2010, p. 265f.).
114 See, e.g., Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 565).

2.1.5.3 The AR(1) Process: A Closer Look

This section concludes with a specific example relevant to DSGE models. In DSGE models, exogenous variables are often assumed to follow an AR(1) process $X = (X_t)_{t\in\mathbb{Z}}$ that may be viewed as the percentage deviation from a variable's steady state value. The white noise process in the corresponding AR(1) representation specifies shocks pushing the variable away from its steady state value.[115] Thus, it seems reasonable to consider the properties of an AR(1) process in more detail.

Let $(X_t)_{t\in\mathbb{Z}}$ be a zero-mean AR(1) process given by

$$(1-\varrho B)X_t = \varepsilon_t \text{ or } X_t = \varrho X_{t-1} + \varepsilon_t, \tag{2.13}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a Gaussian white noise process.

The IRF of the process defined in (2.13) can be derived along two approaches. As the IRF shows the effect of a unit shock on the process, it can be derived recursively from (2.13) by setting $\varepsilon_0 = 1$ and $\varepsilon_t = 0$, for $t \neq 0$. Inserting this into (2.13) and replacing X_t with ψ_t, one finds the IRF of $(X_t)_{t\in\mathbb{Z}}$ to be the solution of the deterministic first-order difference equation

$$\psi_j = \varrho\psi_{j-1} \text{ with initial value } \psi_0 = 1. \tag{2.14}$$

As can be seen easily, the solution to (2.14) is given by $\psi_j = \varrho^j$. In addition, $\sum_{k=0}^{\infty} \psi_k^2 < \infty$ as long as $|\varrho| < 1$.

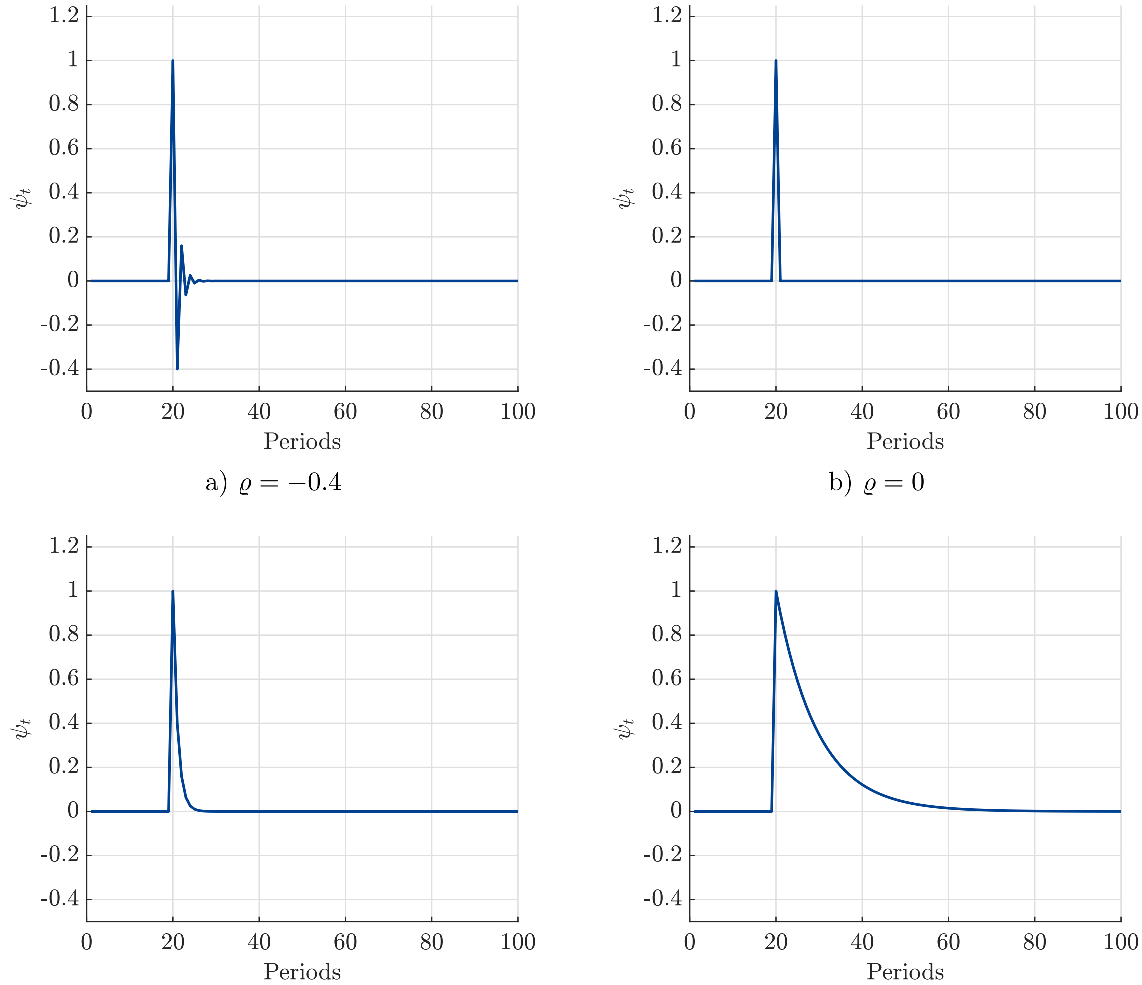
On the other hand, the autoregressive polynomial $\phi(z) = 1 - \varrho z$ of (2.13) has no roots inside the unit circle as long as $|\varrho| < 1$. In this case, $(X_t)_{t\in\mathbb{Z}}$ is stationary and the infinite moving average representation of $(X_t)_{t\in\mathbb{Z}}$ exists and is given by

$$X_t = \frac{1}{1-\varrho B}\varepsilon_t = \sum_{j=0}^{\infty} \varrho^j \varepsilon_{t-j}, \tag{2.15}$$

leading to the same result ($\psi_j = \varrho^j$). Thus, assuming an AR(1) process with $|\varrho| < 1$ implies a shock to have an impact on all subsequent observations of the process, but the impact decays exponentially fast. If $-1 < \varrho < 0$, the IRF decays cyclically to zero.

[115] See Section 4.1.3.1 for details on the specification of the productivity processes in the context of the DSGE models of Chapter 4. See Section 4.3 for a discussion about which processes are typically used in DSGE models.

Figure 2.1 illustrates the IRFs of four different AR(1) processes.[116,117]



a) $\varrho = -0.4$ b) $\varrho = 0$ c) $\varrho = 0.4$ d) $\varrho = 0.9$

Figure 2.1: Impulse-response functions of various AR(1) processes. The shock occurs at period 20. The corresponding autoregressive parameter ϱ is set to $-0.4, 0, 0.4$ and 0.9 in Panels a)-d), respectively. See further Footnote 116.

The higher the parameter ϱ, the slower the decay of the IRF (maybe cyclically if $\varrho < 0$), and the more substantial the impact of the shock on following periods, thus the higher the persistence of the process. Additionally, (2.12) specifies the CIR of an AR(1) process, given by

$$\text{CIR} = \frac{1}{|1-\varrho|}. \tag{2.16}$$

[116] Note that the functions plotted in Figure 2.1 do not solve (2.14). As stated before, the IRF does not depend on the time of the shock occurrence, i.e., (2.14) describes the evolution of the response in the aftermath of the shock. In order to highlight the impulse, the date of the shock was set to 20 in Figure 2.1. Correspondingly, the functions plotted in Figure 2.1 satisfy $\psi_{20+j} = \varrho^j$ with $\psi_{20} = 1$ and $\psi_j = 0$, for $j \leq 20$, i.e. the IRF is shifted by 20 periods ahead.

[117] All figures in this chapter were computed using Matlab code written by the author.

Thus the CIR is a monotone increasing function in ϱ for $|\varrho| < 1$.

In order to calculate the ACF of $(X_t)_{t\in\mathbb{Z}}$, a recursive method can also be applied. Multiplying (2.13) by X_{t-k} and taking expectations results in

$$\mathbb{E}(X_{t-k}X_t) = \varrho\mathbb{E}(X_{t-k}X_{t-1}) + \mathbb{E}(\varepsilon_t X_{t-k}). \tag{2.17}$$

Out of (2.15), X_{t-k} depends only on values ε_j for $j < t - k$. As $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise process, ε_t and X_{t-k} are uncorrelated if $k \geq 1$. Thus, (2.17) becomes

$$\gamma_X(k) = \varrho\gamma_X(k-1) \text{ for } k \geq 1. \tag{2.18}$$

For $k = 0$, (2.17) becomes

$$\begin{aligned}\gamma_X(0) &= \varrho\gamma_X(1) + \mathbb{E}(\varepsilon_t X_t)\\ &= \varrho^2\gamma_X(0) + \mathbb{E}(\varrho\varepsilon_t X_{t-1}) + \mathbb{E}(\varepsilon_t^2),\end{aligned} \tag{2.19}$$

where the last line uses (2.13) and (2.18). Again, ε_t and X_{t-1} are uncorrelated. Consequently, (2.19) pins down the variance of $(X_t)_{t\in\mathbb{Z}}$, i.e.,

$$\operatorname{var}(X_t) = \gamma_X(0) = \sigma_\varepsilon^2/(1-\varrho^2), \tag{2.20}$$

where σ_ε^2 is the variance of the white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$. In summary, (2.20) determines the initial value for the deterministic first-order difference equation (2.18). Dividing (2.18) by $\gamma_X(0)$ finally results in a first-order difference equation that describes the evolution of the ACF of $(X_t)_{t\in\mathbb{Z}}$

$$\rho_X(k) = \varrho\rho_X(k-1) \text{ with initial value } \rho_X(0) = 1$$

It is evident that the ACF and IRF of an AR(1) process solve the same first-order difference equation; hence both functions coincide. Consequently, the parameter ϱ determines not only the impact of a unit shock on future periods but also the degree of correlation within the process. A positive value of ϱ implies positive autocorrelations between all process observations. As outlined above, this translates into an overall smoother path of the process as it tends to lie on the same side of the mean for several periods. The higher the degree of autocorrelation, and thus, the higher the value of ϱ, the greater the smoothing effect. Conversely, if $-1 < \varrho < 0$, the autocorrelations decay cyclically to zero, so that the paths appear coarser, i.e. the paths are characterized by a zigzag behavior. Again, the smaller the value of ϱ, the greater the effect. Figure 2.2 illustrates this on the basis of four paths of AR(1) processes. In the presence of high-order positive autocorrelations (Panel d)), the overall smoother path is evident compared to Panels a)-c).[118]

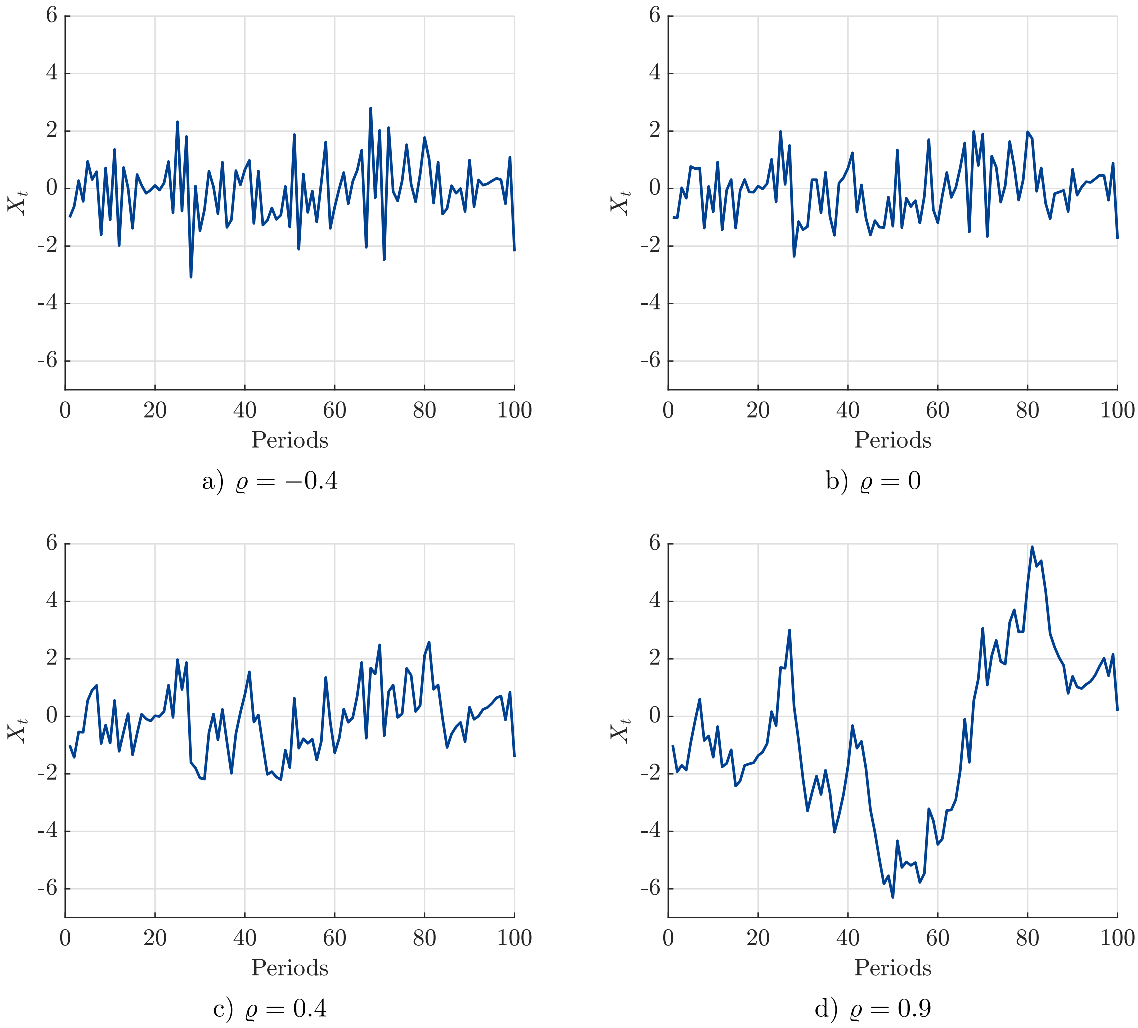


Figure 2.2: Paths of various AR(1) processes. The underlying white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be Gaussian with zero mean and variance 1 and is shown in Panel b) as $X_t = \varepsilon_t$ for $\varrho = 0$. Furthermore, the realization of the white noise process is identical for all panels. The autoregressive parameter is set to -0.4 in Panel a) and 0.4 and 0.9 in Panels c) and d), respectively.

The same conclusions can be drawn from Figure 2.3 which illustrates the corresponding normalized spectral densities of the AR(1) processes of Figure 2.2.[119] Equation (2.11)

118 Note that the range of the path in Panel d) is larger than in the other panels. This is due to the higher overall variance given in (2.20).
119 Recall Footnote 80 for the definition of the normalized spectral density .

indicates the spectral density of X to be[120]

$$f_X(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1 - e^{-i\lambda}|^2} = \frac{\sigma_\varepsilon^2}{2\pi(1 - 2\varrho\cos(\lambda) + \varrho^2)}. \tag{2.21}$$

Dividing (2.21) by (2.20) leads to the normalized spectral density

$$\bar{f}_X(\lambda) = \frac{1 - \varrho^2}{2\pi(1 - 2\varrho\cos(\lambda) + \varrho^2)}.$$

In the case of a white noise process, the spectral density is equal for all frequencies, i.e., each frequency contributes equally to the process's variance (see the light-blue line in Figure 2.3). In the case of negative autocorrelations, there is a higher mass of the density on the cycles with shorter periods (i.e., higher frequencies), highlighting that the process's variance is mainly explained by short cycles. Contrary, positive autocorrelations correspond to a high mass of the spectral density on low frequencies. As expected from (2.7), the higher the parameter ϱ, the higher the value of f_X at the zero frequency, again pointing to a higher degree of persistence.

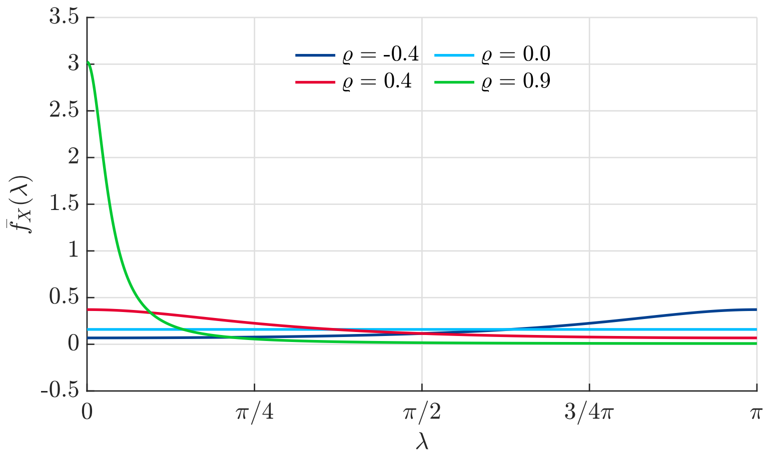


Figure 2.3: Normalized spectral densities of various AR(1) processes. The considered values of ϱ are $-0.4, 0, 0.4$ and 0.9. In order to equalize the area under the spectral density, normalization is performed by dividing the spectral density by the process's variance. Due to the symmetry of the spectral density, $\bar{f}_X(\lambda)$ is only plotted only over the interval $[0, \pi]$.

[120] The second equality follows from $e^{i\lambda} = \cos(\lambda) + i\sin(\lambda)$.

2.2 Mandelbrot's Joseph Effect

2.2.1 The Hurst Phenomenon: A Brief Review

This section briefly reviews the observations made by Harold Edwin Hurst in the 1950s that eventually inspired Mandelbrot to his seminal works on the Joseph and Noah effect, self-similar processes, and fractional Brownian motion. This thesis focuses on the Joseph effect, which will be defined shortly. In Section 2.5, fractional Brownian motion will be picked up again.

Hurst was working on the optimal size of a water reservoir for the Nile River. A reservoir is optimal if it is the smallest reservoir that satisfies the following three properties: the reservoir's outflow is uniform, the reservoir's level at the end of a period is as high as at the beginning of the period, and the reservoir never floods.[121] More precisely, assume that a time series of annual water discharges of a lake or river is given. How much capacity must a water reservoir have had to enable a maximal steady outflow over the recorded period?[122] Hurst relates the ideal size of a water reservoir to the adjusted range statistic, which is given by[123]

$$R(n) = \max_{0\le k\le n}\left\{\sum_{i=1}^{k}(X_i - \bar{X}_n)\right\} - \min_{0\le k\le n}\left\{\sum_{i=1}^{k}(X_i - \bar{X}_n)\right\},$$

where $\bar{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k$ denotes the sample mean of the process X up to period n. In the context of water reservoirs, X_n describes the outflow of a river or lake in period n. If the value of R is divided by the root of the sample variance $S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)$, one obtains the so-called rescaled adjusted range statistic defined by[124,125]

$$R/S(n) := \frac{R(n)}{\sqrt{S_n^2}} = \frac{R(n)}{S_n}.$$

By evaluating the rescaled adjusted range statistic on Nile river data, Hurst founds, by

[121] See Mandelbrot and Wallis (1968/2002, p. 239).

[122] See Hurst (1951, pp. 772f.).

[123] Hurst (1951) does not provide a formula for the adjusted range statistic R. Instead, he describes the calculation of R. The formula was taken from Beran et al. (2013, p. 410) and Graves et al. (2017, p. 4). This formula is a slightly simplified version of the R/S statistic used in Mandelbrot (1972, pp. 281f.).

[124] See Beran et al. (2013, p. 410).

[125] Note that other authors as Mandelbrot (1972, p. 282) and Graves et al. (2017, p. 4) and Beran (1994, p. 33) define S_n^2 as $\frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X}_n)$. The different scaling may not affect the derivations made in this section.

applying a linear regression of $\log(R/S(n))$ against $\log(n)$, that this relationship is nearly linear for large values of n, and further, that the slope of the resulting regression line (denoted by H as an abbreviation for "Hurst exponent") is greater than 1/2.[126] Strikingly, this behavior holds not only for river statistics or rainfalls but also for the size of tree rings, the number of sunspots, and wheat prices.[127] Hurst claimed that

> although in random events groups of high or low values do occur, their tendency to occur in natural events is greater. This is the main difference between natural and random events, an example of which is the discharge of the Nile River at Aswan. The long series of records of flood levels at Cairo shows the same phenomenon. There is no obvious periodicity, but there are long stretches when the floods are generally high, and others when they are generally low. These stretches occur without any regularity either in their time of occurrence or duration.[128]

These empirical observations are often called "Hurst's Law" or "Hurst phenomenon" in the sense of a stylized fact.[129] That these observations are indeed remarkable, and thus justify the term "phenomenon", can be seen from the fact that a wide range of stochastic processes show a scaling of the R/S statistic at the rate 1/2, i.e., they are incompatible with Hurst's observations.[130]

To be more specific, it can be shown for a stationary Gaussian process X with spectral density $f_X(\cdot)$ satisfying $0 < f_X(0) < \infty$ that[131,132]

$$\mathbb{E}\left(R/S(n)\right) \sim C\left(\frac{f_X(0)}{\sigma_X^2}\right)^{1/2} n^{1/2} \text{ as } n \to \infty, \tag{2.22}$$

where σ_X^2 is the variance of X. That is, such a process is incompatible with Hurst's law. In contrast, processes that behave as in (2.22) are sometimes said to follow the "$\sqrt{n}$-law"[133].

To give a concrete example of a class of processes that satisfy the $\sqrt{n}$ law and thus do not satisfy Hurst's law, consider the infinite moving average process $(X_t)_{t\in\mathbb{Z}}$ as in Section 2.1.2, where the innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ are normally distributed. Recall from (2.7) that the spectral

[126] See Hurst (1951, Figure 4 on p. 787). Hurst uses the variable K instead of H.
[127] See Hurst (1951, Figure 4 on p. 787).
[128] Hurst (1951, p. 783).
[129] See Mandelbrot and Wallis (1968/2002, p. 247) and Graves et al. (2017, p. 5).
[130] See Hurst (1956, p. 14).
[131] See Siddiqui (1976, pp. 1274f.), Mandelbrot (1975/2002, p. 524) or Beran et al. (2013, p. 410).
[132] The notation $f(x) \sim g(x), x \to a$ means that $\lim_{x\to a} f(x)/g(x) = 1$, i.e., both functions show the same asymptotic behavior as $x \to a$.
[133] Mandelbrot and Wallis (1968/2002, p. 246).

density of a moderately persistent process ($0 < \text{CIR}^2 < \infty$) yields $0 < f_X(0) < \infty$.

Consequently, all moderately persistent processes fail to satisfy Hurst's law. Especially (as derived in Section 2.1.5) all ARMA processes do so. By noting that $\sigma_X^2 = \gamma_X(0)$, it follows from (2.22) and (2.6) that

$$\mathbb{E}\left(R/S(n)\right) \sim C\left(1 + 2\sum_{k=1}^{\infty}\rho_X(k)\right)n^{1/2}, n \to \infty. \tag{2.23}$$

Thus, all Gaussian processes with a summable ACF also fail to satisfy Hurst's law. This observation suggests that substantial autocorrelation is required to reproduce Hurst's law.

In general, the results found by Hurst can be formalized as[134]

$$\mathbb{E}\left(R/S(n)\right) \sim Cn^H \text{ as } n \to \infty, \tag{2.24}$$

with $H > 1/2$. The letter H is a shorthand notation for "Hurst exponent" and is sometimes referred to as the intensity of the R/S statistic.[135]

In 1968, Mandelbrot and van Ness (1968) introduced a continuous-time stochastic process called fBm $(B_t^H)_{t\geq 0}$ that generalizes the well-known Brownian motion.[136] In addition, if one considers the discrete-time increment process of an fBm called fractional Gaussian noise (fGn), i.e., $X_t^H = B_{t+1}^H - B_t^H$ for $t = 1, 2, 3, \ldots$, Mandelbrot and Wallis (1968/2002) showed that this increment process is stationary and Gaussian, and it is able to replicate Hurst's law.[137] To be more precise, the parameter H of the fBm or fGn turned out to be identical to the intensity of the R/S statistic in (2.24).[138] This was the first time a stationary process satisfying Hurst's law was presented.[139]

Due to (2.23), the question of satisfying Hurst's law is immediately linked to another property, namely the summability of the ACF. That is, (2.23) indicates that processes with a non-summable ACF, i.e., $\sum_{k=0}^{\infty}\rho_X(k) = \infty$, may be able to replicate Hurst's law. Indeed, it can be shown that the ACF of an fGn process is not summable in the case of $H > 1/2$.[140]

[134] See Beran (1994, p. 34).
[135] Mandelbrot (2002, p. 157) preferred the letter J (as a shorthand notation for "Joseph-Exponent") instead of H to disentangle the empirical findings of Hurst from the more general asymptotic behavior of the R/S statistic.
[136] A rigorous treatment of fractional Brownian motion is given in Section 2.5.
[137] Mandelbrot and Wallis (1968/2002, p. 250).
[138] Mandelbrot and Wallis (1968/2002, p. 250).
[139] See Graves et al. (2017, p. 9).
[140] In Lemma 2.5.3 below, it is shown that $\gamma_{X^H}(k) \sim H(2H-1)k^{2H-2}$. Consequently, the autocovariance

In the end, an ACF exists only for stationary processes, but also non-stationary processes can show Hurst's law. For this reason, Mandelbrot introduced a more general concept of "global statistical dependence" and "local statistical dependence" that should apply to all stochastic processes. Ultimately, the presence of global dependence in a stochastic process is why the $\sqrt{n}$-law fails to hold.[141] On the other hand, if a process satisfies the $\sqrt{n}$-law it is called to be locally dependent.

For stationary processes, global dependence may manifest in a non-summable ACF or a spectral density f with $f(\lambda) \to \infty$, as $\lambda \to 0$. For non-stationary processes, global dependence can be manifested by the behavior of the R/S statistic (2.24). Mandelbrot used (2.24) to define global statistical dependence (in the R/S sense): If $H = 1/2$, i.e., the $\sqrt{n}$ law holds, the process shows local dependence. If $H > 1/2$, Hurst's law can be replicated, and the process exhibits global persistent dependence (in the R/S sense). Finally, if $H < 1/2$, the process exhibits global anti-persistent dependence (in the R/S sense).[142] Mandelbrot preferred the R/S statistic to the ACF or spectral density because of its robustness to non-Gaussian distributions and infinite variance processes.[143] Since the latter are not stationary, the ACF cannot be applied to such processes. On the other hand, there may be processes for which an asymptotic relationship, as stated in (2.24), does not exist at all. In such a case, the R/S statistic seems inappropriate to determine the presence of global or local dependence.[144]

However, processes may be labeled as locally dependent according to the R/S statistic, but at the same time, they may be globally dependent in a wider sense.[145] Thus in 2002, Mandelbrot claimed that "the exponent of R/S does *not* suffice to discriminate between local and global dependence."[146]

Dating back to 1968, Mandelbrot and Wallis introduced the colorful name "Joseph Effect" for global dependence in the R/S sense.[147] All processes with $H \neq 1/2$ show the Joseph effect, and the intensity of the R/S statistic may measure how strong the Joseph effect is.[148] The intention for the naming was derived from the biblical story of Joseph, son of Jacob, and especially from Genesis 41, 29-30: "Seven years of great abundance are

function and the autocorrelations are not summable if $H > 1/2$.

141 See Mandelbrot (2002, pp. 156f.).

142 See Mandelbrot (2002, pp. 160 and 167).

143 See Mandelbrot (1972, pp. 286ff.) and Mandelbrot (1975/2002, p. 521).

144 See Mandelbrot (1972, p. 283).

145 See Mandelbrot (2002, pp. 170f.).

146 Mandelbrot (2002, p. 483).

147 See Mandelbrot and Wallis (1968/2002, p. 246).

148 See Mandelbrot (2002, p. 159f) Mandelbrot and Wallis (1969/2002, p. 488).

coming throughout the land of Egypt, but seven years of famine will follow them"[149]. This metaphor nicely describes Hurst's empirical findings cited above. The reasons for the naming are also illustrated in the next section since "the presence of global dependence can often be suspected visually"[150].

2.2.2 A Visual Insight

This section illustrates the differences between a time series showing a type of global dependence and one showing local dependence. Figure 2.4 shows two realizations of stochastic processes X and Y plotted over 3000 periods. Both processes are stationary and have zero means. The process Y depicted in Panel b) shows the tendency to grow for a while, followed by a sudden stop and a reversion of this tendency, which is again followed by expanding periods. The path of X (see Panel a) of Figure 2.4) crosses the zero line more often than the path of Y and seems to show an intensive zigzag behavior. Overall, the path of Y is characterized by longer periods of up- and downswings than the path of X. This observation inspired Mandelbrot to say that Y shows the Joseph effect.[151] This non-period cyclical behavior is typical for processes with global dependence, and indeed, Y in Figure 2.4 is the path of a globally dependent process (in the sense that its ACF is not summable), whereas X is the path of an AR(1) process which is a locally dependent process (having a summable ACF).[152]

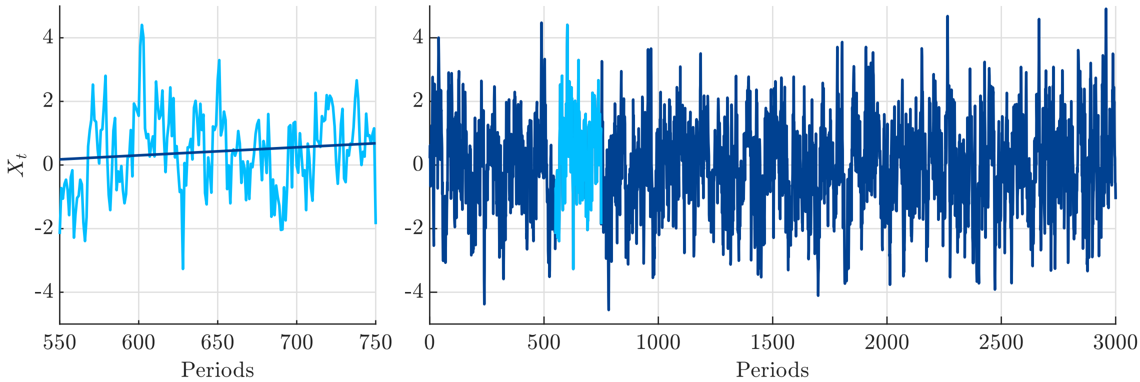
However, the seven fat years followed by seven lean years may indicate some periodicity in the cyclicality described, although Mandelbrot construed the behavior as follows

> clear-cut but not periodic 'cycles' of all conceivable 'periods,' short, medium, and long, where the latter means 'comparable to the length of the total available sample,' and where the distinction between 'long cycles' and 'trends' is very fuzzy.[153]

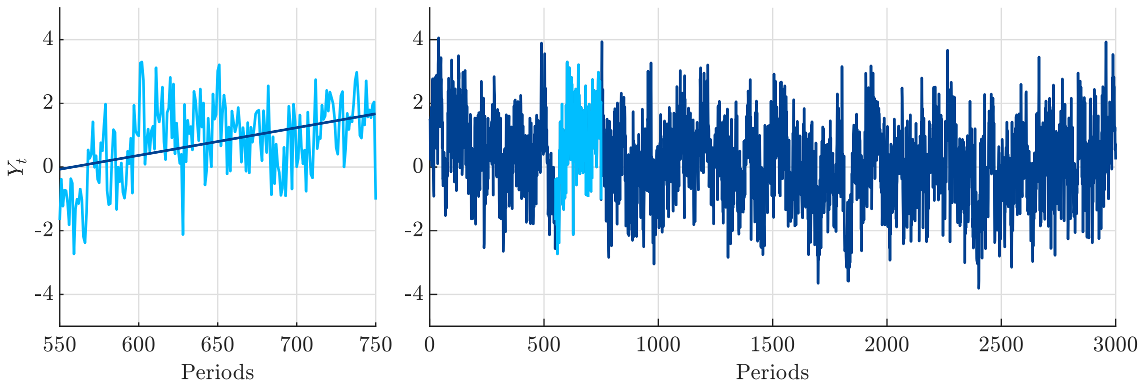
This behavior is visible in Panel b) of Figure 2.4. There, it seems that the cycles are longer in the first half of the sample and shorter in the second half. In contrast, the cyclical behavior of the process X in Panel a) seems to be more regular throughout the sample.

The difficulty in distinguishing longer cycles and trends is illustrated on the left-hand sides

[149] This passage was taken from the new international version of the Bible, see Biblica (2021), URL in list of references.
[150] Mandelbrot (2002, p. 158).
[151] The term Joseph effect was introduced in Mandelbrot and Wallis (1968/2002).
[152] In Figure 2.4, the length of "one period" is not specified, so the reference to the biblical story of Joseph should be treated just in a metaphorical sense.
[153] Mandelbrot (1972, p. 260).



a) Locally dependent process



b) Globally dependent process

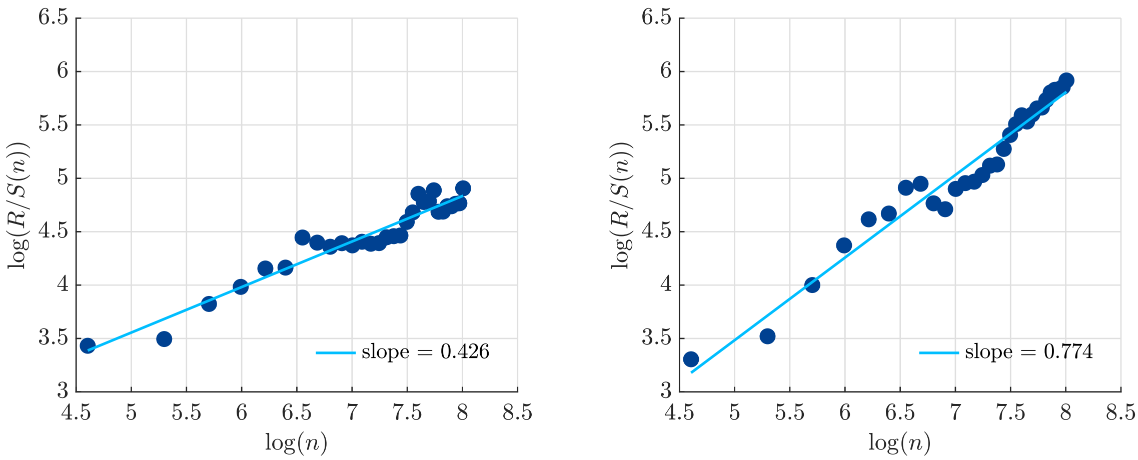
Figure 2.4: Paths of a locally and globally dependent process. Panels a) and b) show the path of a locally and globally dependent process, respectively. The left-hand side shows an extraction from the full sample shown on the right-hand side of each panel. The globally dependent process is represented by a fractionally integrated white noise process (i.e., an ARFIMA$(0, d, 0)$ process) with parameter $d = 0.4$; see Section 2.4 for details. The locally dependent process corresponds to an AR(1) process with parameter $\varrho = 0.719$. The parameter ϱ was chosen to equate the theoretical total process variances (see (A.12) and (2.20)). The underlying white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be Gaussian with zero mean and variance 1, and its realization is identical in both panels.

of Figure 2.4. In these panels, an extraction of the corresponding long time series displayed on the right-hand panels from periods 550 to period 750 is plotted. If this extraction is considered solely, one may find that the globally dependent time series shows an upward trend and looks non-stationary. When considering the extracted periods as a part of the whole sample, there is no trending behavior overall, i.e., the trend disappears and is now part of a longer fluctuation.[154] The process is sometimes called to show "local trends"[155]. The tendency for showing local trends is less pronounced in the locally dependent time series of Panel a). Further, the length of these local trends seems random, and the resulting cyclicity is non-periodic.

[154] See Mandelbrot (2002, p. 158).
[155] Beran (1994, p. 141).

Figure 2.5 shows linear regressions of the corresponding R/S statistics. As one would expect from (2.24), the slope of the regression line is 0.426 and thus nearly 1/2 in the case of the locally dependent process. The corresponding slope of the globally dependent process is 0.774 and thus clearly larger than 1/2. The theoretical value is 0.9.[156]

Therefore, the paths shown in Figure 2.4 belong to classes of stochastic processes for which the global dependence in the R/S sense coincides with the global dependence induced by a non-summable ACF.



a) Locally dependent process

b) Globally dependent process

Figure 2.5: Linear regression of $\log(R/S)$. The regression was performed on the simulated paths shown in Figure 2.4. The R/S statistic is evaluated on a grid with a range of $1, 2, \ldots, 3000$ and a step size of 100; $\log(R/S)$ is regressed on $\log(n)$. Panels a) and b) refer to the locally and globally dependent samples of Figure 2.4, respectively.

The behavior described so far, i.e., the non-periodic cyclicality and the local trending, refer in fact to the case of $H > 1/2$ or $\sum_{k=0}^{\infty} \rho_X(k) = \infty$.[157] If $H < 1/2$ or $\sum_{k=0}^{\infty} \rho_X(k) = 0$, positive and negative correlations will be compensated, meaning that positive values will be followed by negative values, resulting in a more jagged path overall.[158] Although the process is globally dependent if $H < 1/2$ or $\sum_{k=0}^{\infty} \rho_X(k) = 0$, the name Joseph effect seems to be derived from the path properties of a time series with $H > 1/2$. This suggestion may be reasonable since $H > 1/2$ is consistent with Hurst's original findings and thus seems more empirically relevant.

Ultimately, neither the term Joseph effect nor global dependence, nor the analysis of the R/S statistic was entirely adopted by the economics and econometrics profession. The

156 The precision of the estimated values may be increased by Monte Carlo simulations and or by involving more than 3000 periods in each sample. This study is not carried out here, as the aim of Figure 2.5 is to illustrate Hurst's phenomenon and not to give an exact estimator of H.

157 See Mandelbrot (1972, pp. 273ff.).

158 See Mandelbrot (1972, p. 275).

term “Joseph effect” has been translated as “long memory”, “extensive dependence” or “strong dependence”. Nowadays, the R/S statistic serves only as a heuristic estimator of the Hurst exponent H.

The reason for this can be found in the R/S statistic, which comes along with several problems seeming unsatisfactorily from an econometric point of view. So, for example, the R/S statistic cannot distinguish between long memory and a non-stationary stochastic process with a slowly decaying trend.[159] Further, as already mentioned by Mandelbrot, the Hurst phenomenon can be mimicked by locally dependent processes at least over some periods (he called this transient region), i.e., the R/S statistic would indicate a value of $H > 1/2$ in this transient region but values of $H = 1/2$ after this region.[160] When considering real-world data sets with limited length, it may be impossible to decide between a true globally dependent process or a locally dependent process whose transient region corresponds to the whole sample size. In order to exclude such a transient region (given that it does not correspond to the total sample size) from the estimation of H, one may have to drop some values at the beginning of the regression. However, it is difficult to accomplish this task objectively.[161]

Furthermore, an important advantage emphasized by Mandelbrot, namely the robustness of the R/S statistic to infinite variance processes, may be less advantageous from a time series analysis perspective, since it is common practice to explain the data with stationary models.

Therefore, the following section looks at today’s more commonly used definitions of long memory.

2.3 On the Definition of Long Memory

As noted at the end of the last section, the R/S statistic hardly defines long memory anymore. However, it follows from (2.23) that stationary processes with a non-summable ACF may be able to replicate Hurst’s law. From this perspective, it is not surprising that long memory is defined in terms of the ACF or related conditions.

Before discussing different ways of defining long memory, the definition assumed in this thesis is provided[162]

[159] See Beran (1994, pp. 85ff.).
[160] See Mandelbrot and Wallis (1968/2002, p. 248).
[161] See Beran et al. (2013, p. 412).
[162] Definition 2.3.1 is taken from Hassler (2019, Definition 2.4 on p. 22). Definition 2.3.1 is in accordance

Definition 2.3.1

Let $X = (X_t)_{t\in\mathbb{Z}}$ be a stationary stochastic process. Then X is called a long memory process if its autocovariance function is not absolutely summable, i.e.,

$$\sum_{k=0}^{\infty} |\gamma_X(k)| = \infty. \tag{2.25}$$

Otherwise, X is called a short memory process, i.e., if its autocovariance function satisfies $\sum_{k=0}^{\infty} |\gamma_X(k)| < \infty.$ ⋊

Definition 2.3.1 states that decay of the autocovariance is so slow that the sum of its absolute values fails to converge. Obviously, due to Definition 2.1.3, γ_X in Definition 2.3.1 can be replaced with ρ_X, which denotes the ACF of the process X.

Note that Definition 2.3.1 refers to the absolute value of the autocovariance function. From (2.23) one could refer to the values $\gamma_X(\cdot)$ or $\rho_X(\cdot)$ itself.[163]

The reason for referring to the absolute values is as follows. Note that, by (2.6) and (2.7), it yields

$$f_X(0) = \frac{1}{2\pi}\left(\gamma_X(0) + 2\sum_{k=1}^{\infty}\gamma_X(k)\right) = \frac{\sigma_\varepsilon^2}{2\pi}\mathrm{CIR}^2. \tag{2.26}$$

If long memory were defined based on the values of γ_X instead of its absolute values, (2.26) and Definition 2.1.6 would imply that long memory and persistence are the same concepts. That is one would equate strong persistence with long memory and moderate persistence with short memory. However, the dichotomic distinction between short and long memory cannot adequately capture the third case of anti-persistence. Ultimately, one is left with two opportunities: One encloses short memory with anti-persistence or distinguishes between anti-persistence, short and long memory. The former case is not preferable as it is less differentiated.[164] The latter seems less intuitive from a semantic perspective. For these reasons, the concepts of persistence (basically motivated by the cumulative IRF) and memory are considered distinctively.[165] Before the interrelationships between long memory and persistence are considered in more detail, some other definitions

with Guégan (2005, Definition 3.1 on p. 117) and Palma (2007, Equation (3.1) on p. 40).

163 This is done, for example, by Graves et al. (2017, p. 1).

164 From a practical point of view, it seems reasonable to distinguish between anti-persistence and moderate persistence as the former cannot be replicated by traditional time series models such as causal and invertible ARMA processes; see (2.12). This case should be (as for long memory or strongly persistent processes) paid attention to separately.

165 Note that Definition 2.3.1 defines a dichotomy, since the case of $\sum_{k=0}^{\infty}|\gamma_X(k)| = 0$ would imply that $\mathrm{var}(X) = \gamma(0) = 0$ and thus $X_t \equiv \mathit{const}$. A constant process, however, seems less attractive from a probabilistic point of view.

of long memory in the literature are discussed briefly.

It is worth noting that a specific form or rate of decay of the ACF's absolute values is not mentioned in Definition 2.3.1. To circumvent this lack in the specification of γ_X (or ρ_X) in (2.25), some authors propose a definition of long memory by imposing a specific structure on the ACF. On the other hand, (2.26) motivates the definition of long memory along the properties of the spectral density at the zero frequency. For these reasons, two alternative definitions of long memory and their relationships to Definition 2.3.1 are considered in the following.

Beran (1994) specifies a law of how the ACF decays. In his sense, a stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ is a long memory process if its ACF satisfies

$$\rho_X(k) \sim c_\rho k^{2d-1} \text{ as } k \to \infty, \tag{2.27}$$

where $c_\rho > 0$ and $0 < d < 1/2$ are certain constants.[166,167]

It is important to note that (2.27) is an asymptotic property of the ACF, i.e., it states that ρ_X decays asymptotically like k^{1-2d}, but no information about the value of ρ_X can be deduced from this expression. Therefore, the values of $\rho_X(k)$ for small values of k are left unexplained by (2.27), and in addition, the concrete values of $\rho_X(k)$ may be quite small.[168] The important parameter is d as it specifies the rate of the ACF's decay.

One can show that if a process's ACF satisfies (2.27), then its absolute values are not summable; hence, it is a long memory process in the sense of Definition 2.3.1.[169] Thus, (2.27) may be seen as a sufficient condition for a process to show long memory in the sense of Definition 2.3.1.[170]

Due to (2.26), some authors propose a definition of long memory based on a process's spectral density.[171] A stationary stochastic process $(X_t)_{t\in\mathbb{Z}}$ may be called a long memory

166 See Beran (1994, Definition 2.1 on p. 42) and replace his parameter α with $1-2d$. See also Brockwell and Davis (1987, p. 456) and Guégan (2005, Definition 3.6 on p. 119).

167 The notation $f(x) \sim g(x), x \to a$ means that $\lim_{x\to a} f(x)/g(x) = 1$, i.e., both functions show the same asymptotic behavior as $x \to a$.

168 See Beran (1994, p. 43).

169 See Palma (2007, p. 40, especially Theorem 3.1. (b)).

170 There are various expressions such as "long range dependence", or "strong dependence" in the literature to describe processes with a slowly decaying ACF, see, e.g., Beran (1994, Definition 2.1 on p. 42). Apart from the mentioned nuances in the definition of long memory (that have to be treated individually), these terms may thus be used interchangeably. The terms of "a process shows long memory" or "a process is a long memory process" are used synonymously in this thesis.

171 As will turn out in Section 3.1.2, (2.28) is frequently used for estimating the long memory parameter.

process if its spectral density f_X satisfies

$$f_X(\lambda) \sim c_f|\lambda|^{-2d} \text{ as } \lambda \to 0, \tag{2.28}$$

where $c_f > 0$ is a certain constant and $0 < d < 1/2$.[172] Equation (2.28) states that the spectral density has a pole at frequency $\lambda = 0$, i.e., $f_X(\lambda) \to \infty$ as $\lambda \to 0$.

There are further generalizations of (2.27) and (2.28), replacing the constants c_ρ and c_f by functions of k and λ with specific asymptotic properties.[173,174]

The two alternative definitions of long memory stated in (2.27) and (2.28) are, however, equivalent, i.e., a process satisfying (2.27) meets further (2.28) and vice versa.[175,176] Thus, (2.28) may also be seen as a sufficient condition for a process to be a long memory process in the sense of Definition 2.3.1.

To summarize, Definition 2.3.1 is a well-established and quite general definition of long memory. It provides, however, no information about how the autocovariance function or ACF has to decay besides that the decay has to be sufficiently slow.

Furthermore, it is possible to derive closed-form expressions of the ACF or spectral density (at least for its asymptotic behavior) for many stochastic processes, especially those revealing a high empirical relevance.[177] In such a case, (2.27) and (2.28) may be verifiable immediately. Further, the parameter d in these equations offers a more sophisticated way to characterize long memory. To see this, imagine two processes X and Y satisfying (2.27) with corresponding parameters $0 < d_X < d_Y < 1/2$. Concerning Definition 2.3.1, both processes are long memory processes. Regarding (2.27), the decay of the ACF of Y is even slower than that of X; thus, one can say that Y shows long memory in a more pronounced fashion than X. That is, the definition according to (2.27) or (2.28) allows, in contrast to Definition 2.3.1, to distinguish the degree of long memory.

172 See Beran (1994, Definition 2.2) and replace his parameter β with $2d$.

173 To be more precise, such functions are called “slowly varying at infinity”, see Palma (2007, Section 3.1 on pp. 40ff.) or Giraitis et al. (2012, Section 2.3 on pp. 18ff. and Sections 3.1 on pp. 33ff.) for more details.

174 There are further generalizations that allow the pole of the spectral density stated in (2.28) to be at another frequency then zero, see, for example Guégan (2005, Definition 3.5 on p. 119). In addition, Guégan (2005, Section 3 on pp. 116ff.) gives an overview of a variety of long memory characteristics and the relationships among them.

175 See Beran (1994, Theorem 2.1 on p. 43).

176 Note that this equivalence must not hold in the case of non-constant values c_ρ and c_f mentioned above. Then, equivalence may be proofed under additional assumptions about c_ρ and c_f, see Footnote 173 and Palma (2007, Theorem 3.1 on p. 40).

177 See, e.g., the class of fractionally integrated ARMA processes introduced in Section 2.4.

On the other hand, the hyperbolic decay of the ACF stated in (2.27) may be too special. It is conceivable that the absolute values of the ACF are not summable, but the decay is not given to be of this specific form.

The following lemma summarizes the relationships between the Definition 2.3.1, (2.27), (2.28), and the concept of persistence.

Lemma 2.3.2

Let $X = (X_t)_{t\in\mathbb{Z}}$ be an infinite moving average process as in (2.2)*, then it yields*

i) If X is strongly persistent, it is a long memory process.

ii) If X is a short memory process, it is either anti or moderately persistent. Thus, a short memory process cannot be strongly persistent.

iii) If X satisfies (2.27) *or* (2.28)*, it is strongly persistent.*

iv) There are long memory processes not satisfying (2.27)*.*

v) There are anti, moderately, and strongly persistent long memory processes.

Proof

See Appendix A.1.1 □

If long memory were defined by (2.28), the process would also show strong persistence, i.e., (2.28) does not allow to distinguish between strong persistence and long memory. Instead, by using Definition 2.3.1, the concepts of persistence and long memory can be considered separately, see v) of Lemma 2.3.2. At the same time, iv) of the preceding lemma states that Definition 2.3.1 is indeed more general than (2.27) since some long memory processes are ruled out by (2.27) indicating that the law of decay stated in (2.27) may be too restrictive.

The main class of long memory processes considered in the following section (namely fractionally integrated ARMA processes) are highly relevant from an empirical point of view. As will be shown in the following section, they satisfy the conditions (2.27) and (2.28).[178] Thus, Lemma 2.3.2 indicates that these processes are also strongly persistent. This property may be why the distinction between strong persistence and long memory is not rigorous in the literature. However, it should be kept in mind that the two concepts generally refer to different properties of a stochastic process.

[178] Particularly, the parameter d, as a measure to which extend the process shows long memory, builds the focus of many empirical investigations in the literature, see Chapter 3.

2.4 Discrete-Time Long Memory Models: ARFIMA Processes

As seen in Section 2.1.5.1, all causal and invertible ARMA processes are moderately persistent, and it follows directly from (2.10) that

$$\sum_{k=0}^{\infty} |\gamma_X(k)| \le C \sum_{k=0}^{\infty} \beta^k = \frac{C}{1-\beta} < \infty.$$

Consequently, ARMA processes are short memory processes and thus cannot replicate Mandelbrot's Joseph effect mentioned in Section 2.2. Granger and Joyeux (1980/2001) and Hosking (1981) propose a generalization of the ARMA setting to encompass long memory dynamics in the framework of ARMA processes. The resulting fractionally integrated ARMA (ARFIMA) processes are intensively applied in the econometric literature.[179] After the formal definition is provided, the persistence and memory properties of these processes are analyzed in more detail.[180]

Definition 2.4.1 ———

Let p and q be non-negative integers, $d \in \mathbb{R}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ a white noise process with variance σ_ε^2. Then $(X_t)_{t\in\mathbb{Z}}$ is called ARFIMA(p, d, q) process if for all $t \in \mathbb{Z}$,

$$\phi(B)X_t = \theta(B)(1-B)^{-d}\varepsilon_t, \tag{2.29}$$

where the polynomials $\varphi(\cdot)$ and $\theta(\cdot)$ are as in Definition 2.1.8 and the fractional integrating operator $(1-B)^{-d}$ is defined by

$$(1-B)^{-d} = \sum_{k=0}^{\infty} \alpha_k B^k \text{ with } \alpha_k = \binom{k+d-1}{k} = \begin{cases} \dfrac{(k+d-1)!}{(d-1)!k!}, & \text{if } d \in \mathbb{Z}\setminus\{0\} \\ \dfrac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}, & \text{if } d \in \mathbb{R}\setminus\mathbb{Z}, \end{cases} \tag{2.30}$$

and $\Gamma(x)$ denotes the gamma function.[181,182] *For convenience, let $(1-B)^0 \equiv 1$.* ⋊

[179] See Chapter 3.

[180] Definition 2.4.1 is adapted from Palma (2007, p. 43), Hosking (1981, p. 170), Beran (1994, p. 60), Granger and Joyeux (1980/2001, p. 324f.) and Hassler (2019, p. 95f.).

[181] For details on the gamma function, see Appendix A.2.

[182] Palma (2007, p. 43) uses the term "differencing operator" for $(1-B)^{-d}$, but it seems more appropriate to reserve this term for the inverse operator $(1-B)^d$ as carried out, e.g., in Beran (1994, p. 61) or Hosking (1981, p. 166). The expression "integrating operator" for $(1-B)^{-d}$ is in accordance with Granger and Joyeux (1980/2001, p. 324).

In the case of $d = 0$ in Definition 2.4.1, the term $(1 - B)^{-d}$ disappears, and the resulting equation is as in Definition 2.1.8. Thus, ARFIMA processes generalize the class of ARMA processes. If the corresponding ARMA process is stationary and causal, the ARFIMA process is likewise stationary and causal as long as $d < 1/2$.[183] If the corresponding ARMA process is invertible, so is the ARFIMA process as long as $d > -1$.[184] For this reason, the parameter range of d is assumed to be within these two values, i.e., $-1 < d < 1/2$.[185] The corresponding infinite moving average representation (2.2) can be calculated from a time series expansion of the function $\psi(z) = (1 - z)^{-d}\theta(z)/\phi(z)$.[186]

As can be seen from Definition 2.4.1, ARFIMA processes can also be defined for values $d > 1/2$. However, these processes are not stationary anymore. A well-known process that falls within this class of non-stationary processes is an integrated ARMA process (often denoted by ARFIMA$(p, 1, q)$ or ARIMA$(p, 1, q)$), where $d = 1$. Let Y be such a process; then by Definition 2.4.1, one knows that the process $(1 - B)Y_t = Y_t - Y_{t-1}$ is a stationary ARMA process, i.e., by taking first-order differences the process becomes stationary.[187] Since ARMA processes are moderately persistent, their spectral density is positive and bounded at zero, see (2.12). Therefore, ARMA processes are sometimes called $\mathcal{I}(0)$ (i.e., integrated of order 0) processes.[188] In the case of Y, a full difference is needed to obtain a stationary ARMA process. Thus, Y is often called $\mathcal{I}(1)$ process, unit-root process, or difference stationary.[189] ARFIMA processes with $d \notin \{0, 1\}$ generalize this notion and are called $\mathcal{I}(d)$ processes, i.e., one has to apply the fractional difference operator $(1 - B)^d$ to the process in order to obtain a stationary time series with bounded spectral density at the zero frequency.[190]

Unit-root ($\mathcal{I}(1)$) processes are often called to be permanent as their IRF does not decay to zero. To see this, consider the integrated white noise process $X_t = \sum_{k=0}^{t} \varepsilon_k$. Clearly, $\partial X_t/\partial\varepsilon_t \equiv 1$, i.e., the shock ε_t does not dissipate and, thus, affects the evolution of X permanently.[191]

---

[183] See Hassler (2019, Definition 6.1 on p. 104 and p. 104f.) and Palma (2007, Theorem 3.4 on p. 44).
[184] Palma (2007, Theorem 3.4 on p. 44).
[185] Hosking (1981, Theorem 2 on pp. 170f.) and Beran (1994, Definition 2.7 on p. 60) restrict the parameter d to the interval $(-1/2, 1/2)$ in order to ensure stationarity, causality and invertibility. As stated by Hassler (2019, p. 105f. and Proposition 6.2 on p. 106), the condition $d > -1/2$ is, however, not necessary to ensure invertibility. See further Palma (2007, Theorem 3.4 on p. 44 and Remark 3.1 on p. 46).
[186] See Palma (2007, Theorem 3.4 on p. 44).
[187] See Hamilton (1994, pp. 436f. and p. 444).
[188] See Candelon and Gil-Alaña (2004, p. 344).
[189] See Hamilton (1994, p. 436).
[190] See Candelon and Gil-Alaña (2004, p. 344f.).
[191] See, e.g., Hamilton (1994, p. 443).

Often, it is enough to consider stationary ARFIMA processes with $-1 < d < 1/2$, since by applying the difference operator $(1 - B)$, one can shift the order of integration into the stationary region $(-1, 0.5)$.[192] More specifically, let X_t be an ARFIMA$(p, 1.3, q)$, then, the process $(1 - B)X_t$ is a stationary ARFIMA$(p, 0.3, q)$ process.[193] The focus of the remaining chapter lies, therefore, on stationary ARFIMA processes.

2.4.1 Spectral Density, ACF and Persistence

The spectral density of an ARFIMA(p, d, q) process X is given by

$$f_X(\lambda) = f_{ARMA}(\lambda)\,[2\sin(\lambda/2)]^{-2d}, \tag{2.31}$$

where f_{ARMA} is the spectral density of an ARMA process with the same polynomials $\phi(\cdot)$ and $\theta(\cdot)$ as for X given in (2.29).[194] It follows from (2.31) and (2.11) that[195]

$$f_X(\lambda) \sim f_{ARMA}(0)|\lambda|^{-2d} = \frac{\sigma_\varepsilon^2}{2\pi}\frac{|\theta(1)|^2}{|\phi(1)|^2}|\lambda|^{-2d} \text{ as } |\lambda| \to 0. \tag{2.32}$$

Thus, the spectral density of an ARFIMA process satisfies the sufficient condition (2.28) for the process to show long memory if $0 < d < 1/2$. Moreover, by Lemma 2.3.2, the process is also strongly persistent, i.e., the CIR is infinite.

If $d = 0$, the basic ARMA process (a moderately persistent short memory process) is obtained. If $-1 < d < 0$, (2.32) implies that $f_X(0) = 0$. Consequently, according to (2.26) and Definition 2.1.6, the process is anti-persistent as long as $-1 < d < 0$.

To determine the memory properties for an ARFIMA process with $-1 < d < 0$, a closer look at the autocovariance function has to be taken. As with ARMA processes, closed-form expressions of the ACF or autocovariance function of general ARFIMA processes are complicated.[196] For this purpose, an expression analogous to (2.10) is given, which specifies the asymptotic behavior of the autocovariance function of an ARFIMA process as follows.

[192] See Beran (1994, p. 61).
[193] See Beran (1994, p. 61).
[194] See Palma (2007, p. 47).
[195] Note that the sinus function disappears due to a first-order Taylor approximation at zero. More precisely, $\sin(x) \approx x$ for small values of x.
[196] Palma (2007, Section 3.2.4 on pp. 47f.) for a derivation of the ACF of a general ARFIMA(p, d, q) process.

The autocovariance function of ARFIMA process X satisfies[197]

$$|\gamma_X(k)| \sim Ck^{2d-1} \text{ as } k \to \infty, \tag{2.33}$$

where $C \in \mathbb{R}$ and $d \in (-1,0) \cup (0,1/2)$.[198] It follows immediately from (2.33) that $\sum\limits_{k=0}^{\infty} |\gamma_X(k)| < \infty$ for $-1 < d < 0$, i.e., ARFIMA processes are short memory processes if $-1 < d < 0$.[199]

Equation (2.33) highlights that the long-run correlation properties are mainly determined by the parameter d. The parameters p and q and the corresponding autoregressive and moving average coefficients may be used to model the short-term correlation structure of a time series.[200] Table 2.1 summarizes the memory and persistence properties of ARFIMA processes. As this thesis focuses on long memory processes, the parameter range $0 < d < 1/2$ is highly relevant for the considerations of the following sections and chapters.

	Values of d	
$-1 < d < 0$	$d = 0$	$0 < d < 1/2$
anti-persistence	moderate persistence	strong persistence
short memory		long memory

Table 2.1: Persistence and memory properties of ARFIMA processes.

Some authors say that ARFIMA processes with $1/2 < d < 1$ also have long memory, since their IRF still converges to zero.[201] However, since these processes are no longer stationary, Definition 2.3.1 is no longer applicable to such processes. Consequently, the considerations of this thesis, especially those of Chapter 4, concentrate on the stationary long memory parameter range $0 < d < 1/2$.

As mentioned in Section 2.1.5, ARMA processes can approximate every stationary process arbitrarily well. This property contradicts not the need for ARFIMA processes as the order of the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ of the corresponding ARMA model need to be extraordinarily high to mimic the slow decay of an ACF over certain periods.[202] Thus,

[197] See, Hassler (2019, Corollary 6.1 on p. 107 and Proposition 6.3 on p. 108) and Palma (2007, p. 48).
[198] In case of $0 < d < 1/2$, this results follows directly from the equivalence of (2.27) and (2.28) in Section 2.3 together with (2.32). Note that the stated expression also yields if $-1 < d < 0$. The constant C depends on the parameters of the autoregressive and moving average polynomials, the white noise variance, and the parameter d, see, e.g., Palma (2007, p. 48).
[199] Note that $\sum_{k=1}^{\infty} 1/k^p < \infty$ if and only if $p > 1$, see Hassler (2019, Lemma 3.1(c) on p. 29). For $p = 1 - 2d$, this corresponds to $d < 0$.
[200] See Hosking (1981, p. 170) and Beran (1994, p. 61).
[201] See, e.g., P. C. B. Phillips and Xiao (1998, p. 450) or Caporale, Gil-Alaña, and Lovcha (2016, p. 99).
[202] See Brockwell and Davis (1987, p. 465).

from an estimation perspective, the class of ARFIMA processes offers the advantage that the long-run properties of the model can be captured by the single parameter d instead of estimating many possible worse identifiable coefficients of a high-order ARMA process. In practice, however, it seems impossible to decide whether a given finite time series with a slowly decaying ACF is “best” described by a long memory or a possibly high-order short memory process.

2.4.2 The ARFIMA(1,d,0) Process: A Closer Look

This section sheds light on a fractionally integrated AR(1) process. As one target of this thesis is to investigate how model dynamics in a DSGE model change under long memory dynamics, it seems reasonable to consider the ARFIMA$(1, d, 0)$ model as it generalizes directly the standard AR(1) setting which is itself contained as a special case if $d = 0$.[203] By handling the two parameters ϱ and d, different persistence regimes and their impact on the model dynamics can be analyzed by keeping a tractable and quite simple structure of the process contemporaneously. Therefore, the parameter ϱ controls the exponentially fast decaying short-run correlations, whereas the parameter d controls the long-run correlations.

Let $(X_t)_{t\in\mathbb{Z}}$ be a zero-mean ARFIMA$(1, d, 0)$ process given by

$$(1-\varrho B)X_t = (1-B)^{-d}\varepsilon_t \text{ or } X_t = \varrho X_{t-1} + \nu_t \text{ with } \nu_t = (1-B)^{-d}\varepsilon_t \tag{2.34}$$

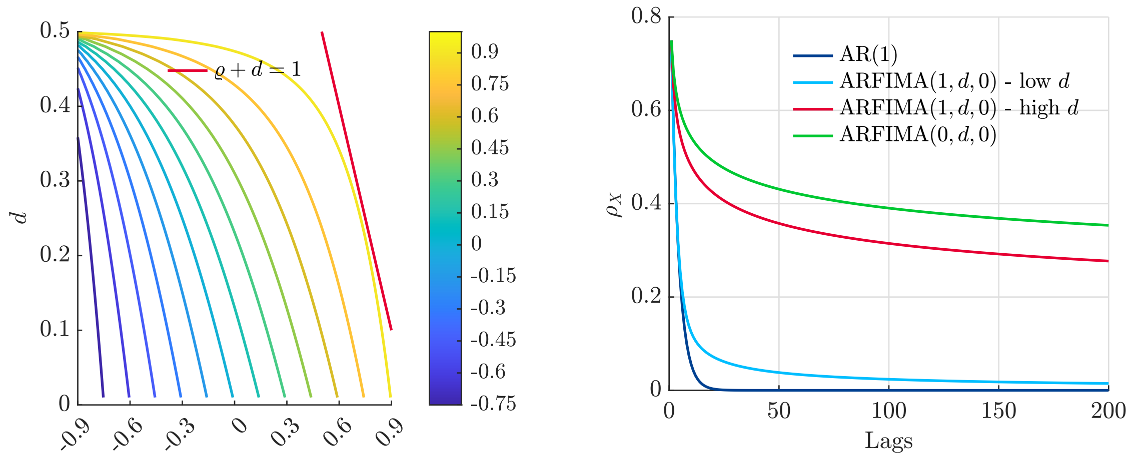
with $|\varrho| < 1$ and $-1 < d < 1/2$. Again, $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a Gaussian white noise process with variance σ^2_ε. The right-hand side of (2.34) states that an ARFIMA$(1, d, 0)$ process can be interpreted as an AR(1) process with respect to the ARFIMA$(0, d, 0)$ process $\nu = (\nu_t)_{t\in\mathbb{Z}}$ instead to the white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$. The long memory properties are, thus, captured fully by ν.[204] The closed-form expressions of the autocovariance function and the ACF are given in (A.10) and (A.11) of Appendix A.3, respectively. They are not stated here, as their representation in terms of the Gaussian hypergeometric function is rather complicated and not enlightening from the perspective of this section.

Instead, Figure 2.6 provides some insights into the autocorrelation structure and the interrelation between the parameters ϱ and d. Panel a) of Figure 2.6 shows how the first-order autocorrelation of X may be determined by the short memory parameter ϱ and the long memory parameter d. In contrast to the AR(1) process considered in Section 2.1.5.3, the first-order autocorrelation of an ARFIMA$(1, d, 0)$ process depends on both parameters,

[203] See Chapter 4 for the implications of an ARFIMA$(1, d, 0)$ technology shock in the context of a DSGE model.
[204] An analogous deduction can be made from (2.29) for the general ARFIMA(p, d, q) process.

i.e., although the parameter d controls for the asymptotic behavior of the ACF, it has an impact on the ACF at small lags as well.

For small values of d ($d < 0.2$), $\rho_X(1)$ depend almost linearly on ϱ and d, i.e., by keeping the value of $\rho_X(1)$ constant, a reduction in the parameter ϱ leads to an almost proportional increase of d. For higher values of $\rho_X(1)$, a strong increase in one parameter is needed to keep $\rho_X(1)$ constant, while, at the same time, the other parameter is reduced by a small amount.



a) Contour plot of $\rho_X(1)$

b) ACFs of various ARFIMA$(1, d, 0)$ processes

Figure 2.6: Autocorrelations of various ARFIMA$(1, d, 0)$ processes. Panel a) shows a contour plot of the first-order autocorrelation $\rho_X(1)$. The red line divides the parameter space according to $\varrho + d = 1$, i.e., all combinations of ϱ and d to the left (right) of the red line yield $\varrho + d < 1 (> 1)$. Panel b) shows four ACFs of various ARFIMA$(1, d, 0)$ specifications. The values of ϱ and d are chosen to yield $\rho_X(1) = 0.75$. The AR(1) process thus has the persistence parameter $\varrho = 0.75$. The pure long memory process has the long memory parameter $d = 0.4286$. The light-blue line refers to an ARFIMA$(1, d, 0)$ specification with $\varrho = 0.6$ and $d = 0.1586$, and the red line refers to an ARFIMA$(1, d, 0)$ specification with $\varrho = 0.1$ and $d = 0.4078$. These combinations of ϱ and d are derived from a numerical solution of (A.13). Note the different scaling of the axes.

The corresponding ACF plotted over time can be seen in Panel b) of Figure 2.6. There, four ACFs of ARFIMA$(1, d, 0)$ processes with $\rho_X(1) = 0.75$ are plotted for various combinations of ϱ and d. The slow decay of the ACF is present in all long memory cases, i.e., for all processes with $d > 0$.

As can be seen from the light-blue line in Panel b) of Figure 2.6, the ACF of an ARFIMA$(1, d, 0)$ process with a low value of d mirrors the exponential decay of an AR(1) process in the first (roughly) 10 periods, but decays much slower afterward. A higher value of d increases the autocorrelations at smaller and higher lags sharply; see the red line in Panel b) of Figure 2.6. The highest autocorrelation at all lags is obtained without the exponentially fast decaying short memory component, see the green line in Panel b) of Figure 2.6.

Thus, combining short and long memory can reproduce various dependence structures between a purely short memory AR(1) process and a purely long memory ARFIMA$(0, d, 0)$ process. This property may corroborate the usefulness of long memory processes. On the other hand, it may be difficult to distinguish between short and long memory from an empirical perspective. Especially for small values of d, the autocorrelations at higher lags may appear insignificant as they are relatively small, see the light-blue line in Panel b) of Figure 2.6.

As for the ACF, the IRF of an ARFIMA$(1, d, 0)$ depends on both parameters. The closed-form expression of the IRF is derived in the following lemma.

Lemma 2.4.2

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARFIMA$(1, d, 0)$ process, as defined in (2.34)*. Then X has the infinite moving average representation $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ with coefficients*

$$\psi_k = \sum_{i=0}^{k} \varrho^i \frac{\Gamma(k-i+d)}{\Gamma(k-i+1)\Gamma(d)} = \sum_{i=0}^{k} \varrho^{k-i} \frac{\Gamma(i+d)}{\Gamma(i+1)\Gamma(d)}.$$

Proof

See Appendix A.1.2 □

By setting $\varrho = 0$ in Lemma 2.4.2, the IRF of an ARFIMA$(0, d, 0)$ process is simply given by the sequence $(\alpha_k)_{k=0}^{\infty}$ given in (2.30).

Figure 2.7 illustrates the shape of various IRFs. Panel a) of Figure 2.7 shows the IRF of various ARFIMA$(0, d, 0)$ processes. Again, in the long run, a rather slow decay can be observed. Compared to an AR(1) process with parameter $\varrho = 0.95$, the decay of the ARFIMA process's IRF is faster in the first periods following the shock (roughly 60 periods after the shock, depending on the concrete value of d). Similar to the AR(1) process, for which a higher value of ϱ increases the shock impact on all subsequent periods, a higher value of d is associated with a more substantial shock impact on all subsequent periods. Recall that, in contrast to an AR(1) process, the CIR of ARFIMA$(0, d, 0)$ process with $d < 1/2$ is infinite. Thus, the main proportion of the IRF is located at higher lags of the IRF. Consequently, the IRF of the AR(1) process in Panel a) of Figure 2.7 eventually dies out faster than the four IRFs of the other four ARFIMA$(0, d, 0)$ processes.

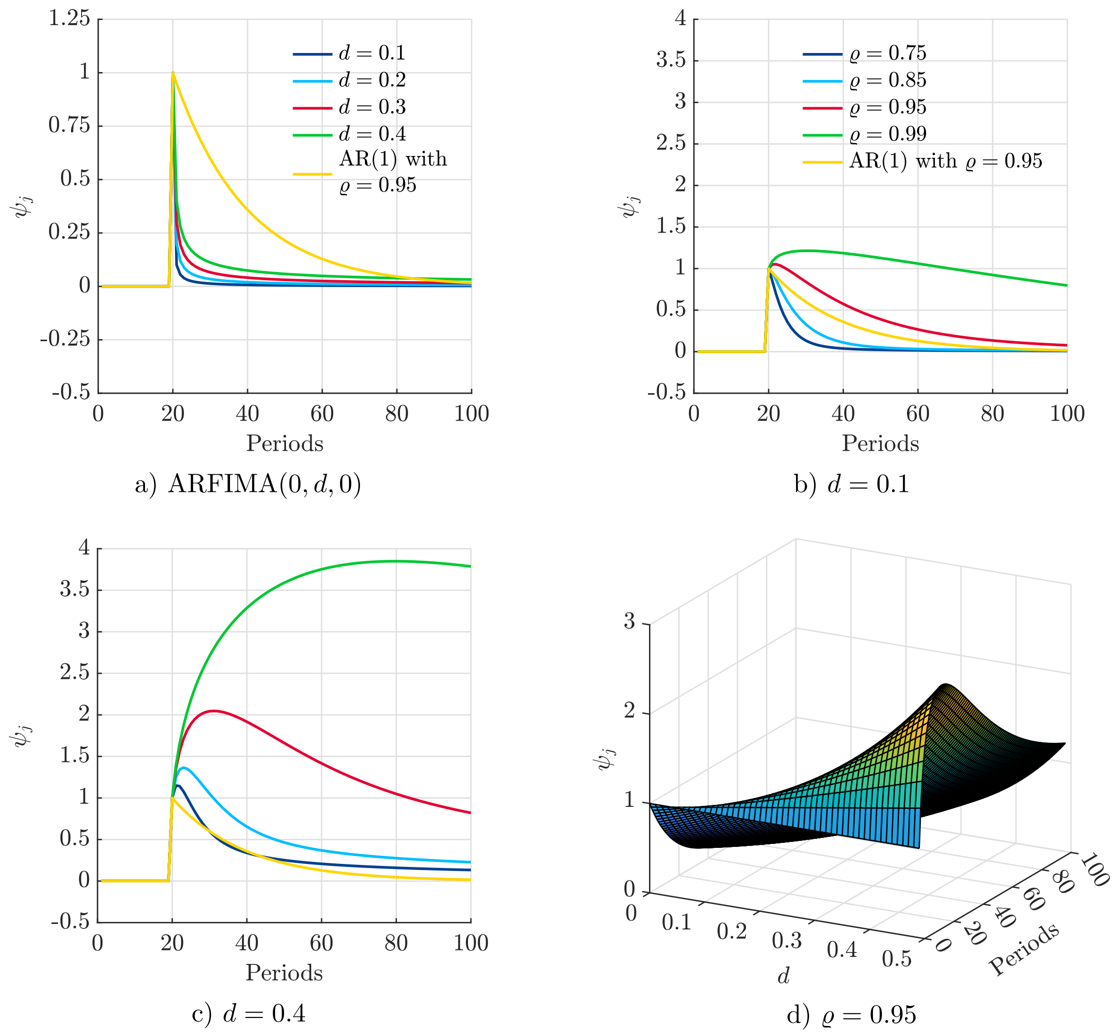


Figure 2.7: Impulse-response functions of various ARFIMA$(1, d, 0)$ processes. The shock occurs at period 20 in panels a) to c), see further Footnote 116. Panel a) shows the IRFs of an ARFIMA$(0, d, 0)$ process for various values of d. Panels b) and c) show IRFs of ARFIMA$(1, d, 0)$ processes for various values of ϱ and $d = 0.1$ in panel b) and $d = 0.4$ in panel c). The legend of panel c) is the same as in panel b) and has been omitted in panel c) for better readability. The yellow lines in panels a) to c) serve as a reference and show the IRF of an AR(1) process with $\varrho = 0.95$. Panel d) shows a three-dimensional plot of the ARFIMA$(1, d, 0)$ process with $\varrho = 0.95$ over a grid of d values, where d takes values from 0 to 0.49 with a step size of 0.01. Note the different scaling of the axes.

Considering short and long memory in combination, see Panels b) and c) of Figure 2.7, in some cases, one can see a hump-shaped IRF, i.e., the impact of a shock increases before it slowly dies out. In other cases, the IRF decreases monotonically, similar to the one of an AR(1) process. However, in the presence of long memory, the decline is slower in the long run than for an AR(1) process.

The larger the value of d, the slower the long-run decay of the IRF for a fixed value of ϱ. The same holds for an increasing ϱ while keeping the parameter d fixed. If a hump shape occurs, it is more distinctive the higher the parameters ϱ and d. The same can be seen in Panel d), where there is no initial increase in the IRF for small values of d, but an even steeper rise with growing values of d.

The graphically illustrated interdependence between the parameters ϱ and d and the relation to the IRF is pinned down more theoretically in the following lemma.

Lemma 2.4.3

Let $|\varrho| < 1$ be the first-order autoregressive parameter and $d \in (-1, 1/2)$ be the fractional order of integration of the ARFIMA$(1, d, 0)$ process $X = (X_t)_{t\in\mathbb{Z}}$. Then,

i) the IRF satisfies $\psi_k = \varrho\psi_{k-1} + \alpha_k$, with initial value $\psi_0 = 1$ and $\alpha_k = \dfrac{\Gamma(d+k)}{\Gamma(k+1)\Gamma(d)}$.

ii) it yields, $\psi_1 = \varrho + d$.

iii) for $d \neq 0$, $\psi_k \sim \dfrac{1}{(1-\varrho)\Gamma(d)} k^{d-1}$ as $k \to \infty$.

iv) the IRF is positive if $0 < \varrho < 1$ and $0 < d < 1/2$. In addition, if $\varrho + d > 1$, the IRF is hump-shaped[205], and if $\varrho + d < 1$, the IRF decays monotonically.

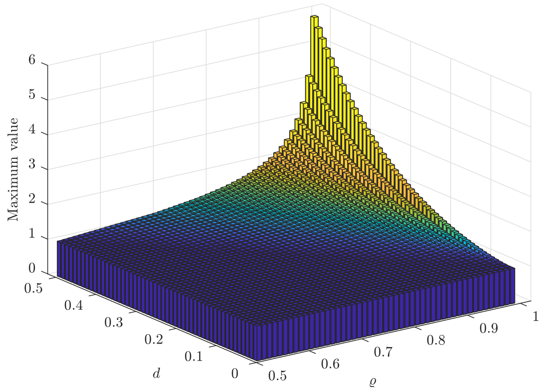
Proof

See Appendix A.1.3 □

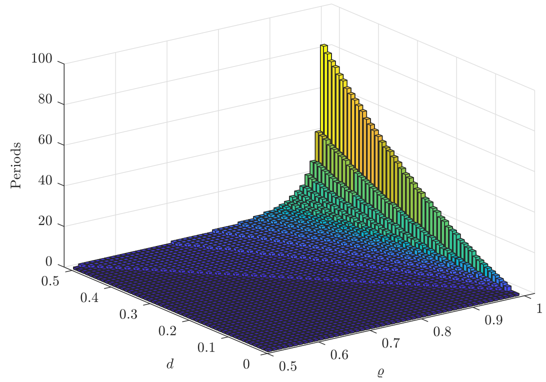
Lemma 2.4.3 and (2.33) highlight that the IRF of an ARFIMA$(1, d, 0)$ process decays asymptotically faster than its ACF. This observation further implies that in contrast to AR(1) processes, the IRF and ACF are generally different. The preceding lemma further highlights that the long memory parameter d controls not only the rate of decay of the ACF and IRF. It has crucial short-run implications worth considering: If there is a unit shock at time zero, Lemma 2.4.3 shows that this shock has an impact of $\varrho + d$ in the following period compared to ϱ in the reference AR(1) case.

[205] In this thesis, a hump-shaped curve or function is characterized by at least one period of increasing values at the beginning, followed by a monotonic decrease after the peak has been reached.

However, as can be seen from Panel a) of Figure 2.6, the condition $\varrho + d > 1$ is associated with a high first-order autocorrelation. In fact, if the first-order autocorrelation is assumed to be less than or equal to 0.9, there is no combination of ϱ and d that satisfies $\varrho + d > 1$.



a) Maximum value of the IRF



b) Number of periods to reach the maximum value

Figure 2.8: Maximum value of an ARFIMA$(1, d, 0)$ process's IRF and the number of periods needed to reach this maximum as a function of d and ϱ. Panel a) shows the maximum value of the IRF, and Panel b) shows the number of periods needed to reach this maximum value, each as a function of the parameters ϱ and d. These values are computed numerically over a grid with a step size of 0.01 and values of d ranging from 0 to 0.49 and ϱ ranging from 0.51 to 0.99.

Panel a) of Figure 2.8 shows that the maximum value of the IRF increases with increasing d and ϱ.[206] The same yields for the number of periods to reach this maximum value, see Panel b) of Figure 2.8. As one would expect from Lemma 2.4.3, the maximum value corresponds to the initial value of $\psi_0 = 1$ until the condition $\varrho + d > 1$ is satisfied. Correspondingly, the number of periods to reach the maximum value is equal to zero as long as $\varrho + d < 1$.[207] The cascaded shape seen in Panel b) implies that an increase in the parameters ϱ and d is not always associated with an extended rise of the corresponding IRF. However, the IRF may reach a higher maximum value, as can be seen from Panel a) of Figure 2.8. Considering the highest parameter constellation on the grid, i.e., $d = 0.49$ and $\varrho = 0.99$, there is a shock amplification up to almost six times higher than the initial shock, and it takes about 80 periods to reach this value.

Figure 2.9 summarizes the persistence properties of an ARFIMA$(1, d, 0)$ process and the findings of Lemma 2.4.3. The dotted right-upper corner is the parameter region with a corresponding hump-shaped IRF. The diagonal red line in Figure 2.9 represents all combinations of ϱ and d for which $\varrho + d = 1$.

In summary, the extension of the short memory and moderately persistent AR(1) process

[206] For the coloring of the three-dimensional bar diagrams shown in Figure 2.8, a Matlab code from Struyf (2014), URL in list of references, was used.

[207] This follows immediately from the IRF's monotonic decrease in the case of $\varrho + d < 1$; see Lemma 2.4.3.

with long memory (and strong persistence) by an ARFIMA$(1, d, 0)$ process can significantly change the shape of the corresponding IRF and ACF. Although the long memory parameter d determines the long-run properties of the process dynamics, it also has impacts on the short-run correlations of the process. In the context of DSGE models, persistence mainly refers to the parameter ϱ and not to d. By including the long memory parameter d into the existing DSGE framework, a more comprehensive range of persistence regimes within the common stationarity assumption can be exploited; see the light-blue area of Figure 2.9. The considerations of this section underline that the parameter d (as ϱ) is a parameter of persistence.[208]

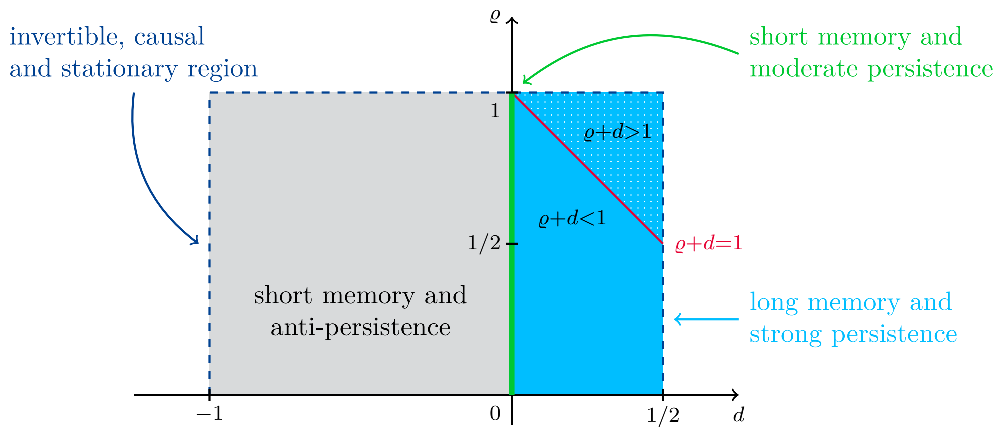


Figure 2.9: Characterization of the persistence and memory properties of an ARFIMA$(1, d, 0)$ process in the d-ϱ plane.

Finally, Figure 2.10 shows some realizations of ARFIMA$(1, d, 0)$ processes for various values of d, keeping the autoregressive parameter constant and equal to $\varrho = 0.75$ in all panels. As the long memory parameter increases, the presence of stronger cyclicality and local trends can be clearly seen from Figure 2.10.[209]

The following section introduces fractional Brownian motion (fBm) as a continuous-time stochastic process with long memory. Some relations to ARFIMA processes are also presented.

208 Hassler (2016, pp. 104 and 108) also refers to d as a measure of persistence, i.e., the larger the value of d, the stronger the persistence (or memory) of the process.

209 The reason why the panels shown in Figure 2.10 are plotted over 1000 periods instead of 100 as in Figure 2.2 is that the long memory features such as local trends and the non-periodic cyclicality are better seen in longer time series.

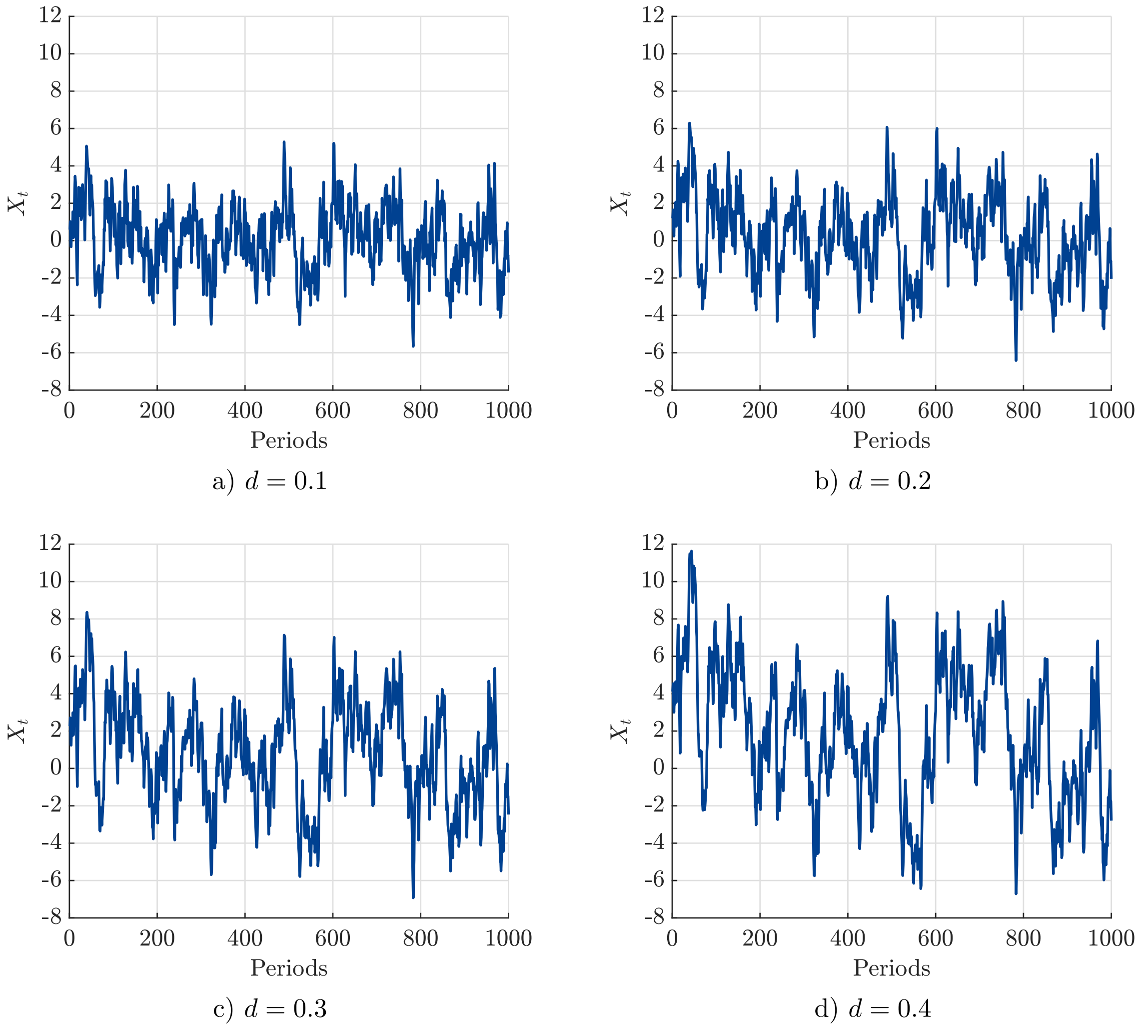


Figure 2.10: Paths of various ARFIMA$(1, d, 0)$ processes. The underlying white noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is assumed to be Gaussian with zero mean and variance 1. Furthermore, the realization of the white noise process is identical for all panels. The autoregressive parameter is set to 0.75 in all panels and the long memory parameter is set to $d = 0.1, 0.2, 0.3, 0.4$ in Panels a) to d), respectively.

2.5 From Discrete-Time to Continuous-Time: Fractional Brownian Motion

In Chapter 3, it is illustrated that ARFIMA processes are widely applied in the econometric literature and in time series analysis. Roughly twenty years before ARFIMA processes were introduced by Granger and Joyeux (1980/2001) and Hosking (1981), Mandelbrot and van Ness (1968) had initially proposed a continuous-time stochastic process called fractional Brownian motion (fBm) to replicate the Joseph effect and Hurst phenomenon mentioned in Section 2.2.[210] In Section 2.2.1, fractional Brownian motion (fBm) was

[210] See Mandelbrot and van Ness (1968, pp. 422f.).

already mentioned, but this section is dedicated to providing an exact definition of fBm and describing its properties more rigorously. Additionally, some relations to ARFIMA processes mentioned in Section 2.4 are provided. This section further builds the basis for the considerations of Chapter 5.

Before the formal definition of fBm is given, recall the definition of a standard Brownian motion[211,212]

Definition 2.5.1

Let $\mathbb{T} = [0, \infty)$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process $W = (W_t)_{t \in \mathbb{T}}$ is called standard Brownian motion or Wiener process if i) $W_0 = 0$ almost surely, ii) the increments $W_{t_1} - W_{t_0}, W_{t_2} - W_{t_1}, \dots W_{t_n} - W_{t_{n-1}}$ are stochastically independent for all $t_0, t_1, t_2, \dots, t_n \in \mathbb{T}$ with $0 = t_0 < t_1 < t_2 < \dots < t_n$, iii) each increment $W_t - W_s$ with $0 < s < t$ is normally distributed with zero expected value and variance $t - s$ iv) almost all paths of W are continuous. ⋊

The restriction to $\mathbb{T} = [0, \infty)$ in Definition 2.5.1 is common in the literature since the index t is often interpreted as time for which it seems natural to start at zero, but sometimes Brownian motions are defined over another index set such as $\mathbb{T} = (-\infty, \infty)$.[213] The increments of a Brownian motion are stationary as their distribution depends only on the time-difference and $t - s$ but not on t and s.[214]

There exist many definitions of a Brownian motion equivalent to Definition 2.5.1.[215] Another yet useful definition refers to the class of Gaussian processes to which the standard Brownian motion belongs.[216] Gaussian processes are well-defined through their expected value function and covariance function.[217] Thus, a Gaussian process $X = (X_t)_{t \geq 0}$ is a Brownian motion if and only if $\mathbb{E}X_t \equiv 0$ for all $t \geq 0$ and $\gamma_X(s,t) = \min\{s,t\}$ for all $s, t \geq 0$.[218]

---

211 Definition 2.5.1 is adapted from Shreve (2004, Definition 3.3.1 on p. 94).

212 The term "$W_0 = 0$ almost surely" used in the definition means that the complementary event (i.e., $W_0 \neq 0$) is contained in a set with probability zero, see Klenke (2013, p. 32). Ultimately, it states that $\mathbb{P}(W_0 = 0) = 1$. Similarly, the term "almost all" means that the probability of the set containing the non-continuous paths has again probability zero.

213 See, e.g., Mandelbrot and van Ness (1968, p. 423) or Doob (1953, p. 392).

214 See Shreve (2004, p. 466).

215 See, e.g., Shreve (2004, Theorem 3.3.2 on pp. 96f.) for different characterizations of a Brownian motion.

216 A process $X = (X_t)_{t \geq 0}$ is said to be a Gaussian process if the vector $(X_{t_1}, \dots, X_{t_n})$ has a multivariate Normal distribution for all $0 < t_1 < t_2 < \dots < t_n$ for every finite $n \in \mathbb{N}$, see Shreve (2004, Definition 4.7.1 on p. 172).

217 See Klenke (2013, Bemerkung 21.10 on p. 475).

218 See Klenke (2013, p. 475).

Fractional Brownian motion appears to be a direct generalization of the standard Brownian motion defined in Definition 2.5.1 and also belongs to the class of Gaussian processes. A definition of fBm, again in terms of the expected value and covariance function, is as follows.[219]

Definition 2.5.2

The continuous and real-valued Gaussian process $B^H = (B_t^H)_{t\geq 0}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called fractional Brownian motion with Hurst index $H \in (0,1)$ if $\mathbb{E}B_t^H = 0$ for all $t \geq 0$ and

$$\gamma_{B^H}(s,t) = \mathbb{E}(B_s^H B_t^H) = \frac{1}{2}\left(s^{2H} + t^{2H} - |s-t|^{2H}\right) \quad \text{for all } s,t \geq 0. \tag{2.35}$$

⋊

The following lemma shows how fBm generalizes the standard Brownian motion defined in Definition 2.5.1. Recall from Section 2.2.1 that the increment process of fBm is called fractional Gaussian noise.

Lemma 2.5.3

Let $B^H = (B_t^H)_{t\geq 0}$ a fractional Brownian motion in accordance with Definition 2.5.2. Define the increment process $X_n^H = \left(X_n^H\right)_{n\in\mathbb{N}_0}$ as $X_n^H = B_{n+1}^H - B_n^H$ for $n = 0,1,2,\ldots$, then,

i) if $H = 1/2$, B^H is a standard Brownian motion.

ii) if $H \neq 1/2$, $\gamma_{X^H}(k) \sim H(2H-1)k^{2H-2}$ as $k \to \infty$. Thus, X^H is a long memory process if $H > 1/2$ and a short memory process if $H \leq 1/2$.

Proof

See Appendix A.1.4 □

Lemma 2.5.3 illustrates that for $H > 1/2$, the increments of an fBm are positively correlated, showing long memory, whereas the increments are negatively correlated, showing short memory if $H < 1/2$. Furthermore, it follows from Part ii) of Lemma 2.5.3 that in the case of $H > 1/2$, the increment process satisfies (2.27) with $d = H - 1/2 \in (0, 1/2)$. This relation hints at how an ARFIMA process with long memory parameter d and fBm with parameter H are interrelated. There is, however, a more elaborate connection between discrete-time ARFIMA processes and a continuous-time fractional Brownian motion in the sense of the following convergence result.

[219] This definition is adapted from Biagini et al. (2008, Definition 1.1.1. on p. 5).

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARFIMA(p,d,q) process with $d = H - 1/2 \in (-1/2, 1/2)$ and let $0 < r \leq 1$ and $[N \cdot r]$ be the integer part of $N \cdot r$. Let c be a certain constant; then, it yields

$$c^{-1/2}N^{-H}\sum_{t=1}^{[N\cdot r]} X_t \Rightarrow B_r^H, \tag{2.36}$$

as $N \to \infty$.[220,221]

Equation (2.36) states that an appropriately scaled sum of an ARFIMA process converges weakly to a fractional Brownian motion.[222]

To gain some further intuition of (2.36), Figure 2.11 depicts the left-hand side of (2.36) (let it denoted by Y_r^N) as a function of r. From Panels a) and b) of Figure 2.11, it can be seen clearly that Y_r^N is a step function whose levels depend on an underlying ARFIMA$(0, d, 0)$ process with $d = 0.25$ that is not depicted. The intervals over which each step is constant have length $1/N$.[223] Consequently, the intervals show a decreasing length with increasing N. From Panel d) of Figure 2.11, it is easily imaginable that the limiting process is indeed continuous and that it resembles visual path properties of an fBm with $H = 0.75$.[224]

Roughly speaking, it follows from (2.36) that fBm may be seen as an integrated ARFIMA process. Thus, an ARFIMA process with long memory parameter $d \in (-1/2, 1/2)$ may be associated with the increments of an fBm with $H = d + 1/2$.

Another perspective on fBm and its relationship to ARFIMA processes offers the following integral representation of fBm originally proposed by Mandelbrot and van Ness (1968)[225]

$$B_t^H = \frac{1}{\Gamma(H+1/2)}\left(\int_{-\infty}^{0}\left[(t-s)^{H-1/2} - (-s)^{H-1/2}\right] dW_s + Z_t^H\right), \tag{2.37}$$

220 See Marinucci and Robinson (1999, p. 115).

221 This convergence result is intended to give the reader an intuition of the relationship between ARFIMA processes and fractional Brownian motion. Therefore, some mathematical details are omitted for convenience. To be more precise, the arrow "⇒" means weak convergence in the space of right-continuous functions whose left-limits exist. For details on this type of convergence, see, e.g., Marinucci and Robinson (1999, pp. 111f.) or Hassler (2019, pp. 132f.). The precise value of c depends on d and can be found in Hassler (2019, p. 133).

222 Note that there are various extensions and variants of this convergence result. For an overview of similar functional central limit theorems in this context, see Marinucci and Robinson (1999, pp. 115f.) and Hassler (2019, pp. 132ff.).

223 See Hassler (2019, p. 132).

224 Due to space limitations, a figure showing some realizations of an fBm is omitted here but can be found, e.g., in Enriquez (2004, pp. 221f.) or Rostek (2009, pp. 8ff.). Since the increments are negatively correlated for $H < 1/2$, the path of a corresponding fBm appears to be jagged, as it is characterized by quick changes in its direction. The greater the value of H, the smoother appears the path of a corresponding fBm as the process tends to keep its direction for a while.

225 See Mandelbrot and van Ness (1968, Definition 2.1 on p. 423).

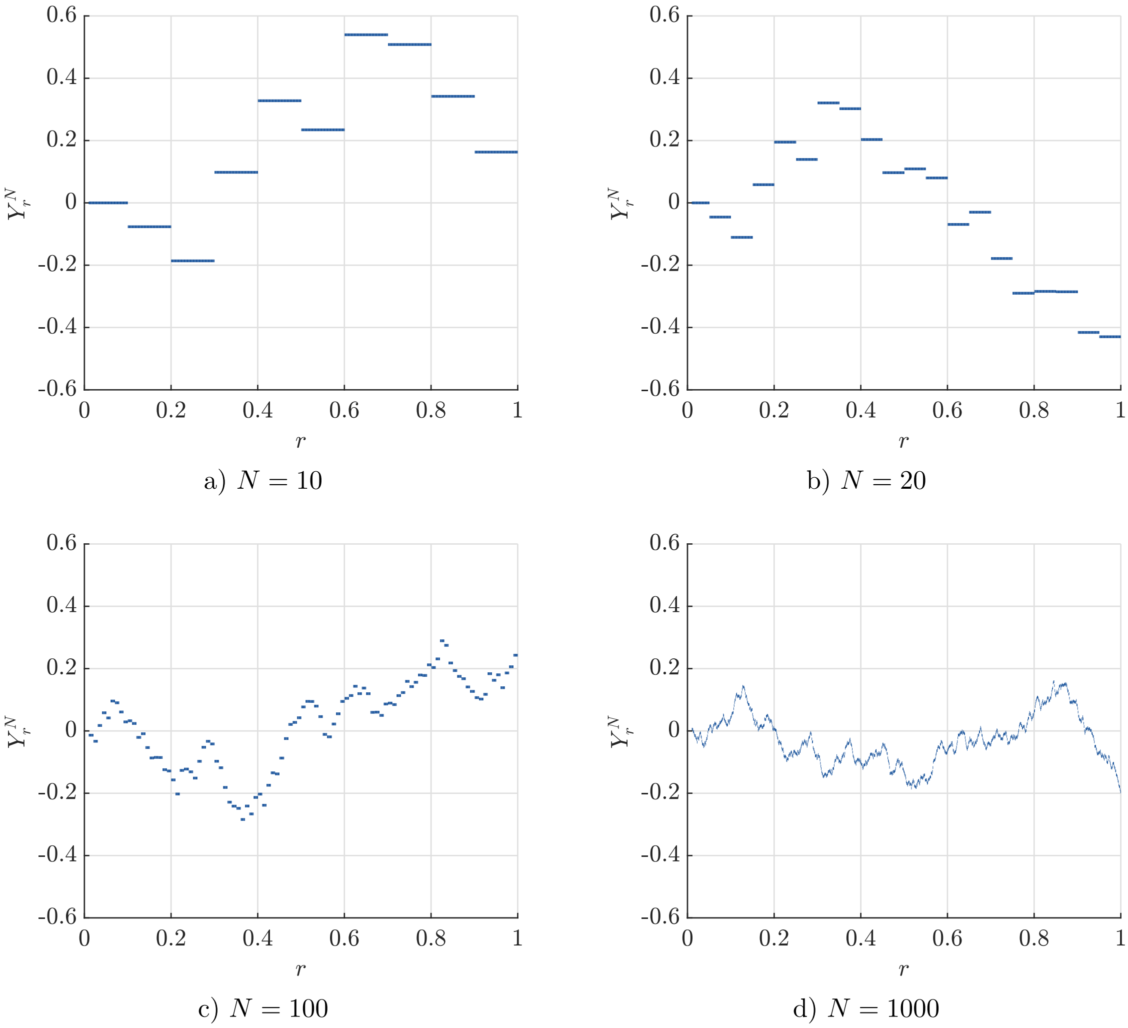


Figure 2.11: Convergence of appropriately scaled sums of an ARFIMA process to fractional Brownian motion. Panels a) to d) show the left-hand side of (2.36) as a function of r for values of $N = 10, 20, 100, 1000$, respectively. The underlying ARFIMA$(0, d, 0)$ process has the long memory parameter $d = 0.25$ and is not shown. For the sake of simplicity, the constant c given in (2.36) is set to 1.

where

$$Z_t^H = \int_0^t (t-s)^{H-1/2}\, dW_s, \tag{2.38}$$

$(W_t)_{t\in\mathbb{R}}$ is a Brownian motion and $\Gamma(\cdot)$ refers again to the gamma function. Sometimes, the process Z^H given in (2.38) is called fractional Brownian motion instead of B^H.[226] Following the notation of Marinucci and Robinson (1999), the process B^H is called type I fractional Brownian motion and Z^H type II fractional Brownian motion.[227] As further pointed out by Marinucci and Robinson (1999), the usage of the process Z^H is more common in the econometric literature, while B^H is often in the focus of the probabilistic literature.[228]

The properties of a type I and II fractional Brownian motion are similar since both processes coincide with a standard Brownian motion if $H = 1/2$. In addition, both processes have zero expected value, and their variances at time t are equal to $|t|^H$.[229] However, the increments of Z^H are, in contrast to the one of B^H, not stationary.[230]

As will turn out in Chapter 5, where long memory is analyzed in the context of a continuous-time macro-financial model, the definition of stochastic differential equations with respect to B^H has drawbacks that seem difficult to overcome. Therefore, Chapter 5 uses an approximation of the process Z^H instead of using B^H directly.[231] Therefore, the rest of this section outlines some properties of the processZ^H and its relations to ARFIMA processes.

Another relation between the two processes B^H and Z^H can be deduced from the following observation. From (2.37), it follows immediately for $t_2 > t_1 > 0$ that[232]

$$B_{t_2}^H - B_{t_1}^H = \frac{1}{\Gamma(H+1/2)} \left(\int_{-\infty}^{t_2} (t_2-s)^{H-1/2}\, dW_s - \int_{-\infty}^{t_1} (t_1-s)^{H-1/2}\, dW_s \right). \tag{2.39}$$

Each integral in (2.39) may be seen as a continuous-time infinite moving average representation.[233,234] A similar reasoning yields for (2.38), but, the continuous-time moving

226 See, e.g., Sowell (1990, p. 498) or Comte and Renault (1996, Definition 1 on p. 105). The process Z^H is also mentioned in Mandelbrot and Wallis (1968/2002, p. 424) and named Holmgren-Riemann-Liouville fractional integral.

227 See Marinucci and Robinson (1999, pp. 113 and 116).

228 See Marinucci and Robinson (1999, pp. 118f.).

229 See Marinucci and Robinson (1999, pp. 144 and 116).

230 See Marinucci and Robinson (1999, p. 117).

231 See Section 5.1 for details.

232 See Mandelbrot and van Ness (1968, Footnote 8 on p. 424).

233 See Comte and Renault (1996, p. 103).

234 The similarities between (2.39) and a discrete-time infinite moving average representation (2.2) become

average representation is truncated at zero (compare the lower integral bounds of (2.39) and (2.38)). Interestingly, a close relationship exists between a truncated discrete-time moving average representation of an ARFIMA process and the process Z^H, as outlined in the following.

Consider an ARFIMA$(0, d, 0)$ process X with its infinite moving average representation and the corresponding truncated-at-zero version X^*, i.e.,

$$X_t = \sum_{k=0}^{\infty} \alpha_k \varepsilon_{t-k} \text{ and } X_t^* = \sum_{k=0}^{t-1} \alpha_k \varepsilon_{t-k}, \tag{2.40}$$

where the α_k's are given in Lemma 2.4.3. By (2.36), an appropriately scaled sum of the full discrete moving average process X converges to B^H. A similar result holds for the truncated version Z^H: Let X^* be as in (2.40), and let $H = d + 1/2$ and $0 \leq r \leq 1$, then, it yields[235]

$$c^{-1/2} N^{-H} \sum_{t=1}^{[N \cdot r]} X_t^* \Rightarrow Z_r^H \text{ as } N \to \infty,$$

where c is a certain constant different from the one given in (2.36).[236,237]

Overall, there is a strong relationship between the discrete-time ARFIMA processes on the one hand and the continuous-time fractional Brownian motion (either type I or II) on the other hand. The parameters d and H are also closely related, since $H = d + 1/2$. A major difference between fBm and ARFIMA processes is that the increments of an fBm are stationary, but the process itself is not. Therefore, an ARFIMA process may rather be associated with the increments of fBm. On the other hand, ARFIMA processes appear to be more flexible than fBm, since they can capture extended short memory dynamics with their additional autoregressive and moving average parameters.

In Chapter 4, the implications of an ARFIMA technology shock in the context of a discrete-time DSGE model are considered in more detail. Fractional Brownian motion is used in Chapter 5 to introduce long memory (technology) shocks in a continuous-time macro-financial model. Before these theoretical implications are derived, Chapter 3 is

apparent.

235 See Marinucci and Robinson (1999, p. 118) or Hassler (2019, Proposition 7.3 on p. 133).

236 The precise value of c can be found in Marinucci and Robinson (1999, p. 118) or Hassler (2019, Proposition 7.3 on p. 133).

237 Note that the convergence results holds not only for ARFIMA$(0, d, 0)$ processes but also for general ARFIMA(p, d, q) processes as can be seen from Hassler (2019, Proposition 7.3 on p. 133). In the case of an ARFIMA(p, d, q) process, one has to replace the white noise process ε in moving average representation of X^* given in (2.40) with the corresponding ARMA(p, q) process as outlined in Hassler (2016, p. 114).

devoted to underlining that long memory processes and ARFIMA processes are highly relevant from an empirical perspective.

3

Long Memory in Economics and Econometrics

After the more theoretical perspective of the last chapter, the focus now shifts to an empirical perspective. The purpose of this chapter is twofold. First, Section 3.1 briefly discusses some estimation methods used in the empirical literature with their strengths and weaknesses. This review of estimation methods makes it possible to evaluate and compare different empirical findings in the literature. Second, the results of various empirical investigations of long memory for selected economic variables are reviewed in Section 3.2. More specifically, Section 3.2 reviews empirical findings of long memory in gross domestic product (GDP) and related time series. Section 3.2.2 then gives three examples of strands of literature that also deal with the presence of long memory.

The final part of the chapter, Section 3.3, provides a rationale for the existence of long memory in economic time series.

This chapter illustrates that fractional integration and long memory are common in some economic time series. Together with plausible mechanisms for why fractional integration and long memory seem reasonable from a macroeconomic perspective, this paves the way for introducing long memory dynamics into a theoretical macroeconomic model in the next chapter.

3.1 On Estimating Long Memory

That long memory may be difficult to be detected in empirical investigations can already be seen from its definition (see Definition 2.3.1). Given a limited data set, the number of calculable lags of the autocorrelation function (ACF) is also limited by the number of observations. Hence, the defining condition of long memory, $\sum_{k=0}^{\infty} |\gamma_X(k)| = \infty$, may not be provable by simply calculating an estimator of the ACF and plugging this estimator into the definition of long memory. Showing the series of the ACF's absolute values to be divergent may, thus, not be possible with any empirical methods at all. Ultimately, an asymptotic property must be estimated along a finite data set. That this is, in general, a challenging task from an empirical point of view seems obvious. For this reason, the estimation techniques discussed in the literature focus on the estimation of a long memory parameter as in (2.27), (2.28) or on the parameter d of an autoregressive fractionally integrated moving average (ARFIMA) or related processes, instead of proving the mentioned long memory definition.[238] Nevertheless, the difficulties in distinguishing clearly from an empirical perspective between various time series concepts such as trends, structural breaks, strong short memory, and long memory remain.

There is a sizeable econometric literature that proposes new methods or evaluates, improves, or revises existing methods for estimating long memory. A detailed review of all of these techniques is beyond the scope of this thesis as the focus lies on the modeling perspective carried out in Chapter 4 and Chapter 5. For this purpose, a general overview of some frequently used estimation techniques in the literature is provided in the following. This brief survey aims to allow for better classification and questioning of the empirical results that will subsequently be presented in Section 3.2. The available estimation techniques can roughly be grouped into heuristic approaches, semiparametric and parametric estimators.

3.1.1 The R/S Statistic as a Heuristic Estimator

The advantage of heuristic approaches is that they, besides stationarity, rarely make additional assumptions regarding the data-generating process. In addition, they can be calculated easily, and often serve as a rough indicator for the presence or absence of long memory. However, as it is already evident from their name, they are less appropriate to derive hard empirical facts from their results. A heuristic approach already mentioned is the R/S statistic of Section 2.2. A value of $H > 1/2$ indicates the presence of long memory, and the value of the long memory parameter d is then given by $d = H - 1/2$. Besides the already mentioned weaknesses of the R/S statistic, Davies and Harte (1987) show that the R/S statistic is very sensitive regarding the presence of short memory. Hence,

[238] In the light of Lemma 2.3.2, this outlines that there is often no clear-cut distinction between strong persistence and long memory in the literature, because the estimation techniques refer to conditions that are sufficient for both strong persistence and long memory.

it is unsuitable for discerning between long and short memory.[239] Lo (1991) suggested the replacement of the denominator (the S part) of the R/S statistic by a statistic that incorporates a sum of weighted autocovariances in order to account for possibly present short memory dynamics.[240]

However, as stated by Teverovsky et al. (1999), the method of Lo (1991) cannot distinguish between short and long memory reliably.[241] They carried out a Monte Carlo analysis and showed that the method of Lo (1991) is systematically biased to indicate short memory, even if the synthetic time series is a "true" long memory process with a high long memory parameter of $d = 0.4$.[242] Further, they argue that with the original R/S statistic, they would have been able to carve out the short and long memory characteristics of the considered processes more appropriately than with Lo's method.[243,244] Ultimately, they advocate not using Lo's method as the sole criterion for determining the presence or absence of long memory.[245] Overall, the suggestions proposed by Lo (1991) turned out not to remedy the weaknesses of the R/S statistic. Contemporaneously, more sophisticated estimation routines have been developed.[246,247]

3.1.2 Semiparametric Approaches: GPH, Local Whittle and Related Estimators

In contrast to heuristic approaches, semiparametric long memory estimators assume a little more structure behind the data-generating process, but they do not estimate a full parametric model.[248] The most discussed semiparametric estimators are the estimator of Geweke and Porter-Hudak (1983) (GPH henceforth) and the Gaussian semiparametric estimator, also called the local Whittle estimator.[249] Both use (2.28) to estimate the long memory parameter of a time series. Hence, only the properties of the spectral density

239 See Davies and Harte (1987, p. 96) and Lo (1991, p. 1288).
240 See Lo (1991, p. 1290).
241 See Teverovsky et al. (1999, Section 3.2 on pp. 219ff.).
242 See Teverovsky et al. (1999, p. 224).
243 See Teverovsky et al. (1999, p. 225).
244 The reason why the original R/S statistic delivers good results in the context of Teverovsky et al. (1999) may be due to the length of the generated processes, which with 10^5 observations is longer than usual economic time series.
245 See Teverovsky et al. (1999, pp. 225f.).
246 A more detailed discussion of the R/S statistic can also be found in Baillie (1996, Section 4.1 on pp. 27f.). There, the Baillie (1996) claims that the R/S statistic does not seem to be a favorable approach for estimating long memory.
247 There are several other heuristic approaches. A comprehensive overview can be found, e.g., in Beran et al. (2013, Section 5.4 on pp. 409ff.).
248 See Diebolt and Guiraud (2005, p. 830).
249 See Diebolt and Guiraud (2005, p. 831) and Beran et al. (2013, Section 5.6.2 and 5.6.3 on pp. 441ff.).

around the zero frequency are used in the estimation. Instead, a parametric approach aims to estimate all parameters of an assumed model (e.g., an ARFIMA(p, d, q) process) based on the data. Therefore, parametric estimators aim to estimate the whole spectral density of the assumed model.

The advantage of the semiparametric estimators is that one obtains an estimator of the long memory parameter without the need for specifying a possibly present complex short memory structure. At the same time, not considering the dynamics of short memory is a drawback, which will be illustrated in the following together with the Geweke-Porter-Hudak (GPH) estimator.

The GPH estimator can be motivated from (2.28). By taking logarithms of both sides, one obtains the asymptotic linear relationship

$$\log\left(f_X(\lambda)\right) \sim \log\left(c_f\right) - 2d\log(\lambda), \text{ as } \lambda \to 0. \tag{3.1}$$

The idea behind the GPH estimator is to replace the spectral density in (3.1) with its empirical counterpart the so-called periodogram. For a stationary time series with values $X_1, \ldots, X_n$, the periodogram $I_{n,X}$ is defined by

$$I_{n,X}(\lambda) := \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{-it\lambda}\right|^2, \tag{3.2}$$

and can be used as an empirical estimator of the spectral density.[250] If the periodogram is evaluated at the Fourier frequencies $\lambda_j = 2\pi j/n$ with $j = 1, \ldots, [(n-1)/2]$, one can derive the following linear regression equation out of (3.1) and (3.2)[251]

$$\log(I_{n,X}(\lambda_j)) = \beta_0 + \hat{d}_{GPH}(-2\log(\lambda_j)) + \xi_j \text{ for } j = 1, \ldots, m, \tag{3.3}$$

where the disturbances $(\xi_j)_{j=1,\ldots,n}$ are independent and identically distributed (i.i.d.).[252]

[250] See Beran et al. (2013, p. 441).

[251] See Beran et al. (2013, p. 441).

[252] Note that Geweke and Porter-Hudak (1983) used an equation similar to (2.31) instead of (2.28) to motivate their estimator. Evidently, by taking logarithms of both sides of (2.31) and letting $\lambda \to 0$, one obtains $\log(f_X(\lambda)) = C(-2d)\log(2\sin(\lambda/2))$, where C is a constant. For this reason, the regression equation originally used in Geweke and Porter-Hudak (1983, Equation 2) is slightly different from (3.3). However, since $\sin(x) \approx x$ for small values of x (see Footnote 195), (3.1) follows immediately. For calculating the GPH estimator in this thesis, (3.3) is used throughout. Doing so is in line with D. W. K. Andrews and Guggenberger (2003, p. 680), Beran et al. (2013, p. 441) and Hassler (2019, p. 172). As pointed out by D. W. K. Andrews and Guggenberger (2003, p. 680) and Hassler (2019, p. 172), the main properties of the estimator remain unchanged whether the original regression equation of Geweke and Porter-Hudak (1983) or (3.3) is used.

The GPH estimator ($\hat{d}_{GPH}$) is then given as the slope of the regression line when $\log(I_{n,X}(\lambda_j))$ is regressed on $(-2\log(\lambda_j))$. The parameter m refers to the smallest m Fourier frequencies that enter the regression and is called bandwidth parameter.[253] Since long memory refers to the property of the spectral density near the origin, it seems reasonable to choose small values of m in order to exclude short memory contamination of the higher frequencies in the regression (3.3).[254]

If one imposes additional regularity conditions as the Gaussianity of the data-generating process and the restriction of the true parameter value d_0 to the interval $(-1/2, 1/2)$, there is a limiting theory for the GPH estimator and the following can be proven to hold.[255,256] If

$$\frac{m}{n^{4/5}} + \frac{(\log(n))^2}{m} \to 0, n \to \infty, \tag{3.4}$$

then, it holds

$$\sqrt{m}(\hat{d}_{GPH} - d_0) \longrightarrow_{\mathcal{D}} N\left(0, \frac{\pi^2}{24}\right) \text{ as } n \to \infty, \tag{3.5}$$

where $N(\mu, \sigma^2)$ refers to the normal distribution with mean μ and variance σ^2. Note that (3.4) states that m has to grow faster than $(\log(n))^2$ as n tends to infinity, but at the same time, m must not grow faster than $n^{4/5}$.[257] Equations (3.4) and (3.5) together with (3.3) uncover a major disadvantage of the GPH estimator: From (3.5), it seems desirable to choose large values of m in order to decrease the variance of the estimator, however, as stated above, it seems desirable to keep the value of m small in order to avoid short term contamination of the regression equation.

Figure 3.1 illustrates this trade-off using a simulated ARFIMA$(1, d, 0)$ process of length $n = 200$ with long and short memory parameters equal to 0.3 and several values of the bandwidth parameter m.[258,259] Unsurprisingly, as the parameter m increases, more frequencies enter the linear regression. Thus, an intensified scatter around the left-hand end of the regression line due to the increasing short memory contamination can be

253 See Beran et al. (2013, p. 442) and Hassler (2019, p. 170).

254 See Geweke and Porter-Hudak (1983, p. 231).

255 See Hurvich, Deo, et al. (1998, Theorem 2 on p. 26) and Hassler (2019, Proposition 9.1 on pp. 170f.).

256 This result holds under a far more general model than $f_X(\lambda) \sim c_f|\lambda|^{-2d}$. Following D. W. K. Andrews and Guggenberger (2003, p.675 and pp. 679f.), the constant c_f can be replaced with an even function $g : [-\pi, \pi] \to \mathbb{R}, \lambda \mapsto g(\lambda)$ covering the short memory properties of the process and it is assumed that g is three times differentiable near zero with $g'(0) = 0$.

257 As stated by Geweke and Porter-Hudak (1983, p. 226), the condition $(\log(n))^2/m \to 0$ is satisfied, e.g., by $m = Cn^{\alpha}$, where C is a constant and $0 < \alpha < 1$. Overall, (3.4) is satisfied, e.g., by values of $m = Cn^{\alpha}$ with $0 < \alpha < 4/5$.

258 The value $n = 200$ was chosen as it is roughly the magnitude of available macroeconomic time series. In quarterly steps, such a time series would cover fifty years.

259 All figures in this chapter were computed using Matlab code written by the author.

observed, resulting in a biased estimator.[260] This bias in the estimator is even stronger the higher the short memory parameter is.[261] The mentioned trade-off can be seen from the lengths of the presented confidence intervals, which show a decreasing length with increasing m, i.e., higher values of m lead to a smaller variance of an increasingly biased estimator.[262] Although the lower limits of all confidence intervals are greater than zero, pointing towards significant long memory, they rarely provide convincing results since all upper bounds exceed the long memory parameter's threshold of 1/2. In addition, selecting the bandwidth parameter m is challenging, and unfortunately, it affects the estimation results in a non-negligible way.

The estimator with the smallest deviation from the true parameter value in Figure 3.1 is the one of Panel c). This value is optimal because it minimizes the mean squared error of the GPH estimator.[263] Note that this value is generally unknown as the whole spectral density of the data-generating process and its second derivative is needed for its calculation.[264] In the end, the selection of m is a priori unclear and, to a certain degree, arbitrary. Hence, the GPH estimator faces similar problems as the R/S analysis.

In order to gain some intuition for plausible values of m from an economic perspective, it may be enlightening to consider which cycle lengths are included in the GPH regression. If one assumes that the ARFIMA realization that is the basis of Figure 3.1 (not depicted) represents an economic time series in quarterly frequency, the value of $m = 14$ in Panel a) of Figure 3.1 is associated with the fourteenth frequency $\lambda_{14} = 2\pi(14)/200$, corresponding to cycles of length $2\pi/\lambda_{14} \approx 14.3$ quarters or 3.6 years. Thus, by fixing $m = 14$, all cycles longer than 3.6 years enter the calculation of the GPH estimator. By doing this, it is assumed that there are no short-run influences on the estimator from all cycles longer than 3.6 years.[265] However, business cycle frequencies discussed in the literature cover time spans ranging from 1.5-2 years up to 8 years.[266] Thus, it may be hard to argue for

[260] The values of m in Panels a),b) and d) are the same as in Geweke and Porter-Hudak (1983, Table 2 on p. 230).

[261] See Hurvich, Deo, et al. (1998, Table 1 and 2 on pp. 32f.).

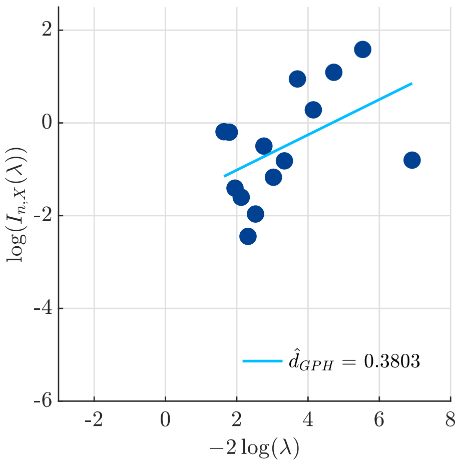
[262] See Beran et al. (2013, p. 442).

[263] The value of $m = 30$ was taken from Hurvich, Deo, et al. (1998, Equation 9 on p. 24 and Table 1 on p. 32).

[264] See Hurvich, Deo, et al. (1998, Equation 9 on p. 24 and p. 33). In Figure 3.1, the theoretical spectral density of an ARFIMA$(1, d, 0)$ process is known from (2.31).

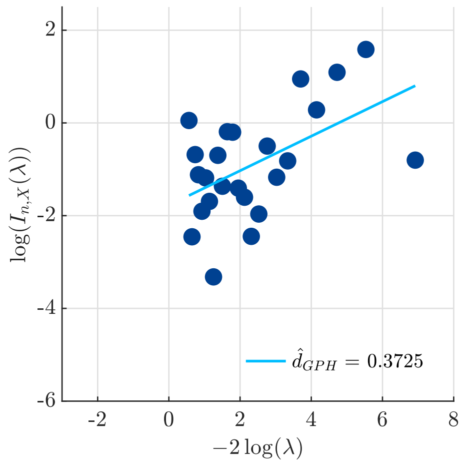
[265] Note that increasing values of m accompany the inclusion of cycles with shorter lengths in the GPH regression. The estimators, postulated in Panels b) to d) in Figure 3.1, hold under the assumption of no short-run influences of even shorter cycles than 3.6 years.

[266] Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 647 and Fig. 21 on p. 647) consider frequencies ranging from $\pi/16 \approx 0.196$ to $\pi/4 \approx 0.785$ corresponding to cycle lengths of 8 and 2 years, respectively. Schüler (2018, p. 14) considers the extended region from 1.5 to 8 years, where a cycle length of 1.5 years corresponds to a frequency of $\pi/3 \approx 1.047$.



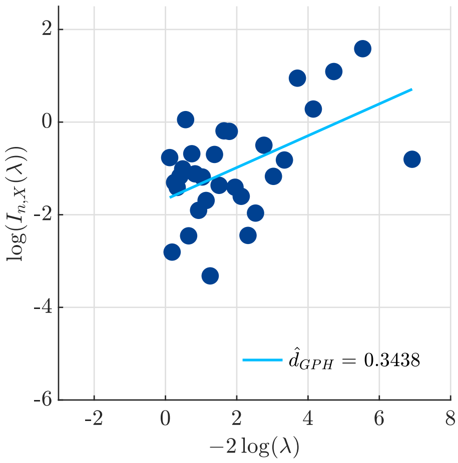
a) $m = \lceil 200^{0.5} \rceil = 14$;

CI for $\hat{d}_{GPH}$: $[0.04, 0.72]$



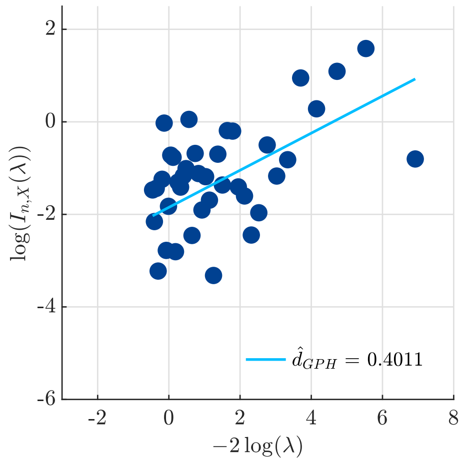
b) $m = \lceil 200^{0.6} \rceil = 24$;

CI for $\hat{d}_{GPH}$: $[0.12, 0.63]$



c) $m = [200^{\alpha}] = 30$, with $\alpha = \frac{\log(30)}{\log(200)}$;

CI for $\hat{d}_{GPH}$: $[0.11, 0.57]$



d) $m = \lceil 200^{0.7} \rceil = 40$;

CI for $\hat{d}_{GPH}$: $[0.2, 0.6]$

Figure 3.1: Geweke, Porter-Hudak estimator of an ARFIMA$(1, d, 0)$ process with autoregressive parameter $\varrho = 0.3$ and long memory parameter $d = 0.3$ for various values of the bandwidth parameter m. These values are $m = 14, 24, 30, 40$ in panels a) to d), respectively. The length of the simulated ARFIMA process is $n = 200$ periods, and the corresponding realization is the same in all panels. The value of the GPH estimator is given as the slope of the regression line. The confidence intervals (CI) are at the 95% level and were calculated from (3.5).

neglecting any short-run influence of cycles longer than 3.6 years.[267] In addition, it may be critical that the estimation of the long memory parameter includes only 14 out of 200 data points.[268] However, to take only cycles longer than 8 years into account, an even smaller number of ordinates has to be used.

There is vast literature suggesting improvements of the GPH estimator, including i) dropping some low frequencies from the regression equation[269], ii) a data-driven choice of the bandwidth parameter[270] or iii) methods of bias reduction.[271,272] Undeniably, these results lead to improvements in the estimation of the long memory parameter; however, a closer look reveals that the advantageous properties of the proposed improvements often only come into play with large sample sizes (at least 500 observations), which are usually not available in macroeconomic contexts.[273] Thus, the difficulties of the GPH estimator arise when it is applied to relatively short time series rather than being inherent in the methodology. For macroeconomic purposes, however, the GPH estimator seems unsuitable.

Another semiparametric estimator proposed by Robinson (1995b) is the so-called Gaussian semiparametric estimator or local Whittle estimator.[274] Like the GPH estimator, the local Whittle estimator uses the smallest m Fourier frequencies for the estimation. However, instead of doing a linear regression, the Whittle estimator minimizes an objective function, which approximates the maximum likelihood function in the frequency domain. To be more precise, the local Whittle estimator $\hat{d}_{LW}$ is defined by[275]

$$\hat{d}_{LW} := \underset{d\in\Theta}{\arg\min} \left(\log\left(\frac{1}{m} \sum_{j=1}^{m} \frac{I_{n,X}(\lambda_j)}{\lambda_j^{-2d}} \right) - \frac{2d}{m} \sum_{j=1}^{m} \log(\lambda_j) \right), \tag{3.6}$$

267 A similar deduction was made by Sowell (1992b, p. 298) when discussing the empirical findings of Diebold and Rudebusch (1989), see Section 3.2.1 for more details. See also Baillie (1996, p. 33).
268 See Beran (1994, p. 98).
269 See, e.g., Robinson (1995a).
270 See, e.g., Hurvich and Beltrao (1994).
271 See, e.g., D. W. K. Andrews and Guggenberger (2003).
272 Note that this list is not exhaustive. Further improvements suggested by Velasco (1999a) extend the interval of the true parameter d to the non-stationary area, i.e., $d > 1/2$. A rich list of related literature can also be found in Beran et al. (2013, p. 440).
273 The results of Hurvich and Beltrao (1994, Table II on p. 298) indicate that the original GPH estimator performs best in terms of a small mean squared error when the true long memory parameter is 0.49 and if there are only minor short memory contaminations, i.e., the autoregressive parameter is 0 or 0.3. Further, their approach outperforms the GPH estimator if the sample size contains 3000 observations, see Hurvich and Beltrao (1994, p. 295). The results of D. W. K. Andrews and Guggenberger (2003, pp. 697f.) indicate that their estimator has a smaller root-mean-square error compared to the GPH estimator for samples of sizes 512 and 2048; the reverse is true for samples of sizes 128.
274 The notion “local Whittle estimator” seems to be more common nowadays.
275 See Robinson (1995b, p. 1632f.) but replace the parameter H in his formulas with $d + 1/2$, Hassler (2019, p. 176), and Beran et al. (2013, p. 445f.).

where $\Theta =\subset (-1/2, 1/2)$ is the range of admissible values of d over which the minimum is taken, and m is the bandwidth parameter with the same meaning as for the GPH estimator. Note that there is no closed-form representation of the local Whittle estimator; thus, the minimization must be done numerically.[276] However, as for the GPH estimator, an asymptotic theory can be derived for $\hat{d}_{LW}$. Especially, given certain regularity conditions, it holds that[277]

$$\sqrt{m}(\hat{d}_{LW} - d_0) \longrightarrow_{\mathcal{D}} N\left(0, \frac{1}{4}\right) \text{ as } n \to \infty. \tag{3.7}$$

By a direct comparison between (3.5) and (3.7), the smaller asymptotic variance and, hence, the higher efficiency of $\hat{d}_{LW}$ compared to $\hat{d}_{GPH}$ becomes evident. However, since the choice of the bandwidth parameter m is again to a large extent arbitrary, and the value of the estimator is quite sensitive to this choice, it is unclear whether $\hat{d}_{LW}$ is preferable to $\hat{d}_{GPH}$.[278]

Similarly, as for the GPH estimator, a vast literature suggests generalizations and improvements of the local Whittle estimator. The range of admissible values of the true long memory parameter can be extended to the interval $(-1/2, 3/4)$ yielding the same limiting distribution as in (3.7).[279] This interval can further be extended to $(-1/2, 1]$, but the limiting distribution changes at values 3/4 and 1.[280] Furthermore, knowledge of the true parameter value is required before applying the aforementioned Whittle estimators, which is overall rather disadvantageous as the estimation aims to carry out knowledge about the true parameter value.[281]

By modifying (3.6), Shimotsu and P. C. B. Phillips (2005) suggested a slightly different

[276] See Robinson (1995b, p. 1632).

[277] See Robinson (1995b, Theorem 2 on p. 1641). For details on the necessary regularity conditions, see Robinson (1995b, Assumption A1' to A4' on pp. 1640f.).

[278] By carrying out a Monte Carlo analysis, Robinson (1995b) compared his estimator with the GPH estimator. Overall the results are mixed. Dependent on the choice of m, he found his estimator to be more biased than the GPH estimator when applied to a purely long memory process, see Robinson (1995b, p. 1654). On the other hand, he found his estimator to be superior if the true parameter value is near the admissible borders of $H = 0.1$ or $d = -0.4$ or $H = 0.9$ or $d = 0.4$, see Robinson (1995b, p. 1654). He further considered the possibility of short-term contamination and found the GPH estimator to be less biased than his estimator in the long memory region ($d > 0$ or $H > 1/2$), see Robinson (1995b, p. 1659).

[279] See Velasco (1999b, Theorem 3 on p. 94). He further suggested a modified estimator applicable to all values of $d > -1/2$. In this modified version, the periodogram $I_{n,X}$ in (3.6) is replaced with a tapered periodogram, see Velasco (1999b, Sections 5 and 6 on pp. 94ff.) for details.

[280] See P. C. B. Phillips and Shimotsu (2004, Theorems 4.1 and 4.2 on pp. 662f.).

[281] Prior knowledge is needed as it is commonly assumed in the derivations of the estimators that the true parameter d_0 belongs to the interval of admissible values Θ, see Robinson (1995b, Assumption A1 on p. 1633 and Assumption A1' on p. 1640),Velasco (1999b, Theorem 2 and 3 on pp. 93f.) and P. C. B. Phillips and Shimotsu (2004, Theorems 3.1 and 3.2 on p. 660 and Theorems 4.1 and 4.2 on pp. 662f.).

estimator and called it exact local Whittle estimator (ELW). For this estimator, the asymptotic relationship (3.7) remains true as long as the interval of admissible values for d has a length smaller than $9/2$.[282] An interval satisfying this condition is for example $\Theta = [-1, 3.5]$.[283] As this interval is rather long and seems to cover the most relevant cases from an empirical perspective, the prior knowledge regarding the true parameter d_0 seems less demanding for the ELW than for the local Whittle estimator. Based on a Monte Carlo simulation, the authors showed that their exact local Whittle estimator outperforms the local Whittle estimator and the tapered version suggested by Velasco (1999b) mentioned in Footnote 279.[284]

On the other hand, applying the ELW requires that the mean of the time series is known, an assumption that is rarely fulfilled in an empirical application.[285] A remedy for this drawback was proposed by Shimotsu (2010). He suggested replacing the ELW with a two-step estimation approach that is able to account for unknown mean and polynomial trends in the time series as well. However, the burden of this higher flexibility is a reduced range of admissible parameter values.[286]

An estimator with a similar mean squared error as the ELW was proposed by Abadir, Distaso, et al. (2007). Their fully extended local Whittle estimator (FELW) is based on a modified periodogram incorporating a certain correction term.[287] Under certain regularity condition, they showed (3.7) to hold as long as the range of admissible values $\Theta \subset [-3/2, \infty)$ and the true parameter d_0 is not a so-called "halfpoint"[288], i.e., $d_0 \neq p - 1/2$, for $p = -1, 0, 1, 2, \ldots$[289] In contrast to the ELW, the length of Θ is not restricted when using the FELW. Consequently, the FELW allows to cover a wide range of admissible values. However, the exclusion of the half-points again requires prior knowledge of the true parameter value, which is generally not available a priori.

Abadir, Distaso, et al. (2007) carried out a Monte Carlo simulation and proposed a value for the bandwidth parameter of $m = n^{0.65}$ for large ($n \geq 500$) sample sizes.[290] This value

[282] See Shimotsu and P. C. B. Phillips (2005, Theorem 2.2 on p. 1897).
[283] See Shimotsu and P. C. B. Phillips (2005, p. 1894).
[284] See Shimotsu and P. C. B. Phillips (2005, Section 3 on pp. 1897ff.). Especially, the ELW outperforms the competing estimators in terms of a significantly lower mean squared error for true parameter values smaller than $-1/2$ and greater than 1, see Shimotsu and P. C. B. Phillips (2005, Tables 1 and 2 on p. 1898).
[285] See Shimotsu and P. C. B. Phillips (2005, Remark 2 on p. 1895).
[286] The range of admissible values is given by $\Theta = (-1/2, 2) \setminus ((-\epsilon, \epsilon) \cup (1 - \epsilon, \epsilon))$ for arbitrarily small values of ϵ, see Shimotsu (2010, Assumption 6c, Theorem 5a and 5b on p. 514).
[287] See Abadir, Distaso, et al. (2007, Section 2.1 on pp. 1355ff.) for details.
[288] Abadir, Distaso, et al. (2007, p. 1366).
[289] See Abadir, Distaso, et al. (2007, Theorem 2.4 and Corollary 2.1 on pp. 1360f.).
[290] See Abadir, Distaso, et al. (2007, p. 1363).

seems to be acceptable for small samples ($n = 125$), too, as it overall produces a smaller mean square error than the value of $m = n^{0.5}$.[291] They further give a warning not to choose the bandwidth parameter too high. More specifically, they do not recommend a value of $m = n^{0.8}$ for the estimation of d.[292] Overall, a suitable region for the bandwidth parameter seems to be $[n^{0.65}, n^{0.8})$. However, this region is again incompatible with the economic perspective for small sample sizes stated above. This problem arises for the FELW, all semiparametric Whittle-based estimators, and the GPH estimator. Recently, Cheung and Hassler (2020) revealed another weakness of the FELW estimator. A Monte Carlo simulation showed that the objective function used by Abadir, Distaso, et al. (2007) is not continuous at the half-point $d_0 = 1/2$, i.e., the estimator is not reliable if the true parameter is near 1/2. Neither are confidence intervals constructed from the estimator reliable if the true parameter is between 1/4 and 3/4.[293,294] The value $d = 1/2$ is of particular interest from an econometric perspective, as it marks the edge point between a stationary process with a high degree of long memory and a non-stationary process.[295] Overall, the authors found tests based on the local Whittle estimator and the two-step ELW not reliable if the true parameter is in a neighborhood of 1/2.[296]

In summary, both estimators, the GPH estimator, and the semiparametric Whittle estimator, with their various extensions and modifications, have significant drawbacks. The undesirable effects of the choice of the bandwidth parameter accompany the advantage of neglecting the short-range influences in the estimation. The choice of the bandwidth parameter seems hardly consistent with the economic perspective of which cycle lengths contribute to long-range effects in time series, especially in short ones common in macroeconomic contexts. Furthermore, not taking the short-range influences into account is a double-edged sword. On the one hand, it simplifies the estimation, but on the other hand, it introduces a substantial bias in the estimated long memory parameter if (substantial) short memory components are present. One possible solution to this problem is to specify

291 See the column "MSE" of Tables 1 and 2 of Abadir, Distaso, et al. (2007, pp. 1362f.).

292 See Abadir, Distaso, et al. (2007, p. 1364).

293 See Cheung and Hassler (2020, pp. 371f. and p.376).

294 A remedy for this weakness of the FELW is provided by Cheung (2020). There, the author proposes to shift the discontinuity of the objective function by fractional differencing the time series away from the minimizer of the objective function, see Cheung (2020, p. 3). Doing so has a major effect on the value of the estimator. As illustrated by Cheung (2020), the point estimators of the five-year treasury yield spread over the federal funds rate and the Chicago Board Options Exchange Standard & Poor's 100 volatility index increase from 0.4999 to 0.6812 and 0.7118, respectively, if the discontinuity is taken into account, see Cheung (2020, p. 3).

295 See Section 2.4.

296 See Cheung and Hassler (2020, p. 381). Further, tests based on the tapered estimator mentioned in Footnote 279 seems to be more reliable in larger sample sizes ($n > 200$), see again Cheung and Hassler (2020, p. 381).

and estimate the process's entire short and long-range components, i.e., to estimate a complete parametric model. Such estimators are called parametric estimators; two of them are discussed in the following section.

3.1.3 Parametric Maximum-Likelihood and Whittle Estimators

A commonly used estimation technique in statistics is maximum likelihood (ML) estimation, which can also be applied in the context of long memory processes. Doing so implies the assumption of a specific parametric model behind the data-generating process. This section discusses the properties of two parametric (ML-based) estimators along with the estimation of an ARFIMA(p, d, q) process. Let $X_1, \ldots, X_n$ be values of a stationary Gaussian time series with zero mean.[297] The target of ML estimation is to estimate an ARFIMA(p, d, q) model with given values p and q, i.e., the order of the autoregressive and moving average polynomial, respectively.[298] Overall, it follows that $p+q+2$ parameters have to be estimated: the parameters of the autoregressive and moving average polynomials $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, the order of fractional integration d and the variance of the innovations σ_ε^2. To simplify notation, the $(p + q + 1)$-dimensional vector ϑ is defined $\vartheta := (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)$. Due to the assumed Gaussianity, the vector $X = (X_1, \ldots, X_n)$ has a common probability density given by[299]

$$p_{X,n}\left(\vartheta, \sigma_\varepsilon^2\right) = (2\pi)^{-n/2}\det\left(\Sigma_X\left(\vartheta, \sigma_\varepsilon^2\right)\right)^{-1/2}\exp\left\{-\frac{1}{2}X^T\Sigma_X\left(\vartheta, \sigma_\varepsilon^2\right)^{-1}X\right\}, \tag{3.8}$$

where $\Sigma_X\left(\vartheta, \sigma_\varepsilon^2\right)$ refers to the covariance matrix depending on the parameters of the assumed model. The ML estimator, denoted by $\hat{\vartheta}_{ML}$ and $\hat{\sigma}_{\varepsilon\,ML}^2$, is the parameter vector that maximizes (3.8). Equivalently, one often minimizes the negative logarithm of (3.8) instead, i.e.,

$$\left(\hat{\vartheta}_{ML}, \hat{\sigma}_{\varepsilon\,ML}^2\right) := \underset{\vartheta, \sigma_\varepsilon^2 \in \Theta}{\arg\min}\, \mathscr{L}_n\left(\vartheta, \sigma_\varepsilon^2\right),$$

where $\mathscr{L}_n\left(\vartheta, \sigma_\varepsilon^2\right) = -\log\left(p_{X,n}\left(\vartheta, \sigma_\varepsilon^2\right)\right)$ is the negative log-likelihood function and the minimum is taken over a certain parameters space $\Theta \subset \mathbb{R}^{p+q+2}$.[300] As for the local Whittle-based estimators mentioned in the previous section, this optimization exercise is carried out numerically. Further, the ML estimator has desirable properties. Given the

[297] The values $X_1, \ldots, X_n$ may be seen as the data, and a possible existing mean can be removed by subtracting the sample mean from each observation.

[298] In practice, the values p and q are unknown a priori. Below, commonly used information criteria are discussed for an objective choice of p and q.

[299] See Beran et al. (2013, p. 417).

[300] One can further drop the constant term $n/2\log(2\pi)$ in $-\log\left(p_{X,n}\left(\vartheta, \sigma_\varepsilon^2\right)\right)$, see Beran et al. (2013, p. 417).

stationarity, Gaussianity, and additional regularity conditions, the estimator converges almost surely to the true parameter value and is asymptotically normally distributed.[301,302]

The ML estimator has been shown to be superior to other semi-parametric estimators, including the GPH and the local Whittle estimator discussed in Section 3.1.2.[303] Major drawbacks of the ML estimator are the restriction of the parameter range of d to the interval $(-1/2, 1/2)$, i.e., in contrast to already discussed local Whittle-based estimators, it covers not the non-stationarity region with $d > 1/2$. Additionally, the demanding computational calculation of the covariance matrix and its inverse in (3.8) and the possible misspecification of the assumed model appear to be rather disadvantageous. Concerning the covariance matrix, (A.10) in Appendix A.3 may underline that the autocovariance function of a quite simple ARFIMA$(1, d, 0)$ process is a rather complex expression. Thus, it is easily imaginable that this is true for higher-order ARFIMA processes as well.[304]

The problem of misspecification arises when the parameters of the "wrong" model are estimated. For example, assume that the data is generated from an ARFIMA$(1, d, 1)$ process, but the ML estimator is applied to estimate an ARFIMA$(1, d, 0)$ process. The obtained estimators of the ARFIMA$(1, d, 0)$ process may be substantially biased compared to the true parameters of the underlying ARFIMA$(1, d, 1)$ process. In order to avoid misspecification of the model in practice, the ML estimator is applied to the data for various combinations of p and q. Then, a score value or information criteria can be calculated for each model, which allows picking out the "best" model according to this value. Two frequently used criteria are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) (or Schwarz information criteria). Such a criterion is needed since, in general, one would expect that a model with more parameters outperforms

[301] This result was proven for the case of $0 < d < 1/2$ by Dahlhaus (1989, Theorem 3.2 on p. 1756) and for the case of $-1/2 < d < 1/2$ by Lieberman et al. (2012, Theorem 1 on pp. 461f.). Note that the results of Lieberman et al. (2012) are proved under more general assumptions. For ARFIMA(p, d, q) processes, Assumption 5 of Lieberman et al. (2012, p. 460) holds only for the case of $-1/2 < d < 1/2$ if the data is demeaned with the sample average. This was stressed by Hassler (2019, p. 152). Details on the regularity conditions can be found in Dahlhaus (1989, pp. 1750f.), Dahlhaus (2006, p. 1045) and Lieberman et al. (2012, pp. 459ff.).

[302] Recall Footnote 212 for a definition of the term "almost surely".

[303] J. Smith et al. (1997, pp. 510ff.) carried out a Monte Carlo simulation and found the ML estimator superior to the GPH estimator for various ARFIMA specifications and under the condition of a known and unknown mean. Baillie and Kapetanios (2016, pp. 371f.) shows the superiority of the ML to the local Whittle estimator.

[304] Sowell (1992a, Section 4 on pp. 173ff.) provided closed-form expressions of the autocovariance function of an ARFIMA(p, d, q) process for his ML estimatior. Hassler (2019, p.110), however, deemed them erroneous as they failed to replicate the formulas for an ARFIMA$(0, d, 1)$ process. Whether this criticism is justified will not be examined in the context of this thesis. Closed-form expressions of moments of ARFIMA processes can additionally be found in Palma (2007, pp. 47f.).

a model with fewer parameters.[305] To be more precise, it is expected that the ML estimator of the innovations' variance ($\hat{\sigma}^2_{\varepsilon ML}$) tends to be smaller the more parameters (degrees of freedom) are available in the model to be estimated.[306] Consequently, one can think of information criteria as introducing punishment terms for models containing too many parameters.[307] Consider, for example, the AIC and BIC, which are defined by[308,309]

$$\text{AIC} := n \log\left(\hat{\sigma}^2_{\varepsilon ML}\right) + 2(p+q+1) \text{ and } \text{BIC} := n \log\left(\hat{\sigma}^2_{\varepsilon ML}\right) + (p+q+1)\log(n),$$

where p and q refer to the orders of the corresponding autoregressive and moving average polynomials, respectively. Thus, model identification is done by estimating an ARFIMA(p, d, q) model for different values of p and q with their corresponding information criteria. The best model from this battery of models is the one with the smallest information criterion.[310] A disadvantage of this procedure is that both criteria may identify different models to be "best".[311] By carrying out a Monte Carlo simulation Bisaglia (2002) founds that in the context of ARFIMA models, the BIC criterion works better in identifying the correct model than the AIC.[312]

Nevertheless, the disadvantageous properties of the ML estimator, such as not covering the non-stationarity region and the difficulties in calculating the covariance matrix $\Sigma_X(\vartheta, \sigma^2_\varepsilon)$ and its inverse matrix in (3.8) remain.[313] Therefore, approximations to (3.8) are proposed in the literature. Such an estimator is the so-called (parametric) Whittle estimator.[314] The idea behind the Whittle estimator is to switch from the time domain of a process expressed in terms of covariances to the spectral domain expressed in terms of the spectral density. The local Whittle estimator discussed in Section 3.1.2 uses only the asymptotic property of the spectral density (2.28) for the m lowest frequencies. Instead, the objective function to be minimized by the parametric Whittle estimator includes the entire spectral density of a fully specified parametric model, thus allowing the simultaneous estimation of any short memory dynamics that may be present.[315] The parametric Whittle estimator

305 See Brockwell and Davis (1987, p. 280).
306 See Brockwell and Davis (1987, p. 280).
307 See Brockwell and Davis (1987, p. 280).
308 See Bisaglia (2002, pp. 40f.).
309 There are different formulations of the AIC and BIC in the literature. For example, Silverberg and Verspagen (2003, Footnote 1 on p. 273) use a formulation regarding the likelihood function.
310 See Bisaglia (2002, p. 40).
311 Examples can be found in the empirical review given in the next section.
312 See Bisaglia (2002, pp. 42f.). This result seems to be in line with a vast literature dealing with model selection criteria in the presence of long memory; a brief literature review on this topic can be found in Bisaglia (2002, p. 34).
313 Hassler (2019, pp. 153f.) lists additional rather disadvantageous properties of the ML estimator.
314 See Beran (1994, p. 109) and Palma (2007, pp. 78f.).
315 The details of the objective function are omitted here. Several variants of Whittle approximations to

shares the nice asymptotic and distributional properties of the ML estimator but can be extended easily to the non-stationary region $d > 1/2$.[316] Thus, the Whittle estimator remedies the major disadvantages of the ML estimator.[317]

3.2 Empirical Findings

3.2.1 Output, GDP and Related Time Series

There has been an intense debate in the macroeconomic literature as to whether macroeconomic time series are best described as stationary time series varying around a deterministic trend or as a unit-root ($\mathcal{I}(1)$) process. Fractionally integrated ($\mathcal{I}(d)$) processes allow for a more nuanced view of this issue. By estimating the fractional order of integration, one can further draw insights into which model may be appropriate for an economic time series. More specifically, let X_t be the logarithm of a quarterly time series of the GDP and $Y_t = (1 - B)X_t = X_t - X_{t-1}$ the corresponding first-order difference process. Note that Y can be regarded as the GDP growth rate series.[318] By estimating the order d_Y of integration of Y, one can draw some insights about the properties of X: If d_Y is about zero this would indicate that X is well-described by a unit-root or $\mathcal{I}(1)$ process.[319] Contrary, if d_Y is near -1, X_t is rather a trend-stationary process.[320] In addition, if the order of integration of X, d_X is between 1 and 1.5, the process Y is likely to have long memory, since $d_Y = d_X - 1$. In such a case, Y would be an $\mathcal{I}(d)$ process with $0 < d < 1/2$.

The empirical investigations in the literature are heterogeneous regarding the employed methods and investigated time series. Therefore, the review is not sorted chronologically but according to the considered time series. From the perspective of a dynamic stochastic general equilibrium (DSGE) model, which is usually estimated with quarterly data, it seems reasonable to start considering the results of empirical investigations based on quarterly data sets. There is also a vast literature investigating annual data, so these will be considered afterward.

Diebold and Rudebusch (1989) focus on data of the United States (of America) (US) and

the likelihood function can be found in ,e.g., Palma (2007, p. 80).

[316] See Palma (2007, Theorem 4.5 on p. 80) for the asymptotic properties of the Whittle estimator and Hassler (2019, p. 157ff.) for a discussion of various extensions of the Whittle estimator and admissible values of d.

[317] See Hassler (2019, p. 160).

[318] Let GDP_t be GDP at time t. Then, by definition of X and Y, one has $X_t = \log(\text{GDP}_t)$ and $Y_t = \log(\text{GDP}_t) - \log(\text{GDP}_{t-1}) = \log(\text{GDP}_t/\text{GDP}_{t-1}) \approx (\text{GDP}_t - \text{GDP}_{t-1})/\text{GDP}_{t-1}$.

[319] See Sowell (1992b, p. 282).

[320] See Sowell (1992b, p. 282).

analyze the quarterly series of the logarithm of the gross national product (GNP) and the corresponding per capita (p.c.) series. Both series are seasonally adjusted and cover the postwar period 1947Q1-1987Q2.[321] Diebold and Rudebusch (1989) use a two-step approach. In the first step, they estimate the parameter d by the methodology of Geweke-Porter-Hudak. For the per capita series, they report an estimator of $d \approx 0.7$.[322] The exact value varies with the number of ordinates incorporated (bandwidth parameter) in the spectral regression of the GPH estimator. If the GPH regression contains $m = n^{0.5}$ ordinates, where n is the total sample size, the GPH estimator is given as $d = 0.68$. The corresponding 95% confidence interval can be calculated from the standard errors reported in Diebold and Rudebusch (1989) and is given as $(0.21, 1.15)$.[323,324] For the per capita series, after estimating d to be about 0.7, the authors take first differences of the series and then apply the operator $(1 - B)^{-0.3}$ in order to obtain an $\mathcal{I}(0)$ series. Then, they specify the short memory dynamics by fitting an autoregressive moving average (ARMA) process to this series. Based on the BIC, they identify an ARFIMA$(1, 0.7, 0)$ model as a parsimonious model among various ARMA specifications.[325] However, as indicated by the authors, the confidence intervals are rather long, and thus, the alternative of a unit root cannot be rejected. Diebold and Rudebusch (1989) also investigated the level series of real GNP. As with the per capita series, the estimated value of d depends on the number of ordinates in the GPH regression. Overall, they found $d \approx 0.9$.[326] By including $n^{0.5}$ ordinates, the estimator of d is exactly 0.9, and the corresponding 95% confidence interval is given as $(0.43, 1.37)$, implying that a unit-root in the GNP series cannot be rejected at the 5% significance level.[327] However, the trend-stationary model for the GNP series lies outside this interval.

Sowell (1992b) uses maximum likelihood estimators to specify various ARFIMA(p, d, q) processes for the same GNP series as in Diebold and Rudebusch (1989) (but extended to 1987Q4). He founds that the differenced series of the log of quarterly seasonally adjusted GNP is best described by an ARFIMA$(3, d, 2)$ process with $d = -0.59$.[328] The provided asymptotic standard error implies a 95% confidence interval of $(-1.27, 0.09)$. Note that Sowell considers first differences; hence the value of the corresponding level series (as

[321] The expression 1947Q1 refers to the first quarter of 1947. This notation is used similarly for other years throughout this thesis.
[322] See Diebold and Rudebusch (1989, Table 2 on p. 200 and p. 201).
[323] See Diebold and Rudebusch (1989, Table 2 on p. 200 and p. 201).
[324] The confidence interval is given by $\hat{d} \pm z_{1-\alpha/2}$ times the standard error, where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution which is 1.96 if $\alpha = 0.05$.
[325] See Diebold and Rudebusch (1989, Table 3 on p. 203 and p. 203).
[326] See Diebold and Rudebusch (1989, Table 2 on p. 200).
[327] See Diebold and Rudebusch (1989, Table 2 on p. 200 and p. 201).
[328] See Sowell (1992b, p. 287).

considered in Diebold and Rudebusch (1989)) is 0.41 with the shifted confidence interval $(-0.27, 1.09)$. This point estimator is smaller than the one reported by Diebold and Rudebusch (1989). In contrast to Diebold and Rudebusch (1989), the results of Sowell (1992b) indicate that neither the alternative of a trend-stationary nor a unit-root process can be rejected from this analysis.[329] Sowell (1992b) points out that these differences between his results and the ones of Diebold and Rudebusch (1989) are due to the bad performance of GPH-estimator in the presence of short memory terms.[330] Further, the chosen number of periodogram ordinates in the analysis of Diebold and Rudebusch (1989) seems restrictive. This choice ruled out short memory effects for cycles longer than 3.1 years.[331] Further, the results obtained by Sowell (1992b) are roughly confirmed from a Bayesian estimation approach of Koop et al. (1997).[332]

The main difference between the maximum likelihood estimator employed by Sowell (1992b), in contrast to the approach followed by Diebold and Rudebusch (1989), is that the former estimates short and long memory parameters simultaneously. In the stepwise approach followed by Diebold and Rudebusch (1989), they first estimate the long memory parameter, which is likely to be biased due to the restrictive choice of ordinates in the regression and the overall sensitivity of the GPH estimator to short memory components. After determining the long memory parameter, they estimate the short memory coefficients based on the filtered time series. Another difficulty of both investigations is the relatively small sample size which implies rather large confidence intervals. If the confidence interval spreads over the range $[0, 1]$ as reported by Sowell (1992b) above, clear support for fractional integration and, in the end, for long memory cannot be drawn.

By considering the logarithm of quarterly seasonally adjusted GDP data for the US and United Kingdom (UK) ranging from 1961Q1 till 2000Q1, Candelon and Gil-Alaña (2004) carried out a similar investigation as Sowell (1992b). By applying a maximum likelihood estimation, they found an ARFIMA$(0, 1.36, 0)$ model for the US GDP series and an ARFIMA$(1, 1.38, 2)$ model for the UK series to best fit the data according to the BIC.[333] They also carried out various integration tests and found that the alternatives of $d = 1$ and $d = 2$ could be rejected.[334] Note that these results would indicate that the quarterly GDP growth rates for both countries would exhibit long memory with $d = 0.36$ and $d = 0.38$ for the US series and for the UK series, respectively. An interesting part of

329 See Sowell (1992b, pp. 292ff.).
330 See Sowell (1992b, pp. 299f.).
331 See Sowell (1992b, p. 298).
332 See Koop et al. (1997, p. 160).
333 See Candelon and Gil-Alaña (2004, p. 354).
334 See Candelon and Gil-Alaña (2004, Table 6 and p. 355).

their analysis is comparing the ability to replicate specific business cycle characteristics of various time series models. They conclude that, in general, ARFIMA processes perform better in replicating business cycle characteristics such as the number of peaks, lengths, and amplitudes of recessions and expansions than their ARIMA counterparts.[335]

More recently, a similar investigation was carried out by Caporale and Gil-Alaña (2013). By investigating quarterly US GDP per capita series ranging from 1948Q1 till 2008Q3 with various semi-parametric and parametric estimation procedures, the authors confirm the already mentioned dependence of the results on the chosen estimation procedure.[336] The non-parametric methods employed by Caporale and Gil-Alaña (2013) induce no evidence of fractional integration in the level series and growth rates.[337] Further, the results indicated by the semi-parametric local Whittle estimator depend to a large extent on the bandwidth parameter that controls the number of frequencies of the periodogram used in the estimation.[338] In addition, the authors also consider the possibility of a linear trend in the GDP per capita and the corresponding growth rate series. By applying a parametric estimation method, they found that the detrended GDP per capita series is best described by an ARFIMA$(1, d, 0)$ process with parameters $d = 0.312$ (the 95%-confidence interval is $(0.196, 0.489)$) and $\varrho = 0.913$.[339] These results indicate that the detrended series is well-described by a long memory process with a strong short memory component. For the US GDP per capita growth rate series, they propose an ARFIMA$(1, d, 0)$ model with parameters $d = -0.622$ and $\varrho = 0.887$.[340] They also identified a structural break in both series at 1978Q2.[341] By repeating their model estimation procedure to the resulting two subsamples (one before and one after the break), they found the long memory parameter d of the GDP per capita series to be 0.49 in the first subsample and $d = 0.314$ in the second one.[342] The corresponding short memory parameters ϱ are 0.774 and 0.875, respectively.[343,344] For the GDP per capita growth series, the parameters in the first subsample are $d = -0.416$ and $\varrho = 0.742$, and for the second subsample $d = -0.666$ and $\varrho = 0.876$.[345]

[335] See Candelon and Gil-Alaña (2004, Section 5 on pp. 356ff.).
[336] See Caporale and Gil-Alaña (2013, p. 608).
[337] See Caporale and Gil-Alaña (2013, Table 1 on p. 599 and p. 600).
[338] See Caporale and Gil-Alaña (2013, Figure 2 and pp. 600ff.).
[339] See Caporale and Gil-Alaña (2013, Table 2 on p. 603 and pp. 603f.).
[340] See Caporale and Gil-Alaña (2013, Table 2 on p. 603 and p. 604).
[341] See Caporale and Gil-Alaña (2013, p. 605).
[342] See Caporale and Gil-Alaña (2013, Table 3 on p. 606 and pp. 605ff.).
[343] See Caporale and Gil-Alaña (2013, Table 3 on p. 606 and pp. 605f.).
[344] Note that the corresponding impulse-response function (IRF) would be expected to be hump-shaped according to Lemma 2.4.3 since $\varrho + d > 1$. This is confirmed by Figure 4 in Caporale and Gil-Alaña (2013).
[345] See Caporale and Gil-Alaña (2013, Table 3 on p. 606 and p. 606).

Table 3.1 summarizes the mentioned empirical results. The estimated values of d depend on the used estimation method. As outlined in Section 3.1.2, the sensitivity of the GPH estimator regarding the incorporated regression ordinates seems problematic. Thus, the parametric approaches used by the other authors seem to be more appropriate. However, there is another lack in the comparability of the results as the considered time series are inconsistent in all studies. Some authors focus on GNP while others analyze GDP, and in addition, the time periods considered differ in almost all studies.

However, not all estimates of d given in Table 3.1 can be used directly to calibrate a linearized DSGE model. Typically, the model variables represent percentage deviations from their steady state values or a balanced growth path. Thus, non-stationary time series have to be made stationary first.[346] This may be done by removing a (linear) trend or seasonal components.[347] Additionally, for a representative agent model such as the one of Chapter 4, a per capita series seems appropriate.[348]

Putting all these aspects together, the results of Caporale and Gil-Alaña (2013) seem best suited in the context of a representative agent DSGE model to describe the evolution of GDP. Therefore, the cyclical component of US GDP per capita that remains after removing a linear trend may be well-described by a long memory process or, more precisely, by an ARFIMA$(1, d, 0)$ process with parameters $\varrho \approx 0.9$ and $d \in (0.2, 0.49)$.[349]

[346] See Section 4.1.4 for details.
[347] See, e.g., Brockwell and Davis (1987, p. 15).
[348] See, e.g., Cooley and Prescott (1995, p. 20) or Lindé et al. (2016, p. 2256).
[349] See Caporale and Gil-Alaña (2013, Table 2 on p. 603 and p. 604).

Authors	Time Series	Estimated value of d	proposed ARFIMA model	Estimator used
Diebold and Rudebusch (1989)	log of US GNP s.a. 1947Q1-1987Q2	≈ 0.9	-	GPH for d, ML for ARMA parts
	log of US GNP p.c. s.a. 1947Q1-1987Q2	≈ 0.7	ARFIMA$(1, 0.7.0)$ or ARFIMA$(1, 0.7, 2)$	
Sowell (1992b)	$(1 - B)$ log of US GNP s.a. 1947Q1-1987Q4	-0.59 (-1.27,0.09)	ARFIMA$(3, -0.59, 2)$	ML
Candelon and Gil-Alaña (2004)	log of US GDP s.a. 1961Q1-2000Q1	1.36	ARFIMA$(0, 1.36, 0)$	ML
	log of UK GDP s.a. 1961Q1-2000Q1	1.38	ARFIMA$(1, 1.38, 2)$	
Caporale and Gil-Alaña (2013)	US GDP p.c. linearly detrended 1948Q1-2008Q3	0.312 $(0.196, 0.489)$	ARFIMA$(1, 0.312, 0)$	parametric approach[350]
	$(1 - B)$ log of US GDP p.c. demeaned, 1948Q1-2008Q3	-0.622 $(-0.812, -0.409)$	ARFIMA$(1, -0.622, 0)$	

Table 3.1: Summary of the empirical studies on quarterly GDP and related time series. All variables are in real terms, p.c. is per capita, and s.a. is seasonally adjusted. Values in parenthesis below the estimates of d indicate 95% confidence intervals. The factor $(1 - B)$ indicates that the first difference of the series was considered.

[350] The authors use a parametric approach based on a test procedure introduced in Robinson (1994). Following Caporale and Gil-Alaña (2013, p. 603), the estimated model should be similar to the one carried out by a maximum likelihood estimation.

The empirical investigations so far have been concentrated on quarterly GDP data and related variables in the post-war period. However, a large strand of the empirical literature also focuses on annual data starting in the 19th century. This literature, governing a rather cliometric perspective, may be less attractive from a quarterly DSGE model perspective. However, it seems reasonable to consider these long-spanning time series as long memory itself is an asymptotic property that would be expected to materialize in the long run. On the other hand, it should be noted that the length of the time series that enters the estimation (which is one limiting factor when calculating, e.g., confidence intervals) need not be longer than their quarterly counterparts due to the lower granularity of the data.

By investigating annual series of GDP per capita growth rates for various countries from 1870 to 2001, Silverberg and Verspagen (2003) found ambiguous results regarding the presence of long memory. By applying a fractional Gaussian noise test, they found strong evidence of long memory for more than half of the considered countries.[351] Contrarily, by applying the maximum likelihood estimator to the data, they found no evidence of long memory for any country.[352] For the US, the best ARFIMA model according to the AIC is given by an ARFIMA$(2, -0.707, 4)$ process.[353] This model would be equivalent to an ARFIMA$(2, 0.293, 4)$ model for the log of GDP per capita series and would induce long memory, at least in the level series, but not in the growth rates.[354] These results are roughly confirmed by the investigation of Gil-Alaña (2020). By analyzing an annual data set ranging from 1870 to 2000 containing the same countries as in Silverberg and Verspagen (2003), Gil-Alaña (2020) found that, for the overwhelming number of countries, the unit-root hypothesis of the log of GDP per capita cannot be rejected.[355] Except for Germany and the US, for which Silverberg and Verspagen (2003) reported values of d significantly different from zero (for the growth rate series), the results are not consistent with those reported by Gil-Alaña (2020). This inconsistency might be due to the different estimation techniques employed by the authors.

The study of Silverberg and Verspagen (2003) and the others mentioned earlier illustrate that the determination of long memory depends crucially on the estimation methods applied.

351 See Silverberg and Verspagen (2003, Table 1 on p. 278).
352 See Silverberg and Verspagen (2003, Table 3 on p. 281 and pp. 280f.).
353 See Silverberg and Verspagen (2003, Table 3 on p. 281).
354 They further investigated various fractionally integrated autoregressive models, i.e., ruling out any moving average terms. They found the best models for the log of US GDP per capita series according to the BIC and AIC to be an ARFIMA$(1, 0.466, 0)$ and ARFIMA$(2, 0.061, 0)$, respectively; see again Silverberg and Verspagen (2003, Table 3 on p. 281).
355 See Gil-Alaña (2020, Table 3 on pp. 86f.). For better comparability between Silverberg and Verspagen (2003) who do not consider the possibility of a trend in the data. The first column of Table 3 in Gil-Alaña (2020, pp. 86f.) should be used.

Consequently, various methods applied to the same series may produce contradictory results. Another problem outlined by Silverberg and Verspagen (2003) is the length of the available data sets in macroeconomic contexts. Their time series contain 133 observations, which is quite short compared to data available in natural sciences. Simultaneously, these time series cover over 100 years, which might be too long for deriving economic insights from stationary time series models.[356] In an earlier study, Silverberg and Verspagen (1999) found (based on the fractional Gaussian noise test) evidence of long memory in the post-war period but weaker evidence of long memory in the pre-war period.[357]

A data set of similar length as the one of Silverberg and Verspagen (2003) and Gil-Alaña (2020) is analyzed by Mayoral (2006). By analyzing annual data of US GNP and GNP per capita (both in logs) dating from 1869 to 2001, she found evidence that both time series are neither trend-stationary nor difference stationary.[358] The maximum likelihood estimator of d ranges between 0.65 and 0.68 for the GNP series and between 0.63 to 0.65 for the GNP per capita series.[359] Compared to the investigations of Silverberg and Verspagen (2003) whose investigations imply a value of $d = 0.293$ for the log of US GDP per capita, the estimator of Mayoral (2006) is larger. This difference may be due to the differences between the GDP and GNP series and the quality of the considered data set, especially for the pre-war periods. The corresponding confidence intervals for the maximum likelihood estimator at the 95%-level provided by Mayoral (2006) lie fully in the interval $(0, 1)$, supporting that GDP and GNP are not unit-root processes.[360] In addition, most point estimators of the employed estimation routines by Mayoral (2006) lie between 1/2 and 1 for both the per capita and the level series.[361] These results are in line with the one derived in Michelacci and Zaffaroni (2000) who analyzed GDP per capita series for various OECD countries.[362] They found values of d between 1/2 and 1 to be typical over their whole set of countries. This observation is in contrast to Mayoral (2006), who only analyzed the US series. However, Silverberg and Verspagen (2000) criticized the results of Michelacci and Zaffaroni (2000) to be an artifact of a non reasonable data filtering.[363]

[356] See Silverberg and Verspagen (2003, p. 283).
[357] See Silverberg and Verspagen (1999, p. 10).
[358] This was carried out by applying various unit-root tests, see Mayoral (2006, Table 1 on p. 906).
[359] These results are in line with the one derived by Diebold and Rudebusch (1989, Table 2 on p. 200) who analyze the same annual data set ranging from 1869 to 1987, in addition to the quarterly data set already mentioned.
[360] She further carried out various tests in order to distinguish between structural breaks in the data and fractional integration. Overall, she concludes that fractional integration models seem preferable to models with structural breaks, see Mayoral (2006, Section 5.1 on pp. 914ff.).
[361] See Mayoral (2006, Table 3 on p. 910).
[362] See Michelacci and Zaffaroni (2000, Table 2 on p. 147).
[363] See Silverberg and Verspagen (2000, pp. 1ff. and p. 6).

Further, they state that the GPH estimation scheme employed by Michelacci and Zaffaroni (2000) is sensitive to the number of ordinates included in the regression and to the presence of short memory components.[364] This criticism is not justified against the methods used by Mayoral (2006). However, since she does not consider countries other than the US, her results may not be generalizable to other countries.

Another investigation covering fourteen European countries was carried out by Caporale, Gil-Alaña, and Monge (2020). Their results may, however, not be directly comparable to the already discussed ones as the logarithms of the GDP per capita series considered by Caporale, Gil-Alaña, and Monge (2020) ranges from 1960 to 2016 only. A significant time trend is detected for the overwhelming part of the countries. However, the values of d depend crucially on the specified parametric model.[365] The estimated values of d obtained from the local Whittle estimator range from 0.959 to 1.242 among the countries, whereby the values of 12 countries lie above 1.[366] These results may indicate weak evidence of long memory in the growth rate series. Interestingly, they found that, among all estimated models, there is a significant negative correlation between the height of GDP per capita and the order of integration.[367] This implies that countries with a high GDP per capita are associated with lower orders of integration.[368] This is argued because the institutional weaknesses in less developed (European) countries reduce the resilience of their economy to shocks by preventing an adequate shock response. Hence, a shock's effects become more persistent in less developed countries.[369] However, this explanation may oversimplify the complexity of the notion of persistence mentioned in Chapter 2. In a similar investigation carried out by Gil-Alaña et al. (2015), the estimates of d for the log of GDP per capita of various African countries ranging from 1960 till 2006 turned out not to be systematically higher than for the European countries as one would expect from a systematic negative correlation between GDP per capita and the order of fractional integration.[370]

A study conducted by Diebolt and Guiraud (2000) examined the GDP growth rates of Germany and France. The data set for Germany was from 1820 to 1989, while France's

[364] See Silverberg and Verspagen (2000, p. 7).
[365] See Caporale, Gil-Alaña, and Monge (2020, Tables 1 to 4 on pp. 2ff.).
[366] See Caporale, Gil-Alaña, and Monge (2020, Table 5 on p. 4). The values given in the text are averages over several bandwidth parameters reported by the authors; see the column "AVG" in the authors' table.
[367] See Caporale, Gil-Alaña, and Monge (2020, p. 8).
[368] See Caporale, Gil-Alaña, and Monge (2020, p. 8).
[369] See Caporale, Gil-Alaña, and Monge (2020, p. 8).
[370] This can be seen from a comparison between Caporale, Gil-Alaña, and Monge (2020, Table 1 on p. 2 and Table 3 on p. 3) and Gil-Alaña et al. (2015, Table 1 on p. 227 and Table 3 on p. 229), respectively.

data set ranged from 1820 to 1996.[371] The study found that an ARFIMA(3, 0.45, 0) model was the best fit for Germany's growth rate series, and an ARFIMA(1, 0.27, 0) model was suitable for France.[372] These results suggest that both countries' GDP growth rates exhibit long memory. The GPH estimator was also employed, and it yielded the same results for France.[373] However, the GPH estimator reported for Germany was smaller (0.21) than the maximum likelihood estimation.[374]

For the United Kingdom, Caporale and Škare (2014) investigated a time series spanning from 1851 to 2013. Based on a maximum likelihood estimation of various ARFIMA models, they found the order of fractional integration d to vary between 0.24 and 1.09 for the real GDP growth rates.[375] The exact values depend on the specification of the ARMA terms, which turned out to be insignificant in all models.[376]

3.2.2 Other Macroeconomic Indicators and Financial Time Series

Besides GDP and the corresponding growth rate series, extensive empirical literature deals with long memory in other macroeconomic and financial time series.[377] This section aims to provide a summary of a part of the existing literature to illustrate that long memory is, indeed, relevant in various economic contexts. More specifically, this section deals with long memory in unemployment rates, inflation rates, and financial time series such as stock market returns and volatility. In these three contexts, long memory is associated with hysteresis, inflation targeting, and the efficiency of financial markets, respectively. The details are considered in the following sections.

3.2.2.1 Unemployment Rates

When analyzing unemployment rates, long memory models are useful for discriminating between different unemployment theories. Following the natural rate theory, one would expect that the unemployment rate fluctuates stationary around a natural rate.[378] In contrast, the hysteresis hypothesis states that the natural rate is not constant but rather time and path dependent.[379] Therefore, past values of the unemployment rate keep

[371] See Diebolt and Guiraud (2000, p. 12).
[372] See Diebolt and Guiraud (2000, Table 1 on p. 14).
[373] See Diebolt and Guiraud (2000, Table 1 on p. 14).
[374] See Diebolt and Guiraud (2000, Table 1 on p. 14).
[375] See Caporale and Škare (2014, Table 3 on p. 6 and p. 6).
[376] See Caporale and Škare (2014, Table 3 on p. 6 and p. 6).
[377] An early contribution reviewing various applications of long memory processes in macroeconomics and finance can be found in Baillie (1996, Section 6 on pp. 43ff.).
[378] See Caporale, Gil-Alaña, and Lovcha (2016, pp. 96f.).
[379] See Hassler and Wolters (2009, p. 119).

influencing the unemployment rate's current "steady state" value.[380] Therefore, shocks to the unemployment rate have a permanent or long-lasting effect given the hysteresis hypothesis, whereas shocks are assumed to die out quickly given the natural rate assumption.[381] Consequently, the natural rate hypothesis is associated with an unemployment rate described by an $\mathcal{I}(0)$ process whereas hysteresis assumes the unemployment rate to be a unit-root ($\mathcal{I}(1)$) process.[382] Caporale, Gil-Alaña, and Lovcha (2016) even go beyond this sharp distinction between $\mathcal{I}(0)$ and $\mathcal{I}(1)$ and find the hysteresis hypothesis already confirmed if the unemployment rate is an $\mathcal{I}(d)$ process with $d > 0$.[383]

Applying various tests and estimators to a monthly US and German unemployment series, Hassler and Wolters (2009) found evidence that the persistence in US unemployment rates is lower than in Germany.[384] In their setting, the $\mathcal{I}(1)$ hypothesis cannot be rejected for Germany, while for the US, it can be rejected at a 95% significance level.[385] Therefore, their results support the hysteresis hypothesis for Germany.

Caporale, Gil-Alaña, and Lovcha (2016) investigated quarterly unemployment series of the United Kingdom, the US, and Japan from 1971Q1-2011Q3 in a univariate and multivariate ARFIMA setting. In a univariate setting, they found that the assumption of a unit root in all employment rates cannot be rejected at a 5% significance level.[386] The point estimators lie, dependent on the employed estimation method, either above or close to 1, where the estimator of the US is lower than the other two.[387] Therefore, supporting the hysteresis hypothesis for all countries.

In addition, the authors employ a multivariate approach where the parameter d of the three time series is estimated jointly in a fractionally integrated vector autoregression.[388] They found that all parameters are smaller than in the univariate approach.[389] For the US series, the point estimator is 0.086 with corresponding 95% confidence interval $[-0.347, 0.519]$.[390] The point estimators with corresponding 95% confidence intervals for the UK and Japan are 0.615 ($[0.28, 0.95]$) and 0.568($[-0.034, 1.17]$), respectively.[391] Given these results, the

[380] See Hassler and Wolters (2009, p. 119).
[381] See Hassler and Wolters (2009, p. 120) and Caporale, Gil-Alaña, and Lovcha (2016, p. 97).
[382] See Hassler and Wolters (2009, p. 120).
[383] See Caporale, Gil-Alaña, and Lovcha (2016, p. 99).
[384] See Hassler and Wolters (2009, p. 127).
[385] See Hassler and Wolters (2009, p. 127).
[386] See Caporale, Gil-Alaña, and Lovcha (2016, p. 104).
[387] See Caporale, Gil-Alaña, and Lovcha (2016, Table 1 on p. 104).
[388] See Caporale, Gil-Alaña, and Lovcha (2016, pp. 14f.).
[389] See Caporale, Gil-Alaña, and Lovcha (2016, Table 1 and 2 on pp. 104f.).
[390] See Caporale, Gil-Alaña, and Lovcha (2016, Table 2 on p. 104). The confidence interval was calculated based on the reported standard errors.
[391] See Caporale, Gil-Alaña, and Lovcha (2016, Table 2 on p. 104).

authors claim that the natural rate hypothesis appears appropriate for the US, while the hysteresis hypothesis seems appropriate for the UK and Japan.[392]

In an earlier study, Gil-Alaña (2001) investigated quarterly unemployment rates of the US and four European countries from the 1960s/1970s to 1998. He used the maximum likelihood estimator to specify a full ARFIMA model for each series and found an ARFIMA$(3, 0.37, 3)$ model best suited for US unemployment rates.[393] Therefore, US unemployment rates would be well-described by a stationary long memory process. Consistent with the results already mentioned, he found the long memory parameter of the US unemployment rates to be lower than that of all other countries in his sample. More specifically, the parameters are 0.67, 0.95, 1.32 and 1.83 for Germany, Italy, France and UK, respectively.[394]

In all of the mentioned works, the parameter d of the US appears to be lower than that of the other countries in the respective study. This evidence may be interpreted in a way that the labor market is more flexible in the US than in other countries.[395] Overall, this short discussion highlights that fractional integration and ARFIMA processes are appropriate tools to uncover various dynamics in the unemployment rate. Additionally, they allow for a more differentiated classification of such series than simply deciding between the presence of a pure hysteresis ($\mathcal{I}(1)$) or a natural rate ($\mathcal{I}(0)$). However, in the context of unemployment rates, the non-stationary parameter range $d > 1/2$ also seems to be important.

3.2.2.2 Inflation Rates

Another frequently investigated macroeconomic indicator is inflation. Whether inflation rates are persistent is of major interest for monetary policy. If inflation is strongly persistent, it requires more time for inflation to adjust in response to a shock. Consequently, the monetary authority (e.g., a central bank) is forced to bring inflation back to its target level by a strong monetary policy.[396] If inflation is less persistent, shocks to inflation will die out quickly. Consequently, inflation will reach its target value quickly, making (strong) monetary policy interventions unnecessary.[397] Therefore, given a positive shock

[392] See Caporale, Gil-Alaña, and Lovcha (2016, p. 106).
[393] See Gil-Alaña (2001, Table 7 on p. 1268).
[394] See Gil-Alaña (2001, Table 7 on p. 1268). This table summarizes each country's best-suited ARFIMA(p, d, q) model.
[395] See Koustas and Veloce (1996, p. 827) for such reasoning in the context of a comparison between USA and Canada. Additionally, see Hassler and Wolters (2009, pp. 119f.).
[396] See Fuhrer (1995, p. 3) and Canarella and S. M. Miller (2017, p. 78).
[397] See Canarella and S. M. Miller (2017, p. 78).

to inflation, the economic costs of disinflation are expected to be higher if inflation is persistent.[398]

On the other hand, it is reasonable that monetary policy itself impacts inflation persistence.[399] More specifically, suppose there is a positive inflation shock to which the central bank responds with a policy intervention that quickly brings inflation back to its target level. Then, along with the policy response, the effect of the shock dissipates fast, inducing less persistence in inflation. Therefore, differences in inflation persistence may be caused by different monetary policy regimes, and an evaluation of inflation persistence allows for conclusions on the effectiveness of monetary policy.[400]

In recent years, the fractional order of integration d became a popular measure for inflation persistence. Early results of Hassler and Wolters (1995) indicate that monthly inflation rates spanning from 1969M1 to 1992M9 for Germany, the US, the UK, France, and Italy have a substantial degree of long memory as their point estimators based on the parametric Whittle estimator lie in the interval $[0.35, 0.5]$.[401]

Kumar and Okimoto (2007) considered monthly inflation rates of the US and the remaining G7 countries from 1960M4 to 2003M4.[402] They used the GPH, ELW, and FELW estimators for three subsamples 1960M5-1975M4, 1974M5-1989M4, and 1988M5-2003M4. The corresponding estimators for the US are $d \approx 0.49, 0.55, 0.27$, respectively.[403] The authors claim that US inflation was strongly persistent during 1960M5 and 1989M4, and that the level of persistence remained almost unchanged during this time.[404] However, the sharp drop in the parameter d in the subsample 1988M5 onward indicates a drop in inflation persistence since then.[405] Additionally, the authors compared their results with the outcomes of short memory persistence models (some of them are mentioned in Section 2.1.5.2) such as the largest autoregressive root or the sum of the autoregressive coefficients. They show that the largest autoregressive root is unable to detect the decline in US inflation persistence adequately.[406] Interestingly, the sum of the autoregressive coefficients shows a similar drop as the parameter d.[407] This illustrates that some persistence measures mentioned

398 See Fuhrer (1995, p. 3).
399 See Fuhrer (2010, p. 450) and Andersson and Li (2020, p.534).
400 See Andersson and Li (2020, p.534).
401 See Hassler and Wolters (1995, Table 7 on p. 42).
402 See Kumar and Okimoto (2007, p. 1461).
403 See Kumar and Okimoto (2007, Table 3 on p. 1468). The values in the text correspond to the means of the three estimators for each subsample.
404 See Kumar and Okimoto (2007, p. 1467).
405 See Kumar and Okimoto (2007, pp. 1467f.).
406 See Kumar and Okimoto (2007, Figure 5 on p. 1473 and p. 1473).
407 See Kumar and Okimoto (2007, Figure 5 on p. 1473 and p. 1473).

in Section 2.1.5.2 lead to contradictory results. Regarding their results for the remaining G7 countries, the authors showed that Germany, Japan, and Canada have lower inflation persistence in terms of the parameter d than the US over the period 1975-2002.[408] The corresponding long memory parameters lie fully in the interval $[0, 0.5]$ for the overwhelming part of the sample.[409] For France, a similar drop in d than for the US can be observed, while Italy's inflation is characterized by a parameter d fluctuating around 0.5.[410] For the UK, there is a slight decline in d from approximately 0.5 in 1978 to approximately 0.3 in 2002.[411]

Canarella and S. M. Miller (2017) investigated inflation persistence of thirteen OECD countries.[412] They focused on whether and how inflation persistence, as measured by the parameter d, has changed following a change in the monetary policy regime. More specifically, the authors address the question of whether an inflation targeting policy affects the parameter d. For their analysis, the authors use a modified version of the GPH estimator.[413] Consequently, the estimators vary strongly with the choice of the bandwidth parameter.[414] However, their results suggest that inflation persistence is lower (i.e., a lower value of d) in the post-inflation targeting period than in the pre-inflation targeting period for most countries in their sample.[415] These results may highlight that monetary policy impacts inflation persistence and that an inflation-targeting policy is associated with lower inflation persistence.

Andersson and Li (2020) go a step beyond this point of view and argue that the parameter d can be regarded as a measure of the inflation target's flexibility.[416] In the authors' opinion, monetary policy controls for the long run persistence.[417] Thus, if a central bank follows a strict inflation target, deviations of inflation from its target are brief and small, indicating a small value of d.[418] Contrary, a more flexible inflation target is associated with higher values of d since the central bank's monetary policy allows for longer deviations of inflation from its target.[419] Given an ARFIMA specification of inflation, Andersson and

408 See Kumar and Okimoto (2007, Figure 6 on p. 1475).
409 See Kumar and Okimoto (2007, Figure 6 on p. 1475).
410 See Kumar and Okimoto (2007, Figure 6 on p. 1475).
411 See Kumar and Okimoto (2007, Figure 6 on p. 1475).
412 A complete list of all countries can be found in Canarella and S. M. Miller (2017, Figure 1 on p. 81). However, the list of countries does not contain any country in the Euro area or the US.
413 See Canarella and S. M. Miller (2017, p. 85).
414 See Canarella and S. M. Miller (2017, Tables 6 and 7 on pp. 92f.).
415 See Canarella and S. M. Miller (2017, p. 93).
416 See Andersson and Li (2020, p. 534).
417 See Andersson and Li (2020, p. 534).
418 See Andersson and Li (2020, p. 534).
419 See Andersson and Li (2020, p. 534).

Li (2020) argue that the short memory parameters are unsuitable for an inflation target's flexibility since the central bank cannot fully control inflation in the medium and short run.[420] They propose a new Bayesian estimator of the parameter d and concentrate on the monthly inflation rates of Canada, the Euro area, Germany, Norway, Sweden, the UK, and the US from 1993 to 2017 (1999-2017 for the Euro area).[421] Interestingly, they report values of d ranging between 0.59 for Sweden and 1.09 for UK.[422] The remaining countries have an estimator between 0.74 and 0.83.[423] In the concept of Andersson and Li (2020), it can be argued that Sweden has the strictest inflation targeting policy among all countries in the sample, while the UK's inflation target is rather flexible.[424] These results contradict partly the findings of Kumar and Okimoto (2007) mentioned above who report values of d near or below to 0.5 for the US and the G7 countries.

Evidence of long memory in US inflation rates is also provided by Boubaker et al. (2021). By analyzing a historical time series of monthly inflation rates ranging from 1871M1 to 2018M4 with various estimators, Boubaker et al. (2021) found evidence of long memory with corresponding parameter $d \approx 0.24$.[425] Additionally, Boubaker et al. (2021) allow for a time-varying long memory parameter. More precisely, they estimate a model allowing for two possible regimes, each associated with its own long memory parameter.[426] For the US series, the authors report significant (at the 1% level) estimators $d_1 = 0.2589$ and $d_2 = 0.2641$ for the two regimes, and the null hypothesis of $d_1 = d_2$ can be rejected.[427] They further show that their time-varying long memory approach outperforms several other time-varying and not time-varying long memory models in terms of the model's forecasting performance.[428] In addition, they divide their time series into five subsamples covering distinct exchange rate regimes (ranging from the gold standard period (1871-1914) to the period of flexible exchange rates (1972-2018)).[429] Overall, their results suggest that inflation became more persistent the more flexible the exchange rate.[430] They further show that the parameter d has become smaller over the period 1983-2018, which is associated with an inflation targeting policy in the US.[431] These results confirm the ones of Canarella

[420] See Andersson and Li (2020, p. 534).
[421] See Andersson and Li (2020, p. 541).
[422] See Andersson and Li (2020, p. 545).
[423] See Andersson and Li (2020, p. 544 and Table 8 on p. 545).
[424] See Andersson and Li (2020, p. 545).
[425] See Boubaker et al. (2021, Table 2 on p. 302).
[426] See Boubaker et al. (2021, p. 303).
[427] See Boubaker et al. (2021, p. 303 and Table 3 on p. 303).
[428] See Boubaker et al. (2021, pp. 304f.).
[429] See Boubaker et al. (2021, pp. 305f.).
[430] See Boubaker et al. (2021, p. 306).
[431] See Boubaker et al. (2021, p. 306).

and S. M. Miller (2017) mentioned above, who show that the adoption of an inflation target is associated with a smaller value of d.

This brief discussion shows that long memory is also interesting from a monetary policy's perspective. Inflation appears to be well-described by long memory processes, and inflation's persistence, in terms of d, may be closely related to the monetary policy and exchange rate regime. The work of Andersson and Li (2020) interprets d as the flexibility of the inflation target. Such a point of view appears to be interesting from a theoretical perspective as it may offer a path for endogenizing long memory in models with inflation. Concerning the ECB's inflation target, which refers to the Euro area's inflation rate, there appears to be a need for a refinement of Andersson and Li (2020) interpretation. This need can be derived from the work of Kumar and Okimoto (2007), who suggest that the countries of the Euro area are characterized by country-specific long memory parameters. Given Andersson and Li's interpretation of the parameter d, it seems unclear how to interpret the country-specific values of d, since the ECB's inflation target only addresses the Euro area's inflation rate. Additionally, Andersson and Li (2020)'s interpretation does not apply in cases without an inflating targeting monetary policy regime.

3.2.2.3 Stock Market Returns and Volatility

In the context of macroeconomic variables, the availability of long time series is limited since many macro indicators are only available quarterly or annually. This data limitation makes it hard to estimate reliable time series models and to sharply decide whether a process has features such as a unit root, structural breaks, time-varying means, etc. Financial data, e.g., stock prices, are available on a daily or even finer granularity. Not surprisingly, there is also extensive literature dealing with long memory in financial time series. The presence of long memory in financial time series may, additionally, motivate the considerations made in Chapter 5, where a continuous-time macro-financial model is considered. There, it is outlined that the evolution of capital in these models is quite similar to the evolution of stock prices in traditional finance models such as the model of Black and Scholes.

Although various stock prices and stock market indices behave differently, there is a set of stylized facts applying to a wide range of financial assets.[432] Two of these stylized facts state that there is almost no autocorrelation in financial assets' return over various time scales (daily, weekly, monthly, etc.).[433,434] At the same time, another stylized fact states

[432] See Cont (2001, p. 224).
[433] See Cont (2001, p. 224).
[434] Following Cont (2001, p. 223), let S_t the price of a financial asset (i.e., a stock or an exchange

the presence of long memory in the absolute values of asset returns with corresponding long memory parameter $d \in [0.3, 0.4]$.[435,436] That the absolute values of the returns show long memory is often interpreted as an indicator for long memory in asset price volatility.[437]

However, the evidence for the absence of long memory in the returns is not pervasive, as the empirical results appear to be mixed.[438] Some authors claim that the presence of long memory in stock market returns contradicts the weak form of the efficient market hypothesis, since the efficient market hypothesis implies that asset returns are stochastically independent.[439]

Cajueiro and Tabak (2004a) suggest using the Hurst exponent of the return series to rank various countries along their market efficiency. More precisely, they consider the return of representative stock market indices from eleven emerging economies, the US and Japan.[440] Then, they employ the R/S statistic on their returns time series to obtain estimators of the Hurst index $H = d + 1/2$. The obtained Hurst estimators serve as a measure of the market efficiency, i.e., a Hurst index $H > 1/2$ indicates market inefficiency, and the closer H to $1/2$, the more efficient the market.[441] For all countries, the reported estimators are higher than $1/2$, indicating long memory in the returns of the representative stock market indices. The most efficient markets appear to be the US and Japan, with an estimated H close to $1/2$.[442] The least efficient market is the one of the Philippines with an estimator $H = 0.64$ ($d = 0.14$).[443] Additionally, the authors show that their estimators are positively correlated with market capitalization and negatively correlated with the average trading costs.[444] This may indicate that larger financial markets tend to be more efficient.

In Cajueiro and Tabak (2005), the authors carried out a similar analysis on the stock market volatility instead of the returns. The authors consider the squared log returns and

rate) or the value of a market index at time t. Then the corresponding return of S is given by $R_{t,\Delta t} = \log(S_{t+\Delta t}) - \log(S_t)$, where Δt is a given time scale. More specifically, if Δt is equal to one day, one week, one month, then the series $R_{t,\Delta t}$ for $t = 1, 2, 3, \ldots$ would correspond to the daily, weekly, monthly log returns of S, respectively.

435 See Cont (2001, p. 224 and p. 230).

436 Note that Cont (2001)'s parameter β is equal to $1 - 2d$ due to (2.27). Consequently, $d = 1/2(1 - \beta)$.

437 See Cont (2001, p. 230).

438 A detailed empirical literature review is beyond the scope of this section but can be found in Lim and Brooks (2011, Section 3.4 on pp. 79f.) and Sewell (2012, Section 6 on pp. 170ff.). A rather rough review is given by Caporale, Gil-Alaña, and Plastun (2019, p. 1763).

439 See Hull and McGroarty (2014, p. 45), Sewell (2012, p. 170) or Lim and Brooks (2011, p. 70).

440 See Cajueiro and Tabak (2004a, p. 350).

441 See Cajueiro and Tabak (2004a, p. 351).

442 See Cajueiro and Tabak (2004a, Table 1 on p. 351).

443 See Cajueiro and Tabak (2004a, Table 1 on p. 351).

444 See Cajueiro and Tabak (2004a, p. 351).

the absolute returns as a measure of volatility.[445] Overall, the estimated Hurst indices vary between 0.68 ($d = 0.18$) and 0.75 ($d = 0.25$) for the squared returns and between 0.7 ($d = 0.2$) and 0.8 ($d = 0.3$) for the absolute returns.[446] Compared to the results of Cajueiro and Tabak (2004a), the Hurst indices of the volatility measures are higher than the ones of the returns series, indicating that there is more evidence of long memory in the returns' volatility than in the returns themselves. Again, the Hurst index of the US volatility measure is lower than that of the other countries in the sample.[447] These results are consistent with Cont (2001)'s stylized facts mentioned above, saying that there is more evidence of long memory in the volatility of stock market returns than in the returns themselves.

In a related paper, the authors carry out a similar study. The authors split their daily data set from 1992 to 2002 into windows of 1008 observations. By shifting this window through the whole data set and calculating the Hurst index for each window, the authors obtain a series of Hurst indices for each country.[448] The authors illustrated that the Hurst index varies over time.[449] For the US and Japan, the Hurst index fluctuates around 1/2.[450] This result confirms the authors' previous reasoning that the US and Japanese markets appear to be more efficient. Additionally, the authors find some negative trends in the Hurst indices for some countries. In their view, this indicates that the financial market has become more efficient in these countries.[451] Again, more developed countries (US and Japan) appear to be more efficient than less developed countries.[452]

A similar investigation of rather developed countries was carried out by Onali and Goddard (2011). The authors investigated returns series from stock market indices of the US and seven European countries. Similar to the works of Cajueiro and Tabak (2004a) and Cajueiro and Tabak (2004b), they calculate Hurst indices for each country and found that larger stock markets appear to be more efficient than smaller ones.[453] Overall, they found strong evidence of long memory (and thus for stock market inefficiency) in the Czech stock market and less evidence of long memory in the Spanish and Swiss stock markets.[454] For the remaining countries containing Germany, Italy, Netherlands, the UK, and the

[445] See Cajueiro and Tabak (2005, p. 672).
[446] See Cajueiro and Tabak (2005, p. 672).
[447] See Cajueiro and Tabak (2005, Table 5 on p. 674).
[448] See Cajueiro and Tabak (2004b, pp. 530f.).
[449] See Cajueiro and Tabak (2004b, pp. 531f.).
[450] See Cajueiro and Tabak (2004b, p. 535).
[451] See Cajueiro and Tabak (2004b, pp. 535f.).
[452] See Cajueiro and Tabak (2004b, p. 535).
[453] See Onali and Goddard (2011, p. 64).
[454] See Onali and Goddard (2011, p. 66).

US, there is no evidence of long memory and, consequently, these markets have to be deemed efficient.[455] These results are in line with the ones of Cajueiro and Tabak (2004a) as smaller capital markets appear less efficient than bigger markets.

In a recent investigation of seven cryptocurrency markets, Assaf et al. (2022) found evidence of long memory in the returns and the volatility of six cryptocurrencies under their consideration.[456] These results point to a high inefficiency in cryptocurrency markets.[457] By estimating the parameter d, they found a higher degree of long memory in the volatility measures than in the returns series.[458]

In a recent study, Vera-Valdés (2022) applied the GPH and ELW to a battery of volatility measures of various stock market indices containing Standard & Poor's 500, Deutscher Aktienindex, Nasdaq 100, and others.[459] The author's focus was to illustrate the impact of the Covid 19 pandemic on stock market volatility. Therefore, he compares estimators of d for a sample before the pandemic (ranging from January 1, 2018, to January 30, 2020) and after the pandemic (ranging from January 31, 2020, to January 15, 2021).[460] His results indicate that almost all stock market volatility measures are characterized by a long memory parameter $0 < d < 0.5$ in the pre-Covid subsample while the parameter d is higher than 0.5 in the post-Covid subsample.[461] Vera-Valdés (2022) claims that the Covid 19 pandemic has a more persistent effect on financial market volatility than other crises before.[462,463]

455 See Onali and Goddard (2011, p. 66).
456 See Assaf et al. (2022, p. 1552).
457 See Assaf et al. (2022, pp. 1559f.).
458 See Assaf et al. (2022, Tables 6 and 7 on pp. 1561f. and pp. 1563f.).
459 See Vera-Valdés (2022, Table 1 on p. 2) for a complete list of all indices investigated by the author.
460 See Vera-Valdés (2022, pp. 2f.) for a description of the data.
461 See Vera-Valdés (2022, Table 2 on p. 4 and p.5).
462 See Vera-Valdés (2022, p. 5).
463 As one would expect from Section 3.1.2, the point estimators reported by Vera-Valdés (2022) depend strongly on the choice of the bandwidth parameter. This can be seen from a comparison of his Table 2 and his Table A4 on p. 4 and pp. 6f., respectively. His main conclusion, that the parameter d is higher in the post-Covid period than in the pre-Covid period, remains valid.

3.3 Possible Explanations for Long Memory in Economic Time Series

On the one hand, empirical evidence of long memory provides convincing arguments for its existence. However, there remains a need for economic intuition where long memory in economic time series comes from. This section is devoted to this question and summarizes various reasons discussed in the literature that provide some theoretical underpinnings for the prevalence of long memory in economic and financial time series. The most frequently mentioned theoretical reason for long memory is cross-sectional aggregation, which will be considered in more detail in the next section. In addition, the following section highlights the unequal implications of cross-sectional and temporal aggregation regarding the aggregate time series's long memory properties. The second subsection summarizes additional rationales for long memory mentioned in the literature.

3.3.1 Cross-Sectional and Temporal Aggregation

From a macroeconomic perspective, the view that aggregation may induce long memory seems appealing as it can be argued that many macroeconomic time series occur as the aggregate of underlying subordinated variables. For example, a consumer price index is measured via a representative basket containing many goods, or an economy's gross value added appears to be an aggregate over various sectors or firms.[464]

A stylized and early explanation of how aggregation induces long memory was given by Granger (1980). He considers N components z_j, for $j = 1, \dots, N$, each assumed to follow a (short memory) AR(1) process, i.e.,

$$z_{j,t} = \varrho_j z_{j,t-1} + \varepsilon_{j,t} \text{ with } |\varrho_j| < 1 \text{ for } j = 1, \dots, N, \tag{3.9}$$

where $(\varepsilon_{j,t})_{t\in\mathbb{Z}}$ for $j = 1, \dots, N$ are Gaussian white noise processes that are assumed to be independent of each other.[465] The linearly aggregated series z_t is given by

$$z_t := \sum_{j=1}^{N} z_{j,t}. \tag{3.10}$$

Granger's result can be summarized as follows. If the ϱ_j^2 follow a beta distribution on

[464] For example Hassler and Wolters (1995, p. 43) use such an aggregation mechanism to argue for long memory in inflation rates.
[465] See Granger (1980, pp. 230f.).

$(0, 1)$ with parameters $p, q > 0$, then, for large N, the ACF of z_t satisfies for large N approximately

$$\gamma_{z_t}(k) \sim Ck^{1-q}, \text{ as } k \to \infty, \tag{3.11}$$

where C is a certain constant.[466,467] Clearly, (3.11) implies (2.27); hence, z_t is a long memory process due to Lemma 2.3.2, and its parameter d is given by $d = 1 - q/2$.[468] Thus, the aggregate series is stationary and shows long memory if $q \in (1, 2)$.

This result of Granger (1980) has been generalized in various ways in the literature. For example, Linden (1999) assumes that the ε_j's in (3.9) are idiosyncratic shocks specific to each micro-unit; in addition, he adds a shock to (3.9) that is common to all micro-units.[469] He then considers the mean of N independent versions of the resulting processes and assumes that the ϱ_j's are uniformly distributed over $(0, 1)$. His aggregated process then again shows long memory.[470,471]

Zaffaroni (2004) goes even beyond the results of Linden (1999). He showed that the sample averages of the idiosyncratic and common components show long memory under far less restrictive assumptions compared to Granger (1980) and Linden (1999). For example, his results hold for more general distributions than the beta or uniform distribution, and the AR(1) process in (3.9) can be replaced with a general ARMA process.[472,473] He further

466 See Granger (1980, p. 233).

467 The density of the beta distribution is given by

$$f(x) = \begin{cases} \left(\dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\right)^{-1} x^{p-1}(1-x)^{q-1}, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

with $p, q > 0$, see, e.g., Czado and Schmidt (2011, p. 14 and Definition 1.17 on p. 18). The distribution is called beta distribution since the factor $\Gamma(p)\Gamma(q)/\Gamma(p+q)$ is also known as the beta function. A figure showing the density of the beta distribution for various parameter combinations of p and q can be found in Czado and Schmidt (2011, p. 18). Note that Granger (1980, Equation (11) on p. 232) states the density of the squared ϱ_j's.

468 This follows immediately from (2.27) by solving $1 - q = 2d - 1$ for d. Additionally, see Granger (1980, p. 233).

469 See Linden (1999, Equation (2.2) on p. 32).

470 See Linden (1999, p. 34).

471 Note that the aggregated process of Linden (1999) has an infinite moving average representation with coefficients $\psi_j = 1/(j + 1)$, see Linden (1999, Equation (3.4) on p. 33). The same process is considered in Hassler (2019, Example 3.5 on pp. 43ff.), who showed that this process is also strongly persistent.

472 See Zaffaroni (2004, Assumption II on p. 78) for the assumption regarding the distribution of the $\varrho_j's$, Zaffaroni (2004, pp. 84f.) for his result concerning the idiosyncratic component, Zaffaroni (2004, pp. 86ff.) for the corresponding results concerning the common component. The extension to general ARMA processes is treated in Zaffaroni (2004, pp. 89f.).

473 A similar analysis considering the aggregation of general AR(p) processes with extensions to continuous-time Ornstein-Uhlenbeck processes is carried out in Oppenheim and Viano (2004).

points out that the mass around 1 of the ϱ_j's distribution determines the degree of long memory in the aggregate.[474] This confirms the results found by Granger (1980) since the mass of the beta distribution increases the closer the parameter q is to 1.[475]

In a theoretical model framework, Haubrich and Lo (2001) use the aggregation argument of Granger (1980) to generate long memory in a model economy's aggregate output. To be more precise, Haubrich and Lo (2001) consider a multi-sector economy with a linear input-output production function, where each sector is assumed to use only its own produced output as input factors.[476] The model economy may be seen as some independent islands producing independently from each other. However, the utility-maximizing agent can consume the output of each island and is resource-constrained by the total output produced by all islands.[477] Given a quadratic utility function and a fully depreciating capital stock, Haubrich and Lo (2001) show that the output of each sector (island) can be described as an AR(1) process, whose autoregressive parameter depends on the island's coefficient of the input-output matrix and the parameters of the utility function.[478] The aggregate output over all islands is then given as the sum of AR(1) processes similarly to (3.10).[479] They finally employ Granger's result by supposing a beta distribution over the island's autoregressive parameters. They show that aggregate output is a long memory process with parameter $d = 1 - q/2$, where q is again the parameter of the beta distribution that shapes its density near 1.[480]

Another model-based approach to illustrate the effect of how aggregating may alter important time series properties was carried out by Abadir and Talmain (2002). The authors argue that aggregation in economic models is unlike (3.11) to a large extent nonlinear.[481] Therefore, the authors set up a more elaborate model than the model of Haubrich and Lo (2001) containing a nonlinear aggregation over AR(1) productivity processes. To be more precise, the authors consider a model containing a monopolistic competitive intermediate goods-producing sector with N firms, each employing a Cobb-Douglas production function with heterogeneous technical efficiency across firms.[482] The economy's aggregate output is then produced by a perfectly competitive final goods-

474 See Zaffaroni (2004, p. 85).
475 See Haldrup and Vera-Valdés (2017, p. 3).
476 See Haubrich and Lo (2001, p. 18).
477 See Haubrich and Lo (2001, p. 18).
478 See Haubrich and Lo (2001, p. 19).
479 See Haubrich and Lo (2001, Equation (17) on p. 19).
480 See Haubrich and Lo (2001, p. 20).
481 See Abadir and Talmain (2002, p. 757).
482 See Abadir and Talmain (2002, p. 751).

producing sector employing a constant-elasticity-of-substitution aggregator.[483] After solving the representative agent's maximization problem, the authors derive a dynamic equation for the economy's GDP per capita (y_t). It is given by[484]

$$y_t = \theta_t C^\gamma y_{t-1}^\gamma, \tag{3.12}$$

where C is a constant equal to the economy's steady state savings rate and $\gamma \in (0, 1)$ is the capital share of the intermediate good producers' production function.[485] A striking feature of (3.12) is the technology process θ, which is given by[486]

$$\theta_t = \left(\frac{1}{N} \sum_{j=1}^{N} \theta_{j,t} \right)^{1/\nu}, \tag{3.13}$$

where ν depends on the parameter of the employed constant-elasticity-of-substitution aggregator, and $\theta_{j,t}$, for $j = 1, \ldots, N$ are the firm-specific intermediate goods producers' technology processes.[487] In order to derive the time series properties of GDP per capita, Abadir and Talmain (2002) assume the firm-specific technology processes (θ_j) to follow AR(1) processes containing common-shock and idiosyncratic-shock components with firm-specific persistence parameters.[488] By supposing concrete probability distributions for the productivity processes' parameters across firms (again a beta distribution for the autoregressive parameters), the authors show that GDP per capita given in (3.12) shows long memory.[489] A major contribution of Abadir and Talmain (2002) may be seen in the nonlinear aggregation mechanism used by the authors. They illustrated that the type of aggregation (linear or nonlinear) mainly affects the time series properties of the aggregate series. Unfortunately, the authors do not define the concept of long memory in their paper. Since they state, at the same time, that the GDP per capita process is not stationary, one can deduce that their notion of long memory differs from the one of Definition 2.3.1 as an autocovariance function is only well-defined for stationary processes.[490] Instead, the authors consider asymptotic laws for the covariances between two instances of time while not making their notion of long memory explicit. Nevertheless, Abadir and Talmain (2002) highlight the role of the aggregation mechanisms and how it shapes the time series

[483] See Abadir and Talmain (2002, p. 752).
[484] See Abadir and Talmain (2002, Equation (18) on p. 756).
[485] See Abadir and Talmain (2002, pp. 751 and 756).
[486] See Abadir and Talmain (2002, Equation (10) on p. 754).
[487] See Abadir and Talmain (2002, pp. 751 and 754). To be more precise, ν^{-1} may be regarded as a measure for monopoly power in the intermediate goods sector, see Abadir and Talmain (2002, p. 754).
[488] See Abadir and Talmain (2002, Equations (20) and (24) on p. 758f.).
[489] See Abadir and Talmain (2002, p. 762).
[490] For the statement of the non-stationarity of GDP per capita, see Abadir and Talmain (2002, p. 762).

properties of the aggregate variable, especially in the presence of nonlinear aggregation. In doing so, they confirm existing literature that aggregation leads to more persistent processes.[491]

Furthermore, it should be noted that long memory arising from the aggregation is generally not of the same type as for the ARFIMA processes considered in Section 2.4.[492] However, as illustrated by Vera-Valdés (2020), ARFIMA processes have been shown to deliver good forecasting performances independent of the underlying long memory generating process, especially for long memory caused by (linear) cross-sectional aggregation.[493] Additionally, by carrying out a simulation study, Vera-Valdés (2021) has shown that key properties such as the periodogram, ACF, and the process's realization of an ARFIMA process and a cross-sectional aggregated process like the one in (3.10) (weighted with $\sqrt{N}$) are quite similar as long as $d = 1 - q/2 \in (0, 1/2)$, where q refers again to the parameter of the beta distribution.[494] That is, ARFIMA processes with long memory ($0 < d < 1/2$) appear to be good time series models to mimic the behavior of aggregated processes. Conversely, the differences between both classes of processes are large if $-1/2 < d < 0$, i.e., if the process is an anti-persistent short memory process according to Table 2.1.[495] Although long memory stemming from linear aggregation differs from long memory implied by ARFIMA processes, Vera-Valdés' results may serve as an argument for using ARFIMA processes.[496]

Furthermore, the literature discussed so far has concentrated on aggregating N AR(1) processes with a random autoregressive parameter drawn from a beta distribution. This assumption appears to be less founded from an empirical perspective but provides manageable closed-form expressions in the first place.[497] Albeit the results remain stable under a

[491] See Abadir and Talmain (2002, p. 757).

[492] In the case of the nonlinear aggregation carried out by Abadir and Talmain (2002) in (3.13), this was already mentioned in the text. Nevertheless, also in the case of linear aggregation of the form (3.10), this fact can be seen clearly from the results of Linden (1999, Equation (3.4) on p. 33) who derived the moving average coefficients of his aggregated process by assuming a uniform distribution. He illustrates that they differ from the ones of ARFIMA processes (see further Footnote 471). More recently, Haldrup and Vera-Valdés (2017, Lemma 1 on p. 3) show a similar result for beta distributed autoregressive coefficients (ϱ_j). To be more precise, they show that the moving average coefficients of the process defined in (3.10) (scaled by $\sqrt{N}$) are different from those of an ARFIMA process stated in Lemma 2.4.2. However, both types of processes share the same asymptotic (hyperbolic) rate of decay.

[493] See Vera-Valdés (2020, p. 817).

[494] See Vera-Valdés (2021, pp. 9 and 13).

[495] See Vera-Valdés (2021, pp. 8ff.).

[496] For example, if $0 < d < 1/2$, ARFIMA processes may be appropriate in a representative agent model context that is aimed to account for long memory without specifying a possibly underlying cross-sectional heterogeneity in the given model context. In such a context, long memory has to be regarded as exogenous to the representative agent.

[497] See Haldrup and Vera-Valdés (2017, p. 3). Especially for their Lemma 1, where the authors derive the coefficients of the aggregated process's IRF, they relate strongly to the assumption of beta distributed

generalized class of distributions containing the beta and uniform distribution as outlined by Zaffaroni (2004), the concrete choice of such a distribution seems overall to be critical.[498] At the same time, the memory properties of the aggregate process depends mainly on the choice of such a distribution. Haubrich and Lo (2001) point out that choosing, for example, a discrete distribution taking on $m < N$ values of the autoregressive parameters implies that the process defined in (3.10) is just an ARMA$(m, m-1)$ process entailing short memory.[499] By assuming N independent AR(1) processes, the aggregated process is an ARMA$(N, N-1)$ process.[500] This result is further generalized to the aggregation of a finite number of discrete stationary processes by Chambers (1998). He showed that the long memory parameter of the aggregate time series corresponds to the largest long memory parameter among all individual time series.[501] Consequently, long memory in an aggregated time series can only occur if at least one individual time series shows long memory.[502] This short discussion illustrates that the aggregation argument for generating long memory appears to be valid only under specific circumstances. A key ingredient is that there is a large number of short memory micro-units over which the aggregation is taken and that the corresponding micro-units' short memory parameters follow a kind of beta distribution.[503] The specification of such a distribution may be one reason why such aggregation mechanisms are often not considered in dynamic macroeconomic models.[504]

A different path for explaining long memory, especially in aggregate output, was taken by Michelacci (2004). The author considers a vintage model with infinitely many firms that move away from the economy's technological frontier as time goes by.[505] Firms in Michelacci's model can decide in each period either to invest in new technologies that bring them back to the economy's technological frontier, or not to do so and move one step away from the frontier in the next period.[506] Thus, the economy's firms are employing different technologies at various distances from the actual technological frontier. Suppose a firm wants to invest in the technology at the frontier. In this case, it faces costs containing a deterministic part dependent on the state of the technology employed and a random

ϱ_j's, see Haldrup and Vera-Valdés (2017, Proof of Lemma 1 on p. 9).

498 See Zaffaroni (2004, Assumption II on p. 78) for the details on the class of possible distribution under which his aggregation results stay valid.

499 See Haubrich and Lo (2001, p. 20).

500 See Granger (1980, p. 231) or Haubrich and Lo (2001, p. 19).

501 See Chambers (1998, Proposition 3 on p. 1061).

502 See Chambers (1998, p. 1061).

503 Many micro-units may be present if one aggregates, e.g., all consumption expenditures of households or the output of all firms in an economy, see Chambers (1998, p. 1060). Fewer micro-units may be given in aggregating, e.g., a price index.

504 See Zaffaroni (2004, p. 77).

505 See Michelacci (2004, p. 1326).

506 See Michelacci (2004, p. 1326).

component independently distributed over all firms reflecting firm-specific opportunity costs of the possible investment.[507] By deriving a Bellmann equation, Michelacci (2004) illustrates that a firm would only invest in the new technology if the expected benefits outweigh the cost of adopting the new technology.[508] Since the cost of adoption is uncertain, one has to specify the probability that a firm employing a certain technology adopts the new technology or not.[509] Michelacci (2004) then analyzes the effect of an exogenous shock to firms' costs of adopting the new technology.

In his setting, aggregate output increases as more firms employ technologies near the frontier.[510] He shows that if the probability of adopting the new technology is increasing in the distance from the actual technological frontier, i.e., firms far apart from the frontier are more likely to adopt the new technology, then, the economy's aggregate output is a short memory process.[511] On the contrary, if the probability depends negatively on the distance from the economy's technological frontier, aggregate output exhibits long memory.[512]

However, it should be noted that Michelacci's notion of long memory differs from the one of Definition 2.3.1. He defines long memory based on the impulse-response function, i.e., in his notation, a process is a long memory process if its IRF decays at a hyperbolic rate similar to iii) of Lemma 2.4.3 with $d > 0$.[513] Conversely, a process is a short memory process if its IRF decays exponentially fast, i.e., if $d = 0$.[514] The definition of Michelacci (2004), thus, covers in contrast to Definition 2.3.1 also non-stationary time series and random walks.[515]

Michelacci (2004) argues that if the probability of adopting the new technology is low for firms far away from the frontier, the propagation of the initial shock is slow, leading to long memory in the aggregated output.[516] This seems reasonable since, in this case, more firms keep producing with obsolete technologies, thereby contributing less to the economy's aggregate output. Instead, suppose it is likely for a low-technology firm to adopt new

507 See Michelacci (2004, p. 1326).
508 See Michelacci (2004, p. 1327).
509 See Michelacci (2004, pp. 1327f.).
510 This follows from Michelacci (2004, Equation (3) on p. 1326). Consequently, the aggregate output would be maximal if all firms produce at the technological frontier.
511 See Michelacci (2004, Proposition 2 on p. 1332).
512 See Michelacci (2004, p. 1332).
513 See Michelacci (2004, p. 1324).
514 See Michelacci (2004, p. 1324).
515 Michelacci's notion of short and long memory can be associated with moderate and strong persistence in the sense of this thesis.
516 See Michelacci (2004, p. 1332).

technologies rapidly. In this case, the shock is absorbed quickly, implying short memory in the aggregate output as more firms tend to employ more productive technologies, thereby increasing aggregate output.[517] He further relates his results to the growth rate of firms. He states that long memory occurs as long as small firms (i.e., firms employing an old technology) grow faster than big firms (i.e., firms already employing a technology near the frontier).[518,519]

To summarize, the model of Michelacci (2004) offers a new perspective on the underlying mechanisms that generate long memory. Specifically, firm heterogeneity and the probabilities of firms investing in state-of-the-art technologies appear to play an essential role in explaining long memory in an economy's aggregate output. On the other hand, the mechanism proposed by Michelacci (2004) depends crucially on how a firm's probability of investing in the new technology evolves with the distance from the economy's frontier. As outlined in Footnote 519, the probabilities have to decrease quite slowly. Furthermore, and as stated by Michelacci (2004) himself, there is a need for firms producing far away from the technological frontier in order to generate long memory.[520] To be more precise, if the distance to the economy's technological frontier is bounded, then the economy's aggregate output is a short memory process.[521]

Moreover, he further illustrates how his approach is related to the aggregation mechanism proposed by Granger (1980) mentioned above. Michelacci (2004) shows that the evolution of a firm with distance i from the technological frontier can be described as an AR(1) process whose autoregressive parameter is equal to the probability of investing in the new technology.[522] Instead of assuming a certain distribution across these autoregressive parameters as done in the papers discussed so far, Michelacci (2004) imposes slowly decaying probabilities along an increasing distance from the economy's technological frontier. These probabilities generate sufficient firms producing far away from the technological frontier needed to generate long memory in the aggregate output.[523]

517 See Michelacci (2004, p. 1332).
518 See Michelacci (2004, p. 1333).
519 To be more precise, let i denote the distance to the economy's technological frontier. Then, Michelacci (2004) requires the probabilities of a firm with distance i to adopt the new technology p_i to be $p_i = h/(i+1)$ for sufficiently large i, see Michelacci (2004, Equation (A4) on p. 1332). The parameter $h > 0$ is the growth rate of a small firm growing faster than the remaining big firms as long as $h > 1$, see Michelacci (2004, p. 1333). The order of integration of the aggregate output is equal to $2 - h$, implying long memory if $h \in (1.5, 2)$, see Michelacci (2004, Proposition 3 on p. 1333).
520 See Michelacci (2004, p. 1335).
521 See Michelacci (2004, p. 1335).
522 See Michelacci (2004, p. 1339).
523 See Michelacci (2004, pp. 1339f.).

Overall, to put the cross-sectional aggregation argument for producing long memory (in aggregate output) in a nutshell, Michelacci (2004), identifies three key ingredients: heterogeneity of the micro-units under consideration, a linkage between actual production and future production for each micro unit, e.g., in the form of an AR(1) process, and a mechanism producing that a sufficiently large amount of the micro-units are highly persistent.[524] The latter can be done by imposing a distribution on the AR(1) coefficients directly as was carried out in the mentioned works by Granger (1980), Haubrich and Lo (2001), Abadir and Talmain (2002) and Haldrup and Vera-Valdés (2017) or by assuming sufficiently slow decreasing probabilities of adopting new technologies in the vintage model of Michelacci (2004).

Another possible path for explaining long memory in an aggregate macroeconomic series resembling some ideas of Michelacci (2004) but, interestingly, not mentioned by him, can be attributed to Parke (1999). He sets up an error-duration model, where the aggregate time series represents the sum of identically, independently distributed white noise errors featuring a stochastic duration.[525] By specifying the survival probabilities, i.e., the probability that a white noise error lasts for k periods, he can precisely recover the autocorrelation function of a long memory ARFIMA$(0, d, 0)$ process.[526] To do so, he requires slowly decaying survival probabilities of order $k^2(1 - d)$.[527] Within a simple model framework, Parke (1999) interprets the errors as a firm's impact on total employment that is assumed to be constant as long as the firm is in business.[528] Thus, the errors' survival probability corresponds to the whole firm's survival probability. Slowly decaying survival probabilities of firms needed for the presence of long memory would then induce that there are some firms lasting for many periods in the market.[529] Explaining long memory in aggregate employment along Parke's approach is then closely related to how long a specific firm acts in the market and how likely it is that the firm will be in the market in the next period. However, Parke's approach cannot be generalized to ARFIMA(p, d, q) processes since, in general, these processes do not satisfy the requirements for the survival probabilities.[530]

Clearly, Parke's approach can also be applied in other contexts. So, it is imaginable that a similar mechanism applies to, e.g., technology. In an economy, many technologies are likely

[524] See Michelacci (2004, pp. 1339f.).
[525] See Parke (1999, p. 632).
[526] See Parke (1999, Proposition 3 on pp. 633f.).
[527] See Parke (1999, p. 633).
[528] See Parke (1999, p. 634).
[529] See Parke (1999, p. 635).
[530] See Parke (1999, p. 636).

used for production. The life spans of such technologies may be regarded as stochastic since permanent research and development bring out new technologies that may supersede old technologies. At the same time, it seems reasonable that if a firm invests in a specific technology, it will use these investments in production for several periods. Consequently, once the technology is applied, the survival probability of whether the technology is used in the next period will decay quite slowly.

A short note on temporal aggregation is given at the end of this section. As illustrated in Section 3.2.1, the estimated parameters of a time series's long memory parameter vary not only across the employed estimators but also across the granularity of the data, especially whether, e.g., GDP is measured quarterly or annually. This dependence on the granularity of a time series may raise the question of to which extent a macroeconomic variable's memory properties also depend on its measurement frequency. For example, consider a time series of a flow variable, say, a quarterly time series of an economy's GDP. The corresponding annual GDP values appear to be the sum of four quarters' GDP values. Is the long memory parameter of the aggregated annual series different from that of the quarterly series from which the annual series was derived? In general, this is not the case, i.e., temporal aggregation leaves the long memory parameter unaltered.[531,532] A similar result is given by Souza (2008) who showed that both estimators, the GPH estimator and Robinson (1995b)'s Gaussian semiparametric estimator discussed in Section 3.1.2 are asymptotically equivalent for the original and the temporally aggregated time series.[533] The reason for the contradicting results presented in Section 3.2.1 may be due to a small sample bias caused by temporal aggregation.[534,535] As outlined in Section 3.1.2, the semiparametric estimators become unreliable if the sample sizes decrease, i.e., the estimator may be reliable for the original series but the aggregated time series appears to be too short for a reliable estimation. This view may be supported by a recent survey of Shi and Sun (2016), who consider daily squared returns of the S& P 500 index from 1928 to 2011 (overall 22,000 observations).[536] They employ the GPH estimator and show that the weekly, biweekly, and monthly series, each aggregated from the daily series, lead to

[531] See Chambers (1998, Proposition 1 on p. 1057).
[532] Chambers (1998, Equations (5) and (7) on pp. 1056f.) further provides closed-form expressions for the spectral density of the aggregate time series. These are corrected in Souza (2005, Equation (3) and (4) on p. 1060). Nevertheless, the main result of Chambers (1998) that long memory is invariant under temporal aggregation remains valid, see Souza (2005, p. 1062).
[533] See Souza (2008, Proposition 1 and 2 on pp. 303f.).
[534] See Hassler (2011, p. 343) and Souza (2005, p. 1062).
[535] Clearly, aggregating a quarterly time series to an annual time series results in an overall shorter time series.
[536] See Shi and Sun (2016, p. 474).

nearly the same long memory estimators as for the daily series.[537] Thus, given this rather long time series, there is empirical evidence that in some cases long memory is invariant under temporal aggregation.

3.3.2 Other Reasons

This section summarizes possible explanations for long memory different from the cross-sectional aggregation mentioned in the previous section. As these arguments are sometimes context-dependent and forced by single authors, the considerations are done without a finer subdivision of this section.

Economic activity sometimes depends on environmental circumstances, e.g., the agricultural production within a year depends on the amount of rain or certain temperatures. As many of these geophysical processes show long memory, it may be conclusive that there is long memory in the corresponding economic data as well.[538,539] Thus, the introduction of long memory shocks in a real business cycle (RBC) model as carried out in Chapter 4 appears to be a practical approach to mimic existing long memory in the data.[540]

Schennach (2018a) proposes a long memory generating mechanism based on networks. To be more precise, she considers a network containing infinitely many agents connected via an input-output relation.[541] She then considers an agent of the network facing a short memory shock. She asks under which conditions the output of another network's participant (or the aggregate output of a group of other participants) shows long memory. She shows that the spectral density of a network participant's output depends on the probability that a certain point in the network is reached after n steps (the sequence c_n in her notation).[542] To be more precise, if the mentioned probabilities decay as $c_n = n^{-\gamma}$ then the spectral density of the destination point's output (or the destination points aggregate output) satisfies (2.28) with $d = 1 - \gamma$.[543] In the next step, Schennach (2018a) relates the parameter γ to the geometric structure of the network. To be more precise,

537 See Shi and Sun (2016, Table 1 on p. 474). Note that the bandwidth parameters for estimating the original and aggregated series's long memory parameters must be the same size, see Shi and Sun (2016, p. 474).
538 See Henry and Zaffaroni (2003, p. 421).
539 This may have been true in ancient Egypt in the time of Joseph, where crop production was highly dependent on the flooding of the Nile. If rainfall has long memory, as Hurst showed (see Section 2.2.1), then crop production is likely to show long memory, too.
540 See Henry and Zaffaroni (2003, p. 421).
541 See Schennach (2018a, pp. 2223f.).
542 See Schennach (2018a, Theorem 1 on p. 2227) and Schennach (2018a, pp. 2227f.) for an interpretation of the coefficients c_n.
543 See Schennach (2018a, Theorem 1 on p. 2227).

she uses a result of fractal geometry, saying that the parameter γ corresponds to half of the network's fractal dimension.[544] Overall, the order of fractional integration d of a destination point's output falls within the range of long memory ($0 < d < 1/2$) as long as the fractal dimension of the network is between one and two.[545] As pointed out by Schennach (2018a), this property is true for hierarchical star-networks.[546]

To summarize, Schennach's results highlight that long memory in a firm's or economy's output series may result from the input-output relations and the interconnectedness between the economy's firms or agents. There is further a one-to-one mapping between the geometrical structure of the underlying network and the long memory parameter.[547]

The question of whether and how rational expectations and long memory are interrelated is addressed by Chevillon and Mavroeidis (2017). Their contribution is discussed in more detail at the end of Section 4.2.2 as there are additional linkages to discrete-time rational expectations models to which the class of DSGE models considered in Chapter 4 belong. In short, they consider a univariate model that includes expectations about the future values of the variable. They show that the solution of the model exhibits short memory when expectations are formed rationally, and long memory when a specific learning algorithm is used.[548]

Another approach for explaining long memory is due to J. I. Miller and Park (2010). They consider a nonlinear transformation of a linear stochastic process exposed to fat-tail distributed shocks.[549,550] Given the properties of the nonlinear function that enters the transformation and the parameters of the shock distribution, their process can mimic the hyperbolic decay (2.27) of the autocorrelation function for $0 < d < 1/4$.[551] Hence, a small degree of long memory might be attributed to existing nonlinearities and exogenous fat-tailed distributed shocks.

[544] See Schennach (2018a, p. 2229).
[545] This is just a summary of the already mentioned results of Schennach (2018a), i.e., $d = 1 - \gamma$ and γ corresponds to one half of the network's fractal dimension.
[546] See Schennach (2018a, p. 2230) and Schennach (2018a, Figure 2 on p. 2230) for a graphical illustration of such a network.
[547] In the supplementary material, Schennach (2018b, Section S2.2 on pp. 8f.) provides an empirical application of her approach. By constructing a network between sectors of the US economy based on input-output accounts, she finds evidence of long memory in the US primary sector. To be more precise, she takes all firms of the primary sector as starting and destination points of the network, i.e., she considers a common shock to this sector. Then, she analyzes this sector's aggregate effect of this shock. The corresponding long memory parameter is $d = 0.42$, see Schennach (2018b, p. 8).
[548] See the end of Section 4.2.2 for a more extended discussion and references.
[549] See J. I. Miller and Park (2010, p. 84) for the definition of their data-generating process.
[550] That such nonlinear transformations may induce long memory behavior was already mentioned by Granger and Ding (1996, Section 6 on pp. 75f.).
[551] See J. I. Miller and Park (2010, p. 86).

As illustrated in Section 2.2.2 (especially in Figure 2.4), Mandelbrot mentioned the presence of "local trends" as a striking feature of a long memory process's realization. From this perspective, it seems not surprising that a large literature postulates that a time series's long memory is caused by a short memory process contaminated by structural breaks, level shifts, a time-varying mean, or regime switches rather than fractional integration. For example, Diebold and Inoue (2001) consider (among other models) a Markov-switching model, where the time series is given by a Gaussian white noise process with a time-varying mean depending on the value of an underlying two-state Markov chain.[552] By allowing the staying probabilities to be dependent on the sample size, they show that the model resembles properties of a long memory process if the Markov chain's staying probabilities increase slowly with the sample size.[553] That is, long memory can be mimicked if the probability of a mean shift decreases with sample size, ensuring that there are few regime switches in the process.[554] By evaluating a Monte Carlo simulation, they show that the GPH estimator leads to positive values of d if the Markov chain's staying probabilities are near unity (about 0.999), i.e., if the probabilities for a regime-switch are low.[555,556]

In a related study, Jensen and Liu (2006) consider a two-state duration-dependent switching model similar to the one of Diebold and Inoue (2001). However, instead of choosing the transition probabilities to depend on the sample size, they assume the lengths of the corresponding regimes to be random.[557] The two regimes they consider are a regime of an expanding economy and a recession regime.[558] They show that if the regime's duration times are drawn from fat-tailed distributions, the process's autocorrelation function satisfies a generalized version of (2.27), and hence, mimics the correlation structure of a long memory process.[559,560] Additionally, the corresponding long memory parameter depends linearly on the tail-index of the underlying fat-tail distribution.[561,562]

---

[552] See Diebold and Inoue (2001, p. 139).
[553] See Diebold and Inoue (2001, Proposition 3 on p. 139).
[554] See Diebold and Inoue (2001, p. 155).
[555] See Diebold and Inoue (2001, p. 152).
[556] Note that Diebold and Inoue (2001, p. 133) use the same definition of long memory as Chevillon and Mavroeidis (2017) given in Footnote 696 below. As stated in Footnote 696, this definition deviates from the one considered in this thesis.
[557] See Jensen and Liu (2006, pp. 600f.).
[558] See Jensen and Liu (2006, pp. 600f.).
[559] See Jensen and Liu (2006, Theorem 3.1 on p. 601).
[560] Davidson and Sibbertsen (2005, Section 2 on pp. 256ff.) consider the same theoretical model as Jensen and Liu (2006) and derive the same result in Davidson and Sibbertsen (2005, Theorem 2.1 on p. 258).
[561] See Jensen and Liu (2006, Theorem 3.1 on p. 601).
[562] By analyzing data on the length of boom and bust cycles in the US economy, Jensen and Liu (2006, Section 4 on pp. 602ff.) found evidence for their approach. The corresponding estimated long memory parameter is consistent with the estimated values of Sowell (1992b) and Diebold and Rudebusch (1989) already mentioned in Section 3.2.1, see Jensen and Liu (2006, p. 603).

The investigation, carried out by Ashley and Patterson (2010), goes beyond a regime-switching model. They argue that long memory in the data is caused by a deterministically and smoothly varying behavior of a time series's mean.[563] Similar to Diebold and Inoue (2001), they employ the GPH estimator to a battery of time series models containing such a trending behavior and structural breaks.[564] They illustrate that the GPH estimator detects long memory for all of these models.[565] Afterward, they propose and apply a detrending mechanism to the considered time series models for which the GPH estimator provides no more evidence of long memory.[566] Overall, their results imply that a time series's long memory may rather be founded in a prevailing smooth varying mean instead of a stationary fractionally integrated process.

In the literature, processes mimicking the autocorrelation function of a long memory process are often labeled as "spurious long memory"[567] processes as they do not appear as a stationary time series showing slowly decaying autocorrelations but rather as short memory processes contaminated by the mentioned level-shifts, regime-switches or smooth trends.[568] It is easily imaginable that from an econometric perspective, it seems difficult to differentiate between these various possible specifications.[569] Empirical implications such as forecasts may, however, differ substantially across a true long memory process such as an ARFIMA process with $0 < d < 1/2$ and a regime-switching or break-contaminated short memory process.[570] Nevertheless, as outlined by Diebold and Inoue (2001), a true long memory process may be a valuable shorthand description of a time series's dependence structure if the true model is given by a break-contaminated short memory process.[571] They further point out that long memory and structural breaks are different concepts but indeed describe the same phenomenon; hence, denoting one concept as "true" and the other as spurious seems doubtful.[572]

This section has presented many possible explanations for the phenomenon of long memory. From an economic perspective, the aggregation mechanisms discussed in Section 3.3.1 and

[563] See Ashley and Patterson (2010, p. 60).
[564] See Ashley and Patterson (2010, Table 1 on p. 67) for an overview of the considered time series models.
[565] See Ashley and Patterson (2010, Table 2 on p. 75 and p. 68) for the postulated evidence of long memory.
[566] See Ashley and Patterson (2010, Table 3 on p. 77 and p. 76) for corresponding results after the trend elimination.
[567] Ohanissian et al. (2008, p. 161).
[568] See Ohanissian et al. (2008, p. 161).
[569] See Ohanissian et al. (2008, p. 161).
[570] See Ohanissian et al. (2008, p. 162) and A. Smith (2005, p. 321).
[571] See Diebold and Inoue (2001, p. 157).
[572] See Diebold and Inoue (2001, p. 157).

the network approach of Schennach (2018a) discussed above seem the most attractive, as they refer to the existing heterogeneity in the real economy. On the other hand, when analyzing theoretical models that abstract from this heterogeneity, for example in a representative agent model, it seems reasonable to treat the phenomenon of long memory exogenously. This approach is taken in the following section, where a stationary ARFIMA process with long memory is used in a DSGE model with rational expectations.

4

Modeling Aggregate Fluctuations in a DSGE-RBC Framework

Dynamic stochastic general equilibrium (DSGE) models have become a commonly used tool in macroeconomics to predict and explain aggregate movements over the business cycle and to analyze monetary and fiscal policy.[573] In addition, they are frequently used as a policy tool by institutions such as central banks or the European Commission.[574] DSGE models combine economic equations derived from agents' or firms' optimization problems, utility functions, and preferences with stochastic shocks such as technology or monetary policy shocks. These shocks are often incorporated into the model via stationary AR(1) processes or small order autoregressive moving average (ARMA)processes[575], where the corresponding autoregressive parameters are often referred to be the persistence parameters.[576]

As mentioned in Section 2.1.5.1 and Section 2.1.5.2, this persistence has to be deemed as moderate persistence according to Definition 2.1.6. In addition, the stochastic processes typically involved are short memory processes. This chapter aims to analyze deviations from the assumption of moderate persistence in the context of a DSGE model by allowing for strong persistence and long memory on the basis of a standard real business cycle (RBC) model. As illustrated in the previous chapter, there is vast literature dealing with

573 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 530).
574 See Lindé et al. (2016, pp. 2189f.).
575 For example, Smets and Wouters (2007, p. 590) use an ARMA(1, 1) process for a price mark-up shock.
576 See, e.g., Cantore, Gabriel, et al. (2013a, p. 427).

long memory in an economic context. Moreover, there is evidence that long memory models perform better in resembling business cycle characteristics than their short memory counterparts.[577] Moreover, there are arguments why aggregate data is likely to show long memory.[578]

These observations motivate this chapter, whose line of reasoning is as follows. To provide the reader with a concrete model structure for what follows, Section 4.1 specifies the model setup. Section 4.2 summarizes the general structure of linear DSGE models. Especially, it illustrates that the solution of a standard linear DSGE model obeys a state-transition equation, which implies that all variables in such a standard DSGE model are short memory processes. This observation provides the basis for including long memory exogenously into the model through fractionally integrated processes. Moreover, this section addresses the concept of stability in linear DSGE models and outlines some relationships with deterministic systems of difference equations.

Section 4.3 provides additional justification for the author's approach by briefly reviewing the existing literature dealing with deviations from the standard AR(1) assumption. Furthermore, it illustrates how the research carried out in this thesis contributes to that literature. In Section 4.4, the results of various model specifications are presented in the form of an impulse-response analysis, including a comparison between the effects of a short memory, long memory, and trend shock to the economy's total factor productivity. An accompanying appendix, Appendix B, provides additional material such as the model summaries and details of the solution method.

The author presented parts of the content in this chapter at the "CIMS DSGE Modelling Conference 2020" during the CIMS Summer School on DSGE Macroeconomic Modelling (Advanced Course) organized by the Center for International Macroeconomic Studies (CIMS) at the University of Surrey.

4.1 The Prototypical DSGE-RBC Economy

4.1.1 General Modeling Assumptions

Before the following section outlines the underlying models for the remaining sections of this chapter, introducing remarks are made, and general modeling assumptions are discussed in this section.

[577] Recall the work of Candelon and Gil-Alaña (2004) mentioned in Section 3.2.1.
[578] Recall the arguments mentioned in Section 3.3.1.

The focus of the research agenda of this thesis lies on long memory in the context of macroeconomic models. Chapter 3 outlines that long memory processes are frequently used in the empirical literature. Despite various shortcomings in the employed estimation methods, they have their justification for being part of the economist's toolbox in their own right. Further, Chapter 3 underlines that long memory is a phenomenon that likely arises in aggregate time series that are often available in macroeconomic contexts.

On the other hand, when stochasticity enters the models, short memory processes (predominantly AR(1) processes) seem to be the method of choice; especially in the context of DSGE models, which are an intensively used class of models in the literature.[579,580]

Further, as the class of DSGE models is well-established with well-known theory and solution methods, it seems appropriate to consider long memory dynamics in this particular class of models, thereby preserving comparability to a vast literature.

In RBC models, economic fluctuations are mainly explained by exogenous technology shocks represented in fluctuations of the Solow residual.[581] On the one hand, these technology shocks appear unsound. Since the Solow residual as a measure of technology is strongly procyclical (with output growth), economic recessions are the consequence of technological decline if technology is the driving force of economic activity.[582,583] However, it may be doubtful that technological decline is a major cause of economic recessions.[584] On the other hand, these technology shocks have to be strongly persistent to resemble the data.[585,586] In short, RBC models appear to be highly stylized and weakly justified from an economic perspective.

579 Again referring to the state of the art DSGE model, e.g., Smets and Wouters (2007), they introduce seven shocks each described by an AR(1) or small order ARMA process, see Smets and Wouters (2007, Section 1 on pp. 588ff.). Additional examples can be found in, e.g., Cantore, Gabriel, et al. (2013a, Equations (18.14) and (18.15) on p. 416 and p. 424), Galí (2008, pp. 22, 30, 50, 54), Aguiar and Gopinath (2007, p. 80).

580 In addition, Blanchard (2009, Footnote 11 on p. 224) states that the assumed shock dynamics are typically left unexplained and are thus not derived from "first principles" (Blanchard (2009, p. 224)).

581 See Mankiw (1989, pp. 82ff.) and Hartley et al. (1997, p. 38).

582 See Mankiw (1989, pp. 83f.), Hartley et al. (1997, p. 45) and Rebelo (2005, pp. 222f.).

583 Additionally, the Solow residual appears to be an inadequate measure of technological movements as its cyclical movement may be explained by labor hoarding (see Mankiw (1989, p. 84) and Hartley et al. (1997, p. 45)) or a varying capital utilization (see Hartley et al. (1997, p. 45)). Following Rebelo (2005, p. 222), this may be a sign that the Solow residual may contain endogenous components. Thus, treating it as an exogenous process may be inadequate.

584 See Rebelo (2005, p. 223).

585 See King, Plosser, et al. (1988, pp. 197f. and p. 231).

586 In addition, Cogley and Nason (1995, pp. 500f.) point out that the dynamics of the shock process determine the dynamics of the model's output; thus, RBC models have weak endogenous propagation mechanisms.

It is well-documented that New Keynesian DSGE models, which unlike RBC models incorporate nominal rigidities such as price and wage stickiness, improve the empirical fit of the model and help explaining the main driving forces of the business cycle.[587] Furthermore, as illustrated by Ireland (2004b), the questionable technology shocks play just a minor role in New Keynesian models to explain the variability of economic variables.[588]

The model in the following section abstracts from the market frictions and rigidities that are present in the literature on New Keynesian DSGE models. Given the growing literature on New Keynesian DSGE models, which seem to be the workhorse models in the DSGE context today, the reader may wonder why the author focuses on an RBC model.

Ultimately, the New Keynesian DSGE models consist of an RBC model as the core around which the New Keynesian framework with market imperfections and rigidities is implemented.[589] Thus, this chapter's model focuses on the core of a wide range of applied DSGE models in the literature.

This focus is in line with the research agenda of this thesis that does not seek to construct and estimate a large-scale DSGE model which explains the data sufficiently well. Instead, as will become clear soon, the introduction of long memory into a DSGE model framework raises various fundamental questions regarding the structure and solvability of the model and the implementation of an impulse-response analysis. To answer these questions, it seems suitable to resort to a prototypical DSGE model, avoiding model-induced complexities such as heterogeneous agents, policy institutions such as governments or central banks, or certain rigidities that may impede the answering of these questions.[590] In addition, the model considered below allows one to derive a recursive algorithm for computing the model's solution, which may not be the case in a more complicated model.[591]

Moreover, it seems common in the literature to use well-understood and relatively simple models for illustrating novel solutions, approximations, or related methods and algorithms.[592] As it is one contribution of this thesis to illustrate how a DSGE model can be solved under wider persistence and memory assumptions, it seems reasonable to do this

[587] For example, Smets and Wouters (2007, p. 597) outline the empirical importance of price and wage stickiness.

[588] See Ireland (2004b, p. 931).

[589] See Cantore, Gabriel, et al. (2013a, p. 411) and Ireland (2004b, p. 923).

[590] The consideration of long memory in more elaborate DSGE model is left for future research. See also the discussion below of arising research questions in the context of a New Keynesian DSGE model.

[591] Many researchers use the software package Dynare to solve and estimate their DSGE models. To the best of the author's knowledge, this is, however, not yet possible for the presented long memory model. Thus, these closed-form expressions facilitate the calculations carried out in this chapter.

[592] For example, Meyer-Gohde and Neuhoff (2018, Section 4 on pp. 14ff.) (discussed in detail in Section 4.3) consider such a RBC model to illustrate their novel testing procedure.

along a simplified RBC-DSGE model. Further, as the effects of long memory are a priori not clear, it seems reasonable to consider an RBC model first. If possibly existing effects due to the implementation of long memory dynamics are not apparent in the RBC model, further considerations in an extended New Keynesian DSGE model may be obsolete.[593] Overall, it seems justified to focus on the RBC core of many DSGE models as a first step, which is carried out in this thesis. The requirement to account for long memory in an New Keynesian DSGE model is warranted. In particular, questions of whether central banks should respond differently to a long memory inflation shock than to a short memory shock appear to be essential from both an empirical and a policymaker's perspective. These issues are left for future research.[594]

Many economic models (including DSGE models) are inherently nonlinear and (possibly) difficult to solve. Therefore, perturbation methods are frequently used to approximate the original model around a specific point. That is, instead of solving the underlying nonlinear model, its Taylor series expansion at a specific point is considered.[595] The specific point is usually the non-stochastic steady state of the underlying nonlinear model.[596] The models presented in this chapter have a unique steady state around which such an approximation can be made.[597] Within the scope of this thesis, the models considered in this chapter are expressed in percentage deviations from their corresponding steady state. Then, a first-order (i.e., a linear) approximation around the steady state is made.[598] The resulting linear system is then solved along the method of Klein (2000).

In light of the growing literature dealing with higher-order approximations, the restriction to a first-order approximation seems restrictive. However, the order of the approximation is closely related to the aim of the model. As, for example, outlined by Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016), for a DSGE model including stochastic shocks with time-varying volatility, at least a third-order approximation is necessary.[599] If such elements are not part of the model, a first- or second-order approximation may be sufficient for modeling purposes. Moreover, in some circumstances, higher-order approximations need not be advantageous over first-order approximations, as they may be more computationally

[593] The resulting effects of introducing a long memory technology shock are discussed in Sections 4.4.2 and 4.4.3.
[594] Clearly, the solution method described in Appendix B appears to be useful to answer these questions.
[595] See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 540).
[596] See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 549).
[597] For details regarding the steady state value, see Section 4.1.4.
[598] To be more precise, the second model with Cobb-Douglas preferences is able to account for growth; thus, there is no steady state in this model. For this reason, the model is made stationary first. The stationarized variables then have a steady state around which the linearization is made.
[599] See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 560).

demanding and may lead to explosive paths even though the steady state is deemed stable in a first-order approximation.[600,601] On the other hand, linear approximations appear only as good approximations in a vicinity of the considered steady state, i.e., the dynamics induced by the linearized model outside of the steady state are only a good approximation for the underlying nonlinear model if the deviations from the steady state (shocks) are not too large.[602] However, the quality of a first-order approximation depends on the extent of the nonlinearity of the original model to be linearized.[603] In general, the error implied by the first-order approximation is small if the nonlinearities stem from concave (e.g., utility functions) or convex (e.g., production functions) functions typically used in DSGE models.[604] Conversely, a linear approximation may perform worse if the underlying model has, e.g., a kink such as that in the Taylor rule nearby the zero lower bound.[605] Since the nonlinearities in this chapter's models come primarily from applied utility and production functions, a first-order approximation seems reasonable. Moreover, a linearized DSGE model seems to be sufficient as the introduction of long memory does not increase the magnitude of the shock per se; instead, long memory primarily refers to a property of the process's autocorrelation function, and thus, long memory primarily controls how quickly the shock dissipates.[606,607] The notion "DSGE model" in this thesis is reserved for models formulated in discrete time. There are, of course, stochastic general equilibrium models incorporating dynamics formulated in continuous-time.[608] These are, however, not addressed in this chapter.[609] The motivation to consider a long memory DSGE model

600 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, pp. 560 and 567).

601 The concept of stability in the context of DSGE models is discussed in more detail in Section 4.2.3.

602 This is, among others, one critique against DSGE models stated, for example, by Stiglitz (2018, p. 75).

603 See Brzoza-Brzezina and Suda (2021, p. 242).

604 See Brzoza-Brzezina and Suda (2021, p. 242).

605 See Brzoza-Brzezina and Suda (2021, p. 242).

606 See Definition 2.3.1 and Lemma 2.3.2.

607 Of course, the consideration of long memory shocks in higher-order approximated DSGE models, may be interesting from the perspective of certainty equivalence to which one is restricted by a first-order approximation, see Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 555). Following Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 555), certainty equivalence refers to the property that a shock's standard deviation does not enter the households' decision rule, although the realization of the shock does. Hence, e.g., time-varying volatility or risk shocks cannot be captured by a first-order approximation. Related problems, whether, for example, households change their precautionary behavior when they face a long memory shock to another shock's volatility, appear to be interesting and are left to future research.

608 An example for an RBC-DSGE model formulated in continuous-time is given in Parra-Alvarez (2018, Section 3 on pp. 1565ff.). In his paper, Parra-Alvarez (2018) compares various solution methods, their computational costs, and their accuracy. A continuous-time New Keynesian DSGE model is considered, e.g., in Fernández-Villaverde, Posch, et al. (2012, Section 2 on pp. 3ff.). General solution methods for continuous-time rational expectations models can be found, e.g., in Sims (2002, Section 3 and 4 on pp. 5ff.) and Buiter (1984).

609 Note that in Chapter 5, long memory is also discussed in a continuous-time macro-financial model.

formulated in discrete time is the class of autoregressive fractionally integrated moving average (ARFIMA) processes described in Section 2.4. Since ARFIMA processes are formulated in discrete time and arise as a generalization of ARMA processes, which, as mentioned above, are incorporated quite frequently in discrete-time DSGE models, the extension of DSGE models with ARFIMA processes lies at hand. Furthermore, ARFIMA processes have shown their empirical importance, which motivates their consideration in the context of a discrete-time DSGE model.[610]

4.1.2 The Structure of the Model

In the following, the prototypical DSGE-RBC model framework is described. The following sections aim to derive the solution of the model under the assumption of a long memory technology shock and to examine the extent to which the presence of long memory changes the decisions of households and their induced effects on the economy compared to a short-memory shock and a trend shock. The model economy is similar to the one discussed in Hansen (1985, Section 3.1 on pp. 313ff.), King, Plosser, et al. (1988, Section 2 on pp. 198ff.) and Aguiar and Gopinath (2007, Section 3 on pp. 78ff.).

There is a closed economy populated by private households and firms. There is no government, no financial sector, and no (central) banks or other financial intermediaries. The households are the owners of the factors of production, namely labor, and capital. They supply labor on the labor market and rent capital to the firms on the market for capital. The firms demand factors of production on these markets and pay wages to the households for the hired labor and a rental rate on the used capital. In contrast to households, firms have access to a production technology that enables them to produce output goods from capital and labor inputs. Firms sell their output on the final goods market to households. Households can use them to satisfy their consumption needs or to invest in capital. An investment in capital increases the stock of capital available for rent to firms in subsequent periods and may (depending on the rental rate) increase the households' income.

Overall, there are three markets: the labor market, the market for capital, and the final goods market. All markets are assumed to be perfectly competitive. The firms' and households' decision problems are outlined in the following.

Firms. Firms produce output goods according to the constant-returns-to-scale Cobb-

[610] See the discussion in Section 3.2.

Douglas production function[611]

$$Y_t = A_t K_t^{\alpha} (\bar{A}_t H_t)^{1-\alpha}, \text{ with } \alpha \in (0,1), \tag{4.1}$$

where Y_t refers to the number of final goods produced, K_t to the amount of capital used in the production, and H_t to labor inputs. The parameter α refers to the output elasticity with respect to capital. There are two sources of technological progress. Purely labor augmenting technological progress is represented by $\bar{A}_t$, whereas productivity affecting both factors is given by A_t.

The reason for including two different types of productivity processes will become apparent soon. The underlying idea is as follows. Below, both productivity processes are specified to follow exogenous stochastic processes exhibiting different properties. Stochastic fluctuations with a transitory character are associated with A_t. To be more precise, A_t will be specified as a stationary stochastic process implying that the linearized model's solution turns out to be stationary as well, i.e., all model variables are stationary processes fluctuating around their non-stochastic steady state values. Thus, there is no growth in either variable. This structure allows one to compare the model's response to short and long memory processes.

In order to incorporate technological growth, $\bar{A}_t$ comes into play. However, introducing growth leads to a non-stationary model. In the absence of any stochastic shocks, $\bar{A}_t$ is assumed to grow at a constant rate.[612] Stochastic shocks are then assumed to alter the growth rate of $\bar{A}_t$. That is, a positive shock to the growth rate will permanently increase $\bar{A}_t$. In order to solve the model, the variables are made stationary by dividing the growing variables by $\bar{A}_t$. However, a balanced growth path is only compatible with purely labor augmenting technological progress.[613,614]

Overall, $\bar{A}_t$ is interpreted as a growing component of technological progress, whereas A_t is associated with a transitory productivity component that fluctuates around its steady state value. This specification offers the possibility for comparing the model's response to shocks with a permanent character and shocks with a transitory (either short or long memory) character. As the model with productivity growth contains the stationary model (by setting $\bar{A}_t \equiv 1$ or the growth rate of $\bar{A}_t$ to zero) as a special case, it seems parsimonious

611 The specification of the production function is adapted from King, Plosser, et al. (1988, Equation (2.6) on p. 200) and Aguiar and Gopinath (2007, Equation (1) on p. 78).

612 As outlined below, some other variables also grow at the same rate implying the existence of a balanced growth path.

613 See King, Plosser, et al. (1988, p. 200).

614 To ensure a balanced growth path, there have to be imposed additional restrictions on the households' utility function, which are discussed below.

to formulate the production function as in (4.1) to cover both cases simultaneously.

Firms demand labor and capital as factors of production. They have to pay (nominal) wages W_t^n and a (nominal) rental rate on capital R_t^n to the households for hiring workforce and renting capital units, respectively. The final goods are sold to households at a price of P_t. Thus, the firm has to decide how much work and capital to demand in order to maximize its profits; thus, they face the following maximization problem

$$\max_{H_t,K_t} P_tY_t - W_t^nH_t - R_t^nK_t = \max_{H_t,K_t} P_tA_tK_t^{\alpha}(\bar{A}_tH_t)^{1-\alpha} - W_t^nH_t - R_t^nK_t. \tag{4.2}$$

Households. Households are confronted with various decisions. First, they face an investment-consumption decision, i.e., they have to decide how much of the final goods are purchased for consumption C_t or investment purposes I_t. Second, they have to decide how much time they supply on the labor market, how much time to devote to leisure L_t, and how much capital to rent to firms. The households' preferences are expressed by a utility function $U(C_t, H_t)$, i.e., households are assumed to gain utility from consumption and suffer from the disutility of supplying labor. Special forms of the utility function are expressed below, but it is assumed that U is twice differentiable satisfying[615]

$$\frac{\partial U}{\partial C_t} > 0, \quad -\frac{\partial U}{\partial H_t} > 0, \quad \frac{\partial^2 U}{\partial^2 C_t} \leq 0, \quad -\frac{\partial^2 U}{\partial^2 H_t} \geq 0, \quad \frac{\partial^2 U}{\partial^2 H_t}\frac{\partial^2 U}{\partial^2 C_t} - \left(\frac{\partial^2 U}{\partial C_t \partial H_t}\right)^2 > 0.$$

These conditions state that households' utility is concave and increasing in consumption, but at a decreasing rate, and that the marginal disutility of labor, given by $-\partial U/\partial H_t$, is positive and non-decreasing.[616]

In addition, households are assumed to make their consumption, investment, labor, and capital supply decisions to maximize their expected discounted future utility. Thus, the households solve the following utility maximization problem

$$\max_{C_t,H_t,K_{t+1}} \mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} U\left(C_s, H_s\right), \tag{4.3}$$

where $0 < \beta < 1$ refers to the households' discount rate or time preference rate. Due to the stochastic processes to be specified for the productivity processes, the future paths of consumption and leisure are uncertain, implying the need for households to form expectations. This need is expressed in the operator $\mathbb{E}_t$, which is a shorthand

615 See Greenwood et al. (1988, p. 404) or Galí (2008, p. 16).
616 See, e.g., Galí (2008, p. 16).

notation for the mathematical expectation operator $\mathbb{E}\left(\cdot|\mathcal{F}_t\right)$, where $\mathcal{F}_t$ refers to the information set available to households at time t. That is, households are assumed to form their expectations rationally.[617] As will turn out in Section 4.2.3 (see (4.33)), the model's solution depends to a large extent on the expectations of the exogenous stochastic processes. Thus, an interesting question to be addressed is whether the households change their consumption and labor supply decision in light of a long memory technology shock. One might expect that the households know the persistence properties of the exogenous process. Thus, it seems likely that they take the slowly dissipating effect of a long memory shock into account and change their behavior accordingly. Such questions are answered in Section 4.4.

Moreover, households face various restrictions and constraints. First, households cannot go into debt; thus, they face a budget restriction limiting their consumption and investment expenditures to the wage and capital income, i.e.,

$$P_t\left(C_t + I_t\right) = W_t^n H_t + R_t^n K_t. \tag{4.4}$$

Second, since households rent their capital goods to firms, it is assumed that there is some wear and tear on the capital goods used in the production process. Thus, households depreciate the capital goods with a constant depreciation rate $\delta > 0$. Depreciated capital cannot be used in the production process of the following periods. However, investment spending increases the available capital goods for rent in the next periods; thus, the accumulation equation of capital is given by

$$K_{t+1} = (1-\delta)K_t + I_t. \tag{4.5}$$

Third, households are time-constraint. They cannot supply arbitrary amounts of labor or consume arbitrary amounts of leisure. Each household has a time endowment that is normalized to one. Time not spent for work is assumed to be spent for leisure, i.e.,

$$H_t = 1 - L_t. \tag{4.6}$$

Although H_t is just proportional to the hours worked by the households (see Footnote 619), during this thesis, it is referred to H_t as hours worked or working hours.[618,619]

617 For more details on the information set and its role in the solution method of Blanchard and Kahn, see Footnote 659.

618 This is in line with, e.g., Lindé (2009, p. 601).

619 To gain some intuition for plausible values of H_t, consider a year with 365 days containing 220 working days (after subtracting weekends, public holidays, etc.). If a household works eight hours daily, this

Market clearing. It is assumed that all three markets, the labor market, the market for capital, and the final goods market, clear at each instance of time. The latter implies

$$Y_t = C_t + I_t. \tag{4.7}$$

The equilibrium wage and the rental rate on capital equate labor and capital demands with their respective supplies.

Equilibrium. In order to consider the equilibrium of the model economy, the price level of final goods P_t is eliminated, and the model's variables are expressed in real terms. Thus, dividing the nominal wage and the nominal rental rate on capital by the price level leads to the definition of the real wage (W_t) and the real rental rate on capital (R_t), i.e.,

$$W_t := \frac{W_t^n}{P_t}, \quad R_t := \frac{R_t^n}{P_t}. \tag{4.8}$$

An equilibrium of the economy is then defined as the set of real factor price processes, the households consumption, investment, and labor plan, the levels of production and capital,

$$\left\{(W_t)_{t\in\mathbb{Z}}, (R_t)_{t\in\mathbb{Z}}, (C_t)_{t\in\mathbb{Z}}, (I_t)_{t\in\mathbb{Z}}, (L_t)_{t\in\mathbb{Z}}, (H_t)_{t\in\mathbb{Z}}, (Y_t)_{t\in\mathbb{Z}}, (K_t)_{t\in\mathbb{Z}}\right\}, \tag{4.9}$$

such that households maximize their discounted future utility, firms maximize their profits, and the markets for goods, labor, and capital clear, given the exogenous productivity processes $(A_t)_{t\in\mathbb{Z}}$ and $(\bar{A}_t)_{t\in\mathbb{Z}}$.

Summary of the equilibrium. The stated model economy satisfies the second welfare theorem; thus, as all assumed households and firms are assumed to be homogeneous and to operate equally on fully competitive markets, the economy's equilibrium can be derived by focusing on a representative household and firm.[620]

Applying a Lagrangian, the solution to the representative household's maximization problem can be summarized by the following two equations. A detailed derivation of the following equations can be found in Appendix B.1. Let $\partial U(C_t, H_t)/\partial X = U_{X,t}$, then

will correspond to $H_t = 220 \cdot 8/(365 \cdot 24) \approx 0.2$. Under the assumption that working and leisure days are equally distributed over a year, this value is also valid if time is measured in quarters. A value of $H_t = 0.2$ was also considered plausible by King, Plosser, et al. (1988, Footnote 28 on p. 215). However, other values are discussed in the literature ranging from 0.2 up to nearly 0.6 depending on how the time endowment is measured and how time for sleep and housework is handled, see, for example, Gomme and Lkhagvasuren (2013, p. 583). If the total amount of hours supplied by a household is restricted to $220 \cdot 24$, then a reasonable value seems to be $H_t = 220 \cdot 8/(220 \cdot 24) = 1/3$.

620 See King, Plosser, et al. (1988, p. 200) and Hansen (1985, p. 313 and p. 314).

optimal labor supply is given by

$$U_{C,t} W_t = -U_{H,t}. \tag{4.10}$$

The Euler equation is given by

$$U_{C,t} = \beta \mathbb{E}_t \left[U_{C,t+1}((1-\delta) + R_{t+1}) \right] \tag{4.11}$$

Solving the profit maximization problem of the representative firm leads to the following equations. Demand for labor is determined by

$$W_t = (1-\alpha)\frac{Y_t}{H_t} \tag{4.12}$$

and demand for capital by

$$R_t = \alpha \frac{Y_t}{K_t}. \tag{4.13}$$

Both equations yield the well-known fact that in competitive markets, marginal factor products equal real factor prices. Equations (4.1), (4.5) to (4.7) and (4.10) to (4.13) correspond to the eight unknowns of the model given in (4.9). To complete the model, the exogenous productivity processes and the specific form of the household's utility function need to be specified.

4.1.3 Functional Forms of the Productivity Processes and Preferences

This section specifies the functional forms of the two exogenous productivity processes A_t and $\bar{A}_t$ as well as the utility function of the representative household.

4.1.3.1 Functional Forms of the Productivity Processes

As already mentioned, exogenous transitory changes of total factor productivity are often described by a stationary AR(1) process with a positive persistence coefficient.[621] As outlined in Section 2.1.5 and Section 2.3, this assumption corresponds to a short memory process and a moderately persistent process. Here, an ARFIMA$(1, d, 0)$ process is used

[621] See, among many others, Hansen (1985, Equation (4) on p. 313), King, Plosser, et al. (1988, Equation (4.1) on p. 212), Aguiar and Gopinath (2007, Equation (2) on p. 78), Smets and Wouters (2007, p. 589), Cantore, Gabriel, et al. (2013a, Equation (18.14) on p. 416), Lindé et al. (2016, Equation (5) on p. 2193).

instead to introduce a long memory and strongly persistent productivity process.[622,623]

Recall that the motivation for using an ARFIMA process is twofold. First, as outlined in Chapter 3, ARFIMA processes are relevant from an empirical perspective. This empirical relevance may facilitate the estimation of a long memory DSGE model in the future, as the methods available for ARFIMA processes may be applicable to the estimation of such a model. Second, since ARFIMA processes include ARMA processes as a special case, one obtains an extension of the existing DSGE model that further allows for better comparability than if this nesting is not preserved.

To be more precise, it is assumed that the transitory component of productivity satisfies

$$(1-\varrho_A B)\left(\log(A_t)-\log(A_{ss})\right)=(1-B)^{-d}\varepsilon_t^A, \tag{4.14}$$

where $|\varrho_A| < 1$ is the short memory parameter and $0 \leq d < 1/2$ is the long memory parameter.[624] The process ε^A is assumed to be a Gaussian white noise process with standard deviation σ_{ε^A}.[625] The expected value or steady state value of $\log(A_t)$ is denoted by $\log(A_{ss})$ and is an exogenously given constant.

Similar to (2.34), (4.14) can be expressed as

$$(1-\varrho_A B)\left(\log(A_t)-\log(A_{ss})\right)=\nu_t^A, \tag{4.15}$$

where $\nu_t^A := (1-B)^{-d}\varepsilon_t^A$ is a fractionally integrated white noise process of order $0 < d < 1/2$. This rearrangement provides some advantages that materialize during the model solution stated in Appendix B.5. It states that one can interpret $\log(A_t)$ as an AR(1) process with

622 Recall from Table 2.1 that an ARFIMA process is a strongly persistent long memory process only if $0 < d < 1/2$.

623 It would further be possible to consider a more complicated structure of short memory components, i.e., to consider an ARFIMA(p,d,q) process with $p > 1$ and $q > 0$. This thesis does not consider such alternatives as they seem uncommon in the literature, where transitory productivity shocks are widely assumed to follow an AR(1) process. Are more detailed discussion regarding this issue is given in Section 4.3. There, it is shown that some authors propose to use more general short memory processes for the exogenous processes in the context of DSGE models. As this thesis aims to investigate the impact of long memory on the model's outcomes, additional investigations of interactions between long memory and a richer short memory structure are left for future research. In addition, a more complex short memory structure of the variables may be implemented endogenously by considering a richer model structure generating persistence such as consumption habits or investment adjustment costs, etc. (see, e.g., Cantore, Gabriel, et al. (2013a, p. 425)). The short memory structure implied by the model contrasts sharply with long memory, which does not appear to be endogenizable in a standard linearized DSGE model. Details on this are carried out in Section 4.2.2.

624 In order to avoid negative values of productivity, (4.14) is written in terms of the logarithm of A_t instead of A_t directly. Doing so is in line with, e.g., Lindé et al. (2016, Equation (5) on p. 2391).

625 Recall Footnote 53 for the definition of a Gaussian white noise process.

respect to the process ν^A instead of ε^A.

The growing component of technological progress is assumed to follow[626]

$$\bar{A}_t = (1 + g_t)\bar{A}_{t-1} = \bar{A}_0 \prod_{u=1}^{t} (1 + g_u), \text{ with initial value } \bar{A}_0, \tag{4.16}$$

where the growth factor $b_t := (1 + g_t)$ evolves according to[627]

$$(1 - \varrho_g B)(\log(b_t) - \log(b_{ss})) = \varepsilon_t^g. \tag{4.17}$$

Similar to (4.14), ε^g is a Gaussian white noise process with standard deviation σ_{ε^g} independent from ε^A.[628] Again, $b_{ss} := \log(1 + g_{ss})$ refers to the expected value of $\log(1 + g_t)$, i.e., according to Footnote 628, g_{ss} is approximately the expected value of the growth rate g_t. As the shocks ε_t^g affect the logarithm of the growth factor of $\bar{A}_t$, they are called growth shocks in this thesis.[629]

Lemma 2.4.3 implies that the impulse-response function (IRF) of an ARFIMA$(1, d, 0)$ process given in (4.14) converges to zero. Hence, shocks to A_t have to be deemed transitory. The same holds for the IRF of the AR(1) process specified in (4.17). Concerning the growth factor b_t, the shocks ε^g are also transitory. The growth shocks, however, alter the level of productivity $\bar{A}_t$ given in (4.16) permanently, i.e., for the level of productivity, the shock ε^g has to be deemed permanent.[630]

4.1.3.2 Functional Forms of the Utility Functions

As can be seen from (4.3), all future periods enter into the household's decision problem raising the presumption that shocks altering the level of productivity permanently will influence the household's decision differently than a quickly decaying shock. That this is indeed the case is already well documented for the comparison between short memory and permanent productivity shocks.[631] In this thesis, the third case, namely long memory

---

626 This specification is inspired by Aguiar and Gopinath (2004, p. 10) and Aguiar and Gopinath (2007, p. 80).

627 Again, the formulation in logarithms is used to avoid negative values of the growth factor $1 + g_t$. However, doing so does not rule out negative values of the growth rate g_t itself. In order to simplify notation and phrasing, the term "growth" is also used in cases of negative growth rates that are associated with a decay of $\bar{A}_t$.

628 Note that $\log(1 + x) \approx x$ for small values of x. Thus, (4.17) states that the growth rate g_t follows approximately an AR(1) process.

629 This is in accordance with Aguiar and Gopinath (2007, p. 80).

630 See Aguiar and Gopinath (2007, p. 80).

631 See, for example, Aguiar and Gopinath (2007, Figure 3 on pp. 88ff.). Part b of Figure 3 in Aguiar and Gopinath (2007) illustrates the different responses of the consumption-to-GDP ratio to a one

productivity shocks to the transitory component A_t, is added to this discussion. It may also be apparent that the household's decision also depends on its preferences. To account for this dependence, two different functional forms of the household's utility function are considered in the following.

In the first scenario, it is abstracted from growth in the model, i.e., $\bar{A}_t \equiv 1$, and the single source of uncertainty in total factor productivity (TFP) is the transitory component A_t specified in (4.14). In this setting, the utility function is assumed to be additive separable in consumption and hours worked, i.e.,[632]

$$U(C_t, H_t) = \frac{C_t^{1-\varsigma}}{1-\varsigma} - \varkappa \frac{H_t^{1+\varphi}}{1+\varphi}, \quad \varsigma, \varkappa, \varphi > 0. \tag{4.18}$$

The specification (4.18) states that the representative household gets utility from consumption but suffers from the disutility of supplying labor. Utility from consumption is given as a standard constant relative risk aversion utility function, where the relative risk aversion (by simultaneously holding H_t constant) is given by the parameter ς.[633]

Additionally, $1/\varsigma$ is the intertemporal elasticity of substitution.[634] The same functional form is assumed for the disutility of labor. Moreover, utility function (4.18) belongs to the family of utility functions with a constant Frisch elasticity of labor supply which can be pinned down by the parameter φ.[635] To be more precise, the Frisch elasticity of labor supply is defined as the real wage elasticity of labor supply by keeping the marginal utility

percent growth and transitory technology shock. Lindé (2009, Section 3 on pp. 600ff.) analyzes the effect of permanent and transitory short memory AR(1) technology shocks in a standard RBC model.

632 This specification is also common in New Keynesian DSGE models, see, e.g., Clarida et al. (2002, Equation (6) on p. 882), Christoffel and Kuester (2008, Equation (2) on p. 867) or Galí (2008, p. 17).

633 Following Chetty (2006, p. 1822), the constant relative risk aversion by holding the labor supply fixed, is given by $-C\,U_{CC}(C,H)/U_C(C,H)$, where U_C and U_{CC} are the first-order and second-order partial derivatives of U with respect to C. Plugging (4.18) into the mentioned definition, it is easy to see that the relative risk aversion is constant and given by ς. However, the risk aversion of the household is smaller than ς if labor supply is assumed to be variable such as in the model considered here, see Chetty (2006, p. 1822 and Appendix A on p. 1831). This point is also highlighted by Swanson (2012, Example 2 on pp. 1675ff.). He showed that the risk aversion of the household with utility function (4.18) is $\varsigma\,(1/(1+\varsigma/\varphi))$ given the additional assumption that consumption is equal to the household's entire labor earnings. This expression is clearly smaller than ς, see Swanson (2012, p. 1676). That a variable labor supply reduces the relative risk aversion of the household seems plausible since the reduction or increase of labor supply provides an additional channel for the household to react to wealth shocks. This additional possibility reduces the household's relative risk aversion; see Chetty (2006, pp. 1823ff.).

634 See, e.g., Swanson (2012, Footnote 1 on p. 1664). It follows from the discussion in Footnote 633 that for the utility function (4.18), the relative risk aversion is not reciprocal to the intertemporal elasticity of substitution when the household's labor supply is variable, see Swanson (2012, Footnote 1 on p. 1664).

635 See R. W. Evans and K. L. Phillips (2018, Section 2.3 on pp. 517f.).

of consumption constant.[636] It is shown in Appendix B.2 that the parameter $\varphi > 0$ is equal to the inverse of the Frisch elasticity of labor supply. Thus, a higher value of φ corresponds to a smaller value of the Frisch elasticity, implying that the household adjusts its labor supply less elastically to an increase in real wages given the assumption of a constant level of consumption.[637,638]

The scaling parameter $\varkappa$ may be seen as a relative measure for the distaste of supplying labor relative to the taste for consumption.[639]

The presence of growing technological progress induces the model to have a balanced growth path. Consequently, utility function (4.18) cannot be used because it is incompatible with such a balanced growth path.[640] Thus, in the second scenario, where growth is considered, it is assumed that the household's utility function is of Cobb-Douglas type, i.e.,[641]

$$U(C_t, H_t) = \frac{(C_t^{1-\gamma}(1-H_t)^{\gamma})^{1-\tau}}{1-\tau}, \quad 0 < \gamma < 1, \quad \tau > 0. \tag{4.19}$$

The parameter γ represents the weight of leisure in the utility function.[642] The parameter $1/\tau$ is the elasticity of intertemporal substitution of the composite good $C_t^{1-\gamma}(1-H_t)^{\gamma}$.[643,644] The Frisch elasticity of labor supply depends, in contrast to the utility function (4.18), on the actual amount of labor supplied by the household.[645]

636 See R. W. Evans and K. L. Phillips (2018, p. 521) and Christiano et al. (2010, p. 299).

637 Since $U_{C,t}(C_t, H_t) = C_t^{-\varsigma}$, one can equally assume that the level of consumption is constant instead of the marginal utility of consumption. Doing so is in line with Christiano et al. (2010, Footnote 6 on p. 299).

638 In light of Footnote 633, it seems not surprising that Swanson's formula for the relative risk aversion depends on the Frisch elasticity $1/\varphi$ as well.

639 See Heathcote et al. (2008, p. 511). Note that this parameter is sometimes normalized to unity as in Heathcote et al. (2008, p. 511) or Galí (2008, p. 17). However, in Section 4.4.1, the parameter is used for calibrating the steady state value of hours worked in the model. Hence, it seems useful to keep this additional degree of freedom.

640 To be more precise, King, Plosser, et al. (1988, Equations (2.8a) and (2.8b) on p. 202) show that the utility function has to be of the following form in order to be compatible with a balanced growth path

$$U(C_t, H_t) = \begin{cases} \dfrac{1}{1-\varsigma} C_t^{1-\varsigma} v(H_t) & \text{if } 0 < \varsigma < 1 \text{ or } \varsigma > 1 \\ \log(C_t) + v(H_t) & \text{if } \varsigma = 1 \end{cases},$$

where some additional restrictions on the function $v(\cdot)$ has to be imposed, King, Plosser, et al. (1988, Footnote 11 on p. 202).

641 See Aguiar and Gopinath (2007, Equation (3) p. 80), Lindé (2009, Equation (7) on p. 600) and Cantore, Gabriel, et al. (2013a, Equation (18.16) on p. 416).

642 See Cooley and Prescott (1995, p. 16).

643 See, e.g., Bechlioulis and Brissimis (2021, p. 107).

644 Note that for the utility function (4.19), τ further refers to the relative risk aversion, see Swanson (2012, p. 1675).

645 To be more precise, the Frisch elasticity is given by $-((1-\gamma)(1-\tau)-1)/\tau)(1-H_t)/H_t$, see Bechlioulis

4.1.4 The Non-Stochastic Steady State and Balanced Growth Path

The model's non-stochastic or deterministic steady state described in the previous section is defined as the model's equilibrium under the assumption that the agents do not anticipate future shocks.[646] Thus, all exogenous stochastic shocks and their expectations are set to zero.[647] Further, the steady state (SS) of the resulting deterministic model can be found by setting $X_{t+1} = X_t = X_{ss}$ for each variable X_t, where X_t is a representative for any of the model's variable. The variable subscript ss refers then to its steady state value.

If technological growth is considered, it follows from (4.17) that labor augmenting technological progress $\bar{A}_t$ will grow at the constant rate g_{ss} implying a balanced growth path, where output Y_t, the capital stock K_t, consumption C_t, investment I_t grow at rate g_{ss} as well.[648] Note that hours worked H_t are bounded by 1, implying that they cannot grow over time; hence their growth rate is zero.[649]. In addition, (4.13) and (4.12) imply that the real rental rate on capital is constant in the steady state, but the real wage W_t grows at rate g_{ss}. Since the condition for a steady state value $X_{t+1} = X_t = X_{ss}$ cannot be satisfied in the presence of growth, the variables have to be rescaled to establish a constant steady state of the rescaled variables.[650]

As the growth in the model comes from the labor augmenting technology process $\bar{A}_t$, the model is made stationary by dividing through $\bar{A}_t$.[651] Let the stationarized variables be denoted by lowercase letters. They are given as[652]

$$y_t \coloneqq \frac{Y_t}{\bar{A}_t}, \quad c_t \coloneqq \frac{C_t}{\bar{A}_t}, \quad i_t \coloneqq \frac{I_t}{\bar{A}_t}, \quad w_t \coloneqq \frac{W_t}{\bar{A}_t}, \quad k_{t+1} \coloneqq \frac{K_{t+1}}{\bar{A}_t}. \tag{4.20}$$

As illustrated in Appendix B.4.2, these stationarized variables have then a constant steady state value.

After the steady state is determined, both models are log-linearized around their deterministic steady state values. Thus, the model equations are expressed in terms of the

and Brissimis (2021, p. 107).

646 See Coeurdacier et al. (2011, p. 398).

647 In the model, this is equal to $\nu_t^A \equiv 0$ and $\varepsilon_t^g \equiv 0$ in (4.15) and (4.17), respectively.

648 See King, Plosser, et al. (1988, p. 201) and, for more details, the derivation in their technical appendix King, Plosser, et al. (2002, p. 92).

649 See King, Plosser, et al. (1988, p. 201).

650 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 544).

651 See Aguiar and Gopinath (2007, p. 80) or Cantore, Gabriel, et al. (2013a, p. 418).

652 Note that the capital stock in period $t+1$, K_{t+1}, is rescaled by $\bar{A}_t$ instead of $\bar{A}_{t+1}$ as it is already determined in period t due to (4.5). Doing so is in line with Lindé (2009, p. 600).

variables' percentage deviation from their respective steady state, i.e., a variable X_t is replaced with $\tilde{X}_t$, where the latter is defined as[653]

$$\tilde{X}_t := \log(X_t) - \log(X_{ss}). \tag{4.21}$$

The linearization was done using the method of Uhlig (1999).[654]

Appendix B.3 summarizes the model without growth and additive separable utility function. More specifically, Appendix B.3.1 contains the nonlinear model equations and Appendix B.3.2 the corresponding steady state values of the model's variables. Appendix B.3.3 summarizes the linearized equations, and Appendix B.3.4 provides the matrices for the model's canonical form needed for solving the model. Similarly, Appendix B.4 summarizes the model with labor augmenting technological growth and Cobb-Douglas (CD) utility function. More specifically, Appendix B.4.1 illustrates the stationarized nonlinear model equations and Appendix B.4.2 the corresponding steady state values of the stationarized variables. Appendix B.4.3 summarizes the linearized equations, and Appendix B.4.4 provides the matrices for the model's canonical form needed for solving the model.

The log-linearized models are solved with the method of Klein (2000) described in Appendix B.5. Before the models' solutions are considered in terms of an impulse-response analysis, the following section summarizes some facts about the solution methodology of linearized DSGE models containing notes on the memory property of a model's solution and its stability.

4.2 On the Structure of Linearized DSGE Models

This section delineates the underlying theoretical background containing the general model structure, the solution's form and memory properties of conventional linear DSGE models, and the concept of stability. This section's character is rather descriptive, illustrating some key concepts of linearized DSGE models. A more rigorous and mathematical consideration is given in Appendix B.5. There, the arguments of this section are seized and reinforced with the necessary equations again.

653 Note that $\tilde{X}_t$ is approximately the percentage deviation of X_t from its steady state value due to $\tilde{X}_t = \log(X_t/X_{ss}) = \log((X_t - X_{ss})/X_{ss} + 1) \approx (X_t - X_{ss})/X_{ss}$.

654 The method is described in Uhlig (1999, Section 3.3 on pp. 32ff.).

4.2.1 The Solution Methodology of Blanchard and Kahn

In the end, the equations of a linearized or linear DSGE model form a set of stochastic difference equations. In order to solve this system, various solution methods are proposed in the literature.[655] A common feature of the solution methods suggested in the literature is that they require the model to be in a particular form, called the canonical form of the DSGE model. This canonical form summarizes the model in a single equation involving specific matrices.

A well-established solution method of linear DSGE models is the one proposed by Blanchard and Kahn (1980). Their method requires the model to fit the following canonical form[656,657]

$$\mathbb{E}_t x_{t+1} = \mathbf{A}_0 x_t + \mathbf{G}_0 z_t, \tag{4.22}$$

where the vector x_t is of size $n \times 1$ and contains all model variables, and n is the number of variables in the model. The matrix $\mathbf{A}_0$ is of size $n \times n$ and the matrix $\mathbf{G}_0$ is of size $n \times n_z$, where n_z is the number of exogenous stochastic shocks specified in the $n_z \times 1$ vector z_t. The elements of the matrices $\mathbf{A}_0$ and $\mathbf{G}_0$ are functions of the model parameters and the variables' steady state values.[658] Again, $\mathbb{E}_t x_{t+1} = \mathbb{E}(X_{t+1}|\mathcal{F}_t)$ refers to the conditional expectation of x_{t+1} formed at time t given the information set $\mathcal{F}_t$ available at time t.[659]

[655] A non-exhaustive list of frequently used methods is given in the following. McCallum (1983, Section 3 on pp. 145ff.) and Uhlig (1999, Section 3.4 on pp. 35ff.) use the method of undetermined coefficients. Moreover, McCallum (1983) uses a set of minimal state variables to determine a unique solution. Binder and Pesaran (1995, Section 2.3 on pp. 149ff.) use a quadratic determinantal equation method. Meyer-Gohde (2010, Section 3 on pp. 986ff.) derives a system of difference equation in the model's infinite moving average coefficients, including models with many expectational lags. The method of Klein (2000) generalizes the method of Blanchard and Kahn (1980) and is discussed in Appendix B.5. The method of Sims (2002, Section 3 and 4 on pp. 5ff.) seems to be more general than the ones of Blanchard and Kahn (1980) and Klein (2000) and applies to continuous-time models as well. Buiter (1984)'s method is an extension of Blanchard and Kahn's solution method to continuous-time rational expectations models.

[656] See Blanchard and Kahn (1980, Equation (1a) on p. 1305).

[657] Other solution methods discussed in the literature may require different but possibly more general canonical forms of the DSGE model, see for example Klein (2000, p. 1408), Sims (2002, p. 1) or Binder and Pesaran (1995, pp. 141f.).

[658] See the examples given by Blanchard and Kahn (1980, pp. 1036f.).

[659] The model includes exogenous stochastic processes, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as described in Definition 2.1.1. These processes are stored in the vector z_t. Consequently, the model solution is generally also a stochastic process. The information set $\mathcal{F}_t$ at time t can be formally defined as a sigma algebra or sigma field over Ω generated by x_t (and possibly z_t), as described by Binder and Pesaran (1995, p. 140 and Footnote 3) or Bårdsen and Fanelli (2015, p. 309). As time progresses, there is a sequence of sigma algebras $(\mathcal{F}_t)_{t\in\mathbb{Z}}$, also called a filtration, which satisfies $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. Common solution methods require that the exogenous stochastic process be adapted to this filtration, meaning that z_t has to be $\mathcal{F}_t$-measurable, as outlined in Binder and Pesaran (1995, p. 140) or Klein (2000, p. 1408). This condition is satisfied if the information set under consideration is generated by the processes x_t and z_t.

A vector x_t satisfying (4.22) at each instance of time is called a solution to the model. Later, it is shown that the models presented in the previous section cannot be cast in the canonical form (4.22). Therefore, the method of Blanchard and Kahn (1980) is not suitable for the purposes of this thesis. Instead, the solution method of Klein (2000) is used and described in more detail in Appendix B.5. However, for the illustrative purpose of this section, it seems appropriate to discuss the key aspects along the simpler structure (4.22) rather than the more general approach of Klein (2000).[660]

Blanchard and Kahn (1980) divide the variables of the model dichotomously in the groups of predetermined and non-predetermined variables.[661] Predetermined variables have an exogenously given initial value x_0^p and, at time $t+1$, they can be expressed as a function of variables "known" at time t, i.e., $\mathbb{E}_t x_{t+1}^p = x_{t+1}^p$, where x_t^p refers to the part of the vector x_t containing the predetermined variables.[662,663]

Let x_t^{np} the part of the vector x_t containing the non-predetermined variables, then (4.22) can equally be written as

$$\begin{pmatrix} x_{t+1}^p \\ \mathbb{E}_t x_{t+1}^{np} \end{pmatrix} = \mathbf{A}_0 \begin{pmatrix} x_t^p \\ x_t^{np} \end{pmatrix} + \mathbf{G}_0 z_t. \tag{4.23}$$

Linear rational expectations models have, in general, many solutions.[664] In order to restrict the set of possible solutions, some additional restrictions on the exogenous stochastic process z_t and the solution of the model have to be imposed to pin down a unique solution process.[665]

Blanchard and Kahn assume that the expectations of a solution x_t and the exogenous stochastic process z_t do not grow without bounds.[666] To be more precise, they impose that for each $t \in \mathbb{Z}$, there are constants $\theta_t^\zeta \in \mathbb{R}$ and $M_t^\zeta \in \mathbb{R}^{n_\zeta}$, where n_ζ is the length of

660 As will turn out in Appendix B.5, the method of Klein (2000) is a direct generalization of the method of Blanchard and Kahn (1980), see Klein (2000, pp. 1406f.).

661 See Blanchard and Kahn (1980, p. 1305).

662 See Blanchard and Kahn (1980, p. 1305).

663 An example of a predetermined variable in the model of the previous section is the capital stock specified in (4.5) or its linearized version given in Table B.3. Given an exogenous initial value, e.g., the steady state value $\tilde{K}_0 = 0$, one has $\mathbb{E}_t \tilde{K}_{t+1} = \mathbb{E}_t \left((1-\delta)\tilde{K}_t + \delta \tilde{I}_t\right) = (1-\delta)\tilde{K}_t + \delta \tilde{I}_t = \tilde{K}_{t+1}$.

664 This is illustrated in Gourieroux, Laffont, et al. (1982) along a simple univariate linear rational expectations model. The authors classify the complete set of solutions in their framework, see Gourieroux, Laffont, et al. (1982, p. 416).

665 See Gourieroux, Laffont, et al. (1982, p. 416) and Gourieroux, Laffont, et al. (1982, Section 3 on pp. 416ff.) for a discussion of various selection criteria in the context of linear univariate rational expectations models.

666 See Blanchard and Kahn (1980, p. 1305).

the vector ζ_t, such that[667]

$$|\mathbb{E}_t \zeta_{t+j}| \leq (1+j)^{\theta_t^\zeta} M_t^\zeta, \text{ for all } j \geq 0 \text{ and } \zeta \in \{x, z\}. \tag{4.24}$$

Equation (4.24) may be considered a kind of stability condition for the model's solution.[668,669] To be more precise, Blanchard and Kahn decouple the system (4.23) into two subsystems. The first subsystem is associated with the stable eigenvalues (i.e., the eigenvalues smaller than one in modulus) of $\mathbf{A}_0$, and the second is associated with the unstable eigenvalues (i.e., the eigenvalues greater than one in modulus) of $\mathbf{A}_0$.[670] Condition (4.24) then enables the determination of a unique solution to the unstable subsystem that satisfies (4.24) again.[671]

It remains is to find a unique solution to the stable subsystem. From a deterministic system of difference equations, one would expect the unique stable solution to be pinned down by certain initial values. The same holds true in this context, but the initial values of the stable subsystem are not given directly. They have to be determined from the given initial values of the predetermined variables. In order to find them, the initial values of the stable subsystem are linked by a system of equations to the initial values of the predetermined variables.[672] It depends on this system of equations whether or not the initial values of the stable subsystem are uniquely determined. In the end, if there are "too many" predetermined variables, the system for determining the initial values of the stable subsystem is overdetermined; thus, it has (in most cases) no solution.[673] On the other hand, if there are "not enough" predetermined variables, the system for determining the initial values of the stable subsystem is underdetermined, and many solutions will exist.[674] If both stable and unstable subsystems are uniquely solved, there is a unique solution to the whole system (4.23).[675]

To be more precise regarding the notion of "too many" or "not enough" predetermined variables, Blanchard and Kahn state that the model (4.22) has a unique solution if

[667] See Blanchard and Kahn (1980, Equation (1c) on p. 1305 and p. 1307).
[668] See King and Watson (1998, p. 1020).
[669] This point will become clear in the context of Klein's solution method discussed in Appendix B.5.1. Klein explicitly defines the notion of a stable solution, and it can be shown that each process satisfying Klein's stability condition also satisfies (4.24).
[670] See Blanchard and Kahn (1980, p. 1307).
[671] See Blanchard and Kahn (1980, p. 1310 and especially Equation (A4)).
[672] This is equation (A5) of Blanchard and Kahn (1980, p. 1310).
[673] See Blanchard and Kahn (1980, p. 1310).
[674] See Blanchard and Kahn (1980, p. 1310).
[675] Note, that in either case, it is (4.24) that allows Blanchard and Kahn to pin down the unique solution to the unstable system, see Blanchard and Kahn (1980, p. 1310).

the number of unstable eigenvalues of $\mathbf{A}_0$ is equal to the number of non-predetermined variables.[676] On the other hand, the model is indeterminate, i.e., there is an infinite number of model solutions if there are more non-predetermined variables than unstable eigenvalues of $\mathbf{A}_0$.[677] Finally, there is no solution to the model satisfying (4.24) if the number of unstable eigenvalues of $\mathbf{A}_0$ exceeds the number of non-predetermined variables.[678] These conditions ensuring the circumstances for which there is a unique solution to the linear DSGE model are sometimes called "Blanchard-Kahn conditions"[679].[680]

Since the matrix $\mathbf{A}_0$ and its eigenvalues depend on the model parameters, the question of the existence of a unique solution is also parameter dependent. Moreover, the existence of the solution is typically proofed numerically for a given set of parameters, as it is, in general, computationally demanding to calculate the eigenvalues of $\mathbf{A}_0$ analytically.[681]

4.2.2 Memory Properties of the DSGE Model's Solution

Assuming a unique solution of the model, the solution depends to a large extent on the conditional expectations of the exogenous stochastic process.[682] The exogenous stochastic process z_t is, however, commonly assumed to be an n_z dimensional white noise process, i.e., $z_t = \varepsilon_t$ with $\mathbb{E}\varepsilon_t = 0_{n_z \times 1}$, $\mathbb{E}\varepsilon_t \varepsilon_t^T = \Sigma_\varepsilon$ and $\mathbb{E}\varepsilon_t \varepsilon_s^T = 0_{n_z \times n_z}$, or a first-order multivariate autoregressive process, i.e., $z_t = \Lambda_z z_{t-1} + \varepsilon_t$.[683,684]

In such cases, the solutions of a wide range of DSGE models can be cast into a state-

[676] See Blanchard and Kahn (1980, Proposition 1 on p. 1308). This condition is equivalent to saying that there are as many predetermined variables as stable eigenvalues. In this case, the model is called saddle-path stable, see Gandolfo (1997, p. 403). A detailed exposition of stability in the context of linearized DSGE models is given in Section 4.2.3.

[677] See Blanchard and Kahn (1980, Proposition 3 on p. 1308). This condition is equivalent to saying that there are fewer predetermined variables than stable eigenvalues.

[678] See Blanchard and Kahn (1980, Proposition 2 on p. 1308). This condition is equivalent to saying that there are more predetermined variables than stable eigenvalues.

[679] Miao (2020, p. 23).

[680] The Blanchard-Kahn conditions should not be confused with the stability condition (4.24).

[681] See Blanchard and Kahn (1980, p. 1309).

[682] See Blanchard and Kahn (1980, Equations (2) and (3) on p. 1308) or (B.31) in the context of Klein (2000)'s solution method.

[683] In the context of the models in the previous section, the exogenous process z_t is simply ε_t^A in the model without growth and long memory, i.e., z_t is a one-dimensional white noise process. In the model with growth but without long memory, one has $z_t = (\varepsilon_t^a, \varepsilon_t^g)^T$, i.e., it is a two-dimensional white noise process. It is also common to specify z_t as a first-order vector autoregressive (VAR(1)) process, see e.g., Klein (2000, p. 1409 and p. 1412), Ravenna (2007, p. 2050) or Chib and Ramamurthy (2014, pp. 154f.).

[684] Note that the restriction to VAR(1) seems not to be restrictive since many processes e.g., vector autoregressive moving average (VARMA) processes can be rewritten as a VAR(1) process, see, e.g., Lütkepohl (2005, pp. 426ff.).

transition equation of the following form[685]

$$x_t = \mathbf{M}x_{t-1} + \mathbf{N}\varepsilon_t, \tag{4.25}$$

where $\mathbf{N}\varepsilon_t$ is again a white noise process with covariance matrix $\mathbf{N}\Sigma_\varepsilon\mathbf{N}^T$. In the end, as can be seen from (4.25), the solution to the model can be represented as a stationary first-order vector autoregressive process similar to the univariate process considered in (2.13). The difference is that the autoregressive parameter is now replaced with the $n \times n$ dimensional matrix $\mathbf{M}$ whose entries are functions of the model's parameters.[686]

Further, the result stated in (2.10) can be generalized to multivariate time series, i.e., the autocovariance function of each component of the DSGE model's solution x_t given in (4.25) is again geometrically bounded and thus absolutely summable.[687] Overall, and similar to a univariate AR(1) or ARMA process, the solution to a standard DSGE model is again a short memory process.[688,689]

The reason why the model solution is a short memory process may lie in the specification of the exogenous processes rather than in the model equations themselves. As already stated and outlined by Blanchard and Kahn (1980) (and in the context of Klein's solution method in Appendix B.5.2), the solution to the model depends on the conditional expectations of the exogenous stochastic process z_t. If this process is assumed to be the VAR(1) process mentioned above, then the conditional expectation of z_{t+j}, given the information set at time t, can be expressed as a linear function of z_t, i.e.,

$$\mathbb{E}_t(z_{t+j}) = \mathbb{E}_t(\Lambda_z z_{t+j-1} + \varepsilon_{t+j}) = \Lambda_z \mathbb{E}_t(z_{t+j-1}) = \Lambda_z^j z_t. \tag{4.26}$$

This recursive dependence of the conditional expectations on the previous values of the

685 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 632, especially Equation (76)), Iskrev (2010, p. 191), Giacomini (2013, p. 6) or Chib and Ramamurthy (2014, pp. 154f.).

686 This form is also derived from Klein's solution method discussed in Appendix B.5.3.

687 See Brockwell and Davis (1987, p. 410).

688 This reasoning holds in a wider context as well. In order to estimate a DSGE model, (4.25) is complemented with a so-called observation or measurement equation which relates the observable variables to the model's variables, see Iskrev (2010, p. 191). The DSGE model's solution can then be written as a state space system that is commonly referred to as the ABCD-representation of the DSGE model, see Fernández-Villaverde, Rubio-Ramírez, Sargent, et al. (2007, p. 1021) or Morris (2016, p. 30). This representation can, however, be cast into a multivariate ARMA process (see Morris (2016, p. 31) or Ravenna (2007, p. 2052)) whose autocovariances and autocorrelations of each component are again geometrically bounded, see Brockwell and Davis (1987, p. 410). Hence, even if (4.25) is augmented with a measurement equation, the solution remains to be a short memory process.

689 Similarly, Davidson and Sibbertsen (2005, p. 254) claim that finite-order difference equations cannot replicate long memory.

process allows the whole solution to be cast into the recursive form stated in (4.25).

If exogenous short memory processes combined with rational expectations lead to a short memory solution of the linearized DSGE model, there seem to be two possible ways to include long memory behavior in the model solution. The first way, considered in this thesis, is to allow for long memory in the exogenous process. This way is carried out in (4.15) where the exogenous shock on technological progress is assumed to follow an ARFIMA$(1, d, 0)$ process.[690]

The second way is to go beyond the rational expectations hypothesis and to allow for other rules on how expectations are built in the model. This path was taken by Chevillon and Mavroeidis (2017). They consider the univariate model[691,692]

$$x_t = \bar{\beta} x_{t+1}^e + z_t \text{ with } \bar{\beta} \in \mathbb{R}, \tag{4.27}$$

where z_t is an exogenous stochastic process, x_t is an endogenous variable and x_{t+1}^e denotes expectations of x_{t+1} conditional on information available at time t.[693] The authors show two things: First, if the exogenous process z_t is a short memory process, $\left|\bar{\beta}\right| < 1$ and expectations are built rationally, i.e., $x_{t+1}^e = \mathbb{E}_t x_{t+1}$, then the solution to (4.27) is a short memory process.[694] Second, if expectations are built according to the following learning algorithm,

$$x_{t+1}^e = a_t, \text{ with } a_t = a_{t-1} + \bar{g}_t(x_t - a_{t-1}),\ \bar{g}_t = \theta/t + f_t, \text{ and } |f_t| \leq M t^{-1-\zeta},$$

where $M, \theta, \zeta > 0$, then the solution to (4.27) shows long memory if $\bar{\beta} > 1 - \dfrac{1}{2\theta}$.[695,696]

Although the results of Chevillon and Mavroeidis (2017) do not apply directly to the

---

690 The details on the solution method in the long memory setting and the conditions under which a unique solution exists are discussed in more detail in Appendices B.5.4 and B.5.5.

691 See Chevillon and Mavroeidis (2017, Equation (1) on p. 2).

692 For convenience, the notation of Chevillon and Mavroeidis (2017) is adopted for the rest of this subsection. Therefore, the symbols used here do not correspond to the meaning introduced in the previous section. Since the formulas stated in this section have only an illustrative character and Chevillon and Mavroedis's formulas are not referred to outside of this section, it seems reasonable not to introduce a specific notation to keep the overall notation in the thesis tractable.

693 See Chevillon and Mavroeidis (2017, p. 2).

694 See Chevillon and Mavroeidis (2017, p. 4, especially Proposition 1).

695 See Chevillon and Mavroeidis (2017, p. 4, especially Theorem 2).

696 Here, it should be highlighted that Chevillon and Mavroeidis (2017) define long memory in a different manner to Definition 2.3.1. In their context a process x_t shows long memory if $[T^{-1}\text{var}(\sum_{t=1}^{T} x_t)]^{1/2} \sim T^d$ as $T \to \infty$ and $d > 0$, see Chevillon and Mavroeidis (2017, p. 3). This definition also applies to non-stationary processes in contrast to Definition 2.3.1, see Chevillon and Mavroeidis (2017, p. 3). Hence, there is a lack of direct comparability between the notions of long memory in this thesis and the one of Chevillon and Mavroeidis (2017).

multivariate models considered in the previous section, they indicate that it may be the combination of short memory in the exogenous process and the assumption of rational expectations that generates short memory in the solution of the considered model. Overall, their results support the procedure followed in this thesis: Generating long memory in a widely applied and well-established DSGE model framework, where expectations are built rationally, calls for long memory in the exogenous stochastic processes.[697]

4.2.3 A Note on Stability

A linearized DSGE model in the canonical form (4.22) forms a set of linear expectational difference equations. This subsection aims to describe the notion of stability in such a context and how it relates to other stability concepts in the context of deterministic models. The discussion builds on the solution method of Blanchard and Kahn (1980) briefly described in Section 4.2.1 but references to the solution method of Klein (2000) outlined in Appendix B.5 are given when they appear appropriate.

In order to develop relations between different notions of stability in deterministic and stochastic model contexts, it seems suitable to consider the corresponding deterministic system to (4.22) first; it is of the following form[698]

$$x_{t+1} = \mathbf{A}_0 x_t + G_t, \tag{4.28}$$

where the sequence $(G_t)_{t\in\mathbb{Z}}$ is assumed to be an exogenous (non-stochastic) sequence not depending on x_t.[699] Such a system is called autonomous if G_t does not depend on t, i.e., if $G_t \equiv G$ for a constant $G \in \mathbb{R}$; otherwise it is called non-autonomous.[700] A solution to this system is again a sequence $(x_t)_{t\in\mathbb{Z}}$ that satisfies (4.28) at each instant of time.[701]

Stability generally refers to a property of equilibrium or steady state points of a dynamical system. Steady states are time-invariant points where the system stays once such a point has been reached, and no additional exogenous shocks pushing the system out of the equilibrium occur.[702] Let $\mathbf{I}_{n\times n}$ be the n-dimensional identity matrix, then, with $G_t \equiv G$, the unique steady state of (4.28) is clearly given by $x_{ss} = (\mathbf{I}_{n\times n} - \mathbf{A}_0)^{-1}G$ if the matrix

[697] This is in line with the statement of Chevillon and Mavroeidis (2017, p. 6) that in case of rational expectations, the memory of the endogenous variables is determined exogenously.
[698] Note that the expectational operator can be omitted if the model is assumed to be deterministic.
[699] The matrix $\mathbf{G}_0$ is omitted here for convenience, but it may be implicitly present in the definition of the exogenous sequence $(G_t)_{t\in\mathbb{Z}}$.
[700] See, e.g., Galor (2007, p. 112).
[701] See, e.g., Galor (2007, p. 16).
[702] See Galor (2007, p. 30).

$(\mathbf{I}_{n\times n} - \mathbf{A}_0)^{-1}$ is invertible.[703]

Let $\|\cdot\|$ denote the Euclidean norm. A steady state point x_{ss} is called locally stable if for every $\epsilon > 0$, one can find $\delta_\epsilon > 0$, such that $\|x_0 - x_{ss}\| < \delta_\epsilon$ implies that $\|x_t - x_{ss}\| < \epsilon$ for all $t \geq 0$; otherwise the steady state is called unstable.[704] Roughly speaking, a steady state value is said to be stable if the solution stays near the steady state whenever the initial value was chosen close enough to the steady state. A more demanding concept of stability is the one of local asymptotic stability. A steady state value is called locally asymptotically stable if it is stable and, in addition, if there is $\mu > 0$ such that if $\|x_0 - x_{ss}\| < \mu$, then $\lim_{t\to\infty} x_t = x_{ss}$.[705] If the convergence to the steady state holds for all initial values $x_0 \in \mathbb{R}^n$, the steady state is said to be globally asymptotically stable.[706,707]

Whether the stability is global or local also depends on the domain over which the system is defined. Of course, a steady state of (4.28) (if existent) can be globally asymptotically stable under certain circumstances. In the context of the DSGE model, however, the linear system was obtained from a linearization around a steady state of an underlying nonlinear system. Thus, the global asymptotic stability of the linear system is merely sufficient for the local asymptotic stability of the corresponding steady state of the nonlinear system.[708]

Again, consider the deterministic autonomous system $x_{t+1} = \mathbf{A}_0 x_t + G$ and assume that the conditions for a unique steady state x_{ss} are met. Then, the stability of the steady state depends solely on the eigenvalues of the matrix $\mathbf{A}_0$, i.e., if the moduli of all eigenvalues are strictly less than one, x_{ss} is globally asymptotically stable.[709] If there is at least one eigenvalue of $\mathbf{A}_0$ with modulus greater than one (i.e., an unstable eigenvalue), the steady state value should be labeled as unstable, as this unstable eigenvalue makes the whole solution explosive.[710] However, if $\mathbf{A}_0$ has stable and unstable eigenvalues, the steady state may also be labeled as "not-wholly-unstable"[711] as there are solutions that converge to the steady state, while other solutions are divergent. Such steady state values are then

703 See Galor (2007, Proposition 2.3 on p. 31).
704 This definition is inspired by Elaydi (2005, Definition 4.2. on p. 176), who provides additional nuances of stability such as uniform stability which are not given here as they go beyond the scope of this section.
705 This definition is again inspired by Elaydi (2005, Definition 4.2. on p. 176) and Galor (2007, Definition 4.2. on p. 95).
706 See Elaydi (2005, p. 177) and Galor (2007, Definition 4.2. on p. 95).
707 Note that there are additional concepts of stability in the context of linear and nonlinear dynamical systems, such as exponential stability, orbital stability, or bounded-input/bounded-output stability, which are far beyond the scope of this section. A short overview can be found in G. Chen (2005).
708 See Galor (2007, p. 98 and Theorem 4.8 on p. 103).
709 See Galor (2007, Corollary 3.6 on p. 86).
710 See Gandolfo (1997, p. 118).
711 Gandolfo (1997, p. 118).

often referred to as "saddle points", or as being "saddle-path stable".[712,713]

If the steady state is a saddle point, one is often interested in finding a solution that converges to the steady state. If $\mathbf{A}_0$ has stable and unstable eigenvalues, such a solution can be found by decoupling the system with certain matrix decomposition in two parts, each associated with the stable and unstable eigenvalues, respectively.[714,715]

After the decoupling, there are two subsystems of the form (4.28) where the corresponding matrices have only stable or unstable eigenvalues, respectively. For the subsystem containing the stable eigenvalues, a solution may be found, e.g., by backward iteration, and for the unstable subsystem, the forward iteration may be appropriate.[716] Furthermore, if the stable subsystem has a steady state, it is globally asymptotically stable.[717]

Since the solution to the stable subsystem appears straightforward, the focus now lies on the unstable subsystem. Let $x^u_{t+1} = \mathbf{A}^u_0 x^u_t + G$ the system associated with the unstable eigenvalues, i.e., the matrix $\mathbf{A}^u_0$ contains all eigenvalues of $\mathbf{A}_0$ that are greater than one in modulus. Moreover, assume that this system has a unique steady state given by $x^u_{ss} = (\mathbf{I}_{n_u \times n_u} - \mathbf{A}^u_0)^{-1} G$. Then, a solution to the unstable subsystem can be found by forward iteration.[718] After k forward iterations, one obtains

$$x^u_t = \left(\mathbf{A}^u_0\right)^{-k} x^u_{t+k} - \sum_{n=1}^{k} \left(\mathbf{A}^u_0\right)^{-n} G. \tag{4.29}$$

Since $\mathbf{A}^u_0$ contains only eigenvalues greater than one in modulus, $\left(\mathbf{A}^u_0\right)^{-k}$ vanishes as $k \to \infty$.[719] Since one is interested in non-explosive and thus bounded solutions of the unstable subsystem, one has that $\left\| x^u_{t+k} \right\|$ is bounded. Hence, as $t \to \infty$, the first term in

712 See Gandolfo (1997, p. 373).

713 Such a situation is likely to occur in DSGE models. As outlined in Section 4.2.1, the Blanchard-Kahn conditions pin down a unique solution when there are as many unstable eigenvalues of $\mathbf{A}_0$ as there are non-predetermined variables.

714 This was already mentioned in Section 4.2.1. In the context of the solution method of Klein (2000), the concrete decoupling is carried out in Appendix B.5.2.

715 For example, Blanchard and Kahn (1980, p. 1307) use a Jordan decomposition. In Appendix B.5.2, a generalized Schur decomposition is used to decouple the system. Both methods can also be applied in a deterministic setting as illustrated by Miao (2020, Section 1.71. on pp. 20ff. and Section 1.72. on pp. 26ff.). If all eigenvalues of $\mathbf{A}_0$ are distinct, $\mathbf{A}_0$ is diagonalizable, and a rather simple decomposition exists; for details, see Gandolfo (1997, p. 123).

716 See Gandolfo (1997, p. 123).

717 See Galor (2007, Corollary 3.6 on p. 86).

718 See Gandolfo (1997, p. 123).

719 The eigenvalues of $\left(\mathbf{A}^u_0\right)^{-1}$ correspond to the inverse eigenvalues of $\mathbf{A}_0$, see, e.g., Gandolfo (1997, p. 123).

(4.29) vanishes, and the solution to the unstable system is given by[720]

$$x_t^u = -\sum_{n=1}^{\infty} \left(\mathbf{A}_0^u\right)^{-n} G = \left[-\left(\mathbf{I}_{n_u \times n_u} - \left(\mathbf{A}_0^u\right)^{-1}\right)^{-1} + \mathbf{I}_{n_u \times n_u}\right] G = \mathbf{A}_0^u x_{ss}^u + G = x_{ss}^u. \quad (4.30)$$

Equation (4.30) states that the unique and bounded solution to the wholly unstable system is precisely the one that is equal to the system's steady state value at each instant of time. Thus, roughly speaking, a convergent solution to the entire system can be found as long as the unstable system is at its steady state value. In contrast, the initial values of the stable system can be chosen freely.[721]

Until now, a saddle-path-stable steady state of a deterministic autonomous system was considered. But how does this saddle-path stability relate to the non-autonomous system and the linearized DSGE models discussed in this chapter? At first glance, one might expect an exogenous shock to push the unstable subsystem out of its steady state, causing the solution to explode. However, this is not the case.

If G_t is not constant but an arbitrary non-stochastic sequence of numbers, the corresponding non-autonomous system does, in general, not have a steady state value.[722,723] Hence, the notion of stability outlined above is not apparent, or instead, it is not applicable in these contexts.[724] However, by focusing on bounded solutions, the forward iteration method still applies and delivers[725]

$$x_t^u = -\sum_{n=1}^{\infty} \left(\mathbf{A}_0^u\right)^{-n} G_{t+n-1}, \quad (4.31)$$

where the sequence $(G_t)_{t\in\mathbb{Z}}$ has to be chosen such that the limit of the series exists.[726,727]

---

720 The second equality follows from an application of the Neumann series (see, e.g., Elaydi (2005, Equation (3.5.19) on p. 167)). The third equality uses some (basic) matrix algebra and the last equality is implied by the definition of a steady state.

721 An two-dimensional example is given, e.g., in Galor (2007, Example 2.4 and 2.5 on pp. 34ff.).

722 See Galor (2007, p. 112f.).

723 A simple univariate example is, e.g., $x_{t+1} = \varrho x_t + (-1)^t$. When $|\varrho| < 1$, a solution by backward iteration can easily be found to be $x_t = \varrho^k \left[x_0 + 1/(1+\varrho)\right] - (-1)^t/(1+\varrho)$. The solution has no limit for all initial values x_0, i.e., no steady state exists. For the corresponding homogeneous system, however, $x_{t+1} = \varrho x_t$, the unique and globally asymptotically stable steady state is given by $x_{ss} = 0$.

724 See Galor (2007, p. 113).

725 Here, the focus lies again on the system associated with the unstable eigenvalues of $\mathbf{A}_0$. Of course, the backward solution also applies to the system associated with the stable eigenvalues. If the stable subsystem is given by $x_{t+1}^s = \mathbf{A}_0^s x_t^s + G_t^s$, then the backward solution delivers $x_t^s = (\mathbf{A}_0^s)^t x_0^s + \sum_{n=0}^{t-1} \mathbf{A}_0^t G_{t-n-1}^s$.

726 The existence of the limit is satisfied if, e.g., the sequence $(G_t)_{t\in\mathbb{Z}}$ is bounded.

727 Recall the univariate example stated in Footnote 723. When $|\varrho| > 1$ the forward solution delivers a bounded solution given by $x_t^u = -(-1)^t/(1+\varrho)$. Thus, by choosing $x_0^u = -1/(1+\varrho)$, the solution remains bounded, but again, it neither converges to a steady state nor is it explosive. The solution

Obviously, if one is only interested in bounded solutions to the non-autonomous system, the corresponding steady state of the autonomous system ($G_t \equiv G$) or homogeneous system ($G_t \equiv 0$) is saddle-path stable by construction, as (4.31) eventually becomes (4.30).

The same methodology applies to the linearized DSGE models considered in this chapter. In the context of the stochastic model (4.22), the forward solution becomes[728]

$$x_t^u = \left[(\mathbf{A}_0^u)^{-1}\right]^k \mathbb{E}_t x_{t+k}^u - \sum_{j=0}^{k-1} \left[(\mathbf{A}_0^u)^{-1}\right]^{j+1} G^u \mathbb{E}_t z_{t+j}, \tag{4.32}$$

where G^u is a certain matrix. The first part of (4.32) on the right-hand side is often referred to as the "bubble solution" and the second term to as the "fundamental solution".[729]

The difference between (4.29) and the stochastic version (4.32) is that the solution to the stochastic model contains expectations about future values of the solution ($\mathbb{E}_t x_{t+k}^u$) instead of future values (x_{t+k}) themselves. Thus, to ensure the solution's boundedness, the expectations of x_{t+k}^u have to be bounded. Here condition (4.24) comes into play. If one is interested in a solution that satisfies (4.24), one obtains

$$\left[(\mathbf{A}_0^u)^{-1}\right]^k \mathbb{E}_t x_{t+k}^u \leq \left[(\mathbf{A}_0^u)^{-1}\right]^k (1+k)^{\theta_t^{x^u}} M_t^{x^u}.$$

As for the deterministic system, the right-hand side goes to zero as $k \to \infty$. Thus, the bubble solution vanishes, and there is a single expectationally bounded solution given by

$$x_t^u = -\sum_{j=0}^{\infty} \left[(\mathbf{A}_0^u)^{-1}\right]^{j+1} G^u \mathbb{E}_t z_{t+j}. \tag{4.33}$$

Overall, by imposing (4.24), one finds a solution to the unstable subsystem whose conditional expectations do not explode. This is the reason why the author called (4.24) a stability condition.

As can be seen from (4.33), the model solution depends to a large extent on the conditional expectations of the exogenous stochastic process. If they are of the recursive type mentioned in (4.26), the solution (4.33) can be simplified dramatically, and it depends linearly on z_t. In the case of an exogenous long memory process, the conditional expectations are not of this recursive type. Much more effort has to be undertaken to evaluate the series in (4.33). As for the exogenous sequence $(G_t)_{t\in\mathbb{Z}}$ of the deterministic model, the stochastic

fluctuates symmetrically around the steady state of the corresponding homogeneous system.

728 See the derivation of (B.29) in Appendix B.5.2.

729 See Blanchard and Fischer (1993, p. 221), Miao (2020, pp. 7f.) and Turnovsky (2000, pp. 91f.).

process $(z_t)_{t\in\mathbb{Z}}$ has to be chosen such that the series (4.33) is well defined. For this reason, Blanchard and Kahn (1980) assume the boundedness restriction (4.24) to hold for the exogenous process as well.[730]

As outlined in Appendix B.5.1, Klein (2000) uses a slightly different condition to pin down a non-explosive solution. He defines a stochastic process to be stable if the unconditional mean of a stochastic process is uniformly bounded, see (B.21).[731] The difference between Klein's condition and Blanchard and Kahn's condition is that Blanchard and Kahn (1980) refer to the conditional expectations which are allowed to grow at a polynomial rate by (4.24), whereas Klein (2000) refers to the unconditional mean that must not grow at all.

However, it should be noted that stability in the sense of Klein (2000) or Blanchard and Kahn (1980) is a property of the model's solution, i.e., it is a property of the whole trajectory of the solution $(x_t)_{t\in\mathbb{Z}}$. In contrast, the stability concept in the deterministic setup is a property of a steady state.

The idea of exogenous shocks pushing the unstable system on an explosive path appears to be misleading. The key ingredient in all considered models (deterministic autonomous and non-autonomous or stochastic) of this section is the solution's boundedness (in a certain sense). In the presence of exogenous shocks, the system (4.22) does, in general, not have a steady state. However, focusing on the solution that rules out the bubble term keeps the solution bounded in a sense that the expectations do not explode.[732] In the end, the paths of a bounded solution of a DSGE model behave similarly to the ones of an AR(1) (or ARFIMA$(1,d,0)$) process depicted in Figure 2.2 (or Figure 2.10), i.e., the solution fluctuates regularly around its mean value and shows no explosive behavior. However, if the exogenous shocks are chosen such that the system has a steady state, e.g., $z_t \equiv G$, then this steady state is by construction saddle-path stable in the sense mentioned above. Moreover, the derived solution is on the stable arm and converges to the steady steady state value.

These similarities may be why linearized DSGE models are often said to be saddle-path stable.[733] Given the arguments of this section, one might rather say that the steady state of the corresponding deterministic system is saddle-path stable and that the solution to the DSGE model is bounded in mean.

730 See Blanchard and Kahn (1980, p. 1305 and p. 1307).
731 See Klein (2000, Definition 4.1 on p. 1411).
732 The reason for focusing on only bounded solutions of an economic model may be justified with certain transversality conditions that the solution must fulfill, see Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 553 and p. 604).
733 See Gandolfo (1997, p. 403) or Cantore, Gabriel, et al. (2013b, p. 452).

Two remarks are made at the end of this section. The first concerns the first-order approximation of the underlying nonlinear system, and the second concerns the rational expectations hypothesis and determinacy.

The linearized DSGE models considered in this chapter are first-order approximations of an underlying nonlinear model. Hence, only minor conclusions on the dynamic properties of the nonlinear model can be drawn from the linear approximation. To be more precise and as outlined by Galizia (2021), deriving a bounded solution along Blanchard and Kahn's or Klein's solution method may be inappropriate if the underlying nonlinear model features a limit cycle.[734] In such cases, these methods may indicate that there is no solution (because the Blanchard and Kahn conditions are not satisfied), even though there may be a limit cycle in the nonlinear model.[735] Since the method of Klein (2000) described in Appendix B.5 delivers a unique and bounded solution to the models considered in this chapter, the phenomena described by Galizia (2021) are not present in the context of this thesis.

The posited saddle-path property of linearized DSGE models may be due to the assumption of rational expectations, as many models in which expectations are built rationally exhibit such a saddle-point property.[736] Another concept of stability in the context of stochastic models is the one of expectational stability or E-stability, which addresses the question of whether, given small deviations from the rational expectations hypothesis, e.g., when expectations are built based on a learning algorithm, the system eventually converges to a rational expectation equilibrium.[737] The solution method of Blanchard and Kahn mentioned in Section 4.2.1 and the method of Klein (2000) presented in Appendix B.5 are designed to find the unique (bounded) solution in the case of model determinacy, i.e., if a unique solution actually exists. In this case, the solution is stable in the sense that the conditional expectations do not explode. Unfortunately, this may not imply that the

[734] See Galizia (2021, p. 871).
[735] See Galizia (2021, pp. 871 and 877).
[736] See Burmeister (1980, p. 804) or Turnovsky (2000, p. 187).
[737] See G. W. Evans (1985, p. 1218). By investigating E-stability, to be more precise, the model's agents are supposed not to know the true parameter values of the model's rational expectations solutions; instead, they try to estimate these parameters, see G. W. Evans and Honkapohja (2001, p. 40). E-stability then investigates whether these estimated parameters converge (in a certain sense) to the parameters of the rational expectations solution, G. W. Evans and Honkapohja (2001, p. 40). Formally, a (deterministic) map between the perceived and true model's solution law of motion is constructed that contains the rational expectations solution as a fixpoint, see G. W. Evans and Honkapohja (2001, pp. 40f.). The rational expectations solution is then called E-stable if the fixpoint of this mapping is locally asymptotically stable in a sense mentioned at the beginning of the section, see G. W. Evans and Honkapohja (2001, pp. 41). For details and additional refinements of the concept of E-stability, see G. W. Evans and Honkapohja (2001, Section 2.9 on pp. 39ff. and pp. 140f.).

solution is also expectationally stable.[738] On the other hand, if the model is indeterminate, i.e., if many solutions exit, some solutions may appear to be expectationally stable.[739]

This chapter focuses on models in which expectations are built rationally. Thus, deviations from the rational expectations hypothesis go beyond the scope of this thesis. Therefore, whether the solutions to the models treated in this chapter are expectationally stable is left for future research. Instead, for all parameter constellations considered in this thesis, the method of Klein (2000) delivers a unique, stable solution in the sense that its unconditional mean is bounded.[740,741] The analysis of this thesis focuses on this particular solution.

4.3 Rationales for Long Memory in a DSGE Context

Before considering the implications of long memory technology shocks in the next section, this section is devoted to justifying the author's approach to considering long memory technology shocks in the context of a DSGE model.

Chapter 3 highlights that long memory processes are frequently analyzed in the economics and econometrics literature, and there is a large and still growing literature dealing with the presence of long memory in economic time series, or proposing refined estimation and testing techniques to detect or reject the long memory hypothesis. The empirical evidence of long memory in macroeconomic time series, especially for GDP and related variables, given in Section 3.2.1 is mixed and highly dependent on the estimation method employed. However, the results indicate that long memory can contribute to explaining the data as it covers richer dynamics regarding the autocorrelations and impulse-response functions than traditional ARMA processes often used in DSGE models for the specification of exogenous processes. As pointed out by Schorfheide (2011), the choice of the frequently used AR(1) specification for the exogenous processes in the model is more or less arbitrary.[742] He further points out that incorporating richer dynamics in the exogenous process may increase the models' empirical fit and help to prevent misspecification.[743]

Allowing for a more elaborate correlation and dependence structure in the model that goes beyond the AR(1) assumption is not new. Ireland (2004a) argues that there might

[738] See Bullard and Eusepi (2014, p. 8).
[739] See G. W. Evans (1985, p. 1218).
[740] See Appendix B.5.1 for details. As outlined there, the boundedness of the unconditional mean further implies condition (4.24).
[741] Theorem B.5.2 summarizes all conditions that ensure the existence of a unique and stable solution of the models in this thesis.
[742] See Schorfheide (2011, p. 22) and, additionally, Cúrdia and Reis (2010, p. 3).
[743] See Schorfheide (2011, p. 22).

be correlations in the data that cannot be captured by the simple structure of the DSGE model.[744] He argues that incorporating correlated measurement errors may improve the model fit when bringing the DSGE model to the data.[745] Contrary to the approach in this thesis, he assumes an exogenous AR(1) process for technology, solves the model, and then adds measurement errors for the estimation.[746] Since these errors are added after the model is solved, they cannot enter the agents' expectations and behavioral responses in the model, thereby neglecting the effects on the model's endogenous propagation mechanisms.[747]

For this reason, Cúrdia and Reis (2010) proposes to incorporate correlated disturbances directly into the model equations by allowing the exogenous processes to follow a VAR(k) process.[748] To be more precise, Cúrdia and Reis (2010) consider a government expenditure shock and a technology shock and allow them to follow a VAR(k) process.[749] In the context of the model given in the previous sections, this assumption would be equivalent to specifying the transitory technology process $\log(A_t)$ given in (4.14) as an AR(k) process. By doing so, Cúrdia and Reis (2010) preserve the state space representation of the model that allows the estimation of the model.[750] Overall, they show that correlated disturbances can help explain/resolve existing puzzles in the RBC theory as the model's responses change qualitatively given the more general correlation structure.[751]

Another related work to the approach followed in this thesis is Meyer-Gohde and Neuhoff (2018). They generalize the AR(1) assumption to allow for a general ARMA specification of the technology shock in a neoclassical growth model.[752] Interestingly, they do not specify the order of the ARMA process a priori but rather let the data decide which order fits the data best.[753] Although there is some uncertainty regarding the exact ARMA specification, their approach identifies an ARMA$(3, 0)$ model for the technology process to fit the data best, thereby clearly rejecting the commonly used AR(1) assumption.[754] Furthermore, by taking the model uncertainty into account, they derive IRFs of the model and found evidence that the technology's IRF is hump-shaped.[755]

---

744 See Ireland (2004a, p. 1210).
745 See Ireland (2004a, p. 1210).
746 See Ireland (2004a, Equation (3) on p. 1208 and Equations (8)-(11) on pp. 1208f.).
747 See Cúrdia and Reis (2010, p. 5).
748 See Cúrdia and Reis (2010, Equation (4) on p. 10).
749 See Cúrdia and Reis (2010, Equation (4) on p. 10).
750 See Cúrdia and Reis (2010, p. 12).
751 See Cúrdia and Reis (2010, p. 18f.). Cúrdia and Reis (2010, Figure 2) further illustrates how the model's IRF changes under the more general correlation structure.
752 See Meyer-Gohde and Neuhoff (2018, p. 15).
753 See Meyer-Gohde and Neuhoff (2018, p. 18).
754 See Meyer-Gohde and Neuhoff (2018, p. 18).
755 See Meyer-Gohde and Neuhoff (2018, p. 22 and Figure 10 and 11 on pp. 23 and 26, respectively).

Therefore, the approach followed in this thesis is a direct generalization of the work of Meyer-Gohde and Neuhoff (2018) and Cúrdia and Reis (2010), as both consider only generalized short memory processes as alternatives for the technology shock. Instead, as outlined in (4.14), the transitory technology shock is allowed to follow a long memory ARFIMA process in this thesis.

The cost of this generality is the loss of the model's state space representation. Chan and Palma (1998) show that an ARFIMA process with $d \neq 0$ has no such finite-dimensional state space representation.[756] Since the transitory technology shock given in (4.15) is such a process and part of the model, the model does not have a finite-dimensional state space representation neither. However, this may become an issue primarily when trying to estimate the corresponding model, not when solving the model.[757]

In the study of Meyer-Gohde and Neuhoff (2018), they used a fully calibrated neoclassical growth model. They concentrated on estimating the order and the parameters of the ARMA process.[758] A desirable procedure would, instead, simultaneously estimate the deep model parameters and the exogenous processes. To the best of the author's knowledge, such a procedure is not available yet.

The inclusion of long memory dynamics in a DSGE framework seems empirically reasonable and necessary and is further illustrated in the work of Moretti and Nicoletti (2010). To the best of the author's knowledge, this is the only study that pays attention to long memory dynamics in a DSGE framework. Their simulation study shows that standard estimation techniques produce a substantial bias in the model's deep parameters if the data-generating process has long memory.[759]

To be more precise, they consider a simple RBC-DSGE model and solve it given the assumption that technology follows an AR(1) process.[760] Then, they simulate the model by feeding ARFIMA$(0, d, 0)$ shocks into the AR(1) process.[761] From the simulated model, they then try to re-estimate the true parameter values originally used for the simulation.[762] The reported bias in the estimates occurs as the model tries to replicate the strong persistence in the data-generating process via its endogenous propagation mechanisms.[763] Since the DSGE model produces indeed only short memory dynamics (see Section 4.2.2), the deep

[756] See Chan and Palma (1998, Corollary 2.1. on p. 722).
[757] Some promising ways to estimate a long memory DSGE model are pointed out in Chapter 6.
[758] See Meyer-Gohde and Neuhoff (2018, p. 15).
[759] See Moretti and Nicoletti (2010, pp. 21f.).
[760] See Moretti and Nicoletti (2010, Equation (49) on p. 18).
[761] See Moretti and Nicoletti (2010, p. 17).
[762] See Moretti and Nicoletti (2010, p. 17).
[763] See Moretti and Nicoletti (2010, p. 22).

model parameters show bias towards more persistent endogenous dynamics.[764]

However, instead of incorporating long memory into the model equation as carried out in this thesis (see again (4.14)), they propose a generalized version of the Kalman filter to remove the persistence from the data before the estimation.[765] Essentially, they remove the persistence from the data that cannot be explained by their DSGE model.[766] By applying their generalized estimation routine, the formerly reported bias disappears, and the estimated values are quite close to the true model parameters.[767] Furthermore, they applied their generalized estimation procedure to US data. They found parameter estimates closer to parameter values consistent with those implied by national accounts such as the capital share or the depreciation rate.[768]

Overall, their approach looks similar to the one of Ireland (2004a) mentioned above, as they try to specify an exogenous process outside of the actual model to capture the data features that remain unexplained by the model. By doing so, they neglect the effects that long memory in the exogenous process can have on the endogenous propagation mechanisms of the model.[769] On the other hand, Moretti and Nicoletti (2010) argue in favor of their approach against the usage of more general exogenous processes in the model itself precisely because they affect the endogenous propagation mechanisms.[770] In their opinion, when specifying data-dependent exogenous stochastic processes as it is done, e.g., by Meyer-Gohde and Neuhoff (2018), the way of how agents build their expectations is also data-dependent.[771] Thus, different data sets would imply different ways of forming expectations (albeit expectations are built rationally) and thus opens the door for a type of Lucas critique.[772] Furthermore, they question a lack of economic interpretation and parameter inflation when higher-order exogenous processes are specified in the model.[773] Similarly, Cúrdia and Reis (2010) point to the trade-off between exogenous shock's simplicity and interpretability and the model's produced biases and misspecifications.[774]

From this perspective, low-order exogenous processes providing convincing parameter interpretations seem overall to be preferable to a fully flexible approach such as the one of

[764] See Moretti and Nicoletti (2010, p. 22).
[765] See Moretti and Nicoletti (2010, p. 21).
[766] See Moretti and Nicoletti (2010, p. 21).
[767] See Moretti and Nicoletti (2010, Table 3 on p. 23).
[768] See Moretti and Nicoletti (2010, p. 27).
[769] In the end, this is the same argument from Cúrdia and Reis (2010, p. 5) mentioned above.
[770] See Moretti and Nicoletti (2010, pp. 9f.).
[771] See Moretti and Nicoletti (2010, pp. 9f.).
[772] See Moretti and Nicoletti (2010, pp. 9f.).
[773] See Moretti and Nicoletti (2010, p. 6).
[774] See Cúrdia and Reis (2010, p. 7).

Meyer-Gohde and Neuhoff (2018). Considering an ARFIMA$(1, d, 0)$ as a two-parameter process, where one parameter deals with the short memory dynamics while the other tackles the long memory dynamics, may provide a reasonable contribution to the approaches discussed so far. The long memory parameter may then absorb excess persistence possibly present in the data that cannot be explained with the remaining short memory structure of the model while keeping the number of parameters small.

Additionally, in the light of the discussion in Section 3.3, there are convincing arguments for the presence of long memory in the data that further call for taking long memory into account in economic models as well. In the model introduced in Section 4.1, total factor productivity is the single source of uncertainty. Consequently, to include long memory dynamics in the model, TFP is the single possible way to do so. Nevertheless, the arguments given in Section 3.3 apply directly to TFP and provide, besides the mechanical introduction of long memory in an RBC model, some TFP-related rationales for introducing long memory this way. To illustrate this point intuitively recall the error-duration model of Parke (1999) mentioned in Section 3.3.1, which might be interpreted in a technological context as well: It is easily imaginable that there are various technologies employed in the economy each contributing to the economy's level of technology that may be measured by TFP. In Parke's error duration model, TFP can be viewed as the aggregate value over all technologies used in the economy with stochastic lifetimes. Thus, long memory in TFP is likely to occur if the single technologies have slowly decaying survival probabilities, i.e., if the probability that a technology exists for k periods decays slowly to zero as k tends to infinity. Following Parke (1999), this would imply that there are many short-lived technologies, but at the same time, some technologies have survived over a long time.[775] This may be plausible with many concurring technologies, where some are successful and widely accepted in the economy, whereas others are driven out. Once a technology is established, it may be reasonable that it is employed for a while since it might be costly for firms to change production technologies frequently.

This section has provided convincing arguments for introducing long memory in DSGE models and has shown how this thesis contributes to an ongoing discussion in the literature. The following section provides a detailed analysis of the model-implied impulse-response functions and shows the effect of introducing long memory in TFP.

[775] See Parke (1999, Exmaple A on pp. 635f.) in the context of his model for aggregate employment. To build the bridge to TFP, replace "employment" and "firm" in his context with "TFP" and "technology", respectively.

4.4 Model Comparison and Impulse-Response Analysis

4.4.1 Model Parameters

In order to carry out an impulse-response analysis of the models presented in Section 4.1, the models' parameters have to be pinned down. In order to do so, it is resorted to common values in the literature. An overview of the parameters used in the following section is given in Table 4.1. Recall that time is measured in quarters.

Param.	Description	add. sep. utility	CD utility
A_{ss}	steady state transitory TFP	1	1
g_{ss}	steady state growth rate	0	0.005
β	time preference rate	0.98	0.98
δ	depreciation rate	0.025	0.025
α	output elasticity with respect to capital	0.33	0.33
φ	inverse of the Frisch elasticity of labor supply	2	–
ς	inverse of the intertemporal elasticity of substitution	2	–
$\varkappa$	relative distaste of supplying labor	31.76	–
γ	exponent of leisure	–	0.6243
τ	inverse of the intertemporal elasticity of substitution for the composite good $C_t^{1-\gamma}(1-H_t)^\gamma$	–	2

Table 4.1: Benchmark parameter values. The table reports the benchmark parameters for the models with additive separable utility function (4.18) and Cobb-Douglas utility function (4.19). These parameters are used for the impulse-response analysis in the next section. The values of the remaining parameters $\varrho_A, d, \varrho_g, \sigma_{\varepsilon^A}$, and σ_{ε^g} will vary throughout the analysis and are reported in the corresponding figure captions.

The steady state value of the exogenous transitory productivity process is normalized to 1.[776] In addition, it is assumed that the steady state growth rate $g_{ss} = 0.005$ of the labor

[776] This is in line with Aguiar and Gopinath (2007, Equation (2) on p. 78) or Lindé (2009, p. 600). It is common in the literature to model log deviations of transitory productivity processes as zero-mean AR(1) processes. Normalizing A_{ss} to 1 ensures that the process in (4.14) has a zero mean. See

augmenting technological progress corresponds to an annual growth rate of approximately 2%.[777] Recall from (4.20) that the model is assumed to have a balanced growth path, i.e., the variables given (4.20) such as GDP and consumption expenditures grow at the same rate.

For the time preference rate β, a value of 0.98 is assigned.[778] The quarterly depreciation rate is assumed to be $\delta = 0.025$.[779] A value of $\alpha = 0.33$ is assumed for the output elasticity of capital.[780]

For both utility functions, the inverses of the intertemporal elasticities of substitution are set to $\tau = \varsigma = 2$.[781] For the inverse of the Frisch elasticity a value of $\varphi = 2$ is assigned.[782]

A value of $\varphi = 2$ corresponds to a Frisch elasticity of 0.5. However, empirical estimates of the Frisch elasticity are inconclusive. As stressed by Fiorito and Zanella (2012), microdata-based estimates of the Frisch elasticity may, in general, not be taken for the calibration of macroeconomic models since aggregation increases the estimates of the Frisch elasticity substantially.[783] Fiorito and Zanella (2012) report values of an aggregate Frisch elasticity between 0.6 and 1.7.[784] Peterman (2016) reports even higher estimates of the macro Frisch elasticity in the range of 2.9 and 3.1.[785] Chetty et al. (2011), on the other hand,

---

Appendix B.3.2 for more details.

777 This is in line with recent empirical findings that the annual US GDP per capita growth rate is about 2%. This value seems relatively stable over time, see Kohlscheen and Nakajima (2021, Table 4 on p. 49). Lindé (2009, p. 600) and Cantore, Levine, et al. (2015, Table 2 on p. 141), instead, assume slightly higher steady state growth rates of about 2.5% p.a. and 3% p.a., respectively.

778 This is in line with Aguiar and Gopinath (2007, Table 3 on p. 86). Note that Cantore, Gabriel, et al. (2013a, Table 18.2 on p. 432) and Cooley and Prescott (1995, p. 22) use a slightly higher value of 0.987. Lindé (2009, p. 600) and Cantore, Levine, et al. (2015, Table 2 on p. 141) specify $\beta = 0.99$.

779 This is in line with Smets and Wouters (2007, p. 592), Lindé (2009, p. 600), Cantore, Gabriel, et al. (2013a, Table 18.2 on p. 432), Cantore, Levine, et al. (2015, Table 2 on p. 141). For the annual depreciation rate, one arrives at approximately 10%. Also common values of the quarterly depreciation rate are $\delta = 0.012$ or $\delta = 0.005$, see Cooley and Prescott (1995, p. 22) and Aguiar and Gopinath (2007, Table 3 on p. 86), respectively.

780 This value is in line with Brzoza-Brzezina, Kolasa, et al. (2013, Table 1 on p. 40) and Mitra et al. (2013, p. 1953). A value of $\alpha = 0.33$ is further in the range for this parameter frequently used in the literature, e.g., Aguiar and Gopinath (2007, Table 2 on p. 86) specify $\alpha = 0.32$, Lindé (2009, p. 600) uses a value of $\alpha = 0.36$, and Cooley and Prescott (1995, p. 22) set $\alpha = 0.4$.

781 This is in line with, e.g., Lindé (2009, p. 600), Aguiar and Gopinath (2007, Table 3 on p. 86), Brzoza-Brzezina, Kolasa, et al. (2013, Table 1 on p. 40) and Gomme and Lkhagvasuren (2013, Table 24.2 on p. 586).

782 This value is in accordance with Domeij and Flodén (2006, p.250), Christoffel and Kuester (2008, p. 872) and Brzoza-Brzezina, Kolasa, et al. (2013, Table 1 on p. 40).

783 See Fiorito and Zanella (2012, p. 185).

784 See Fiorito and Zanella (2012, p. 185). They also indicate that the corresponding micro estimates based on panel data are between 0.08 and 0.12.

785 See Peterman (2016, Table 3 on p. 108). These estimates are higher than the values of Fiorito and Zanella (2012), see also the discussion in Peterman (2016, pp. 110ff.).

recommend a value for the Frisch elasticity in a representative agent model as the ones considered in this thesis, of 0.75.[786] In stark contrast to Peterman (2016), Chetty et al. (2011) underlines that values above 1 are inconsistent with the data.[787]

Furthermore, as pointed out by Gomme and Lkhagvasuren (2013), micro-founded estimates of the Frisch elasticity may lead to unrealistic steady state values of hours worked in the model under consideration.[788] For this reason, the suggestion made by Gomme and Lkhagvasuren (2013) is followed, and the parameter $\varkappa$ is calibrated to match a steady state of hours worked $H_{ss} = 0.33$, i.e., the representative household supplies roughly a third of its discretionary time on the labor market in the steady state.[789] Using the steady state values in Table B.2, one arrives at a value of $\varkappa = 31.76$.[790]

In order to keep both models comparable, the remaining parameter γ in the model with Cobb-Douglas preferences is chosen such that the steady state values of both models coincide. Since the model with CD preferences has a balanced growth path, in contrast to the model with additive separable utility, this is done by assuming that the steady state growth rate is zero ($g_{ss} = 0$). Overall, $\gamma = 0.6243$ ensures that both models without growth have the same steady state values.[791]

4.4.2 Short and Long Memory Technology Shocks in the Model with Additive Separable Utility

Before the model's response to a long memory technology shock is analyzed, the model's response to the standard transitory AR(1) technology shock is illustrated. The methodology how to derive the impulse-response functions is outlined in the context of the model with Cobb-Douglas utility in Appendix B.5.5.[792]

Recall that in this section it is abstracted from growth of the economy, i.e., $\bar{A}_t \equiv 1$ and

[786] See Chetty et al. (2011, p. 474).
[787] See Chetty et al. (2011, p. 474).
[788] See Gomme and Lkhagvasuren (2013, p. 583).
[789] See Gomme and Lkhagvasuren (2013, p. 583), Domeij and Flodén (2006, p. 246) and Christoffel and Kuester (2008, p. 872).
[790] This value is roughly of the same magnitude as in Domeij and Flodén (2006, Table 1 on p. 246) who report a value of 30 in their model.
[791] This value is of the same magnitude as those used in the literature. Lindé (2009, p. 600), for example, uses a value of 0.67, Aguiar and Gopinath (2007, Table 3 on p. 86) choose a value of 0.64 and Cantore, Gabriel, et al. (2013a, Table 18.1 on p. 420) choose a value of 0.69.
[792] The procedure for deriving the IRFs in the model with additive utility (but without growth) is quite similar to the one for the model with CD utility. The procedure is illustrated along the model with CD utility since the CD model involves two shocks and is, thus, more complex than the model with additive separable utility. The derivation of the IRF to the transitory TFP shock and the growth shock is given in Section B.5.5.1 and Section B.5.5.2, respectively.

$g_{ss} \equiv 0$. In addition, TFP is equal to the transitory productivity component A_t.

Figure 4.1 illustrates the IRFs of the model's variables for various values of the autoregressive parameter ϱ_A over 140 quarters. The IRF of the total factor productivity is depicted in the upper-left-hand panel of Figure 4.1. These are just the IRFs of an AR(1) process similar to Figure 2.1. In all cases, the shock dissipates at an exponential rate, more slowly the higher the parameter ϱ_A.

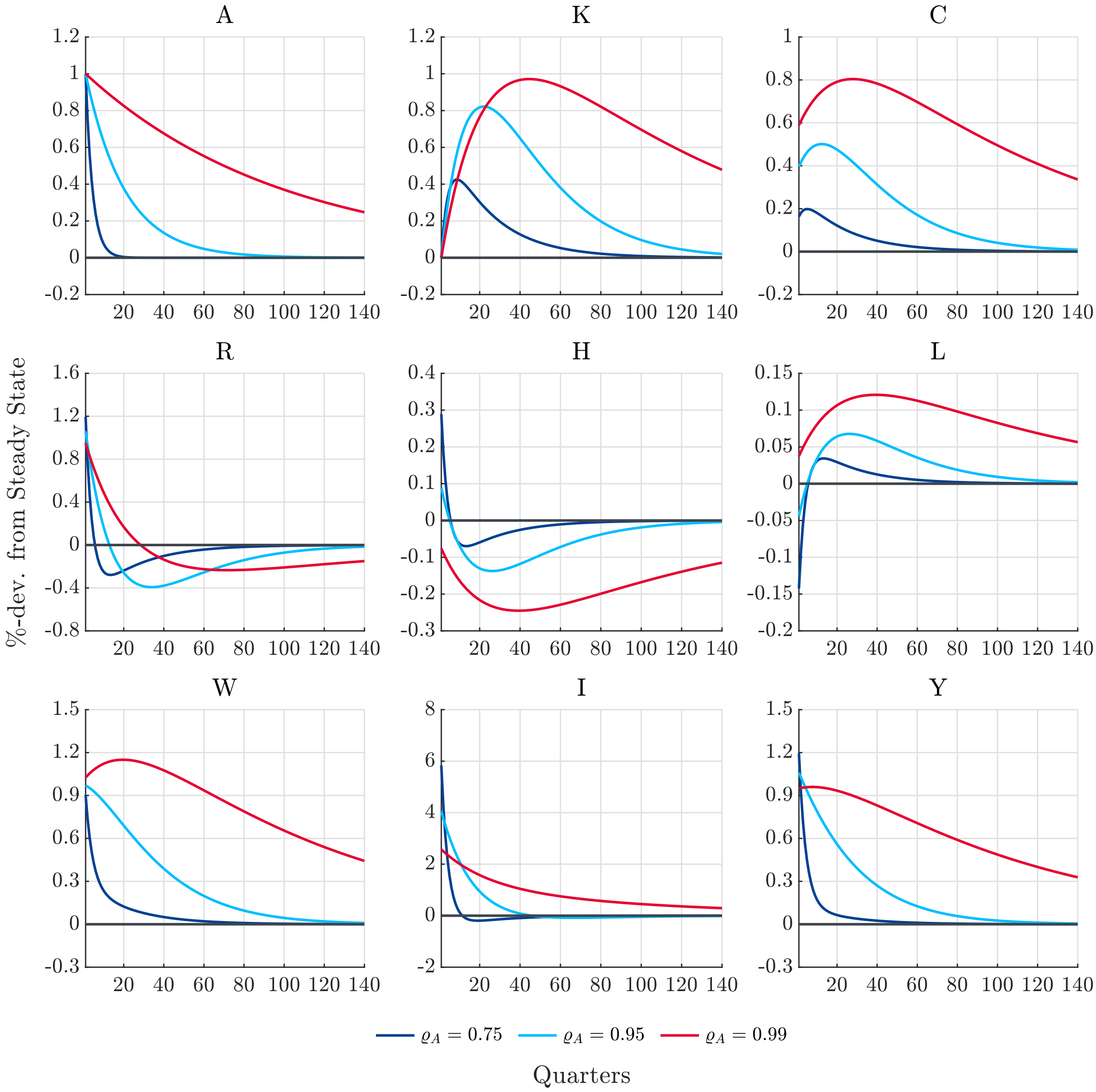


Figure 4.1: Responses of the model with additive separable utility function to a 1% transitory short memory technology shock for various values of ϱ_A. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Note the different scaling of the vertical axis.

An initial increase in total factor productivity increases the marginal productivities of labor and capital. Thus, the representative firm demands more factors of production, see (4.13) and (4.12).[793] Since the capital stock is predetermined, it cannot respond immediately to the positive technology shock; hence the rental rate on capital raises initially as can be seen from the left-hand panel in the second row of Figure 4.1.[794]

In the labor market, both the demand and supply sides are affected. By plugging the production function (4.1) into (4.12) or by considering their respective linearized counterparts stated in Table B.3, the labor demand curve is given by

$$\tilde{W}_t = \tilde{A}_t + \alpha \tilde{K}_t - \alpha \tilde{H}_t. \tag{4.34}$$

Clearly, a positive technology shock shifts (4.34) upwards. The linearized version of labor supply curve (4.10) is given by

$$\tilde{W}_t = \varsigma \tilde{C}_t + \varphi \tilde{H}_t, \tag{4.35}$$

see again Table B.3. Since consumption rises initially, as illustrated in the upper-right-hand panel of Figure 4.1, the labor supply curve (4.35) is also shifted upwards. Therefore, the real wage increases, but the effect on employment is ambiguous. It can be seen from the center panel of Figure 4.1 that the initial response of employment depends (among others) on ϱ_A. A high value of the autoregressive parameter provokes a large initial response of consumption that shifts the labor supply curve substantially upwards, leading to an overall

793 Note that by Table B.3, one has that $\tilde{R}_t = \tilde{Y}_t - \tilde{K}_t$ and $\tilde{W}_t = \tilde{Y}_t - \tilde{H}_t$. Thus, the panels of the rental rate on capital R_t and the real wage W_t in Figure 4.1 additionally show the responses of the capital and labor productivity from their corresponding steady state values, respectively.

794 All figures in this chapter were computed using Matlab code written by the author on the basis of the solution methodology described in detail in Appendix B. In practice, DSGE models are often solved numerically with the Matlab software package Dynare; see Adjemian et al. (2022, Sections 1 and 2) for a short introduction and technical description of Dynare. To the best of the author's knowledge, Dynare cannot be applied to the long memory model yet. For the pure short memory case, i.e., when TFP is an AR(1) process, Dynare can be applied, and the author's code delivers (fortunately) concurrent results to Dynare's ones. The presentation of the IRFs of the capital stock in this and the following section differs slightly from the one of Dynare. The reason is that Dynare computes IRFs of predetermined variables with a "stock at the end of the period"-concept, i.e., Dynare's IRF of the capital stock shows the response of the capital stock that is used in the following production period, see Adjemian et al. (2022, pp. 22f.). In this thesis, the author decides to follow a "stock at the beginning of the period"-concept, i.e., the IRF of the capital stock in this thesis shows the response of the capital stock used in the current production period. Assume, for example, that the economy is in its steady state at period 0, i.e., $\tilde{K}_0, \tilde{I}_0 = 0$. With the "stock at the beginning of the period"-concept followed by the author, one has $\tilde{K}_1 = (1-\delta)\tilde{K}_0 + \delta \tilde{I}_0 = 0$. By following the "stock at the end of the period"-concept like Dynare, one has $\tilde{K}_1 = (1-\delta)\tilde{K}_0 + \delta \tilde{I}_1 = \delta \tilde{I}_1$ which reflects the capital stock used for production in period 2. This differing concept is why the capital stock's IRF computed by Dynare jumps initially, although the capital stock is actually a predetermined variable, see Adjemian et al. (2022, p. 22). In the end, one can recover the "stock at the end of the period"-concept from the "stock at the beginning of the period"-concept by shifting the capital stock's IRF calculated by the author one period backward.

negative initial response of employment. Obviously, the opposite behavior is mirrored in the response of leisure; see the right-hand panel in the second row of Figure 4.1

For all values of ϱ_A, labor income $\tilde{W}_t + \tilde{H}_t$ increases initially, and consequently, total income has to rise as illustrated in the lower-right-hand panel of Figure 4.1.[795] Since the initial response of employment depends negatively on ϱ_A, the positive response of Y is smaller for higher values of ϱ_A. Since consumption does not increase as much as total income, investment expenditures have to rise, too.

In the periods following the shock, consumption rises for some periods for all values of ϱ_A. The length of this rising period further depends positively on ϱ_A. Additionally, the capital stock increases as long as investment expenditures compensate for depreciation, and it declines if investment expenditures are smaller than the depreciation of the capital stock.[796] The capital stock expands in the periods following the shock due to the initial positive response of investment expenditures. The increasing capital stock and consumption expenditures affect the labor market adversely. It follows directly from (4.34) that the increasing capital stock shifts the labor demand curve upward due to its positive effect on labor productivity. However, at the same time, the initial TFP shock dissipates, which has a dampening effect on the labor demand.

On the other hand, the labor supply curve is also pushed upwards as long as consumption expenditures are rising, see (4.35), i.e., the effects on employment and the real wage depend on the relative shifts in the labor supply and demand curves. In the case of $\varrho_A = 0.99$, the labor demand curve is shifted upwards for some periods following the shock, i.e., the positive effect of the expanding capital stock outweighs the dampening effect of dissipating TFP. Together with the upward shifted labor supply curve, this leads to increasing wages; see the lower-left-hand panel of Figure 4.3. In contrast, the labor demand curve shifts downwards in the cases of $\varrho_A = 0.75$ and $\varrho_A = 0.95$. This shift unambiguously reduces employment and, given the parameters in Figure 4.3, lets the real wage fall.

The effect on employment could thus be either positive or negative. Given the parameters

[795] It follows from the Cobb-Douglas production function that the share of labor income in total income $W_t H_t / Y_t$ is constant, see (4.12). Thus the percentage deviation of total income from its steady state is equal to the one of labor income, i.e., $\tilde{Y} = \tilde{W} + \tilde{H}$, see Table B.3.

[796] Note that investment expenditures need not fall below their steady value to induce a decline in the capital stock. If investment expenditures are equal to their steady state value, they compensate for the depreciation of the steady state capital stock, i.e., $I_{ss} = \delta K_{ss}$. Consequently, if the capital stock is much larger than its steady state value, a small positive deviation of investment expenditures from their steady state cannot compensate for the large depreciation associated with the high capital stock. That is, investment expenditures are positive and may be above their steady state value, but the capital stock is already beginning to depreciate.

in Table 4.1, employment decreases in the periods following the shock for all depicted values of ϱ_A. Again, the behavior of leisure is opposite to the behavior of employment. Overall, decreasing wages and employment imply a falling labor income and thus falling total income or output, as can be seen from the lower-right-hand panel of Figure 4.1. In the case of $\varrho = 0.99$, the initial wage increase can compensate for the reduced employment by keeping the labor income nearly constant for about ten periods.

Since consumption expenditures grow for some periods after the shock and total income decreases, investment expenditures must also decrease. Since consumption increases at a decreasing rate, it follows from the Euler equation (4.11) that the rental rate on capital decreases. Note that the rental rate on capital crosses its steady state line at the same time as consumption peaks.

After the peak, consumption returns to its steady state value, and thus the labor supply curves shift downwards. So does the labor demand curve, at least after the capital stock has peaked. This downward shift implies falling real wages. The effect on employment depends again on the relative shifts in the demand and supply curves. Figure 4.1 shows that employment returns to its steady state level after reaching a negative peak. Overall, output and investment converge steadily to their steady state levels; the higher the autoregressive parameter, the slower the convergence. The rental rate on capital converges back to its steady state value from below. This behavior follows again from the Euler equation as the rate at which consumption expenditures decrease becomes less negative with time.

That income Y and consumption expenditures C are above their respective steady state values even if the initial shock has dissipated illustrates how the household manages the positive productivity shock intertemporally and how the household smooths its consumption path.[797]

As the labor market is adversely affected during the dissipation of the technology shock, Figure 4.2 illustrates how the labor market equilibrium, i.e., wages and employment behave while the TFP shock dies out. There, the percentage deviation of the real wage from the steady state is plotted against the percentage deviation of hours worked, each divided by the initial shock size, i.e., the vertical and horizontal axes in Figure 4.2 show the normalized percentage deviation of the real wage W_t and hours worked H_t from their steady states, respectively.[798] The resulting graphs are wage-employment loci that illustrate how wages

797 Consider, for example, the case of $\varrho_A = 0.75$ in Figure 4.1. At period 20, the initial shock is almost fully dissipated, but income and consumption are clearly above their respective steady state values.

798 Figure 4.2 is derived from Figure 4.1 by plotting the IRF of W_t against the IRF of H_t each divided by the initial shock size 0.01. To be more precise, each panel of Figure 4.2 show the pairs of $100((\log(W_t) - \log(W_{ss}))/0.01)$ and $100((\log(H_t) - \log(H_{ss}))/0.01)$ for $t = 1, \ldots, 140$. Figure 4.2 was

and the hours worked interact in response to an initial transitory short memory TFP shock. The filled dot in each panel of Figure 4.2 marks the combination at the period of shock occurrence ($t = 1$) and the filled square marks the variable combination at period $t = 140$, which is equal to the last period in Figure 4.1. As can be seen from the lower-left-hand and the center panel of Figure 4.1, the initial shock is almost fully dissipated at $t = 140$ for the cases of $\varrho_A = 0.75$ and $\varrho_A = 0.95$. Thus, the square in the left-hand and middle panel of Figure 4.2 are close to the point $(0, 0)$, where the real wage and hours worked equal their steady state values. In the case of $\varrho_A = 0.99$, the square is far from this point as the initial shock still has a substantial effect at $t = 140$ due to a higher autoregressive parameter.[799]

The solid lines in Figure 4.2 indicate that wages and employment move in the same direction, i.e., either are both increasing or decreasing. Correspondingly, dashed lines in Figure 4.2 refer to periods where employment and the real wage behave unequally, i.e., employment rises with falling wages or vice versa. An asterisk marks an inflection point, where the behavior of the real wage or employment changes. As already mentioned, the response of consumption pushes the labor supply curve upwards. The labor demand curve is initially pushed upwards as the technology shock positively affects labor's marginal productivity. Then, the labor demand curve's movement depends on whether the capital stock grows faster than the technology shock dissipates. As already mentioned, in the cases of $\varrho_A = 0.75$ and $\varrho_A = 0.95$, the decay of the technology shock is faster than the accumulation of capital; thus, the labor demand curve is shifted downwards in all periods following the shock. As shown in the left-hand and middle panels of Figure 4.2, this leads to falling wages and employment immediately after the shock until employment reaches its minimum value. Afterward, wages continue to decrease, but employment begins to rise again. Overall, a higher value of the autoregressive parameter slows down the fading out of the technology shock; hence, by (4.34), the downward shift in the labor demand curve is also slowed down and the inflection point in the left-hand and middle panels of Figure 4.2 is pushed upwards. In the case of $\varrho_A = 0.99$, the fading out of the technology shock is so slow that the labor demand curve shifts upwards right after the shock. This leads to periods of rising wages by keeping employment decreasing; see the right-hand panel of Figure 4.2.

The rate ϱ_A at which the initial shock dissipates affects the household's consumption decision, labor supply, and (among others) labor demand. The technology shock is,

inspired by Costa Junior (2016, Figure 2.14 on p. 54).

799 Of course, the endpoint of the locus can be brought closer to $(0, 0)$ by plotting the locus over more periods. This has not been done here to keep Figure 4.1 and Figure 4.2 comparable.

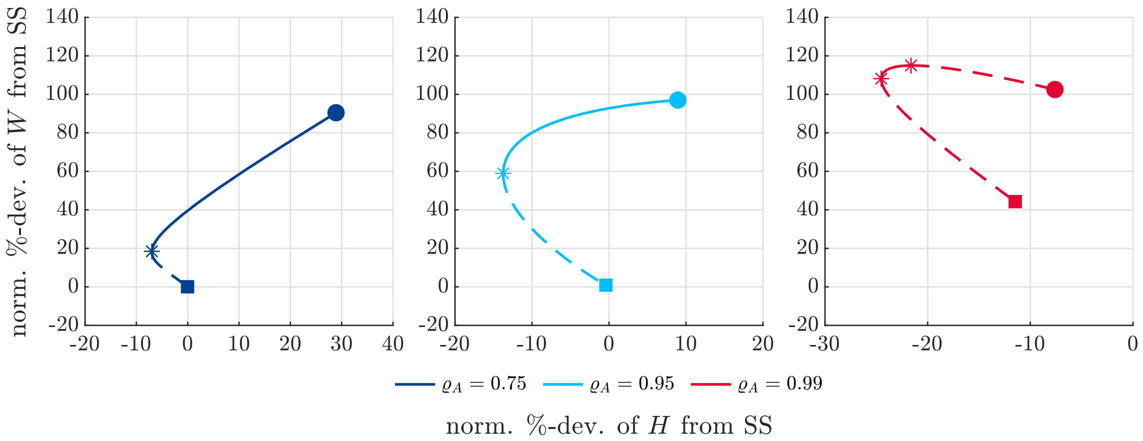


Figure 4.2: Wage-employment loci for various values of ϱ_A. The vertical and horizontal axes report percentage deviations of the real wage and hours worked from their respective steady states divided by the initial shock size. The filled dot marks the variables' combination at the period of shock occurrence ($t = 1$) and the filled square at ($t = 140$). Solid lines indicate periods in which wages and employment behave in the same way, either increasing or decreasing; dashed lines indicate periods in which employment and the real wage behave in different ways, i.e., employment is rising with falling wages or vice versa. An asterisk marks inflection points where the behavior of the real wage or employment changes. Note the different scaling of the horizontal axis.

however, for all values of ϱ_A in Figure 4.1 and Figure 4.2 a short memory process whose cumulative impulse response (CIR) is given by (2.16). According to Definition 2.1.6, these processes are also moderately persistent. The question immediately arising is whether a long memory technology shock affects the model economy differently than a short memory shock. Therefore, Figure 4.3 illustrates the model responses to a 1% transitory pure long memory shock for various values of d. At the same time, ϱ_A is set to 0 for all panels in Figure 4.3 to illustrate the effects that stem from the long memory parameter alone. The technology process is thus an ARFIMA$(0, d, 0)$ process and hence by Table 2.1 also strongly persistent.

As above, the upper-left-hand panel of Figure 4.3 shows just the IRF of an ARFIMA$(0, d, 0)$ process such as the ones in Panel a) of Figure 2.7. For all values of d, there is a significant drop in the IRF in the periods following the shock before the slow decay sets in. Not surprisingly, the decay is slower the higher the parameter d is. In contrast to Figure 4.1, the decay is now at a hypergeometric rate, i.e., the decay is described by Ck^{d-1} as $k \to \infty$, a constant $C > 0$, and k is the number of periods after the shock; for details see Lemma 2.4.3.[800]

However, the main mechanisms are similar to the ones already described around Figure 4.1. The positive technology shock increases marginal productivities of labor and capital and

[800] Recall that the decay is simply ϱ_A^k in Figure 4.1.

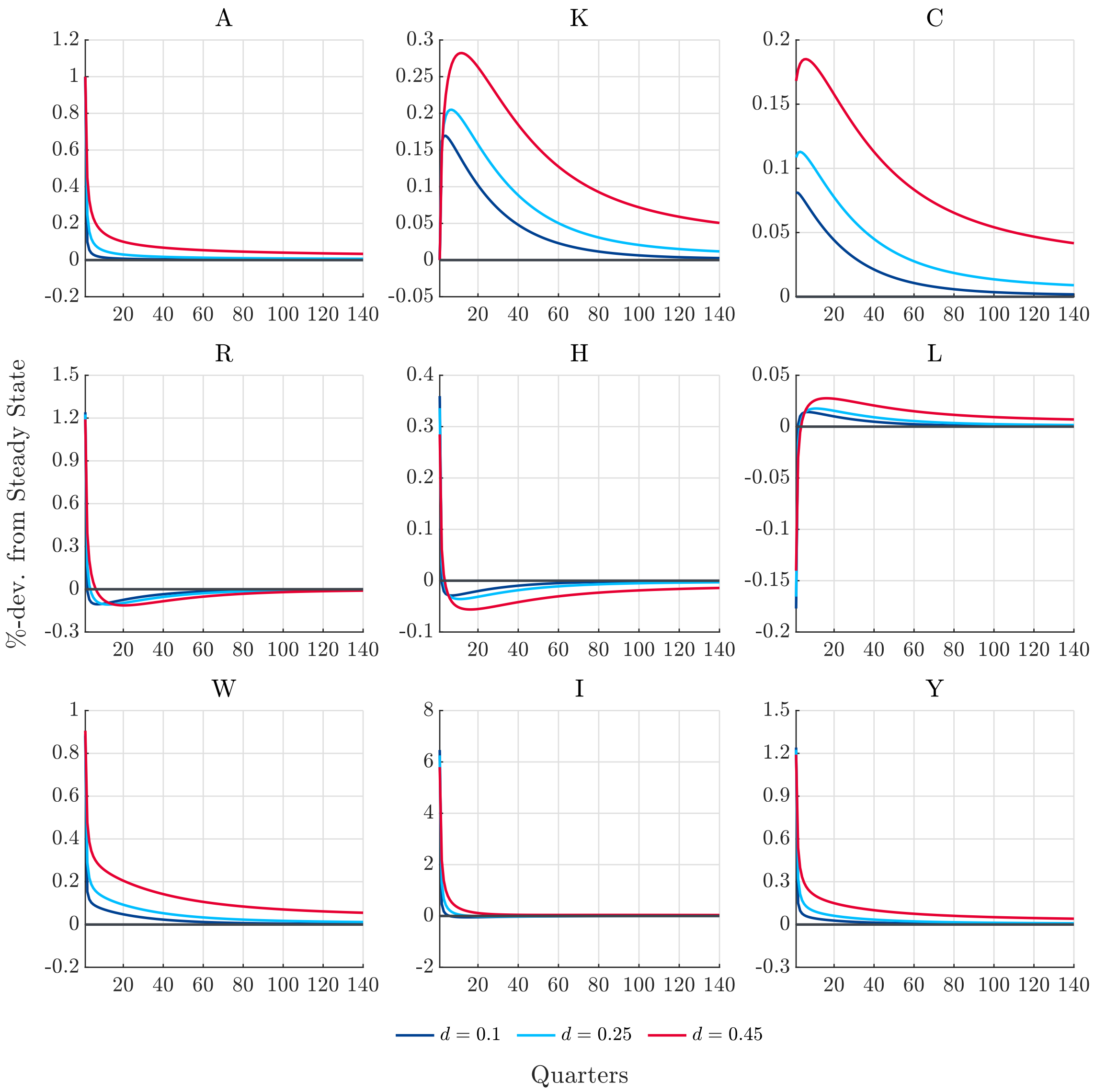


Figure 4.3: Responses of the model with additive separable utility function to a 1% transitory long memory technology shock for various values of d. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Note the different scaling of the vertical axis.

thus increases the demand for production factors. The rental rate on capital rises initially. Similar to the response to a short memory AR(1) process, the initial response of the rental rate on capital depends negatively on the value of d, i.e., an increasing value of d leads to a less pronounced positive response of R_t. This negative dependence seems overall to be less distinctive than in Figure 4.1. The initial response is roughly 1.2% for all values of d, but the initial response of R_t varies between 1% and 1.2% in Figure 4.1.

Similar to Figure 4.1, consumption's initial response depends positively on d. For all values of d, however, the magnitude of the initial response is smaller than the one in Figure 4.1 with $\varrho_A = 0.75$. For this reason, the labor supply curve is shifted less upwards than in all cases depicted in Figure 4.1. Thus, the real wage increases in all cases and the effect on employment is still ambiguous. However, in all cases shown in Figure 4.3 employment initially rises. This increase indicates that the shift in the labor supply curve is too small to generate an initial negative response in hours worked, even in the case of strong long memory ($d = 0.45$), see the center panel of Figure 4.3. The effect on output is again positive as factor prices and inputs increase initially. Due to the small response of consumption, investment expenditures increase at a similar magnitude as in Figure 4.1 with $\varrho_A = 0.75$. The initial response of investment expenditures decreases with an increasing value of d but to a much lesser extent than in Figure 4.1 with an increasing value of ϱ_A.

Due to the positive response of investment expenditures, the capital stock begins to accumulate, which in turn has a positive effect on labor demand. However, this positive effect on the labor demand does not materialize due to the significant drop in TFP, which affects labor demand negatively and, thus, outweighs the capital's positive effect. Overall, there is a large downward shift in the labor demand curve. On the other hand, the labor supply curve shifts upwards by a small amount when d is equal to 0.25 and 0.45, due to rising consumption expenditures, and downwards when d is equal to 0.1 Compared to the shift in the labor demand curve, these shifts are negligible as the overall response of consumption is of a small magnitude. Therefore, the labor market's response is mainly determined by the significant downward shift in the demand curve, resulting in decreasing wages and employment as seen from the center and the lower-left-hand panel of Figure 4.3.

The effect of a pure long memory TFP shock can also be seen from Figure 4.4, which shows the wage-employment loci for the model responses of Figure 4.3. The shape of all loci is essentially the same. In each case, there is one inflection point that marks the transition between an equal behavior of wages and employment (both either increasing or decreasing) to an unequal behavior (in all panels of Figure 4.4 employment begins to rise at the inflection point, and wages keep falling). Overall, the loci are similarly shaped as the ones in Figure 4.2 with $\varrho_A = 0.75$. Furthermore, an increase in the parameter d shifts the inflection points upwards, similar to how an increase in the value of ϱ_A does in Figure 4.2, but to a much lesser extent. In contrast to Figure 4.2, where there is a period of increasing wages with $\varrho_A = 0.99$, there is no such period in Figure 4.4, even for large values of d since TFP drops faster than the capital stock expands.

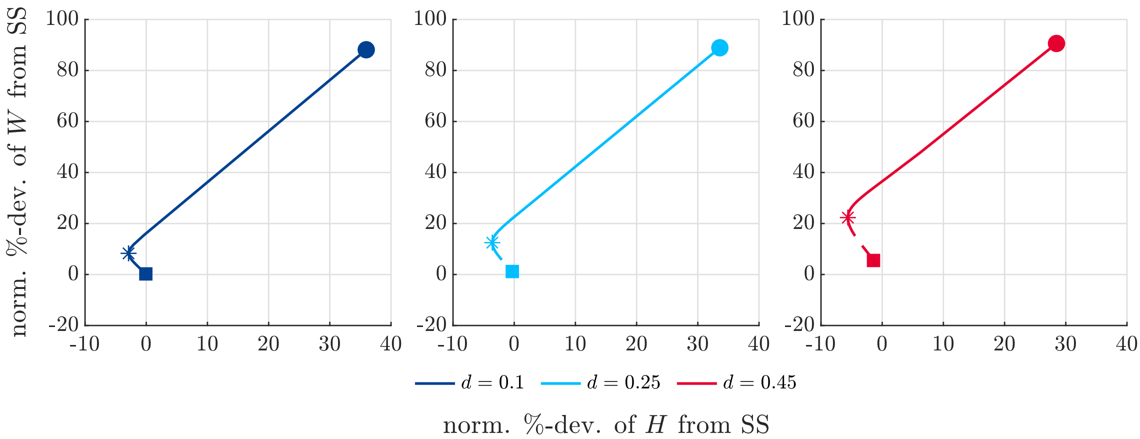


Figure 4.4: Wage-employment loci for various values of d. The vertical and horizontal axes report percentage deviations of the real wage and hours worked from their respective steady states divided by the initial shock size. The filled dot marks the variables' combination at the period of shock occurrence ($t = 1$) and the filled square at ($t = 140$). Solid lines indicate periods in which wages and employment behave in the same way, either increasing or decreasing; dashed lines indicate periods in which employment and the real wage behave in different ways, i.e., employment is rising with falling wages or vice versa. An asterisk marks inflection points where the behavior of the real wage or employment changes.

In summary, increasing values of d shifts the IRFs in the same direction as increasing values of ϱ_A in Figure 4.1. Consequently, increasing the degree of persistence (either moderate or strong persistence) has similar effects regarding the direction of the shifts, but the magnitude may differ substantially. Not surprisingly, long memory is a property that determines the asymptotic behavior of the dynamics. For all variables in Figure 4.3, there is a slow return to the steady state values. Compare, for example, the response of consumption with $d = 0.45$ in Figure 4.3 with the one of in Figure 4.1 with $\varrho_A = 0.75$. The initial response is comparable in magnitude, but the value of consumption at period 140 is higher in the long memory case than that in the short memory case, reflecting the hypergeometric decay of the initial TFP shock.

By comparing the upper-right-hand panels of Figure 4.1 and Figure 4.3, it seems surprising that a pure long memory shock is not able to change the consumption decision of the household to the same extent as a short memory shock. Recall from Table 2.1 that technology in Figure 4.3 is not only a long memory process but also a strongly persistent process, i.e., according to Definition 2.1.7 the cumulative sum of the IRF (CIR) is infinitely large. In Figure 4.1, however, the cumulative effect of the shock is given as $1/(1-\varrho_A)$, see (2.16). Thus, the cumulative impulse response of the technology shock in Figure 4.1 ranges between 4 ($\varrho_A = 0.75$) and 100 ($\varrho_A = 0.99$) and is thus negligible compared to that in Figure 4.3 which are all infinite. By the assumption of rational expectations, the household knows the stochastic nature of the shock, i.e., the household knows that its income will be above its steady state value for very long periods following a positive long memory technology shock. Due to an intertemporal income effect, i.e., a higher income tomorrow already increases consumption expenditures today, one might have expected a more significant initial response of consumption in the long memory case than in the short memory case. So why is the household so unaffected while facing such a massive cumulative effect on TFP and on its income?

The reason may be found in the permanent income hypothesis, which states that households make their consumption decisions based on their permanent income, where the permanent income at time t is defined as the sum of total income at time t, net capital assets at time t and expected discounted labor income, i.e.,[801]

$$\begin{aligned} Y_t^p &:= (R_t + (1-\delta))K_t + W_t H_t + \mathbb{E}_t \left(\sum_{j=1}^{\infty} \prod_{i=1}^{j} (R_{t+i} + (1-\delta))^{-1} W_{t+j} H_{t+j} \right) \\ &= Y_t + (1-\delta)K_t + (1-\alpha)\mathbb{E}_t \left(\sum_{j=1}^{\infty} \prod_{i=1}^{j} (R_{t+i} + (1-\delta))^{-1} Y_{t+j} \right), \end{aligned} \tag{4.36}$$

where Y_t^p refers to the household's permanent income.[802] The effects of the long memory shock on income and the capital stock can be seen from the lower-right-hand and the middle panel in the first row of Figure 4.3, respectively. Since labor income is a fixed fraction of total income due to (4.12), the percentage deviation of labor income from its steady state is equal to that of total income Y.[803]

The difference between the IRFs of Y in Figure 4.1 and Figure 4.3 is the distribution of the "mass" of the IRF over the periods. In the short memory case, a substantial part of the cumulative IRF is allocated to the first 140 periods (roughly 74% with $\varrho_A = 0.99$ and more than 99% with $\varrho_A = 0.95$) while the substantial part of the IRF in the long memory case is allocated to the periods after period 140.[804] These long-run effects of the technology shock affect the household's consumption decision at period t through the future income Y_{t+j} in (4.36). The effect of a positive technology shock on the household's

[801] See Wen (2001, pp. 1224f.).
[802] The second line uses (4.4), (4.7) and (4.12).
[803] Recall from Table B.3 that $\tilde{Y}_t = \tilde{H}_t + \tilde{W}_t$.
[804] The mass after period 140 is infinite, like the whole cumulative impulse-response.

permanent income, however, cannot be assessed easily from (4.36), the IRF depicted in Figure 4.1 and Figure 4.3 since a technology shock affects the return on capital, income, wages and hours worked simultaneously.

To gain some intuition, assume, for the moment, a constant rental rate on capital in (4.36), which is used by the household to discount its future income. Then, the discount factor on future labor income decreases exponentially. Related to the long memory case, this implies that the periods in which the mass of the IRF lies are associated with such small discount factors that the present value of the household's future labor income, and consequently its consumption expenditures at time t, are barely affected. Conversely, in the short memory case, the periods with a significant mass of the IRF correspond to those with high discount factors, resulting in an overall higher present value of future labor income.

That this mechanism also holds in the case of a time-varying rental rate on capital can be seen from Figure 4.5, which shows the percentage deviation of the household's permanent income and the respective contributions of the current income, current capital stock, and the discounted labor income in response to a 1% transitory TFP shock. For a precise definition of these shares and their calculation, see Appendix B.3.5. Panels a) and b) of Figure 4.5 refer to the IRF of permanent income with the same parameters as in the light-blue lines of Figure 4.1 and Figure 4.3. As expected from Figure 4.1 and Figure 4.3, the contribution of current income and capital stock to the percentage deviation of the household's permanent income is smaller in the pure long memory case than in the short memory case. It is further illustrated that in both cases, permanent income is primarily influenced by the discounted future labor income. In the pure long memory case, the contribution of the discounted labor income is substantially smaller than in the short memory case. This observation confirms the reasoning drawn from above, i.e., the household discounts away the sizable cumulative effect of a long memory technology shock. Thus its consumption decision in the periods after the shock occurrence is relatively unaffected by the pure long memory shock.

Overall, the considerations so far illustrate that a pure long memory technology shock leads to qualitatively similar model responses as a pure short memory shock, but quantitatively they differ substantially. Furthermore, for the household's consumption and labor supply decision, the dynamics immediately after the shock seem more decisive than the long-run asymptotic behavior of the technology shock's IRF. From this perspective, one is tempted to say that long memory, which primarily determines the asymptotic behavior of the IRF and the autocorrelation function (ACF), plays at least a minor role in linear DSGE models. Instead, the focus should be on the correct specification of the short term dynamics of the

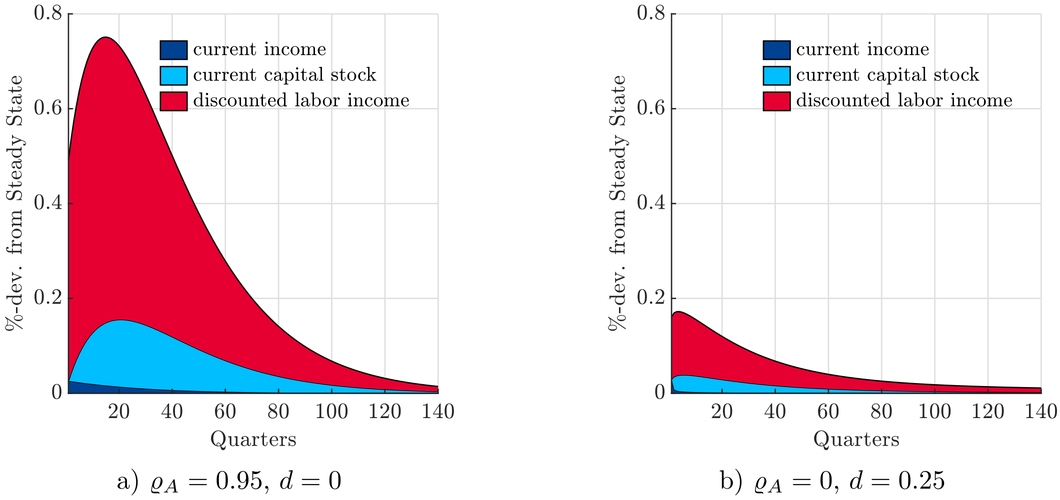


Figure 4.5: Response of the household's permanent income to a 1% transitory long and short memory technology shock in the model with additive separable utility function. The colored areas highlight how the three components, current income, current capital stock, and discounted labor income contribute to the percentage deviation of the permanent income. Panels a) and b) correspond to the parameter constellation of the light-blue line in Figure 4.1 and Figure 4.3, respectively. For a precise definition of these shares, see Appendix B.3.5.

exogenous processes.

However, as illustrated in Section 2.4.2, long memory may also affect the response to a shock in the short run, especially in the presence of additional short memory dynamics. To be more precise, in Panels b) and c) of Figure 2.7, it is illustrated that the IRF of an ARFIMA$(1, d, 0)$ process depends on both the short memory and long memory parameter. This can be seen from a visual comparison between the red and yellow lines in Panels b) and c) in Figure 2.7, the introduction of an additional long memory component to an AR(1) process does not only affect the long-run behavior of the IRF. Long memory also has an effect on the periods right after the shock. Lemma 2.4.3 tries to formalize this. There, it is illustrated that the TFP's IRF is hump-shaped if $\varrho_A + d > 1$, see part iv) of Lemma 2.4.3. That the long memory parameter affects the periods immediately after the shock can further be seen from part ii) of Lemma 2.4.3, which states that, given a 1% transitory TFP shock, TFP deviates about $(\varrho_A + d)$ % from its steady state in the first period following the shock.

However, the reverse is also true, i.e., the short memory part of the ARFIMA$(1, d, 0)$ process has long-run effects on the IRF of the whole process as illustrated in part iii) of Lemma 2.4.3. This part states that the higher ϱ_A, the higher the impulse-response function for a given value of d. Thus, the autoregressive parameter does not affect the rate at which the IRF converges to zero but affects the level from which the hypergeometric decay sets in. From the findings above, the short-run implications of the long memory component seem to be more relevant for the household's decision than the long-run effects of the short memory component.

Figure 4.6 seizes on some parameter combinations of ϱ_A and d depicted in Figure 2.7 and illustrates the model's response to the corresponding TFP shock. In the case of $\varrho_A = 0.75$ and $d = 0.1$, the model responses are qualitatively similar to the dark-blue lines ($\varrho_A = 0.75$ and $d = 0$) in Figure 4.1. The IRF of TFP is dominated by a quick decay of similar order as the purely exponential one. The quantitative outcomes differ slightly since the decay immediately after the shock is less pronounced than in the pure AR(1) case with $\varrho_A = 0.75$. This can also be seen from the left-hand panel of Figure 4.7, where the wage-employment locus along the dissipating TFP shock is depicted. Overall, the locus is quite similar to that shown in the left-hand panel of Figure 4.2.

By increasing the degree of long memory to $d = 0.4$ and leaving $\varrho_A = 0.75$ unchanged, there is the expected hump shape in the IRF of TFP. As stated above, the rental rate on capital rises initially due to the positive effect on the marginal productivity of capital. Consumption increases about 0.4 %, which is quite similar to the initial response of consumption in Figure 4.1 with $\varrho_A = 0.95$. Thus, the labor supply curve is shifted upwards. The initial positive response of the real wage results from the upward shift in the labor demand curve caused by the increase in the marginal productivity of labor. The effect on employment is still positive given the parameters in Figure 4.6. As the capital stock begins to accumulate and there are some periods of increasing TFP, the labor demand curve is shifted upwards for some periods following the shock. The same holds true for the labor supply curve due to increasing consumption expenditures. As can be seen from the center panel of Figure 4.6, the shift in the demand curve seems to be more prominent in the first periods, resulting in periods of rising employment and wages. After these periods of increasing wages and employment, there is one period with still increasing wages but falling employment. This period of increasing wages and employment is reflected in the short distance between the two asterisks at the point $(10, 120)$ in the middle panel of Figure 4.7. Overall, there are three inflection points in the wage-employment locus depicted in the middle panel of Figure 4.7, two of them separated by one period.

In Figure 4.6, output responds positively and shows a small hump shape comparable in size with the one of TFP; see the lower-right-hand panel of Figure 4.6. Since output or total income increases more rapidly than consumption, investment expenditures also have to rise during some periods. These investment expenditures then accelerate capital accumulation.

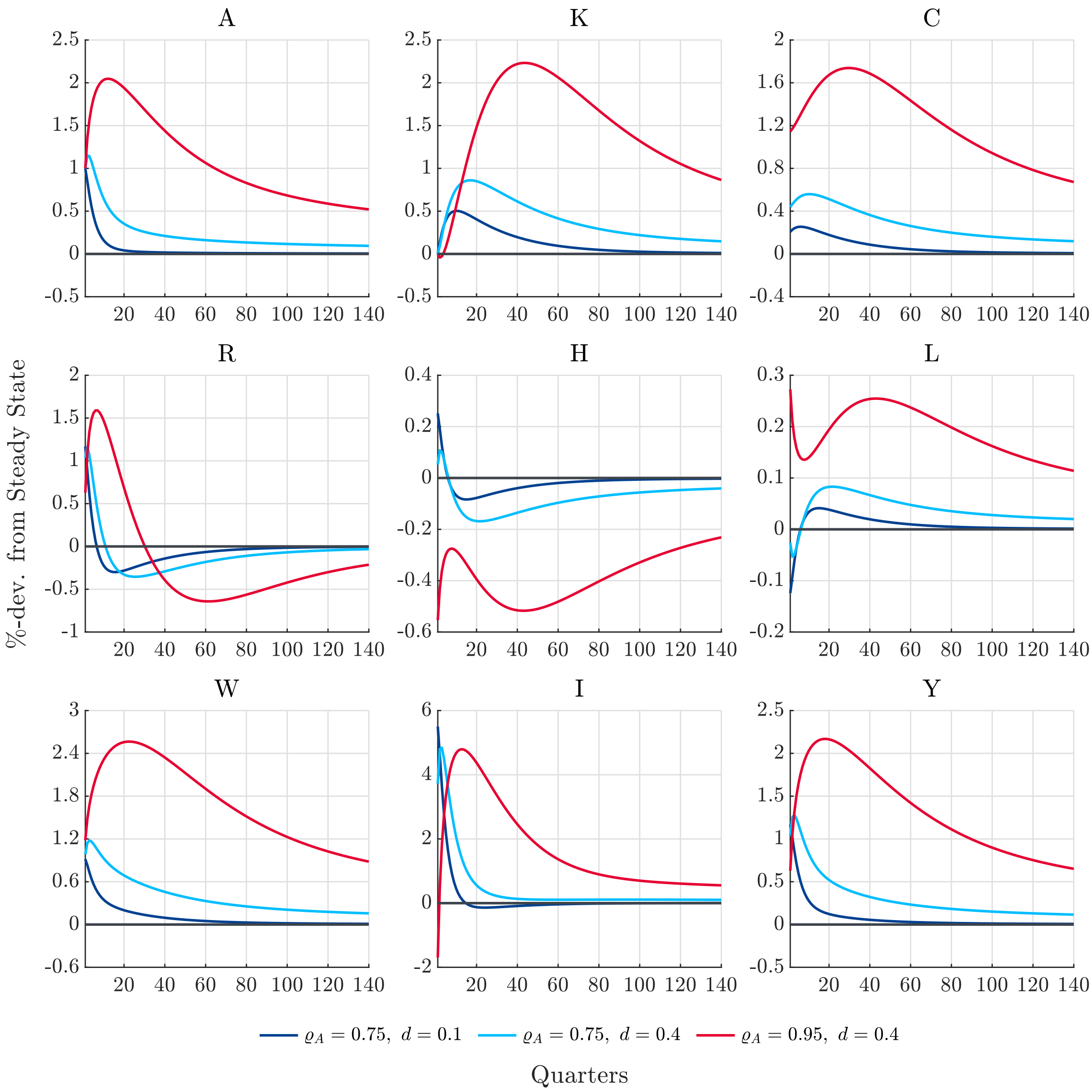


Figure 4.6: Responses of the model with additive separable utility function to a 1% transitory long memory technology shock for various values of ϱ_A and d. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Note the different scaling of the vertical axis.

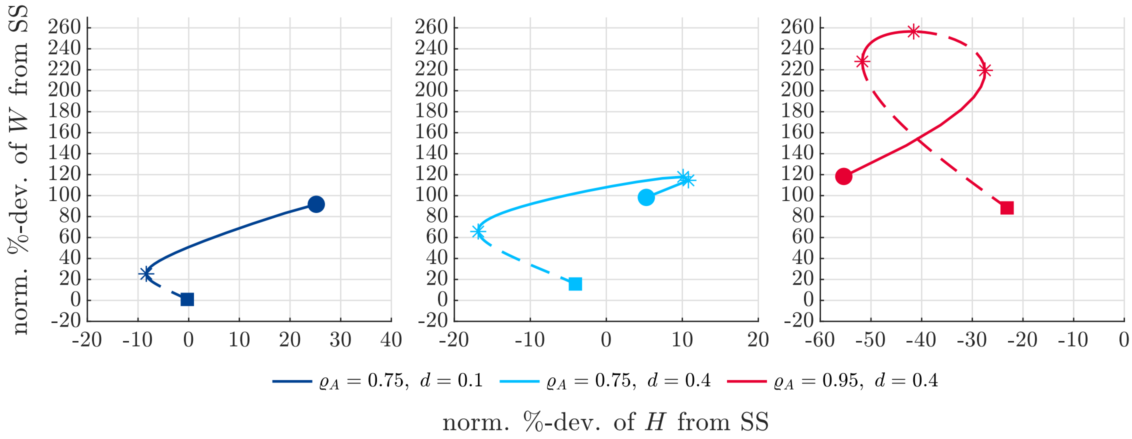


Figure 4.7: Wage-employment loci for various values of ϱ_A and d. The vertical and horizontal axes report percentage deviations of the real wage and hours worked from their respective steady states divided by the initial shock size. The filled dot marks the variables' combination at the period of shock occurrence ($t = 1$) and the filled square at ($t = 140$). Solid lines indicate periods in which wages and employment behave in the same way, either increasing or decreasing; dashed lines indicate periods in which employment and the real wage behave in different ways, i.e., employment is rising with falling wages or vice versa. An asterisk marks inflection points where the behavior of the real wage or employment changes.

In the end, the small initial increase of TFP is mirrored in the rental rate on capital, wages, employment, investment, and income. Consumption expenditures and the capital shock, whose responses were already hump-shaped in the cases before, show a more distinctive hump shape. All variables show the slow decay in later periods that characterizes the presence of long memory.

In the last case, the model responses are somewhat more extreme, i.e., by keeping the long memory parameter at $d = 0.4$ but raising the short memory parameter ϱ_A to 0.95. As expected from the condition $\varrho_A + d > 1$, there is a massive hump in TFP which roughly doubles over the first twelve periods following the shock. Afterward, the shock dissipates slowly.

Consequently, a massive initial positive response of consumption pushes the labor supply curve substantially upwards. Similar to the case of $\varrho_A = 0.99$ in Figure 4.1, the shift in the supply curve outweighs the upward shift in the labor demand curve, and employment falls initially. Due to the significant adverse effect on employment, the initial response of output is still positive but smaller than in the other two cases. Furthermore, and in contrast to all other cases considered so far, the response of consumption is greater than the one of total income, i.e., the consumption-to-income ratio rises initially.[805] Consequently, investment expenditures drop initially, and they cannot compensate for capital depreciation anymore,

[805] Note that the percentage deviation of the consumption to income ratio from its steady state value is $\tilde{C}_t - \tilde{Y}_t$. Therefore, its IRF is obtained by simply subtracting the IRF of total income from that of consumption.

pushing the capital stock below its steady state value. As household income rises faster than consumption expenditures, investment expenditures rise rapidly after the initial decline. This increase in investment then leads to accelerated capital accumulation.

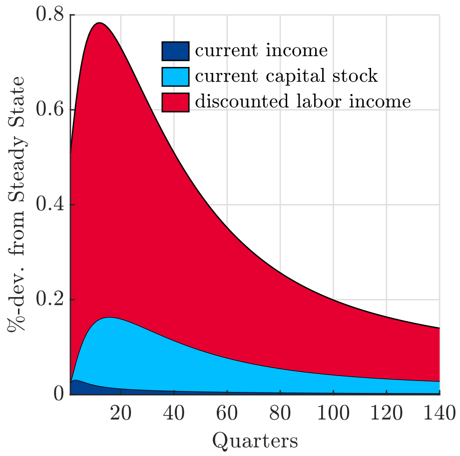
Overall, the initial shock is still significant after 140 periods. At $t = 140$, it is about half its initial size and thus higher than in the case of $\varrho_A = 0.99$ of Figure 4.1.

Again, the large response of consumption may be explained by the response of the household's permanent income. Panel b) of Figure 4.8 shows the IRF of the household's permanent income for the same parameters as for the red line in Figure 4.6. It can be seen that through the large response of wages and income, the discounted future labor income is substantial. At each instant of time, the percentage deviation of the household's permanent income is larger than the maximum value for the parameter constellation $\varrho_A = 0.75$ and $d = 0.4$, illustrated in Panel a) of Figure 4.8. In the end, the substantial effect on permanent income causes a large positive response in consumption, which in turn triggers a large negative response in labor supply, and thus in employment.

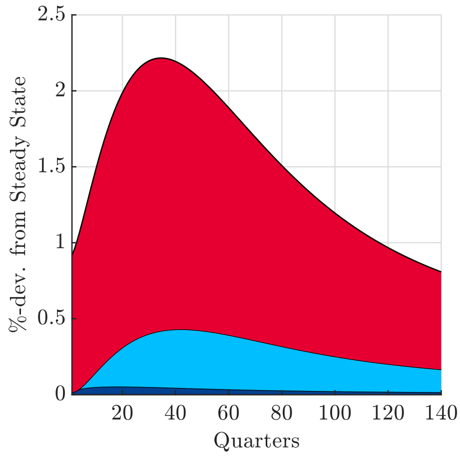
The effects on the labor market and the net effects on the simultaneously shifted labor demand and supply curves during the TFP shock's dissipation are given in the right-hand panel of Figure 4.7. As already mentioned, strong persistence indicates a negative response of employment similar to the case of $\varrho_A = 0.99$ in Figure 4.1. As illustrated in the right-hand panel of Figure 4.7, the relative shifts in the labor demand and supply curves generate periods of rising wages and employment before employment begins to fall. Similar to the case of $\varrho_A = 0.75$ and $d = 0.4$, there are three inflection points in the right-hand panel of Figure 4.7 caused by the hump-shaped IRF of TFP.

Note that the condition $\varrho_A + d > 1$ is not sufficient to create such a cyclical structure with three inflection points. To understand this, recall that the periods of increasing wages and employment right after the shock ($t = 2, 3, \ldots$) are caused by upward shifting labor demand and supply curves. From the considerations made so far, an increasing labor supply curve will likely occur due to expanding consumption expenditures right after the shock occurrence.[806] To generate periods of increasing wages and employment, the labor demand curve has to shift upwards by a sufficiently large amount, at least in periods 2 and 3. From (4.34), it follows that the intercept of the labor demand curve ($\tilde{A}_t + \alpha\tilde{K}_t$) depends to a large extent on the evolution of TFP. Hence, if $\varrho_A + d > 1$, TFP increases to the level of $(\varrho_A + d)\%$ in period 2. However, if $\varrho_A + d$ is just slightly above 1, it can be seen from Panel b) of Figure 2.8 that $\varrho_A + d$ is the maximum value of TFP's IRF. Thus, there is no

[806] Increasing consumption expenditures at the beginning were obtained in all cases of Figure 4.1, Figure 4.3 and Figure 4.7 except for a pure low level ($d = 0.1$) long memory technology shock.



a) $\varrho_A = 0.75$, $d = 0.4$



b) $\varrho_A = 0.95$, $d = 0.4$

Figure 4.8: Responses of the household's permanent income to a 1% transitory long memory technology shock in the model with additive separable utility function for various values of ϱ_A. The colored areas highlight how the three components, current income, current capital stock, and discounted labor income contribute to the percentage deviation of the permanent income. Panels a) and b) correspond to the parameter constellation of the light-blue and red lines in Figure 4.6, respectively. For a precise definition of these shares, see Appendix B.3.5. The legend of Panel b) is the same as in Panel a). Note the different scaling of the vertical axis.

additional upward shift in the labor demand curve for such combinations of ϱ_A and d.[807] In short, the hump in the IRF of TFP must be sufficiently large to generate a cyclical response in the labor market as the one shown in the right-hand panel of Figure 4.7. Panel b) of Figure 2.8 shows that such a pronounced hump occurs if $\varrho_A + d$ is well above 1.

Evidently, the response of the labor market depends crucially on the parameters of the utility function. From (4.35) follows that φ (the inverse of the Frisch elasticity) governs the slope of the labor supply curve and ς (the inverse of the intertemporal elasticity of substitution) governs the impact of consumption on the labor supply. If labor supply were independent of consumption expenditures, a positive technology shock would shift the labor demand curve upwards by keeping the labor supply curve unchanged. Hence, effects on employment and wages would unambiguously be positive.[808] Following Aguiar and Gopinath (2004), these positive employment effects are damped by the income effect on labor supply, which is reflected in the dependence of labor supply on consumption

[807] To be more precise, the intercept of the labor demand curve is given by $\tilde{A}_t + \alpha\tilde{K}_t$. Suppose TFP's IRF reaches its maximum in periods 2, i.e., $\tilde{A}_2$. In this case, it further depends on the response of the capital stock whether there is an additional upward shift in the labor demand curve in periods 3, i.e., whether $\tilde{A}_3 + \alpha\tilde{K}_3 > \tilde{A}_2 + \alpha\tilde{K}_2$. In the middle panel of Figure 4.7, TFP's IRF peaks at period 2, but the response of the capital stock is large enough to induce an additional upward shift in the labor demand curve. This, however, need not be true for all combinations of ϱ_A and d with $\varrho_A + d > 1$.

[808] See Aguiar and Gopinath (2004, p. 14).

expenditure.[809] Thus, the strong negative response of employment with $\varrho_A = 0.95$ and $d = 0.4$ seen in Figure 4.8, reflects the predominance of a strong income effect on labor supply.

The following section considers the Cobb-Douglas utility specified in (4.19). It turns out that the dependence of the labor supply on consumption is reduced and that negative effects on employment are not apparent or less pronounced given these preferences.

4.4.3 Short Memory, Long Memory, and Growth Shocks in the Model with Cobb-Douglas Utility

This section compares the model's responses to short memory, long memory, and trend shocks. Positive trend shocks as specified in (4.17) alter the level of labor augmenting technology $\bar{A}_t$ permanently by inducing a new balanced growth path. This impact contrasts short and long memory technology shocks that fade out with time; thus, the model returns to its steady state value. The findings from Section 4.4.2 indicate that for the household, the behavior of TFP in the periods immediately after the shock seems decisive. This reasoning is founded in the permanent income hypothesis, indicating that the household's consumption decision is based on its permanent income. In Section 4.4.2, it is further illustrated that the response of the permanent income is dominated by the discounted future labor income, which (by nature of discounting) values a higher labor income immediately after the shock to a larger extent than a higher labor income in the long run. Especially the comparison between a trend shock and a transitory long memory shock with $\varrho_A + d$ sufficiently larger than one seems interesting, as in both cases, technology increases for some periods while attaining a new balanced growth path in the former case and returning to the old steady state/balanced growth in the latter case.

Before the effects of trend shocks are analyzed in more detail, it is shortly highlighted which effects on the model's responses to transitory shocks are induced by the change of the utility function. Recall, from Section 4.1.3.2, that the additive separable utility function considered in Section 4.4.2 is not compatible with a balanced growth path. Therefore, a Cobb-Douglas utility function specified in (4.19) is considered to analyze trend shocks.

To analyze the effects of the change in the utility function, both models have to be comparable, i.e., as there is no growth (for the moment) in the model with additive separable utility, there should also be abstracted from growth in the model with CD utility. Consequently, the steady state growth rate in the model with Cobb-Douglas utility g_{ss}

[809] See Aguiar and Gopinath (2004, p. 13).

should be set to zero. In the absence of growth shocks, the purely labor augmenting technological progress is $\bar{A}_t \equiv 1$, see (4.16). Thus, there is no balanced growth path in the model with CD utility and due to (4.20) the meaning of the variables is the same in both models, i.e., the stationarized variables defined in (4.20) equal their non-stationarized counterparts.[810]

It seems intuitive that the main model responses should be similar in sign and magnitude in both models, as there are the same channels through which the technology shocks affect the economy. Thus, the focus lies on the household's consumption and labor supply decision which depend to a large extent on the assumed utility function. For example, in the model with CD preferences, the linearized version of the labor supply curve is given by[811]

$$\tilde{w}_t = \tilde{c}_t + \frac{H_{ss}}{L_{ss}} \tilde{H}_t. \tag{4.37}$$

The labor demand curves are equal in both models.[812] Evidently, the labor supply curve (4.37) is less sensitive to an increase in consumption compared to the model with additive utility function specified in (4.35). Hence, given the identical upward shift in the labor demand curve due to a positive technology shock, the same increase in consumption shifts the labor supply curve (4.37) less upwards than in the model with additive separable utility. Thus, negative effects on total employment, i.e., employment falls below its steady state value, seem less likely in the model with CD preferences.[813]

Furthermore, the slope of the labor supply curve (4.37) is roughly 1/2 and thus smaller than in the model with additive separable utility, where the slope is given by $\varphi = 2$ for the standard parameters.[814]

The overall response of the labor market to a transitory long memory technology shock is illustrated in Figure 4.9, which is the analogous figure to Figure 4.7 with CD utility function. The gray lines in Figure 4.9 correspond to the respective loci of Figure 4.7 and are plotted for better comparability. The movement of employment and wages along the technology shock have similar shapes for both utility functions. For the CD utility function, they appear to be shifted rightwards, indicating that negative employment effects

[810] Recall from Section 4.4.1 that the model parameters of the model with CD utility were calibrated such that both models without growth have the same steady state values.

[811] The equation for the labor supply curve follows directly from the first and fourth row of Table B.6.

[812] See (4.34), the fifth and seventh row of Table B.6. Note that in the absence of growth $\tilde{b}_t \equiv 0$.

[813] Of course, falling employment depends on the response of consumption, which is generally different in the two models.

[814] Recall from Section 4.4.1, that steady state hours worked were calibrated to $H_{ss} = 0.33$, thus steady state leisure is given by $L_{ss} = 0.67$.

are indeed less pronounced in the model with CD utility. Especially this can be seen from the right-hand panel of Figure 4.9, where the main part of the locus is located above the steady state value of hours worked. However, the range of wages is nearly identical in all panels of Figure 4.9, indicating that the wage response is quite similar for both models.[815]

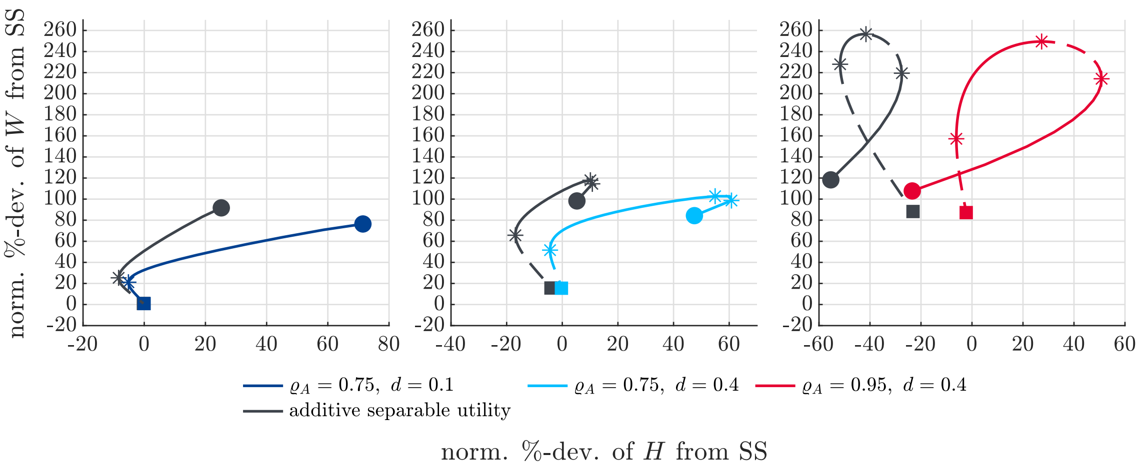


Figure 4.9: Wage-employment loci in the model with CD utility without growth for various values of ϱ_A and d. The vertical and horizontal axes report percentage deviations of the real wage and hours worked from their respective steady states divided by the initial shock size. The filled dot marks the variables' combination at the period of shock occurrence ($t = 1$) and the filled square at ($t = 140$). Solid lines indicate periods in which wages and employment behave in the same way, either increasing or decreasing; dashed lines indicate periods in which employment and the real wage behave in different ways, i.e., employment is rising with falling wages or vice versa. An asterisk marks inflection points where the behavior of the real wage or employment changes. Gray lines refer to the loci arising from the additive separable utility function given the same values with ϱ_A and d, see Figure 4.7. Note the different scaling of the horizontal axis.

Since the responses of the other variables are indeed quite similar in sign and magnitude to those in the additive separable utility function model, a detailed graphical comparison of the models' responses to a transitory short or long memory technology shock for various values of ϱ_A and d is given in Appendix B.6. There, Figures B.1 to B.3 compare the model responses of both models to purely short memory, purely long memory and ARFIMA$(1, d, 0)$ transitory TFP shocks.

Overall, the differences in the models' responses due to the two different utility functions can be summarized as follows. In the CD utility model, the household's labor supply is not as negatively affected by an increase in consumption expenditures as in the additive separable utility model. On the other hand, consumption and income initially increase more in the CD utility model than in the additive separable utility model. Overall, negative

[815] The analogous figures to Figure 4.2 and Figure 4.4 show a similar pattern, i.e., the locus with CD utility appears to be shifted rightwards compared to the model with additive separable utility. For this reason, these figures are not given in the text.

effects on total employment are not apparent or less pronounced in the model with CD utility than in the one with additive separable utility. This depends, however, on the persistence of the shock. The models' responses to the remaining variables are quite similar in sign and magnitude in both models. The differences in the responses for each variable in the model seem to be larger the more persistent the initial shock was. This observation highlights that the results presented in the previous section are, to a certain degree, robust against the choice of a specific utility function.[816]

After comparing the models' responses without growth, the focus is now on the model with CD utility function incorporating growth. At first, comparing the effects of short memory and long memory transitory technology shocks in the model with CD utility seems plausible. That is, instead of transitory deviations from a steady state, transitory deviations from the balanced growth path have to be considered. Formally this can be achieved by plotting the respective IRFs of the stationarized variables specified in (4.20) with $g_{ss} = 0$ and $g_{ss} = 0.005$, which is the calibrated standard parameter for the expected growth rate, see Table 4.1.[817]

The IRFs with $g_{ss} = 0$ are already part of the figures in Appendix B.6. It appears that an increase in the growth rate from $g_{ss} = 0$ to $g_{ss} = 0.005$ has negligible effects on the IRFs of a transitory technology shock over all considered combinations of the persistence parameters ϱ_A and d. Therefore, a comprehensive graphical comparison such as the one of Section 4.4.2 is omitted here. Instead, Figure 4.10 compares the wage-employment loci for the two cases $g_{ss} = 0$ and $g_{ss} = 0.005$ given the same parameters of ϱ_A and g_{ss} as in Figure 4.9. The loci from Figure 4.9 appear as gray lines in Figure 4.10 for better comparability of the two cases. The overall small deviations are apparent. Compared to the loci with a growth rate of zero (gray lines), the loci appear to be slightly shifted leftwards in the model with a small positive growth rate $g_{ss} = 0.005$.

816 As already mentioned, qualitative differences between both models are obtained in employment's response to a technology shock caused by the extent to which the labor supply depends on the household's consumption expenditures. Thus, a natural generalization would be to consider a utility function that removes the dependence of the labor supply on consumption. Such a class of utility function would be, for example, the class of GHH preferences, see Greenwood et al. (1988, p. 404) and Aguiar and Gopinath (2004, p. 13). Such considerations go beyond the scope of this thesis and are, therefore, left for future research.

817 Recall that the stationarized model given in Table B.4 is further linearized around its steady state value and that the linearized model equations are given in Table B.6. Further, recall that variables with a tilde refer to the variable's percentage deviation from its steady state value; see (4.21). Thus, for a trending variable (e.g., output Y_t) it yields, $\tilde{y}_t = \log(y_t/y_{ss}) = \log\left(Y_t/\left(\bar{A}_t y_{ss}\right)\right) \approx \left(Y_t - \bar{A}_t y_{ss}\right)/\left(\bar{A}_t y_{ss}\right)$. In the absence of growth shocks it follows from (4.16) that $\bar{A}_t = (1+g_{ss})^t$. Hence, $\bar{A}_t y_{ss}$ refers to the balanced growth path of Y_t, and evidently, $\tilde{y}_t$ refers to the percentage deviation of output from its balanced growth path.

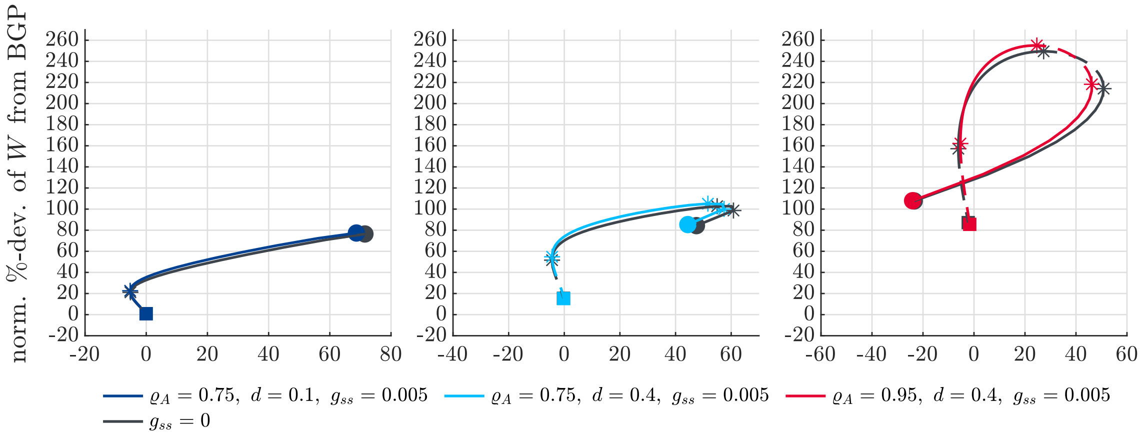


Figure 4.10: Wage-employment loci in the model with CD utility for various values of ϱ_A, d and g_{ss}. The vertical and horizontal axes report percentage deviations of the real wage and hours worked from their respective steady states divided by the initial shock size. The filled dot marks the variables' combination at the period of shock occurrence ($t = 1$) and the filled square at ($t = 140$). Solid lines indicate periods in which wages and employment behave in the same way, either increasing or decreasing; dashed lines indicate periods in which employment and the real wage behave in different ways, i.e., employment is rising with falling wages or vice versa. An asterisk marks inflection points where the behavior of the real wage or employment changes. Gray lines refer to the loci arising from the model without growth given the same values of ϱ_A and d, see Figure 4.9. Note the different scaling of the horizontal axis.

So far, it was illustrated that the change from the additive separable utility function defined in (4.18) to the CD utility function specified in (4.19) reduces the dependence of the household's labor supply decision on the household's consumption expenditures. Furthermore, introducing a small positive growth rate leaves the model responses to a transitory technology shock qualitatively and quantitatively unchanged. What remains to compare are the model responses to a trend shock specified in (4.17) and the model's responses to a transitory technology shock.

Recall that the growing component of technology $\bar{A}_t$ is assumed to be labor augmenting, thereby ensuring the existence of a balanced growth path. In contrast, the transitory component of technology A_t was assumed to make the capital stock and hours worked more productive. From the production function (4.1) follows that

$$Y_t = A_t K_t^{\alpha} (\bar{A}_t H_t)^{1-\alpha} = A_t \bar{A}_t^{1-\alpha} K_t^{\alpha} H_t^{1-\alpha}.$$

Thus, one can define TFP as the product of both productivity processes, i.e., $\text{TFP}_t := A_t \bar{A}_t^{1-\alpha}$. Instead of considering impulse-response functions of A_t or $\bar{A}_t$, the impulse-response function of TFP is considered to ensure better comparability between both kinds of shocks.[818] Thus, a growth shock of $\varepsilon_t^g = 0.01$ is not sufficient to ensure a one percent

[818] Note that this procedure is in line with the one carried out in Section 4.4.2, where $\bar{A}_t \equiv 1$ due to the

increase in TFP.[819]

Before the effects of growth and transitory shocks are contrasted, the model response to a one percent increase in TFP stemming from a growth shock is illustrated. Figure 4.11 shows the model responses for various values of ϱ_g, which is the short memory parameter of the growth shock. For the trending variables, the IRFs in Figure 4.11 show percentage deviations from the "old" balanced growth path, i.e., deviations are expressed from the balanced growth path on which the variable would grow if there were no growth shocks.[820,821]

Given the trend shock, TFP initially rises by one percent and positively affects the marginal productivities of capital and labor. Along the fading out of the trend shock, TFP reaches its new balanced growth path from below as illustrated in the upper-left-hand panel of Figure 4.11. The higher the value of ϱ_g, i.e., the slower the dissipation of the trend shock, the higher the level of the new balanced growth path.[822] As for the transitory technology shock considered in Section 4.4.2 and Appendix B.6, consumption expenditures rise. These higher consumption expenditures again shift the labor supply curve upwards. With increasing TFP and the implied push of labor demand, the real wage increases initially; see the lower-left-hand panel of Figure 4.11. The response of hours worked is again ambiguous, but given the parameter values of Figure 4.11, the initial response of hours worked appears to be negative in all cases, as can be seen from the center panel

absence of growth. In Section 4.4.2, a one percent increase in the transitory technology component corresponds to a one percent increase in TFP.

819 Total factor productivity grows along $(1+g_{ss})^{(1-\alpha)t}$ in absence of growth shocks. Thus the percentage deviation of TFP from its old balanced growth path is given by $\log\left(\text{TFP}_t/(1+g_{ss})^{(1-\alpha)t}\right) = \log(A_t) + (1-\alpha)\sum_{s=1}^{t}\tilde{b}_s$. Correspondingly, for the period of shock occurrence $t=1$, it yields $\log\left(\text{TFP}_1/(1+g_{ss})^{(1-\alpha)}\right) = \log(A_1) + (1-\alpha)\tilde{b}_1 = (1-\alpha)\varepsilon_1^g$. The latter equation holds since A_t is not affected by a growth shock, hence $A_t \equiv A = 1$. Evidently, a growth shock of size $\varepsilon_1^g = 0.01$ is not sufficient to increase TFP by one percent. Instead, the growth shock has to be of size $\varepsilon_1^g = 0.01/(1-\alpha)$. Therefore, Figure 4.11 shows the IRFs of a growth shock of size $\varepsilon_1^g = 0.01/(1-\alpha) \approx 0.0149$.

820 For the trending variables $Y_t, C_t, I_t, W_t, K_{t+1}$, the old balanced growth path is given by $(1+g_{ss})^t$, where g_{ss} is the steady state growth rate, see (4.20).

821 Note that not all curves plotted in Figure 4.11 are, strictly speaking, IRFs in the sense of Definition 2.1.5. The trending variables are not stationary, but Definition 2.1.5 refers only to stationary processes. However, the functions shown Figure 4.11 are computed from the IRFs of the stationarized variables specified in (4.20); these are in accordance with Definition 2.1.5 as they are stationary. The percentage deviation for, say, consumption from its old balanced growth path follows directly from (4.20), (4.16) and (4.17); it is given by $\log\left(C_t/(c_{ss}(1+g_{ss})^t)\right) = \tilde{c}_t + \sum_{s=1}^{t}\tilde{b}_s$. For better comparability between the figures of this chapter and between the trending and non-trending variables, the author finds it helpful to plot the percentage deviation from the old balanced growth path instead of the IRFs of the stationarized variables.

822 Given an initial trend shock of size $\varepsilon_1^g = \sigma_{\varepsilon^g}$, it follows from Footnote 819 that the long run percentage deviation of TFP from its old balanced growth path is given by $\sigma_{\varepsilon^g}(1-\alpha)\sum_{s=1}^{\infty}\varrho_g^{s-1} = \sigma_{\varepsilon^g}(1-\alpha)/(1-\varrho_g)$. For the other trending variables, the long run percentage deviation corresponds to $\sigma_{\varepsilon^g}/(1-\varrho_g)$.

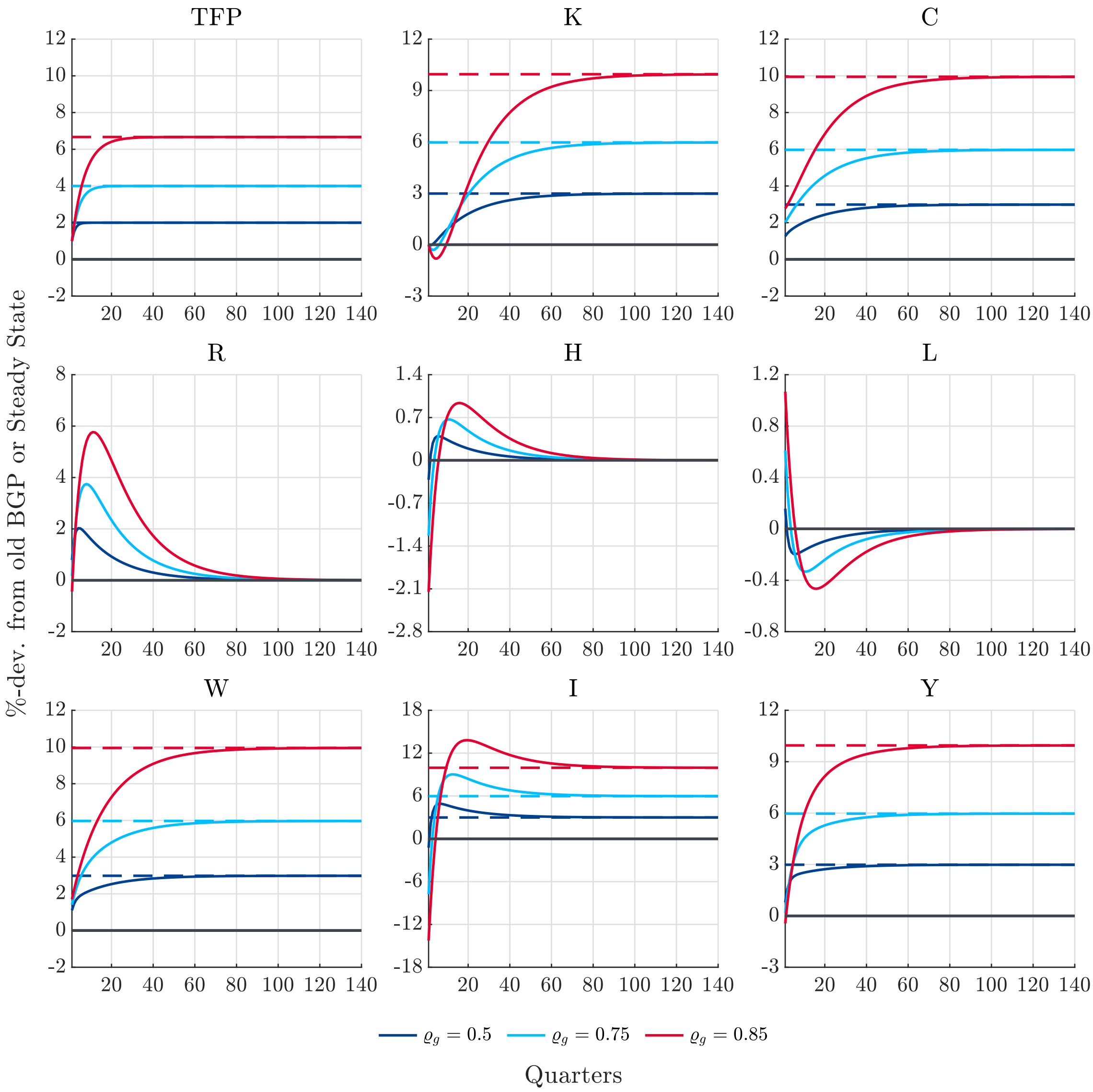


Figure 4.11: Responses of the model with CD utility function to a 1% TFP trend shock for various values of ϱ_g. The vertical axes report percentage deviations from the variables' old balanced growth path or steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their old balanced growth path (if existing) or steady state values. Dashed lines refer to the new balanced growth path (if existing). Note the different scaling of the vertical axis.

of Figure 4.11.[823] Correspondingly, with decreasing employment, the household's leisure consumption rises initially; see the right-hand panel in the second row of Figure 4.11.

The capital stock is again a predetermined variable and does not initially respond to the TFP shock. On the other hand, capital demand is negatively affected by the drop in employment. With $\varrho_g = 0.85$, the positive effect on capital demand due to higher TFP is outweighed by the negative response of employment. As a consequence, the rental rate on capital initially falls below its steady state value in this case; see the left-hand panel in the second row of Figure 4.11. The reverse is true for the other values of ϱ_g.

Furthermore, consumption rises more than the household's income or output, leading to an overall positive initial response of the consumption-to-income ratio. Thus, the household's investment expenditures fall initially and cannot compensate for the capital stock's depreciation. Consequently, the capital stock falls below its old balanced growth path for some periods; see the center panel in the first row of Figure 4.11. With growing TFP, consumption expenditures and output grow. Since the latter grows faster than the former, investment expenditures depicted in the center panel of the third row in Figure 4.11 rise quickly, and the accumulation of the capital stock sets in.

The increasing consumption expenditures continuously shift the labor supply curve upwards. At the same time, positive TFP and the beginning accumulation of the capital stock shift the labor demand curve upwards. The resulting steadily increasing real wage rate can be seen from the lower-left-hand panel of Figure 4.11. For several periods, the shift in the labor demand curve seems to exceed the shift in the supply curve resulting in periods of increasing employment. Afterward, the shift in the supply curve outweighs the shift in the demand curve, and employment decreases steadily and reaches its steady state value. Capital demand is continuously pushed upwards with increasing TFP and employment. Together with the decreasing capital stock, this causes a steep increase in the rental rate on capital in the first periods. The accumulation of the capital stock then gradually enlarges the capital supply in the economy, and the rental rate on capital returns to its steady state value, as can be seen from the left-hand panel in the second row of Figure 4.11.

Overall, the mechanisms driving the model responses are similar to those described in Section 4.4.2 and Appendix B.6. However, the initial consumption expenditures of the household (with all of their implied consequences for the other model variables) react more strongly to the trend shocks depicted in Figure 4.11 than to the short memory transitory shock shown for example in Appendix B.6. Before these differences in the household's

[823] For smaller values of ϱ_g, an initial positive response of hours worked is possible by leaving the other parameters unchanged.

consumption decision are considered in more detail, Figure 4.12 finally compares the model responses to a one percent short memory, long memory, and trend TFP shock.[824]

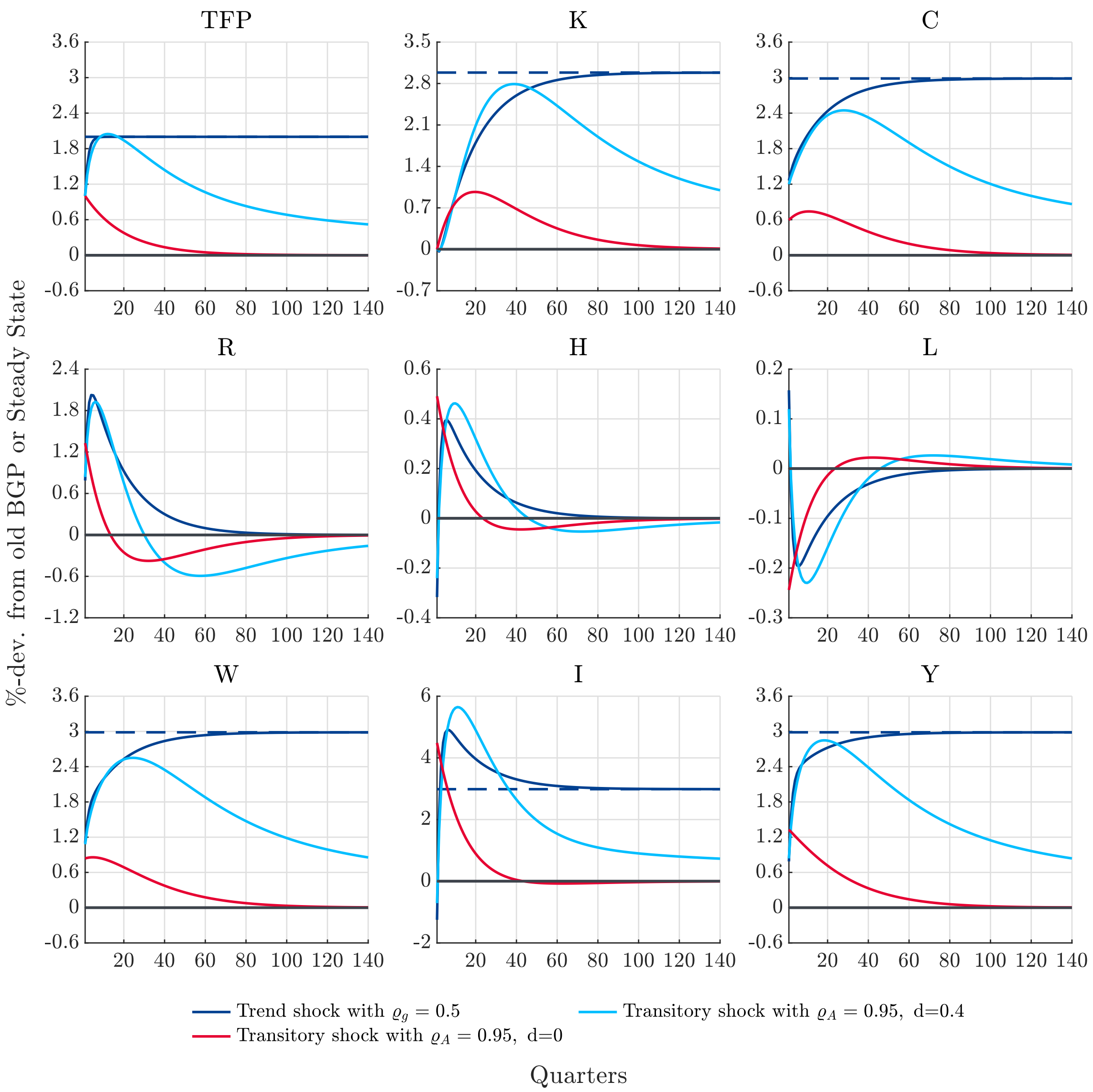


Figure 4.12: Comparison of the model's response to a 1% TFP short memory, long memory, and trend shock. The vertical axes report percentage deviations from the variables' old balanced growth path (if existing) or steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their old balanced growth path (if existing) or steady state values. Note the different scaling of the vertical axis.

A direct comparison between a trend shock and transitory short memory TFP shock with $\varrho_A = 0.95$ (dark-blue and red lines in Figure 4.12) reveal significant differences. As can be seen from the upper-right-hand panel of Figure 4.12, the initial response of consumption expenditures to a trend shock is nearly twice as high as the one to

[824] Recall that the trend shock ε_g^1 has to be of size $0.01/(1-\alpha)$ to ensure an increase in TFP of 0.01, while a transitory TFP shock of size $\varepsilon_1^A = 0.01$ is sufficient to do so.

a transitory short memory technology shock. Consequently, the upward shift in the household's labor supply curve induced by its consumption decision is more pronounced for a trend shock than for a transitory short memory technology shock. Given the same initial upward shift in the labor demand curve, employment reacts positively to a short memory technology shock and negatively to a trend technology shock. Hence, income and investment expenditures increase more in response to a transitory technology shock than to a trend shock. Accordingly, the consumption-to-income ratio drops initially in response to a trend shock but rises in response to a transitory technology shock.

These different responses in consumption are again typically explained by the permanent income hypothesis. For example, Aguiar and Gopinath (2007) compare the effects of a trend and transitory short memory technology shock and ascertain:

> As a trend shock implies a greater increase in permanent income, consumption will respond more to such shocks. [...] In response to a growth shock, consumption responds more than income given the anticipation of even higher income in the future. The higher future income follows from the fact that the innovation to productivity is not expected to die out and capital adjusts gradually.[825]

Figure 4.12 highlights how a transitory long memory TFP shock with $\varrho_A = 0.95$ and $d = 0.4$ affects the economy in comparison to the other two kinds of shocks already discussed. TFP rises for about 12 periods and peaks roughly at the level of the new balanced growth path induced by the reference trend shock. Afterward, the shock dies out slowly, and TFP returns to its old balanced growth path. The response of TFP to the long memory shock is very similar to the response to the trend shock in the first ten periods. Interestingly, the responses of the household's consumption expenditures to both shocks are also quantitatively and qualitatively similar in the first periods. The same is true for the responses of the remaining variables, which are very similar to the response to a trend shock in the first periods, even though the household anticipates the long memory technology shock to die out.[826]

[825] Aguiar and Gopinath (2007, p. 87).

[826] Clearly, the concrete model responses depend on the values of ϱ_g, ϱ_A and d. For Figure 4.12, it needs high values of $\varrho_A = 0.95$ and $d = 0.4$ to mimic the response of a trend shock with $\varrho_g = 0.5$. By considering panel a) of Figure 2.8, it appears that even with $\varrho_A = 0.99$ and $d = 0.49$, the maximal value of an ARFIMA$(1, d, 0)$ process's IRF is roughly 6, i.e., a one percent TFP shock accelerates to a level of six percent, before it returns to the old balanced growth path. On the other hand, given a one percent trend shock with $\varrho_g = 0.84$, TFP reaches a new balanced growth path that lies 6.25% above the old one. For higher values of ϱ_g, the new balanced growth path reaches even higher values. Thus, the possibilities of ARFIMA processes to mimic the responses of a trend shock seem to be bounded.

Ultimately, these findings confirm the ones made in Section 4.4.2. For the household, the long-run effects of a technology shock are less important for its consumption decision in the period of the shock occurrence, i.e., whether a higher balanced growth path of TFP is reached or whether TFP finally returns to its old balanced growth path is secondary. The household increases its consumption expenditures when it expects TFP to rise, at least for some periods. Later, when the responses of TFP start to diverge due to the different sources of the shock, the household adjusts its consumption expenditures accordingly. Some periods after the shock, the shape of the impulse-response functions of the long memory technology shock looks similar to those of a transitory short memory shock. Compare, for example, the hump in the IRF of the capital stock or the overshooting of the rental rate on capital and hours worked, which are not visible for the growth shock. Overall, the relationship between the three considered kinds of shocks may be summarized as follows.

Whenever $\varrho_A + d > 1$, a transitory long memory technology shock has an increasing impact for some periods similar to those of a dissipating trend shock. While the trending variables reach a new balanced growth path with the dissipating trend shock, they eventually return to their old balanced growth path in response to a transitory short or long memory technology shock. Therefore, the model's IRF to a short memory technology shock with $d = 0$ and an ARFIMA$(1, d, 0)$ technology shock with $\varrho_A + d > 1$ share similar shapes after the peak of TFP's response has been reached. Instead, when $\varrho_A + d < 1$, the qualitative model responses are similar to those of an AR(1) process since the asymptotic characteristics of the long memory shock materialize outside the household's decision-making horizon. Although the slow decay in the IRF is visible, it does not cause a significant change in the household's consumption behavior at the time of shock occurrence. On the one hand, the household is assumed to be infinitely-lived and to incorporate all future periods in its utility maximization decision (see (4.3)). On the other hand, however, the household discounts the future at an exponential rate, making an infinite cumulative technology shock negligibly small.

5

Continuous-Time Macro-Financial Model

In Chapter 4, long memory in a discrete-time dynamic stochastic general equilibrium (DSGE) model has been considered. Given the canonical form of a DSGE model, it can be clearly seen that these models are, in the end, systems of difference equations involving expectations of certain variables.[827]

Similar to the deterministic dichotomous distinction between difference and differential equations, there are, besides the stochastic difference equations treated in Chapter 4, continuous-time stochastic models called stochastic differential equations. In this chapter, a similar question as in Chapter 4 is addressed, i.e., whether and how can long memory be introduced in a continuous-time stochastic framework and if it can, how sensitive are the model's implications regarding the presence of long memory. The context of these considerations builds a model from the emerging literature on continuous-time macro-financial models. To be more precise, the model considered in this chapter is taken from Brunnermeier and Sannikov (2016, Section 2 on pp. 1504ff.).

The following section introduces the shock-generating process to incorporate long memory dynamics into the model. This process is then considered in more detail in the following sections. Section 5.2 introduces the model setup, and in Section 5.3, the model's equilibrium is defined, and its solution is derived. Section 5.4 contains the results, and a comparison with the benchmark model is carried out. The corresponding Appendix C contains in Appendix C.1 a summary of the Itô calculus for dealing with stochastic integrals and

[827] Recall the DSGE model's canonical form (4.22) in the context of Blanchard and Kahn (1980)'s solution method and (B.20) in the context of Klein (2000)'s solution method.

stochastic differential equations. Appendix C.2 contains the proofs of the lemmas and propositions stated in this chapter, and Appendix C.3 is a note on the derivation of the model's equilibrium price.

An earlier version of the content in this chapter is available as a working paper titled "The Impact of Long-Range Dependence in the Capital Stock on Interest Rates and Wealth Distribution". This paper is also included in the conference paper series "Beiträge zur Jahrestagung des Vereins für Socialpolitik 2019 - 30 Jahre Mauerfall - Demokratie und Marktwirtschaft" under the session titled "A01 Macroeconomics - Financial Markets". A revised version of this working paper was also presented at the World Congress of the Econometric Society in 2020.

5.1 Definition of the Shock-Generating Process

A milestone in financial modeling was built by Black and Scholes in 1973 with their financial market model in which the stock's price is given by a geometric Brownian motion whose dynamics are described by the stochastic differential equation:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t, \tag{5.1}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are given constants and $(W_t)_{t \geq 0}$ is a Brownian motion on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$.[828,829] Note that W_t is a Brownian motion within this chapter, not the real wage W_t as in the previous chapter. Since the two model contexts are unambiguously different, the excessive use of the letter W should not cause much confusion. Instead, it allows for a standard notation in both research areas.

Not only in finance but also in macroeconomics, modeling in a continuous-time stochastic framework became more popular. There have been arising classes of macroeconomic models based on stochastic differential equations in the recent literature.[830]

These types of models combine two concepts. The first is the idea of stochastic modeling, which is already realized and well-understood in discrete-time settings via DSGE models, to consider the effects of random shocks on the economy.[831] The second concept is to

[828] Recall Definition 2.5.1 for the definition of a Brownian motion.

[829] Black and Scholes (1973) describe the stochastic properties of the underlying stock price more verbally. Merton provides in Merton (1973, Section 6 on pp. 162ff.) a derivation of Black's and Scholes's model based on (5.1), see Merton (1973, Equation (23) on p. 162).

[830] See Brunnermeier and Sannikov (2016, Section 1.3 on pp. 1502ff.). Additional literature is mentioned later in this chapter.

[831] Brunnermeier and Sannikov (2016, Figure 1 on p. 1500) illustrates nicely how the continuous-time

model in continuous-time frameworks to describe the whole dynamics of a variable instead of linear approximations around a deterministic steady state.[832]

So the idea behind these new macro models is to apply the well-developed continuous-time stochastic framework from financial mathematics and finance to macroeconomic problems.[833] For example, Brunnermeier and Sannikov investigate in Brunnermeier and Sannikov (2016) the effects of financial frictions on the wealth distribution within the economy and go thereby further than classical finance issues such as option pricing.[834] To realize this, Brunnermeier and Sannikov describe the evolution of capital in the economy with the following stochastic differential equation

$$\frac{dk_t}{k_t} = (\Phi(\iota_t) - \delta)\ dt + \sigma\, dW_t. \tag{5.2}$$

where capital grows with investment minus depreciation $(\Phi(\iota_t) - \delta)$.[835] The term dW_t is interpreted as a macroeconomic shock with volatility σ modeled as stochastic disturbance generated by a Brownian motion $W = (W_t)_{t \geq 0}$.[836] This stochastic differential equation is very similar to (5.1), which Black, Scholes, and Merton used to describe the evolution of stock prices.

From the empirical side, there is evidence that (5.1) is unsuitable for modeling the dynamics of stock prices since some stylized facts about stock prices cannot be captured by (5.1).[837] As outlined in Section 3.2.2.3, there is empirical evidence that log returns of stocks are correlated, also suggesting a failure of the efficient market hypothesis. However, they appear stochastically independent in (5.1) since the increments of Brownian motions are stochastically independent. In the light of Section 2.5, fractional Brownian motion (fBm) appears to be a natural generalization of a Brownian motion. Thus, it seems natural to replace (5.1) with

$$dS_t = \mu S_t\, dt + \sigma S_t\, dB_t^H, \tag{5.3}$$

where $H \in (0, 1)$ denotes the Hurst index.[838]

macro-financial literature combines the two strands of macroeconomics and finance.

832 See Brunnermeier and Sannikov (2016, pp. 1498 and 1502).

833 See Brunnermeier and Sannikov (2016, p. 1501).

834 See Brunnermeier and Sannikov (2016, p. 1516) for the introduction of financial frictions in their extended model and Brunnermeier and Sannikov (2016, Section 3.4.2 on pp. 1523) for a discussion of the frictions' implications.

835 See Section 5.2 for more details.

836 The term σdW_t is also referred to as aggregate risk in Brunnermeier and Sannikov (2016, p. 1534).

837 See Section 3.2.2.3.

838 Recall the definition of fBm, Definition 2.5.2.

Admittedly, this idea is not entirely new. There was a detailed discussion in the financial mathematics literature in the 2000s about whether or not it is helpful to replace the Brownian motion in (5.1) with a fractional Brownian motion. On the one hand, using fBm instead of Brownian motion opens up the possibility to model autocorrelation in stock returns since the parameter H controls for the correlation of the increment process. On the other hand, the generalization from (5.1) to (5.3) appears not as straightforward as it seems.

From a mathematical perspective, the stochastic differential equation (5.3) has to be well-defined before any implications for stock prices or other variables can be derived. In the end, each stochastic differential equation is always a shorthand notation of a corresponding integral equation, i.e., (5.3) essentially stands for[839]

$$S_t = S_0 + \int_0^t \mu S_s \, ds + \int_0^t \sigma S_s \, dB_s^H. \tag{5.4}$$

Thus, for a stochastic differential equation like (5.3) to be well-defined, the corresponding integral equation (5.4) has to be well-defined. To be more precise, the critical part of (5.4) is the stochastic integral with respect to the fractional Brownian motion, i.e., a stochastic integral of the form

$$\int_0^t \lambda_s \, dB_s^H, \tag{5.5}$$

where $(\lambda_t)_{t\geq 0}$ is a given deterministic function or a stochastic process. With $H = 1/2$, it yields that $B_t^{1/2} = W_t$ is a Brownian motion. In this case, the corresponding stochastic integral (5.5) is often interpreted in the Itô sense that offers nice analytic tools to deal with the corresponding stochastic differential equations.[840,841] Extensions of the Itô integral also exist when the integrator is replaced with a combination of a Brownian motion and a jump process such as a Poisson process.[842] The most general class of stochastic processes for which the Itô integral (5.5) is well-defined appears to be the group of so-called semimartingales.[843] Due to its technicality, a precise definition of a semimartingale is not given in this thesis, but it can be found, e.g., in Rogers and Williams (1987, p. 23). Important processes that are indeed semimartingales are Brownian motions, Poisson

839 For the case of a Brownian motion, i.e., $H = 1/2$, see Øksendal (2013, p. 65).

840 Appendix C.1 provides a brief review of the Itô calculus.

841 Another possible way to interpret stochastic integrals with respect to Browinan motion is the so-called Stratonovich integral. Without going much into the details, the Stratonovich integral seems to be unsuitable for applications in finance. A brief discussion of this issue can be found, e.g., in Shreve (2004, p. 136). For this reason, a detailed discussion of the Stratonovich integral is omitted here.

842 See Shreve (2004, Definition 11.4.3 on pp. 475f.).

843 See Rogers and Williams (1987, p. 391) and Rogers and Williams (1987, p. 394) for Itô's formula if X is a semimartingale.

processes, or, more generally, so-called Lévy processes i.e., processes with stationary and independent increments, and diffusion processes, i.e., solutions to stochastic differential equations with respect to a Brownian motion.[844]

Unfortunately, fBm is not a semimartingale for $H \neq 1/2$.[845] Thus, the Itô integration calculus cannot be applied to the stochastic integral in (5.3).[846] Therefore, other attempts (mainly in the financial mathematics literature) have been made to define stochastic integrals and related stochastic differential equations such as (5.3).

Early approaches attempted to interpret a stochastic integral like the one in (5.5) in a pathwise sense.[847] With such an interpretation of (5.5), the corresponding Black-Scholes model (5.4) is not free of arbitrage.[848] However, the absence of arbitrage appears to be a necessary equilibrium condition in a financial market.[849] These problems with the pathwise definition of the integral can be solved if additional assumptions on the model are proposed, e.g., if transaction costs are introduced.[850] Overall, using an integral calculus that introduces arbitrage possibilities into the financial market seems questionable, and a carryover to a macro-financial model seems, thus, less promising.

In order to remedy the drawbacks of the pathwise integration calculus, various authors have proposed another mathematical setting to introduce fBm in a Black-Scholes model. Biagini et al. (2008) consider a model based on the so-called Wick-Itô-Skorohod integration

[844] See Rogers and Williams (1987, p. 24).

[845] See, e.g., Biagini et al. (2008, Section 1.8 on pp. 12f.).

[846] See Biagini et al. (2008, p.13).

[847] Recall that for each fixed $\omega \in \Omega$, the map $t \mapsto B_t^H(\omega)$ is called the path of the fractional Brownian motion, i.e., the path is a function of t; see also Section 2.1.1. Without going much into the details, the idea behind a pathwise integration is the following. Consider two deterministic functions g and F defined on an interval $[a, b]$. The Riemann-Stieltjes integral $\int_a^b g(s)\, dF(s)$ can be defined given some regularity conditions for the functions g and F, see, e.g., Salopek (1998, Section 2 on pp. 218ff.) for a summary of properties of the Riemann-Stieltjes integration calculus. If, for example, F is differentiable with derivative f, the mentioned integral simply corresponds to $\int_a^b g(s)f(s)\, ds$, see, e.g., Salopek (1998, Proposition 2.1 on p. 221). The integral, however, may exist even if F is not differentiable. The idea for defining a pathwise integral with respect to fractional Brownian motion is to replace the function F with the path of a fractional Brownian motion, which is indeed not differentiable, see, e.g., Biagini et al. (2008, Proposition 1.7.1 on p. 12).

[848] This result was proven in various contexts. An argument in a general framework is given by Rogers (1997, pp. 100f.). Salopek (1998, pp. 225 and 228) showed that there is an arbitrage opportunity if $H > 1/2$ and Biagini et al. (2008, Section 7.1 on p. 170f.) constructs a concrete arbitrage strategy in a fractional Black-Scholes market when stochastic integrals with respect to fBm are defined in a pathwise sense.

[849] See, e.g., Øksendal (2013, p. 273), Biagini et al. (2008, Remark 7.1.2. on p. 171) or Karatzas and Shreve (1998, pp. 11f.).

[850] Guasoni (2006, Proposition 5.1 on p. 578) proves that a model incorporating transaction costs is free of arbitrage in a setting with pathwise defined integrals. Wang (2010, Equation (2.19) on p. 442) derives the pricing formula for a European call option in the presence of transaction costs under the additional assumption that trading is allowed only at discrete instances of time.

calculus.[851] The approach of Biagini et al. (2008) covers the whole range of $H \in (0,1)$ and, thus, extends the approach of Hu and Øksendal (2003), who considered the Wick-Itô-Skorohod integration calculus only for $H \in (1/2, 1)$.[852].

The resulting Black-Scholes model in the context of the Wick-Itô-Skorohod calculus is arbitrage-free but suffers from a lack of adequate economic interpretation.[853] As outlined by Björk and Hult (2005), the model introduced in Hu and Øksendal (2003) is indeed free of arbitrage in a mathematical manner, but the definition of self-financing portfolios in the model has no relation to economic reality.[854] They construct a portfolio in the Wick-Itô-Skohorod based Black-Scholes model of Hu and Øksendal (2003), which has a positive amount of shares of the risky asset and zero amount of the risk-free asset.[855] Clearly, the value of this portfolio should be non-negative in reality. However, this portfolio indeed has a negative value on a set with positive probability in the context of Hu and Øksendal (2003).[856]

In addition, within the Wick-Itô-Skorohod integration framework, the value of a European call option at time $t = 0$ is derived by Hu and Øksendal (2003).[857] Based on a preprint version of Hu and Øksendal (2003), Necula (2002) extends this formula for arbitrary instances of time.[858] In contrast to the original Black-Scholes formula for the valuation of a European call option, the formula of Necula (2002) depends not only on the time difference between the time of valuation and maturity-time but also on the valuation-time itself.[859] This was strongly criticized as being counter-intuitive by Rostek and Schöbel (2013).[860] Additionally, the authors show that by correcting for the economic misinterpretation that comes along with the Wick-Itô-Skorohod calculus, the resulting value of a European call option no longer depends on the stock's volatility, and the price of a call option eventually shows a deterministic character.[861]

The reasons for these differences between a Black-Scholes model driven by fBm and a Brownian motion may be caused by the correlation of an fBm's increments. These

851 See, especially, Biagini et al. (2008, Chapter 4 on pp. 99ff.) for a formal definition and Biagini et al. (2008, Section 7.2 on pp. 172ff.) for an application in a fractional Black-Scholes setting.
852 See Hu and Øksendal (2003, p. 2).
853 See Biagini et al. (2008, Theorem 7.2.6 on p. 176) for the statement that the Black-Scholes market is arbitrage-free.
854 See Björk and Hult (2005, p. 200).
855 See Björk and Hult (2005, p. 208).
856 See Björk and Hult (2005, p. 208).
857 See Hu and Øksendal (2003, Corollary 5.5 on pp. 27f.).
858 See Necula (2002, Theorem 4.2 on p.13), URL in list of references.
859 See Necula (2002, Theorem 4.2 on p.13), URL in list of references.
860 See Rostek and Schöbel (2013, p. 34).
861 See Rostek and Schöbel (2013, p. 34).

correlations allow for a certain degree of predictability as the history enters into the forecasts of the future, thus, removing the random character from a call option's price.[862]

Thus, both approaches of Biagini et al. (2008) and Hu and Øksendal (2003) appear unsatisfactory from an economic perspective. Biagini et al. (2008) suggest interpreting S_t in (5.4) as a total-company value that is not directly observed by the market participants rather than the actual price of a company's stock, but this point of view seems to be far from the spirit of the original Black-Scholes model.[863]

This short discussion raises doubts that implementing fractional Brownian motion along pathwise defined integrals or the Wick-Itô-Skorohod calculus into a macro-financial model delivers promising results. This discussion further illustrates that changing the underlying mathematical framework may introduce difficulties in the interpretation and comparability of existing models. Thus, a desirable approach would be to combine long memory and the existing Itô integration calculus. Although such a path cannot be taken directly with fBm, a solution was proposed by Rogers (1997):

> Formally, the fractional Brownian motion is the convolution of Brownian increments with a power-law kernel, and the arbitrage is happening because of the behavior of that kernel near zero. Long-range dependence [i.e, long memory] is happening because of the behavior of the kernel at infinity, so the remedy is clear; we convolute the Brownian increments with some kernel which has the same behavior at infinity but a more orderly behavior at zero, and everyone will be happy![864]

The convolution mentioned in the quotation refers to the continuous-time moving average representation of a type I and II fBm given in (2.39) and (2.38), respectively. While Rogers (1997) proposes an adjusted kernel for a type I fBm, Thao and Thomas-Agnan (2003) propose the usage of a type II fBm.[865]

Thao and Thomas-Agnan (2003) argue that the process Z^H is appropriate to model the noise in financial markets.[866] Since, again, Z^H is not a semimartingale, they proposed an approximation of this process which is defined as follows.[867]

Definition 5.1.1 ———

[862] See Rostek and Schöbel (2013, pp. 30 and 34).
[863] See Biagini et al. (2008, p. 172).
[864] Rogers (1997, p. 95).
[865] See Rogers (1997, Section 5 on p. 104) and Thao and Thomas-Agnan (2003, p. 109), respectively.
[866] See Thao and Thomas-Agnan (2003, p. 109).
[867] See Thao (2003, p. 256) for the statement that Z_t^H is not a semimartingale.

For $H \in (0,1)$ and $\varepsilon > 0$, let the stochastic process $Z^{H,\varepsilon} = (Z_t^{H,\varepsilon})_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be defined by

$$Z_t^{H,\varepsilon} = \int_0^t (t - s + \varepsilon)^{H-1/2} \, dW_s, \tag{5.6}$$

where $(W_t)_{t\geq 0}$ denotes a standard Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P})$. ⋊

Definition 5.1.1 is more general than in Thao and Thomas-Agnan (2003), since they consider only the case of $H > 1/2$. The more general case of $H \in (0,1)$ is considered in Thao (2006).[868] From Definition 5.1.1 the suggestion of Rogers (1997) cited above becomes clear. The process $Z^{H,\varepsilon}$ can be regarded as an aggregation of Brownian shocks weighted by the integral kernel $K(t+\varepsilon, s)$, which is given by $K(t,s) = (t-s)^{H-1/2}$. By introducing the parameter ε, the behavior of this kernel at zero is influenced by keeping the structure of the kernel in the long run almost unchanged.[869]

The process defined in Definition 5.1.1 generalizes a Brownian motion in a similar fashion than an fBm. More specifically, with $H = 1/2$, the process $Z^{1/2,\varepsilon}$ is again a Brownian motion for every $\varepsilon > 0$. Another striking feature of the process $Z^{H,\varepsilon}$ is that as $\varepsilon \to 0$, it converges in the mean-square sense to the process Z^H defined in (2.38).[870,871] Thus, for small values of ε, $Z^{H,\varepsilon}$ can be viewed as an approximation of Z^H.

In the following, some properties of the process $Z^{H,\varepsilon}$ are considered. Ultimately, The aim is to define a stochastic differential equation with respect to the process $Z^{H,\varepsilon}$. As mentioned above, this leads to difficulties whenever the underlying stochastic process is not a semimartingale; since then, the Itô integration calculus cannot be applied. Both processes, B^H and Z^H, are not semimartingales, but the process $Z^{H,\varepsilon}$ is indeed one.[872]

Lemma 5.1.2

The process $Z^{H,\varepsilon}$ defined by (5.6) is a semimartingale and has the following decomposition

$$Z_t^{H,\varepsilon} = \int_0^t \int_0^s (H - 1/2)(s - u + \varepsilon)^{H-3/2} \, dW_u \, ds + \varepsilon^{H-1/2} W_t = \int_0^t \varphi_s^{H,\varepsilon} \, ds + \varepsilon^{H-1/2} W_t$$

868 See Thao (2006, Equation (2.1) on p. 125).

869 A similar approach based on a slightly different integral representation is given by Rogers (1997, Section 5 on p. 104). He also imposes sufficient conditions on his integral kernel for which the corresponding process becomes a semimartingale.

870 See Thao (2006, Theorem 2.1 on p. 126). This result is extended for another weighting kernel by Dung (2011, Proposition 2.1 on p. 1846).

871 Mean-square convergence means that, $\mathbb{E}(|Z_t^{H,\varepsilon} - H_t^H|^2) \to 0$, as $\varepsilon \to 0$. Additionally, Dung (2011, Proposition 2.1 on p. 1846) showed that $\mathbb{E}(|Z_t^{H,\varepsilon} - H_t^H|^p) \to 0$ as $\varepsilon \to 0$, for all $p > 0$. That is, all moments of $Z^{H,\varepsilon}$ converge to those of Z^H as $\varepsilon \to 0$.

872 This result, stated in the following lemma, was also proved by Dung (2011, Proposition 2.1 on p. 1846) in a more general setting.

with $\varphi_s^{H,\varepsilon} = \int_0^s (H-1/2)(s-u+\varepsilon)^{H-3/2}\,dW_u$. *Thus, one can write*

$$dZ_t^{H,\varepsilon} = \varphi_t^{H,\varepsilon}\,dt + \varepsilon^{H-1/2}\,dW_t.$$

Proof

See Thao (2006, Lemma 2.1 on p. 125). □

With Lemma 5.1.2, it is possible to consider a stochastic differential equation with respect to the process $Z^{H,\varepsilon}$ in the classical Itô sense. For this reason, the process $Z_t^{H,\varepsilon}$ seems appropriate as stochastic noise in a financial model.[873] Consequently, the process $Z_t^{H,\varepsilon}$ is used in the context of a macro-financial model in the following sections.

The main reason for considering a model driven by the process $Z^{H,\varepsilon}$ is the property of long memory. Recall from Definition 2.3.1 that long memory is a property of stationary processes, i.e., a process is a long memory process if the sum of the absolute values of its autocorrelation (or autocovariance) function diverges. Unfortunately, the increments of the process $Z^{H,\varepsilon}$ are unlike the one of fBm, not stationary.[874,875]

Nevertheless, it may be reasonable to call the increment process a long memory process in the following sense. Dung (2013), defines the autovariance function of $Z^{H,\varepsilon}$ as[876]

$$\rho^{H,\varepsilon}(n) := \mathbb{E}\left(Z_1^{H,\varepsilon}(Z_{n+1}^{H,\varepsilon} - Z_n^{H,\varepsilon})\right) \text{ for } n \in \mathbb{N}. \tag{5.7}$$

In addition, he proved the following proposition.

Proposition 5.1.3

The autovariance function defined in (5.7) *satisfies*

$$\sum_{n=1}^{\infty} \rho^{H,\varepsilon}(n) = \infty, \text{ if } H > 1/2 \text{ and } \sum_{n=1}^{\infty} \rho^{H,\varepsilon}(n) < \infty, \text{ if } H < 1/2.$$

Proof

See Dung (2013, Proposition 4.1 on p. 343) □

[873] Dung (2011, Remark 4.1 on p. 1854) derives a pricing formula for a European call option in the corresponding Black-Scholes model. Mrázek et al. (2016, p. 1038) and Yang et al. (2016, p. 7), URL in list of references, use this process as stochastic disturbance in a stochastic volatility model.

[874] See Dung (2013, p. 343).

[875] Recall from Section 2.5 that the process Z^H has non-stationary increments as well.

[876] See Dung (2013, p. 342).

Proposition 5.1.3 states that the process $Z^{H,\varepsilon}$ can be viewed as a long memory process in terms of the autovariance function instead of the autocovariance function.[877] Obviously, the non-stationarity of the increment process is disadvantageous. However, compared to the drawbacks of another integration calculus as described above, this drawback seems bearable. Moreover, as will turn out in the next section (especially Figure 5.1), the sample ACF of the economy's output shows similarities to the theoretical ACF of an ARFIMA$(0, d, 0)$ process. Consequently, the non-stationarity of the increments may not be seen as a major drawback.

The aim of this section was manifold. It was highlighted that there was (and still is) a discussion in the literature dealing with incorporating long memory dynamics into a continuous-time financial model, especially into models similar to the seminal one of Black and Scholes. Then, some appearing challenges in the definition of stochastic differential equations with respect to an fBm were discussed, and a possible solution (the process $Z^{H,\varepsilon}$) was provided. Since continuous-time macro-financial models connect both strands of literature (macroeconomics and finance), the following sections are devoted to analyzing the effects when the process $Z^{H,\varepsilon}$ is plugged into a macro-financial model.

5.2 The Model Setup

This section introduces the model setup. The model is mainly based on Brunnermeier and Sannikov (2016, Section 2 on pp. 1504ff.), which in turn is an extension of the model presented by Basak and Cuoco (1998).[878] In the latter model, Basak and Cuoco (1998) suppose a finite time horizon and consider a financial market with two assets, a risky asset (namely a stock depending on an exogenous dividend process) and a riskless bond in zero net supply.[879]

Brunnermeier and Sannikov (2016) link the risky asset in the model of Basak and Cuoco (1998) to a production function in the sense that the risky asset of Basak and Cuoco (1998) is explicitly employed in this production function and produces output goods.[880] This output corresponds to the dividend payment of the stock in the setting of Basak and

[877] The autocovariance functions of $Z^{H,\varepsilon}$ measures the covariance between the first increment $Z_1^{H,\varepsilon} - Z_0^{H,\varepsilon}$ and the increment n periods later, i.e., $Z_{n+1}^{H,\varepsilon} - Z_n^{H,\varepsilon}$. If the increment process were stationary like the increment process X^H of an fBm (see Lemma 2.5.3), the autovariance function would correspond to the autocovariance function since $\operatorname{cov}\left(X_1^H, X_n^H\right) = \operatorname{cov}\left(X_{t+1}^H, X_{t+n}^H\right) = \gamma_{X^H}(n)$. Thus, the condition $\sum_{n=1}^{\infty} \rho^{H,\varepsilon}(n) = \infty$ generalizes Definition 2.3.1 correspondingly.

[878] See Brunnermeier and Sannikov (2016, p. 1504).

[879] See Basak and Cuoco (1998, p. 310).

[880] See the next section for details.

Cuoco (1998).

As outlined in (5.2), Brunnermeier and Sannikov (2016) use a Brownian motion for generating their macro shock. Here, this approach is generalized, and it is assumed that the macro shocks are generated by the process given in Definition 5.1.1 to account for long memory in the dynamics of the risky asset. Since the risky asset is employed with a linear production technology, the assumption of correlated macro shocks corresponds to long memory in the growth rates of output in the model. As in Brunnermeier and Sannikov (2016), an infinite time horizon is assumed, and their notation is adapted.

The economy is assumed to be populated by two kinds of agents called experts and households. Experts are allowed to hold (risky) capital to produce output. Furthermore, they can lend or borrow money at a risk-free interest rate r_t, which is determined in equilibrium. The financial friction for experts comes into effect as they only have the possibility to borrow money from households to finance their capital investments, i.e., experts have to issue a risk-free asset at an interest rate r_t. They cannot issue equity. Households are constrained as they cannot invest in the risky asset, i.e., they can only hold the risk-free asset, which is in zero net supply in the economy. As in the original model, it is assumed that all agents are small and behave as price-takers. In the following, the economy is characterized in more detail.

5.2.1 The Production Technology

Assume that there is an infinite number of experts and households, each with a total mass of one. To be more precise, let $\mathbb{I} = [0, 1]$ and $\mathbb{J} = (1, 2]$ denote the index set of experts and households, respectively. Each expert i produces output y_t^i by using his capital holdings k_t^i in the production function

$$y_t^i = a k_t^i, \tag{5.8}$$

where $a > 0$ is a productivity parameter that is assumed to be equal among all experts, and k_t^i is the amount of capital at time t held by expert i.[881] As in the discrete-time DSGE model treated in Chapter 4, the price of output is normalized to one and is treated as the numeraire.

Capital is held by experts, and it is assumed that capital evolves according to the stochastic differential equation:

$$\frac{dk_t^i}{k_t^i} = \left(\Phi(\iota_t^i) - \delta\right) dt + \sigma\, dZ_t^{H,\varepsilon}, \tag{5.9}$$

[881] See Brunnermeier and Sannikov (2016, p. 1504).

where $(\iota_t^i)_{t\geq 0}$ is expert i's re-investment rate process. That is, ι_t^i describes the amount of output that is re-invested by expert i into his capital stock expressed as investment rate per unit of capital.[882,883] After re-investment into capital, there are $(a-\iota_t^i)k_t^i$ units of output left for consumption. As in the original model the function $\Phi : [0,1] \longrightarrow [0,\infty), x \mapsto \Phi(x)$ is assumed to be twice differentiable and strictly concave with $\Phi(0) = 0, \Phi'(0) = 1, \Phi' > 0$ and $\Phi'' < 0$. The function Φ represents the investment adjustment costs of transforming output into capital.[884,885] The formulas derived in the following section hold for general functions Φ satisfying the mentioned assumptions. For the derivation of the equilibrium processes, the concrete form of Φ is assumed to be[886]

$$\Phi(\iota) = \frac{\log(\kappa\iota + 1)}{\kappa},$$

where, κ refers to the adjustment cost parameter.[887] Note that κ controls for the concavity of Φ, i.e., larger values of κ make Φ more concave while $\Phi(\iota) \to \iota$ as $\kappa \to 0$.[888] Therefore, the smaller κ, the smaller the investment adjustment costs.[889] The depreciation rate of capital is given by δ and is assumed to be non-negative. The volatility parameter σ is assumed to be positive. As can be seen from (5.9), the function Φ, the depreciation rate δ and the volatility parameter σ are equal among all experts $i \in \mathbb{I}$.

Following Brunnermeier and Sannikov (2014), capital k_t^i may be understood as capital in efficiency units rather than physical capital.[890] In their view, capital measures the future production potential of capital instead of today's physical capital.[891] Therefore, the notion of capital is rather broad in this chapter's model context and broader than in the previous chapter. There is, however, an alternative interpretation of (5.9) that illustrates some similarities between the model of this chapter and the discrete-time model of Chapter 4. To be more precise, the shocks dW_t or $dZ_t^{H,\varepsilon}$ may be interpreted as aggregate shocks to total factor productivity (TFP) as outlined in the sequel.[892]

882 See Brunnermeier and Sannikov (2016, p. 1504).
883 Throughout it is assumed that all stated stochastic processes are restrictedly progressively measurable, see Karatzas and Shreve (1998, Definition 1.7.1 on p. 28), to ensure that they are adapted to the filtration generated by the Brownian motion. This is rather a technical standard assumption, which is also imposed on the consumption process in the model of Basak and Cuoco (1998, Section 1.2 on p. 312). For technical details on the filtration, see Karatzas and Shreve (1998, p. 2).
884 See Brunnermeier and Sannikov (2016, p. 1504) and Brunnermeier and Sannikov (2014, p. 384).
885 Here, $\Phi'(\cdot)$ and $\Phi''(\cdot)$ refer to the first and second derivatives of Φ, respectively.
886 This is in line with Brunnermeier and Sannikov (2016, p. 1507).
887 See Brunnermeier and Sannikov (2016, p. 1507).
888 See Brunnermeier and Sannikov (2016, p. 1507).
889 See Brunnermeier and Sannikov (2016, p. 1507).
890 See Brunnermeier and Sannikov (2014, p. 385).
891 See Brunnermeier and Sannikov (2014, p. 385).
892 See Di Tella (2017, p. 2044) and Brunnermeier and Sannikov (2014, Footnote 2 on p. 385).

Let TFP_t be TFP at time t which is assumed to evolve according to

$$d\mathrm{TFP}_t = \sigma \mathrm{TFP}_t \, dZ_t^{H,\varepsilon} \text{ with } \mathrm{TFP}_0 = a.$$

Then one may rewrite expert i's capital holdings as capital in efficiency units. More specifically, let $k_t^{i,p}$ denote expert i's capital holdings measured in physical capital. Then, $k_t^i = \mathrm{TFP}_t k_t^{i,p}$ corresponds to expert i's capital holdings measured in efficiency units. Instead of imposing (5.9) for the evolution of capital in efficiency units, one can equally assume that physical capital evolves according to $dk_t^{i,p} = \left(\Phi(\iota_t^{i,p}/\mathrm{TFP}_t) - \delta\right) k_t^{i,p} \, dt$ and $\iota_t^{i,p}$ denotes the re-investment-rate per unit of physical capital.[893] Therefore, one can interpret the stochastic shocks in (5.9) as TFP shocks, i.e., positive shocks increase the effective capital stock while negative shocks erode the effective capital stock.[894]

Overall, the approach followed in this chapter is similar to the one of Chapter 4: Instead of assuming uncorrelated macro or TFP shocks dW_t as in (5.2), correlated shocks $dZ_t^{H,\varepsilon}$ are involved in the dynamics of the capital stock (5.9).[895]

By Lemma 5.1.2, one can rewrite (5.9) and obtains

$$\frac{dk_t^i}{k_t^i} = \left(\Phi(\iota_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon}\right) dt + \sigma\varepsilon^{H-1/2} \, dW_t \tag{5.10}$$

with $\varphi_t^{H,\varepsilon} = \int_0^t (H - 1/2)(t - u + \varepsilon)^{H-3/2} \, dW_u$ and $(H, \varepsilon) \in (0, 1) \times (0, \infty)$.

Before introducing the rest of the model structure in more detail, it seems helpful to clarify the main differences (5.2) and (5.9) or (5.10) and their implications for the economy's aggregate capital stock and output given these two different specifications of the law of motion of experts' capital stocks.

For the sake of simplicity, assume a constant re-investment rate $\iota_t \equiv \iota$ for the moment.[896] Since the re-investment rates are then equal among all experts, (5.9) and (5.8) hold for the economy's aggregate capital stock K_t and output Y_t as well. Formally, the economy's

893 See Brunnermeier and Sannikov (2014, Footnote 2 on p. 385) and Di Tella (2017, Footnote 7 on p. 2044).

894 Another related specification is considered in Adrian and Boyarchenko (2012). The authors propose a continuous-time model in which productivity evolves according to a geometric Brownian motion. Rather than directly specifying the evolution of capital in efficiency units, they specify the evolution of physical capital and productivity. For more details, see Adrian and Boyarchenko (2012, p. 5).

895 Recall that in Chapter 4, the white noise process ε_t^A in the evolution of the transitory TFP component was replaced with the fractionally integrated white noise process ν_t^A, see (4.15). A similar replacement is made here, as the uncorrelated increments dW_t are replaced with the increments of the $dZ_t^{H,\varepsilon}$.

896 It will turn out in Section 5.3, that ι_t is indeed constant in equilibrium.

aggregate capital stock and output are given by aggregating over all experts, i.e.,

$$K_t = \int_{\mathbb{I}} k_t^i \, di, \quad Y_t = \int_{\mathbb{I}} y_t^i \, di.$$

To distinguish the two capital processes of the benchmark model and the one considered here, let K_t^{BS} be the aggregate value of the capital stock of Brunnermeier and Sannikov (2016) given in (5.2) and K_t be the capital process specified in (5.10). The solutions to the corresponding stochastic differential equations are given in the following lemma.

Lemma 5.2.1
Let $\iota_t \equiv \iota$, then the solution to (5.9) *and* (5.10) *in terms of the aggregate capital stock is given by*

$$K_t = K_0 \exp\left(\left(\Phi(\iota) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2}\right) t + \sigma Z_t^{H,\varepsilon}\right). \tag{5.11}$$

Consequently, the capital stock in the benchmark model is given by

$$K_t^{BS} = K_0^{BS} \exp\left(\left(\Phi(\iota) - \delta - \frac{\sigma^2}{2}\right) t + \sigma W_t\right). \tag{5.12}$$

Proof
See Appendix C.2.1 □

Assume that one can observe the amount of output Y_t at $N+1$ discrete points in time, i.e., one observes $Y_0, \ldots, Y_N$. Then, the log differenced times series (or the growth rates of output and capital) in the setting of Brunnermeier and Sannikov (2016) are given by

$$R_{Y,n}^{\mathrm{BS}} := \log\left(\frac{Y_n^{\mathrm{BS}}}{Y_{n-1}^{\mathrm{BS}}}\right) = \log\left(\frac{K_n^{\mathrm{BS}}}{K_{n-1}^{\mathrm{BS}}}\right) = \Phi(\iota) - \delta - \frac{\sigma^2}{2} + \sigma(W_n - W_{n-1}).$$

Instead, by assuming the capital dynamics (5.11), output's and capital's growth rates become

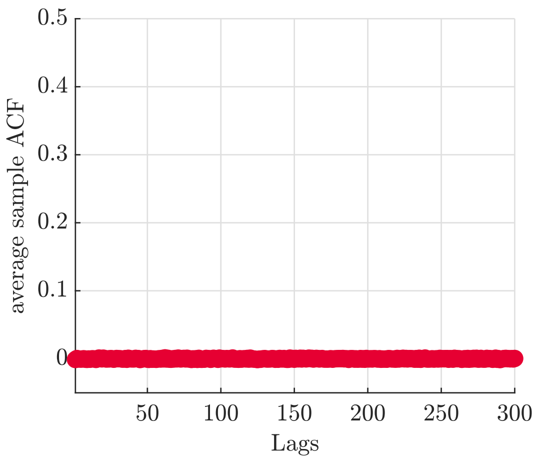
$$R_{Y,n} := \log\left(\frac{Y_n}{Y_{n-1}}\right) = \log\left(\frac{K_n}{K_{n-1}}\right) = \Phi(\iota) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2} + \sigma\left(Z_n^{H,\varepsilon} - Z_{n-1}^{H,\varepsilon}\right). \tag{5.13}$$

From a time series perspective, one can argue that Brunnermeier and Sannikov (2016) the log output series is modeled as a unit root process, since its first difference behaves as white noise with expected value $\Phi(\iota) - \delta - \sigma^2/2$.[897] By (5.13), one assumes that the growth

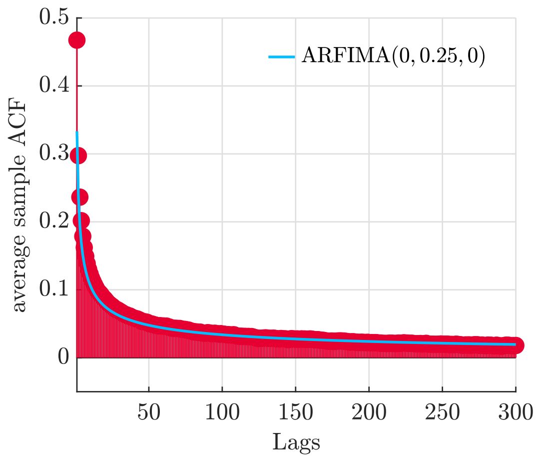
[897] Recall from Definition 2.5.1 that the increments of a Brownian motion are independent and identically

rates of output behave like a fractionally integrated white noise or an ARFIMA$(0, d, 0)$ process of order $d = H - 1/2$.[898]

The implications of these different approaches can be seen in Figure 5.1.[899] In Panel a), there is a white noise behavior of the autocorrelation function of the output's growth rates series. In contrast, one can observe in Panel b) a slowly decaying autocorrelation function similar to the one of an ARFIMA$(0, d, 0)$ process indicating long memory.[900]



a) ACF of $R^{\mathrm{BS}}_{Y,n}$



b) ACF of $R_{Y,n}$ with $H = 0.75$ and $\varepsilon = 10^{-3}$

Figure 5.1: Average sample autocorrelation function of the model-implied output growth rates. Panels a) and b) show the average sample autocorrelation function of the sequences $(R^{\mathrm{BS}}_{Y,n})_{n=1,\ldots,360}$ and $(R_{Y,n})_{n=1,\ldots,360}$, respectively. In Panel b), the light-blue line refers to the theoretical autocorrelation function (ACF) of an ARFIMA$(0, 0.25, 0)$ process. The average was taken over 100 seeds of each series. The corresponding Brownian shocks coincide in both cases. The parameters are $\iota = 0.04$, $\Phi(\iota) = \kappa^{-1}\log(\kappa\iota + 1)$ with $\kappa = 10$, $\sigma = 0.1$, $\delta = 0$, and $\varepsilon = 10^{-3}$.

Note that the single source of randomness in both cases is the Brownian motion $(W_t)_{t\geq 0}$. The autocorrelation enters the model via a different aggregation approach of past Brownian shocks. To be more illustrative, consider the following identities

$$W_t = \int_0^t 1\, dW_S \text{ and } Z_t^{H,\varepsilon} = \int_0^t (t - s + \varepsilon)^{H-1/2}\, dW_s.$$

distributed (i.i.d.). Therefore, the sequence $R^{\mathrm{BS}}_{Y,n}$, for $n = 1, \ldots, N$ is also i.i.d.

898 Note that, in contrast to Chapter 4, the Hurst index H is used instead of d to measure long memory in the exogenous process. This notation helps to get not confused about d with the differential terms such as dt, dW_t and $dZ_t^{H,\varepsilon}$. Moreover, the usage of H seems to be more common in a finance context. However, it should be kept in mind that both values are related to each other by $d = H - 1/2$, see Section 2.5.

899 As in the previous chapters, all figures in this chapter were computed using Matlab code written by the author.

900 At the end of Section 5.4 the evolution of the economy's aggregate capital stock is considered in more detail. More specifically, Lemma 5.4.1 derives the evolution of the capital stock's expected value. Some implications of long memory for this value are also discussed there.

Hence, Brunnermeier and Sannikov (2016) assume through (5.12) that all Brownian shocks enter into the development of the capital stock with a constant weight of 1. Instead, when assuming (5.11), one imposes that the weight of the Brownian shock depends on the time when it happened. That is, if one considers the impact of the Brownian shock that happened at time $s < t$, it is weighted with $(t - s + \varepsilon)^{H-1/2}$. The impact of the past shock is, thus, mainly determined by the time difference $(t - s)$ between the points of time t and s. Figure 5.2 shows the weighting kernel $K(t, s) = (t - s + \varepsilon)^{H-1/2}$ as a function of the time difference $(t - s)$ for various values of H. If $H > 1/2$, it can be seen that the weight of a shock increases with the amount of time that has passed since the shock occurred. Here, Rogers (1997)'s quotation given in Section 5.1 becomes clear again. He states that the long-run behavior of the weighting kernel determines long memory. Figure 5.2 illustrates that the higher the value of H, the larger the weighting kernel for long time differences and, thus, the higher the degree of long memory.

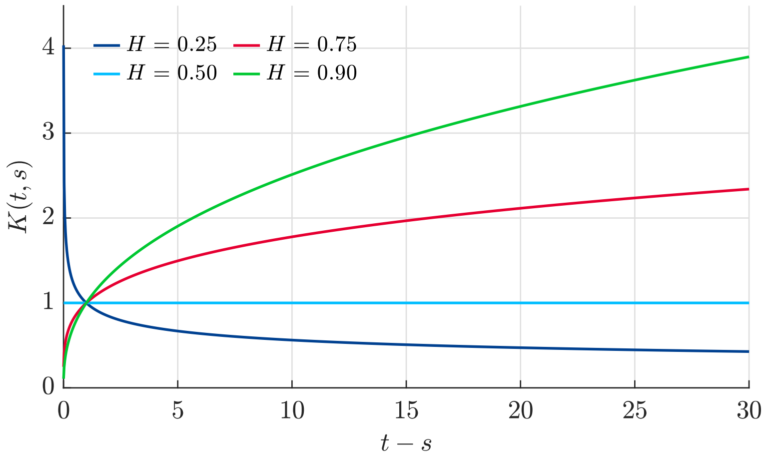


Figure 5.2: Weighting kernel of $Z^{H,\varepsilon}$ for various values of H. The plot shows the kernel $K(t, s) = (t - s + \varepsilon)^{H-1/2}$ as function of the time difference $(t - s)$ for different values of H with $\varepsilon = 10^{-3}$.

Consequently, including the exogenous shock process $Z^{H,\varepsilon}$ in the dynamics of the experts' capital holdings introduces long memory in the economy's GDP growth rates. A fact supported by empirical evidence outlined in Section 3.2.1. As shown in (5.10), the correlation of the $Z^{H,\varepsilon}$ shocks expressed by the parameters ε and H affects the dynamics of experts' capital holdings in two ways: First, there is the additional drift component $\sigma\varphi_t^{H,\varepsilon}$ (drift effect), which represents the weighted influence of past Brownian shocks. If there is a sequence of negative (positive) shocks, these negative (positive) shocks will be accumulated in $\varphi_t^{H,\varepsilon}$ and keep influencing the evolution of capital negatively (positively). Second, there is an additional risk component $\varepsilon^{H-1/2}$ occurring as a multiplier of the volatility σ (volatility effect). If $\varepsilon < 1$, this multiplier is smaller than one in the presence of long memory $(H > 1/2)$. Therefore, introducing the process $Z^{H,\varepsilon}$ reduces aggregate risk.

Note that the capital dynamics proposed in the original model neglect those effects. However, it is possible to recover the original setting by setting $H = 1/2$ in (5.10). Therefore, the dynamics given in (5.9) or (5.10) are more general than those in the original model.

The following sections provide more details on the model economy and examine the implications of a long memory TFP shock on the wealth distribution between households and experts and the equilibrium interest rate. Before the agents are introduced in more detail, let the price process of the capital be denoted by $q = (q_t)_{t\geq 0}$, i.e., q_t represents the price of one unit of capital at time t. Furthermore, assume that the price process is given by[901,902]

$$\frac{dq_t}{q_t} = \mu_t^q dt + \sigma_t^q \, dW_t. \tag{5.14}$$

5.2.2 Agents and Preferences

Assume that there is an infinite number of experts and households, each with a total mass of one. To be more precise, let $\mathbb{I} = [0,1]$ and $\mathbb{J} = (1,2]$ denote the index set of experts and households, respectively. The consumption process of expert $i \in \mathbb{I}$ is given by $(c_t^i)_{t\geq 0}$ and the consumption process of household $j \in \mathbb{J}$ is denoted by $(\tilde{c}_t^j)_{t\geq 0}$.[903]

Both types of agents are assumed to have logarithmic utility and a constant time preference rate of $\rho > 0$. Both types of agents are assumed to maximize the expected present value of total utility given by

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t} \log\left(c_t^i\right) dt\right] \text{ and } \mathbb{E}\left[\int_0^\infty e^{-\rho t} \log\left(\tilde{c}_t^j\right) dt\right]$$

for expert $i \in \mathbb{I}$ and household $j \in \mathbb{J}$, respectively.[904]

The following section treats the agents' utility maximization problems in more detail.

901 At first glance, this may seem arbitrary because one might expect that the dynamics of q are affected by the same macro shocks $dZ_t^{H,\varepsilon}$ as the capital. If one assumes that $dq_t/q_t = \hat{\mu}_t^q dt + \hat{\sigma}_t^q \, dZ^{H,\varepsilon}$ for some processes $(\hat{\mu}_t^q)_{t\geq 0}$ and $(\hat{\sigma}_t^q)_{t\geq 0}$, one obtains $dq_t/q_t = (\hat{\mu}_t^q + \hat{\sigma}_t^q \varphi_t^{H,\varepsilon})dt + \varepsilon^{H-1/2}\hat{\sigma}_t^q \, dZ^{H,\varepsilon}$ by using Lemma 5.1.2. By defining $\mu_t^q = (\hat{\mu}_t^q + \hat{\sigma}_t^q \varphi_t^{H,\varepsilon})$ and $\sigma_t^q = \varepsilon^{H-1/2}\hat{\sigma}_t^q$, one obtains the stated expression. Hence μ_t^q and σ_t^q can depend on H and ε. As will be demonstrated in the following sections, this issue does not matter in equilibrium.

902 This assumption is in accordance with Brunnermeier and Sannikov (2016, Equation (2) on p. 1505).

903 In the following, all individual specific processes for households are additionally marked with $\sim$ to distinguish them from the individual processes of experts.

904 It is assumed that these expected values exist.

5.2.2.1 Experts

Let k_t^i denote the capital holdings of expert $i \in \mathbb{I}$. Applying Lemma 5.1.2, one can rewrite (5.9) to obtain the evolution of the capital holdings of an individual expert $i \in \mathbb{I}$, i.e.,

$$\frac{dk_t^i}{k_t^i} = \left(\Phi(\iota_t^i) - \delta\right) dt + \sigma\, dZ_t^{H,\varepsilon} = \left(\Phi(\iota_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon}\right) dt + \sigma\varepsilon^{H-1/2}\, dW_t \text{ with } k_0^i = \bar{k}_0^i, \tag{5.15}$$

where $\bar{k}_0^i$ is the initial amount of capital held by expert i which is assumed to be given as a non-negative constant, and $(\iota_t^i)_{t\geq 0}$ is the process of re-investment rates chosen by expert $i \in \mathbb{I}$.

At time t, expert i has to decide how much of the generated output he should re-invest into the risky capital, i.e., he has to choose ι_t and how much he wants to consume. Furthermore, he has to decide how much of his wealth is invested in capital or is held/lent at the interest rate r_t, i.e., he has to make the portfolio choice between capital and the risk-free asset. Let $n^i = (n_t^i)_{t\geq 0}$ be the wealth process of expert i and π_t^i the amount of risk-free assets held by expert i at time t. Thus, the wealth of expert i is given by

$$n_t^i = q_t k_t^i + \pi_t^i B_t, \tag{5.16}$$

where $dB_t = r_t B_t dt$ describes the dynamics of the risk-free asset B_t with interest rate process $r = (r_t)_{t\geq 0}$. The initial wealth is given by $n_0^i = q_0 \bar{k}_0^i$. The evolution of the wealth of expert i is affected by four sources:

- After choosing ι_t^i the capital produces $(a - \iota_t^i)k_t^i$ units of output, see Section 5.2.1.
- The capital's value varies with an amount of $d(q_t k_t^i)$.
- The expert earns (or has to pay) the interest rate on risk-free assets, i.e., $\pi_t^i\, dB_t = \pi_t^i B_t r_t\, dt$.
- The expert has consumption expenditures of c_t^i.

Thus, one can express the dynamics of the wealth process n^i by

$$dn_t^i = k_t^i(a - \iota_t^i)\, dt + d(q_t k_t^i) + \pi_t^i B_t r_t\, dt - c_t^i\, dt. \tag{5.17}$$

Moreover, one can express capital holdings and holdings of risk-free assets as shares of wealth, i.e., define the part of the wealth of expert i invested in capital as

$$x_t^i := \frac{q_t k_t^i}{n_t^i}. \tag{5.18}$$

Combining (5.16) and (5.18), one can rewrite (5.17) yielding in[905]

$$dn_t^i = x_t^i n_t^i \frac{(a - \iota_t^i)}{q_t} \, dt + \, d(q_t k_t^i) + (1 - x_t^i) n_t^i r_t \, dt - c_t^i \, dt. \tag{5.19}$$

The differential $d(q_t k_t^i)$ can be calculated by using Itô product formula (see (C.6) in Appendix C.1). The dynamics of the price process q are given in (5.14) and the dynamics of the capital holdings in (5.15). Overall, one obtains

$$\frac{d(q_t k_t^i)}{q_t k_t^i} = \left(\Phi(\iota_t^i) - \delta + \sigma \varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2} \sigma \sigma_t^q\right) \, dt + (\sigma \varepsilon^{H-1/2} + \sigma_t^q) \, dW_t. \tag{5.20}$$

Plugging this into (5.19), one obtains the dynamics of the wealth process

$$\begin{aligned} dn_t^i = x_t^i n_t^i \frac{(a - \iota_t^i)}{q_t} \, dt + \Big[&\left(\Phi(\iota_t^i) - \delta + \sigma \varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2} \sigma \sigma_t^q\right) \, dt \\ &+ (\sigma \varepsilon^{H-1/2} + \sigma_t^q) \, dW_t \Big] \, q_t k_t^i + (1 - x_t^i) n_t^i r_t \, dt - c_t^i \, dt. \end{aligned}$$

After some rearrangements, one finally has

$$\begin{aligned} dn_t^i = &\left[\frac{(a - \iota_t^i)}{q_t} + \left(\Phi(\iota_t^i) - \delta + \sigma \varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2} \sigma \sigma_t^q\right) \right] x_t^i n_t^i \, dt \\ &+ (1 - x_t^i) r_t n_t^i \, dt - c_t^i \, dt + (\sigma \varepsilon^{H-1/2} + \sigma_t^q) x_t^i n_t^i \, dW_t. \end{aligned} \tag{5.21}$$

Given the price process $(q_t)_{t \geq 0}$ and the risk-free rate $(r_t)_{t \geq 0}$, one is now able to state the utility maximization problem of expert i given by

$$\max_{x^i, c^i, \iota^i} \mathbb{E} \left[\int_0^\infty e^{-\rho t} \log\left(c_t^i\right) dt \right] \text{ such that } \begin{cases} x_t^i \geq 0 \text{ for } t \geq 0 \\ c_t^i \geq 0 \text{ for } t \geq 0 \\ n^i \text{ follows (5.21) with } n_0^i = q_0 \bar{k}_0^i \\ n_t^i \geq 0 \text{ for } t \geq 0 \end{cases}. \tag{5.22}$$

This maximization problem is solved in Section 5.3.

[905] Here it is assumed that there are no other sources of income or expenditures, i.e., the share of wealth invested into the risk-free assets corresponds to one minus the wealth share invested into capital.

5.2.2.2 Households

This section derives the maximization problem of an individual household $j \in \mathbb{J}$. His wealth process is denoted by $\tilde{n}^j = (\tilde{n}^j_t)_{t\geq 0}$. Since households can hold risk-free assets only, the wealth process is given by

$$\tilde{n}^j_t = \tilde{\pi}^j_t B_t, \tag{5.23}$$

where $\tilde{\pi}^j_t$ denotes the amount of risk-free assets held by household j at time t. The household earns the interest rate from holding the risk-free asset and has expenditures for consumption.[906] Therefore, the wealth dynamics of household j is given by

$$d\tilde{n}^j_t = \tilde{\pi}^j_t \, dB_t - \tilde{c}^j_t \, dt = \left(\tilde{\pi}^j_t B_t r_t - \tilde{c}^j_t\right) \, dt. \tag{5.24}$$

Inserting (5.23) in (5.24) leads to

$$d\tilde{n}^j_t = \left(\tilde{n}^j_t r_t - \tilde{c}^j_t\right) \, dt. \tag{5.25}$$

Assume that the initial wealth $\tilde{n}^j_0$ of household j is given by $\bar{\tilde{n}}^j_0 > 0$. As the risk-free asset is in zero net supply in the economy, the economy's total wealth is given by $q_t K_t$, i.e., total wealth is equal to the value of the aggregate capital stock.[907] Consequently, in order to ensure a positive initial wealth of households, they have to hold a certain amount of capital at $t = 0$. Since households are not allowed to hold capital in this model economy, it is assumed that they sell their initial capital holdings immediately to experts.[908] That is, the initial wealth of household j can equally be written as $\bar{\tilde{n}}^j_0 = q_0 \bar{\tilde{k}}^j_0$, where $\bar{\tilde{k}}^j_0$ refers to the household j's initial capital holdings. The utility maximization problem of household $j \in \mathbb{J}$ can now be formally stated as

$$\max_{\tilde{c}^j} \mathbb{E}\left[\int_0^\infty e^{-\rho t} \log\left(\tilde{c}^j_t\right) dt\right] \text{ such that } \begin{cases} \tilde{c}^j_t \geq 0 \text{ for } t \geq 0 \\ \tilde{n}^j \text{ follows (5.25) with } \tilde{n}^j_0 = \bar{\tilde{n}}^j_0 \, . \\ \tilde{n}^i_t \geq 0 \text{ for } t \geq 0 \end{cases} \tag{5.26}$$

The following section provides a formal definition of the equilibrium in the economy, and the solution to the agents' maximization problem is derived.

[906] Like the experts, households are assumed to have non-negative wealth. So (5.23) implies that households hold a non-negative amount of risk-free assets.
[907] See Brunnermeier and Sannikov (2016, p. 1504).
[908] This view is in line with Di Tella (2017, p. 2046).

5.3 Definition and Solution of the Equilibrium

The formal definition of the economy's equilibrium reads as follows[909]

Definition 5.3.1 ———

Let $(K_t)_{t\geq 0}$ be the aggregate capital stock in the economy given by

$$K_t = \int_{\mathbb{I}} k_t^i \, di \text{ with initial value } K_0 = \int_{\mathbb{I}} \bar{k}_0^i \, di + \int_{\mathbb{J}} \bar{\tilde{k}}_0^j \, dj.$$

An equilibrium is defined as the following families of stochastic processes. Price of capital and the risk-free rate $\{(q_t)_{t\geq 0}, (r_t)_{t\geq 0}\}$; experts' investment and consumption decisions, capital holdings and wealth processes $\{(\iota_t^i)_{t\geq 0}, (c_t^i)_{t\geq 0}, (k_t^i)_{t\geq 0}, (n_t^i)_{t\geq 0}\}_{i\in\mathbb{I}}$; households' consumption decisions and wealth processes $\{(\tilde{c}_t^j)_{t\geq 0}, (\tilde{n}_t^j)_{t\geq 0}\}_{j\in\mathbb{J}}$; such that i) experts solve their maximization problem (5.22)*, ii) households solve their maximization problem* (5.26)*, iii) the market of consumption goods clears, i.e.,*

$$\int_{\mathbb{I}} c_t^i \, di + \int_{\mathbb{J}} \tilde{c}_t^j dj = \int_{\mathbb{I}} (a - \iota_t^i) k_t^i \, di \text{ for all } t \geq 0,$$

and iv) the risk-free asset is in zero net supply, i.e.,

$$\int_{\mathbb{I}} \pi_t^i \, di + \int_{\mathbb{J}} \tilde{\pi}_t^j dj = 0 \text{ for all } t \geq 0.$$

⋊

Before the maximization problem of expert $i \in \mathbb{I}$ is solved, the market price of risk process of expert i, $(\vartheta_t^i)_{t\geq 0}$, is introduced as follows[910]

$$\vartheta_t^i := \frac{\dfrac{a - \iota_t^i}{q_t} + \Phi(\iota_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2}\sigma\sigma_t^q - r_t}{\varepsilon^{H-1/2}\sigma + \sigma_t^q}. \tag{5.27}$$

The market price of risk process describes the quotient of the excess return of capital over the risk-free rate divided by the volatility of capital.[911] The numerator can be interpreted as the expected excess return of (the value of) capital $q_t k_t^i$ over the interest rate r_t, see (5.20), since $(a - \iota_t^i)/q_t$ is the output left after investment expressed in $q_t k_t^i$ units, see

[909] Definition 5.3.1 is adapted from Brunnermeier and Sannikov (2014, pp. 388f.) and Di Tella (2017, p. 2048).

[910] This market price of risk process corresponds to the Sharpe ratio of Brunnermeier and Sannikov (2016, p. 1508).

[911] See Brunnermeier and Sannikov (2016, p. 1507), Karatzas and Shreve (1998, p. 27) or Shreve (2004, p. 216).

(5.21). Thus it can be interpreted as an additional dividend of the capital.[912,913]

Proposition 5.3.2

Consider the problem stated in (5.22) *and define the stochastic process* $\xi^i = (\xi^i_t)_{t\geq 0}$ *as*

$$\xi^i_t = \exp\Big(-\int_0^t r_s\,ds - \int_0^t \vartheta^i_s\,dW_s - \frac{1}{2}\int_0^t (\vartheta^i_s)^2\,ds\Big) \tag{5.28}$$

which follows the dynamics

$$d\xi^i_t = -r_t\xi^i_t\,dt - \vartheta^i_t\xi^i_t\,dW_t \text{ with } \xi^i_0 = 1. \tag{5.29}$$

Then,

i) the optimal re-investment rate process $(\hat{\iota}^i_t)_{t\geq 0}$ *satisfies* $\Phi'(\hat{\iota}^i_t) = \frac{1}{q_t}$ *for all* $t \geq 0$ *as long as* $\hat{\vartheta}^i_t > 0$, *where* $\hat{\vartheta}^i = (\hat{\vartheta}^i_t)_{t\geq 0}$ *is the market price of risk process corresponding to* $(\hat{\iota}^i_t)_{t\geq 0}$.

ii) the optimal consumption process is given by $\hat{c}^i_t = \rho e^{-\rho t}\frac{1}{\hat{\xi}^i_t}$, *where* $\hat{\xi}^i_t$ *is given by* (5.28) *corresponding to* $\hat{\vartheta}^i$.

iii) the optimal wealth process is given by $\hat{n}^i_t = e^{-\rho t}\frac{1}{\hat{\xi}^i_t}$.

iv) The optimal fraction of wealth to invest into capital is given by $\hat{x}^i_t = \dfrac{\hat{\vartheta}^i_t}{\varepsilon^{H-1/2}\sigma + \sigma^q_t}$.

Proof

See Appendix C.2.2 □

The choice of the optimal re-investment rate $\hat{\iota}^i_t$ which satisfies $\Phi'(\hat{\iota}^i_t) = 1/q_t$ if $\hat{\vartheta}^i_t > 0$ states that experts do not re-invest into capital if there is a negative excess return of capital over the interest rate r_t. Equation (C.20) in Appendix C.2.2 highlights that experts have to claim a positive excess return over the risk-free rate to invest at least a positive amount of wealth into capital. Since experts have to hold all the capital in the economy, (C.20) in Appendix C.2.2 implies that the market price of risk is indeed positive.

The following proposition solves the households' maximization problem.

[912] Brunnermeier and Sannikov (2016, p. 1506) name the expression $(a - \iota^i_t)/q_t$ "dividend yield" of the capital.

[913] In the following, all variables marked with ˆ indicate optimal values, i.e., values that solve agents' maximization problems.

Proposition 5.3.3

Consider the maximization problem stated in (5.26)*. The optimal wealth process* $(\hat{\tilde{n}}_t^j)_{t\geq 0}$ *of household* $j \in \mathbb{J}$ *is given by*

$$\hat{\tilde{n}}_t^j = \bar{\tilde{n}}_0^j e^{-\rho t} \exp\left(\int_0^t r_u \, du\right)$$

and the corresponding consumption process $(\tilde{c}_t^j)_{t\geq 0}$ *is given by* $\hat{\tilde{c}}_t^j = \rho \hat{\tilde{n}}_t^j$.

Proof

See Appendix C.2.3 □

Now one can solve for the model equilibrium and find the equilibrium risk-free rate r_t as well as the equilibrium price of capital q_t.

Optimal re-investment rate. From Proposition 5.3.2, one knows that each expert $i \in \mathbb{I}$ chooses the re-investment rate process $(\hat{\iota}_t^i)_{t\geq 0}$ such that

$$\Phi'(\hat{\iota}_t^i) = \frac{1}{q_t} \text{ or } \hat{\iota}_t^i = \Psi\left(\frac{1}{q_t}\right) \text{ for all } t \geq 0, \tag{5.30}$$

as long as $\hat{\vartheta}_t^i$ is positive. Note that Ψ is the inverse function of Φ'. As stated above, the market price of risk $\hat{\vartheta}_t^i$ is indeed positive. The optimal re-investment rate does not depend on i, so one can drop the index i, i.e., $\hat{\iota}_t \equiv \hat{\iota}_t^i$. That is, all experts choose the identical re-investment rate. The same holds true for the market price of risk defined in (5.27), i.e., $\hat{\vartheta}_t \equiv \hat{\vartheta}_t^i$.

Experts' and households' optimal wealth and consumption. The optimal wealth and consumption processes of experts are given by

$$d\hat{n}_t^i = \left[(\hat{\vartheta}_t)^2 + r_t - \rho\right] \hat{n}_t^i \, dt + \hat{\vartheta}_t \hat{n}_t^i \, dW_t \text{ with } \hat{n}_0^i = q_0 \bar{k}_0^i$$

and $\hat{c}_t^i = \rho \hat{n}_t^i$, respectively.[914] By inserting $\hat{\iota}_t$ into (5.15), one obtains the equilibrium process of capital holdings $(k_t^i)_{t\geq 0}$ as

$$\frac{dk_t^i}{k_t^i} = \left(\Phi(\hat{\iota}_t) - \delta + \sigma \varphi_t^{H,\varepsilon}\right) dt + \sigma \varepsilon^{H-1/2} \, dW_t \text{ with } k_0^i = \bar{k}_0^i.$$

[914] It follows directly from ii) and iii) of Proposition 5.3.2 that experts consume a fixed fraction of their wealth.

This implies that the aggregate capital stock K_t follows the dynamics

$$\frac{dK_t}{K_t} = \left(\Phi(\hat{\iota}_t) - \delta + \sigma\varphi_t^{H,\varepsilon}\right) dt + \sigma\varepsilon^{H-1/2}\, dW_t \text{ with } K_0 = \int_{\mathbb{I}} \bar{k}_0^i\, di. \tag{5.31}$$

For household $j \in \mathbb{J}$, the equilibrium wealth and consumption processes are given by Proposition 5.3.3.

Goods market clearing. Since the risk-free asset is in zero net supply in the economy, the economy's net worth is given by the total value of capital, i.e., $q_t K_t$. Plugging this into the goods market clearing condition, see iii) of Definition 5.3.1, one obtains

$$\int_{\mathbb{I}} \hat{c}_t^i\, di + \int_{\mathbb{J}} \hat{\hat{c}}_t^j dj = \int_{\mathbb{I}} (a - \hat{\iota}_t) k_t^i\, di \text{ and thus } \rho \left(\int_{\mathbb{I}} \hat{n}_t^i\, di + \int_{\mathbb{J}} \hat{\hat{n}}_t^j dj\right) = (a - \hat{\iota}_t) K_t.$$

This implies[915]

$$\rho q_t = (a - \hat{\iota}_t) \text{ or } \rho q_t = \left(a - \Psi\left(\frac{1}{q_t}\right)\right). \tag{5.32}$$

By the monotonicity of Ψ, recall that $\Phi'' < 0$, and since $a, \rho > 0$, the intermediate value theorem implies that the price $q_t \equiv q$ is uniquely determined by this equation.[916] Additionally, the price q_t has to be constant, implying that $\mu_t^q = \sigma_t^q \equiv 0$.[917] Moreover, the re-investment rate process $\hat{\iota}_t$ must also be constant due to (5.30) and the constant price of capital, i.e., $\hat{\iota}_t \equiv \hat{\iota}$.

With $\Phi(\hat{\iota}) = \log(\kappa\hat{\iota} + 1)/\kappa$, the equilibrium value of q can be derived as follows. From $\Phi'(\hat{\iota}) = (\kappa\hat{\iota} + 1)^{-1}$, follows that $\Psi(x) = \kappa^{-1}(x^{-1} - 1)$. Plugging these expressions into (5.32), leads to

$$\rho q = a - \frac{1}{\kappa}(q - 1)$$

and, thus, $q = \dfrac{a\kappa + 1}{\rho\kappa + 1}$.[918]

In addition, the consumption rate ρ can also be expressed as

$$\rho = \frac{a - \hat{\iota}}{q} = \frac{a - \Psi(1/q)}{q},$$

915 Recall that $\int_{\mathbb{I}} \hat{n}_t^i\, di + \int_{\mathbb{J}} \hat{\hat{n}}_t^j dj = q_t K_t$, i.e., the aggregate wealth in the economy is equal to the value of the aggregate capital stock $q_t K_t$.

916 The idea behind this statement is as follows. Since, $\Phi'' < 0$ one knows that Φ' is monotonically decreasing. Since Ψ is the inverse function of Φ', it is also monotonically decreasing. Therefore, the left-hand side of (5.32) is monotonically increasing in q_t, while the right-hand side is monotonically decreasing in q_t, thus, establishing a unique intersection point. A more rigorous argument is given in Appendix C.3.

917 This is similar to the benchmark model, see Brunnermeier and Sannikov (2016, p. 1507).

918 See Brunnermeier and Sannikov (2016, p. 1507).

where the second equality uses (5.30). Inserting this into $\hat{\vartheta}_t$, leads to

$$\hat{\vartheta}_t = \frac{\rho + \Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon} - r_t}{\varepsilon^{H-1/2}\sigma}.$$

The optimal fraction of wealth that expert i invests into capital is given by iv) of Proposition 5.3.2 and does not depend on i in the equilibrium. Thus, it follows from (5.18) that

$$\frac{qK_t}{N_t} = \frac{\hat{\vartheta}_t}{\varepsilon^{H-1/2}\sigma}, \tag{5.33}$$

where $N_t = \int_{\mathbb{J}} n_t^j \, dj$ denotes the aggregate wealth of experts.[919] Then,

$$\eta_t := \frac{N_t}{qK_t} \in [0,1] \tag{5.34}$$

denotes the wealth share of experts. Thus, η_t is a measure of the wealth distribution within the economy. The higher η_t, the higher the wealth share of experts, and the lower the wealth share of households.

By inserting the definition of η_t into (5.33), one obtains

$$\frac{1}{\eta_t} = \frac{\hat{\vartheta}_t}{\varepsilon^{H-1/2}\sigma} = \frac{\rho + \Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon} - r_t}{(\varepsilon^{H-1/2}\sigma)^2}. \tag{5.35}$$

Equation (5.35) finally determines the equilibrium risk-free rate

$$r_t = \rho + \Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon} - \frac{(\sigma\varepsilon^{H-1/2})^2}{\eta_t}. \tag{5.36}$$

As in the original model, one can determine a law of motion of η_t given in the next lemma.

Lemma 5.3.4

The law of motion of η is given by

$$\frac{d\eta_t}{\eta_t} = \left(\frac{1-\eta_t}{\eta_t}\right)^2 (\varepsilon^{H-1/2}\sigma)^2 \, dt + \frac{(1-\eta_t)}{\eta_t}\varepsilon^{H-1/2}\sigma \, dW_t.$$

Proof

See Appendix C.2.4 □

[919] See Brunnermeier and Sannikov (2016, p. 1504).

Equation (5.33) clarifies that $\hat{\vartheta}_t$ is positive if and only if $\eta_t > 0$. The corresponding equation for an individual expert $i \in \mathbb{I}$ is (C.20) in Appendix C.2.2 or iv) of Proposition 5.3.2, i.e., an expert holds capital only if he expects a positive excess return over the risk-free rate. If the risk of capital investments is not compensated with positive expected returns, experts will not be incentivized to invest or hold capital. Interestingly, the market price of risk tends to infinity as $\eta_t \to 0$. This behavior seems degenerative, but it is rather founded in the restrictive structure of the model. Since experts are imposed on holding the total capital stock (they cannot sell capital to households), they must be convinced to hold it by even higher risk premia. From an economic perspective, it seems, thus, reasonable to assume that $\eta_t > 1/C > 0$ for a constant C in order to ensure that $\hat{\vartheta}_t < C\varepsilon^{H-1/2}\sigma$. Clearly, since η_t refers to expert's wealth share, reasonable values of η_t would be smaller than or equal to 1, implying that $\hat{\vartheta}_t \geq \sigma\varepsilon^{H-1/2}$.

5.4 Results and Comparison with the Benchmark Model

This section presents the results of introducing long memory shocks into the continuous-time macro-financial model given in the previous section and highlights the differences to the benchmark model. It turns out that different variables are affected in different ways by long memory of experts' capital dynamics and hence by long memory of the growth rates of the economy's output and the aggregate capital stock. Since the approach outlined in the preceding sections generalizes the original model, the focus lies on the effects of changes in the parameters H and ε. Recall that the original model corresponds to $H = 1/2$ independent of the value of ε. Values of $H < 1/2$ indicate negatively correlated shocks with short memory, whereas values of $H > 1/2$ indicate positively correlated shocks with long memory in the sense of Proposition 5.1.3.

Let the drift and the volatility of η_t be denoted by $\mu_\eta^{H,\varepsilon}\eta_t$ and $\sigma_\eta^{H,\varepsilon}\eta_t$, respectively.[920] Regarding Lemma 5.3.4, it yields

$$\begin{aligned} d\eta_t &= \mu_\eta^{H,\varepsilon}\eta_t\,dt + \sigma_\eta^{H,\varepsilon}\eta_t\,dW_t \\ &= \frac{(1-\eta_t)^2}{\eta_t}(\varepsilon^{H-1/2}\sigma)^2\,dt + (1-\eta_t)\varepsilon^{H-1/2}\sigma\,dW_t. \end{aligned} \tag{5.37}$$

Note that η_t does not depend on the history of shocks, i.e., the experts' wealth share is

[920] This is in accordance with Brunnermeier and Sannikov (2016, Figure 2 on p. 1510).

independent of the drift effect, expressed by the term $\varphi_t^{H,\varepsilon}$. This is because the net worth N_t and the value of the capital stock q_tK_t depend in the same way on the drift effect. Consequently, the drift effect cancels in η_t, which is simply the fraction of N_t and q_tK_t, see (5.34).[921] However, η_t depends on the correlation of shocks expressed by the volatility effect $\varepsilon^{H-1/2}$. The interest rate depends on both the volatility and the drift effect.

Let η^{BS} and r^{BS} be the processes of experts' wealth share and the interest rate in the model of Brunnermeier and Sannikov (2016), respectively. They are given by[922]

$$d\eta_t^{\mathrm{BS}} = \frac{(1-\eta_t^{\mathrm{BS}})^2}{\eta_t^{\mathrm{BS}}}\sigma^2\,dt + (1-\eta_t^{\mathrm{BS}})\sigma\,dW_t \tag{5.38}$$

and

$$r_t^{\mathrm{BS}} = \rho + \Phi(\Psi(1/q)) - \delta - \frac{\sigma^2}{\eta_t^{\mathrm{BS}}}. \tag{5.39}$$

A direct comparison between (5.38) and (5.37) reveals how the correlations of shocks impact the dynamics of experts' wealth share. The dynamics of η and η^{BS} coincide for $H = 1/2$ or $\varepsilon = 1$. In the case of $H = 1/2$, both models are identical. In the case of $\varepsilon = 1$, the interest rate differs from the one in the original model since it depends on the drift effect $\varphi^{H,\varepsilon}$ (compare (5.36) with (5.39)).

To illustrate the results and to obtain good comparability, the same parameter values as in the original model are chosen. The effect of a variation of H on the drift and the volatility of η are presented in Figure 5.3 for a fixed value of $\varepsilon = 10^{-3}$. The effects of a variation of ε on η is shown in Figure 5.4.

From Figure 5.3, it can be seen that the positive correlation of shocks and long memory ($H > 1/2$) implies a lower drift and a lower volatility of the wealth share of experts. This observation appear plausible, since as aggregate shocks have long memory, shocks tend to be followed by shocks of the same sign, implying that the behavior of the exogenous process is to certain degree predictable. This predictability has a risk reducing effect (volatility effect, see Section 5.2.1). As a consequence, the experts' sector grows more slowly, on average, than in the benchmark model.

The reverse is true in the case of $H < 1/2$. In this case, the volatility is higher than in the original model, which reflects the negative correlations indicating less predicable movements in the exogenous process. In this case, the drift is higher as well. So, on average,

921 Recall from (C.11) of Appendix C.1 that the drift of a quotient of two Itô processes depends on the difference of the two corresponding drift processes.

922 See Brunnermeier and Sannikov (2016, Equations (9) and (11) on pp. 1508f.).

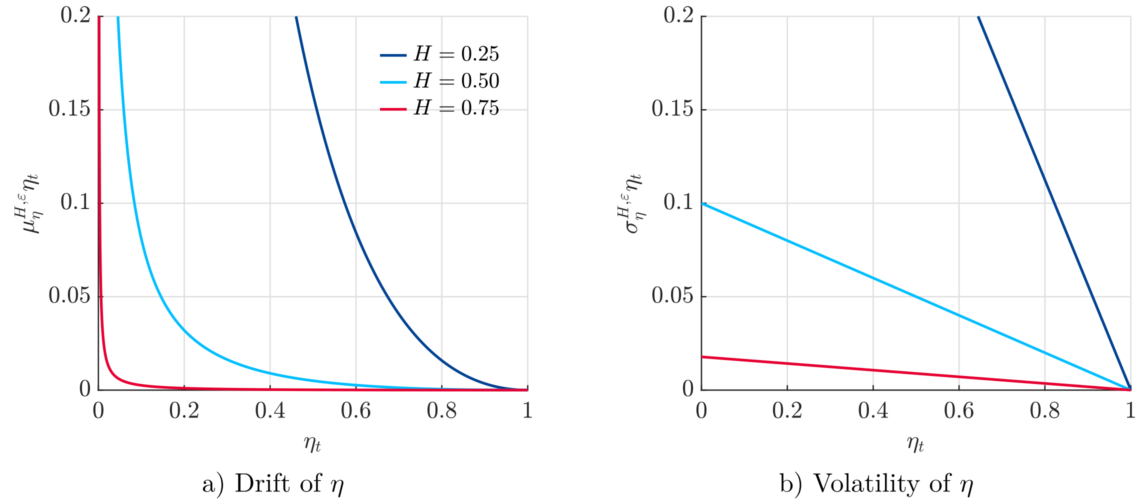


Figure 5.3: Drift and volatility of η as functions of η_t for various values of H. The remaining parameters in (5.37) are $\sigma = 0.1$ and $\varepsilon = 10^{-3}$. The legend is the same in both panels.

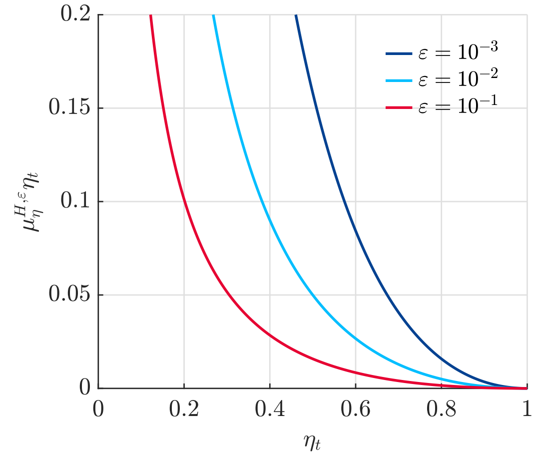
the experts' sector grows faster if the shocks to the economy are negatively correlated than if the shocks are positively correlated.

From Panels a) and b) of Figure 5.4, it can be seen that in the case of $H < 1/2$, an increase in ε leads to a decline in both the drift and the volatility of η. The reverse is true for $H > 1/2$ as can be seen from Panels c) and d) of Figure 5.4, i.e., an increase in ε also rises the drift and the volatility of η if $H > 1/2$. Although the choice of the parameter ε seems to be arbitrary, recall from Section 5.1 that the process $Z^{H,\varepsilon}$ was motivated by an approximation argument, i.e., $Z^{H,\varepsilon}$ converges to a type II fBm as $\varepsilon \to 0$. From this perspective, "small" values of ε seem preferable.

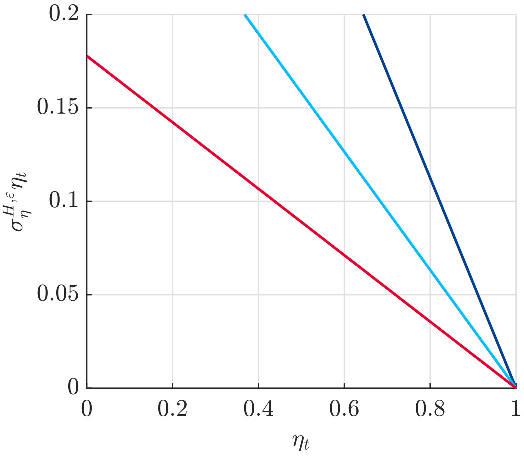
As in the original model, the experts' sector tends to overwhelm the households' sector since the drift rate is positive for all values of η_t, see (5.37), Figure 5.3 and Figure 5.4.[923] Brunnermeier and Sannikov (2016) argue that this is because experts have an advantage over households because they can hold capital, which allows them to earn risk premia.[924] In addition, it follows from Lemma 5.3.4 that if the experts' wealth share reaches 1, it remains at this level.[925] Hence, the state $\eta_t = 1$ may be regarded as a steady state value of η.

To illustrate the convergence to the steady state of $\eta = 1$, Figure 5.5 shows six sample paths of η plotted over time.[926]

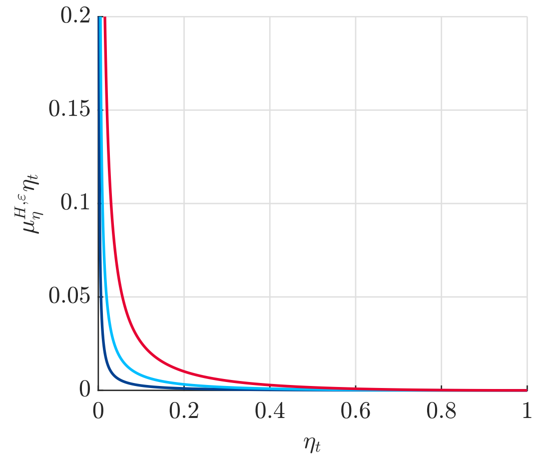
923 See Brunnermeier and Sannikov (2016, pp. 1509f.).
924 See Brunnermeier and Sannikov (2016, p. 1510).
925 Note that $d\eta_t \equiv 0$ for $\eta \equiv 1$.
926 It is assumed that time is measured in years and that there is one Brownian shock each day, i.e., $dt = 1/360$. Doing so is in line with Di Tella (2017, p. 2079) with the difference that the Brownian



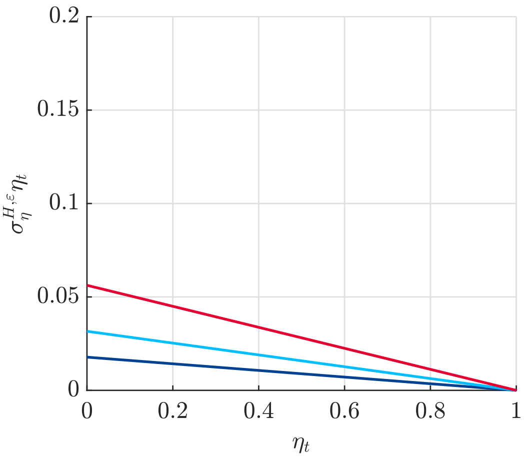
a) Drift of η with $H = 0.25$



b) Volatility of η with $H = 0.25$



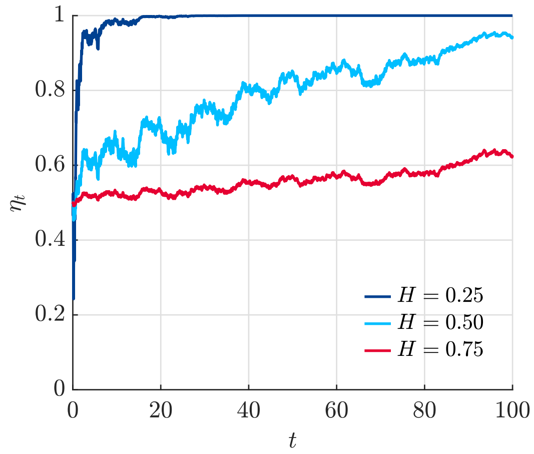
c) Drift of η with $H = 0.75$



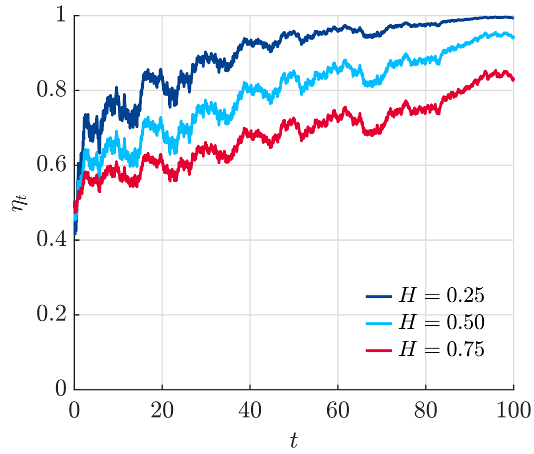
d) Volatility of η with $H = 0.75$

Figure 5.4: Drift and volatility of η as functions of η_t for various values of ε. The remaining parameters in (5.37) are $\sigma = 0.1$ and $H = 0.75$ in Panels a) and b) and $H = 0.25$ in Panels c) and d). The legend is the same in all panels.

As one would expect from Figure 5.3, all paths of η show a growing tendency due to the overall positive drift. From Panel a) of Figure 5.3, one can see that with $H < 1/2$, the specific path reaches the steady state more quickly than if $H > 1/2$. This mirrors the situation of Figure 5.3, where small values of H are associated with a large positive drift. By comparing Panels a) and b) of Figure 5.5, one can see that higher values of ε speed up the convergence to the steady state if $H > 1/2$ and slows down the convergence if $H < 1/2$, whereas the path of η is unaffected by changes of ε in the case of $H = 1/2$.



a) $\varepsilon = 10^{-3}$



b) $\varepsilon = 10^{-1}$

Figure 5.5: Sample paths of η for various values of H and ε. The paths of η are plotted over one hundred years, assuming that there is one daily shock, i.e., $dt = 1/360$. The remaining parameters are $\sigma = 0.1$, initial value $\eta_0 = 0.5$ and $\varepsilon = 10^{-3}$ in Panel a) and $\varepsilon = 10^{-1}$ in Panel b). In order to highlight the effects of parameter variation and, therefore, to exclude random effects, all six paths are generated by the same Brownian shock.

Regarding η, i.e., the wealth distribution of this model economy, one can see that long memory in the growth rates of output does not change the model outcomes qualitatively compared to the benchmark model with uncorrelated growth rates. In both cases, the steady state of η is equal to 1. However, in the case of positively correlated shocks, the tendency to the steady state is slowed down, i.e., the time needed for the experts to overwhelm the economy is increased.[927]

The reason for a more slowly growing experts' sector in the presence of long memory is that long memory appears to be risk-reducing. This property follows immediately from (5.10), where one can see that aggregate risk ($\sigma\varepsilon^{H-1/2}$) is smaller than in the Benchmark model (σ). Furthermore, if $\varepsilon \to 0$, i.e., the closer the process $Z^{H,\varepsilon}$ is a type II fBm, aggregate

shocks dW_t are drawn from a centered normal distribution with variance dt instead of a binomial distribution.

927 This can also be seen in Panel a) of Figure 5.6 (below) that plots the average of 10^5 paths of η over time. There, it can be seen clearly that η grows much slower on average if $H > 1/2$.

risk vanishes and, according to (5.37), experts' wealth share would remain constant.

Such behavior is obtained in various continuous-time model contexts involving long memory shocks stemming from fractional Brownian motion or related processes. For example, Dung (2013) considers a Black-Scholes market in which the stock price evolves like (5.9). In his fractional Black-Scholes model, he derives the pricing formula for a European call option. He shows that it corresponds to the original formula when the volatility σ is replaced with $\sigma\varepsilon^{H-1/2}$.[928] The same result was derived by Cheridito (2001), who uses a mixed process of a Brownian motion and a fractional Brownian motion in his Black-Scholes model.[929] To be more precise, he uses the process $\sigma(B_t^H + \varepsilon W_t)$ as noise in his model, where B_t^H is a type I fractional Brownian motion and W_t is an ordinary Brownian motion independent of B_t^H. Like the process $Z^{H,\varepsilon}$, Cheridito's process is closer to a fractional Brownian motion the closer ε is to zero.[930] A reason for this risk-reducing behavior may be that through the correlations of the shocks, the evolution of the capital stock (or the stock price in the mentioned Black-Scholes model) becomes predictable to a certain degree.[931] By exploiting this predictability, experts face smaller risks from holding the capital stock, thereby reducing the risk premia they can earn from their capital holdings.[932]

By rearranging (5.35), one obtains

$$\frac{\left(\varepsilon^{H-1/2}\sigma\right)^2}{\eta_t} = \rho + \Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon} - r_t. \tag{5.40}$$

The right-hand side of (5.40) is the excess return of capital over the risk-free rate, which is, for each value of η, smaller than in the benchmark model if there is long memory in the growth rates of the output. Consequently, the rewards from holding capital decrease in the presence of long memory, and experts' wealth share η tends to 1 much slower than in the benchmark model.

One might expect that the price of capital depresses when its excess return goes down. However, in this simple model, where experts hold the complete capital stock, capital is not traded; thus, the price of capital remains constant. Therefore, the risk-free rate has to rise to generate a decreasing excess return on capital.

928 See Dung (2013, Theorem 4.1 on pp. 343f.).
929 See Cheridito (2001, p. 933).
930 See Cheridito (2001, p. 933) for more details.
931 See Rostek and Schöbel (2013, p. 30).
932 See Rostek and Schöbel (2013, p. 31).

In the benchmark model where $H = 1/2$, (5.40) becomes

$$\frac{\sigma^2}{\eta_t^{\mathrm{BS}}} = \rho + \Phi(\Psi(1/q)) - \delta - r_t^{\mathrm{BS}}. \tag{5.41}$$

Thus, as in the long memory model, a reduced excess return can only be achieved by increasing the interest rate, since the price of capital and thus the optimal re-investment rate and the return on capital are constant.

With $H \neq 1/2$, the drift of capital depends directly on t and on the shocks up to time t through $\varphi_t^{H,\varepsilon}$. That is, the value of the risk-free rate depends on the concrete realization of the stochastic shocks. So, for example, assume that a sequence of positive shocks hit the economy. These shocks induce a rising wealth share of experts η_t. An increase in η_t, in turn, has an increasing effect on the interest rate r_t as can be seen from (5.36). At the same time, the positive shocks accumulate in $\varphi^{H,\varepsilon}$ and raise the drift and the return of capital that, in turn, affects the interest rate positively. Overall, there has to be a decline in the excess return since η increases through the series of positive shocks.

Given the sequence of positive shocks, one might expect the risk-free rate to be higher than in the benchmark model to compensate for the increasing return on capital caused by an increase in $\varphi^{H,\varepsilon}$. However, the concrete value of the risk-free rate depends on the concrete realization of the shocks, i.e., even if two different shock realizations would lead to the same value of η_t, the corresponding value of the risk-free rate may be different because the history of shocks accumulates differently in $\varphi^{H,\varepsilon}$. Thus, in contrast to the benchmark model, an increasing risk-free rate cannot be deduced from an increasing η.

Since $\mathbb{E}\varphi_t^{H,\varepsilon} = 0$, one obtains from (5.36) that the expected value of r_t is given by[933]

$$\mathbb{E}r_t = \rho + \Phi(\Psi(1/q)) - \delta - (\sigma\varepsilon^{H-1/2})^2\, \mathbb{E}\left(1/\eta_t\right). \tag{5.42}$$

From Panel a) of Figure 5.5, one would expect that in the presence of long memory, η is smaller on average than in the benchmark model. On the one hand, one would expect the term $\mathbb{E}\left(1/\eta_t\right)$ in (5.42) to be higher in the long memory model. However, due to $\sigma\varepsilon^{H-1/2} < \sigma$ in the case of $H > 1/2$, η only slightly affects the expected interest rate in (5.42). That is, the smaller multiplier $\sigma\varepsilon^{H-1/2}$ tends to increase the interest rate in (5.42).[934] In addition, for the same reason, one would expect the interest rate to be less

933 Note that $\mathbb{E}\varphi_t^{H,\varepsilon} = 0$ from the properties of the Itô integral. A similar argument is used for $\mathbb{E}Z_t^{H,\varepsilon} = 0$ in Appendix C.2.5.

934 More precisely, given the parameters of Figure 5.6, the term $\sigma\varepsilon^{H-1/2}$ is approximately equal to 3×10^{-4}, and thus rather small.

volatile on average in the long memory model since the expected value of r_t is less affected by variations in η.

Figure 5.6 shows the results of a Monte Carlo simulation illustrating that the interest rate increasing effect of the lower multiplier outweighs the interest rate reducing effect of an overall lower value of η_t. Panel a) of Figure 5.6 shows an approximation of $\mathbb{E}\eta_t$ as a function of time. As one would expect from the single realization of Panel a) of Figure 5.5 and the overall depressed drift shown in Panel a) of Figure 5.3, η reaches its steady state value more slowly in the presence of long memory. Furthermore, the reduced variability of the risk-free rate r_t can be seen clearly in Panel b) of Figure 5.6, which shows that the volatility effect $\sigma\varepsilon^{H-1/2}$ eliminates almost all of the variation in the expected interest rate caused by η.

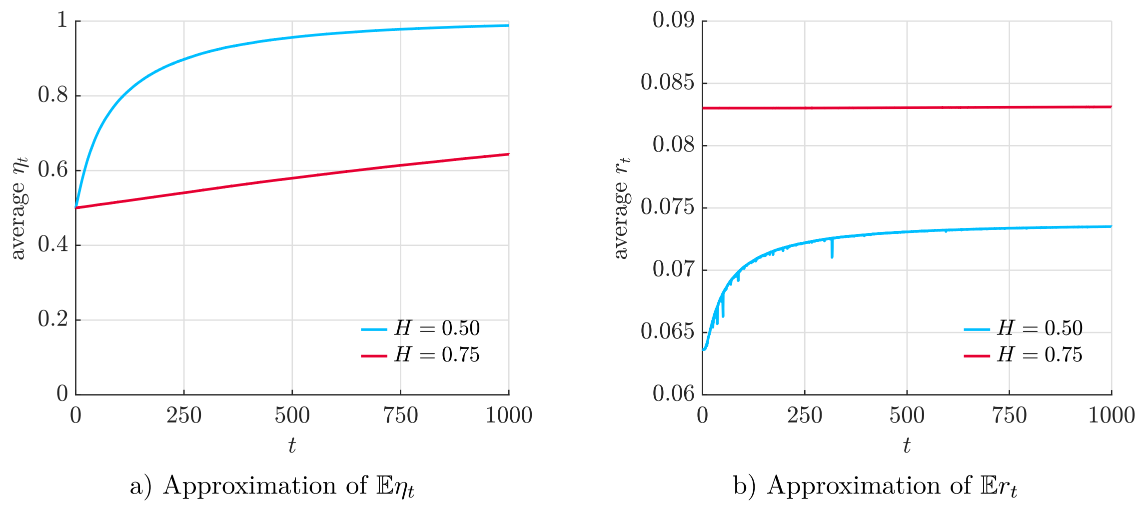


Figure 5.6: Average of η_t and r_t as function of t. In order to approximate $\mathbb{E}1/\eta_t$, the average of 10^5 simulations of (5.37) over 1000 years was taken. Then, by averaging over the $1/\eta$ and plugging it into (5.42), one obtains an approximation of $\mathbb{E}r_t$. Some outlier realizations for which η became negative (36 in the case of $H = 0.5$ and 3 in the case of $H = 0.75$) were dropped from the averaging. The initial value $\eta_0 = 0.5$ is identical for all realizations and $\varepsilon = 10^{-3}$. The remaining parameters are $\sigma = 0.1$, $\rho = 0.05$, $\Phi(\iota) = \log(\kappa\iota + 1)/\kappa$ with $\kappa = 10$. For the optimal re-investment rate yields $\iota = \Psi(1/q) = 0.04$, since $q = (a\kappa + 1)/(\rho\kappa + 1)$ with productivity parameter $a = 0.11$. Note the different scaling of the vertical axis.

The differences between the benchmark model and the model entailing long memory in the growth rates of output may further be seen in the special case of $\varepsilon = 1$.[935] In this case, the evolution of η and η^{BS} coincide, and the excess return from holding capital in both models is the same for each value of η.[936] Equating (5.40) with (5.41) yields

$$r_t = r_t^{\text{BS}} + \sigma\varphi_t^{H,\varepsilon} \text{ or } \mathbb{E}r_t = \mathbb{E}r_t^{\text{BS}}, \tag{5.43}$$

935 Recall that this case is far from choosing small values of ε in order to view $Z^{H,\varepsilon}$ as an approximation of a type II fBm.

936 This follows from (5.35), which states that $\hat{\vartheta}_t = \sigma/\eta_t$ in both models.

i.e., the risk-free interest rate of the model with long memory fluctuates around that of the benchmark model while maintaining the mean value. Equation (5.43) underlines that, even in the case of identical wealth distribution, the interest rates between the two models can be different. Ultimately, these differences are caused by the different evolution of aggregate capital stocks in the two models.

The differences in the evolution of the aggregate capital stock are not present in the evolution of the wealth distribution but affect the evolution of the interest rate as it depends closely on the return on capital. At the end of this section, the evolution of the aggregate capital stock is considered in more detail. As will turn out, introducing long memory into the growth rates of the capital stock leads, on average, to a more rapidly accumulating capital stock. Moreover, it follows from Lemma 5.2.1 that the aggregate capital stock in the equilibrium evolves according to

$$K_t = K_0 \exp\left(\left(\Phi\left(\Psi(1/q)\right) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2}\right)t + \sigma Z_t^{H,\varepsilon}\right). \tag{5.44}$$

The following lemma derives the expected value of the aggregate capital stock.

Lemma 5.4.1

Let the aggregate capital stock be given by (5.44)*, where* K_0 *is its given and positive initial value, then*

$$\mathbb{E}K_t = K_0 \exp\left(\left(\Phi\left(\Psi(1/q)\right) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2}\right)t + \frac{\sigma^2}{4H}\left((t+\varepsilon)^{2H} - \varepsilon^{2H}\right)\right) \tag{5.45}$$

Proof

See Appendix C.2.5 □

With $H = 1/2$, (5.45) reduces to $\mathbb{E}K_t^{\text{BS}} = K_0^{\text{BS}}\exp\left(\left(\Phi\left(\Psi(1/q)\right) - \delta\right)t\right)$, i.e., the capital stocks' expected value grows at rate $\left(\Phi\left(\Psi(1/q)\right) - \delta\right)$ which is the net investment rate minus the depreciation rate. In the case of long memory ($H > 1/2$), the capital stock grows faster on average, implying that the total wealth in the economy qK_t also grows faster on average.

Furthermore, in the case of $H < 1/2$, the asymptotic properties of (5.45) as $t \to \infty$ are determined by the term

$$\Phi\left(\Psi(1/q)\right) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2}. \tag{5.46}$$

That is, with $H < 1/2$, the expected value of the capital stock will grow as long as (5.46) is positive. On the other hand, $\mathbb{E}K_t \to 0$ as $t \to \infty$ if (5.46) is negative.

One can argue that shocks can erode the capital stock and output and, thus, create (temporarily) a decreasing total wealth in the economy. However, the evolution of capital and production should, from an economic perspective, not be constructed as having a decreasing tendency on average. Therefore, by imposing a non-negative condition on (5.46), one obtains

$$H \geq \frac{\log\left(\sigma^{-1}\sqrt{2\left(\Phi\left(\Psi(1/q)\right) - \delta\right)}\right)}{\log(\varepsilon)} + \frac{1}{2}. \tag{5.47}$$

Given the parameter values of Figure 5.5, (5.47) implies that $H \geq 0.36$ in the case of $\varepsilon = 10^{-3}$ and $H \geq 0.09$ in the case of $\varepsilon = 10^{-1}$, i.e., condition (5.47) is not satisfied in Panel a) of Figure 5.5 for the case of $H = 0.25$.

In the case of $H < 1/2$, the excess return on capital would be higher for each value of η as can be seen from (5.40), and, as Panel a) of Figure 5.5 suggests, the wealth share of η reaches its steady state value very quickly. However, since (5.47) is not satisfied in this case, one arrives at a somewhat counter-intuitive situation: On the one hand, experts receive high risk premia for their capital holdings, thereby capturing the whole economy quickly, but on the other hand, the capital stock tends to erode over time. At the end of this process, one has a situation where experts own everything, but everything is equal to nothing since, on average, the capital stock and the economy's total wealth erode completely. Figure 5.7 illustrates this result. The expected values of the corresponding capital stocks given the same parameters as in Panel a) of Figure 5.5 are plotted over time. It can be seen that the capital stock grows faster on average in the case of long memory ($H > 1/2$) and erodes completely if $H = 0.25$ and $\varepsilon = 10^{-3}$ not satisfying (5.47).

The faster increase of the capital stock in the case of long memory is indeed substantial. From Figure 5.7, it can be seen that after fifty years, the final value of the capital stock with $H = 0.75$ is nearly three times higher than the corresponding value with $H = 1/2$.

Additionally, by letting $\varepsilon \to 0$, (5.47) implies that $H \geq 1/2$. That is, with $H < 1/2$, the capital stock decreases on average the closer the process $Z^{H,\varepsilon}$ is to a type II fBm. This relationship does not only hold in the context of this model but also, e.g., in a fractional Black-Scholes model, where the stock price dynamics are described by a stochastic differential equation similar to (5.9).[937] This raises concerns about whether

[937] Dung (2013, Theorem 4.1 on pp. 343f.) considers a Black-Scholes market in which the stock evolves like (5.9). Additionally, he derives the pricing formula for a European call option in his fractional Black-Scholes model. He shows that it corresponds to the original formula when the volatility σ is replaced with $\sigma\varepsilon^{H-1/2}$. This appears plausible as the original pricing formula does not depend on

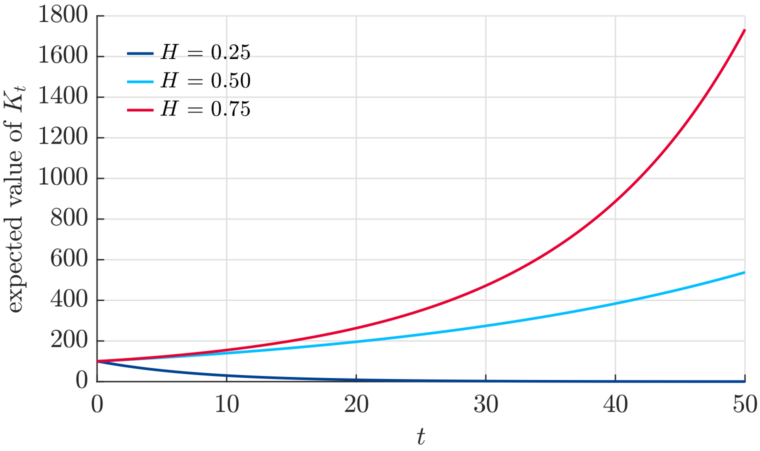


Figure 5.7: Expected value of the aggregate capital stock as a function of t for various values of H. The figure plots (5.45) as a function of time over 50 years for various values of H. The initial value K_0 is set to 100, the remaining parameters are as in Figure 5.6.

the process $Z^{H,\varepsilon}$ with $H < 1/2$ leads to meaningful results when used to model noise in economic variables that are typically assumed to grow on average. However, the parameter range of $H > 1/2$ appears to be more relevant than $H < 1/2$. In this case, the expected capital stock grows over time.

In summary, it was shown that long memory in the growth rates of output affects the variables in the model economy differently. Long memory affects the return on capital via the term $\varphi_t^{H,\varepsilon}$ (drift effect). In addition, long memory reduces the aggregate risk in the economy since for small values of ε and $H > 1/2$, $\sigma\varepsilon^{H-1/2} < \sigma$ (volatility effect). The wealth distribution in the economy depends only on the latter. Consequently, the benchmark model with volatility $\tilde{\sigma} = \sigma\varepsilon^{H-1/2}$ leads to the same wealth dynamics as in the model with $H \neq 1/2$, but, the evolution of the economy's total wealth may differ substantially. In the extreme case with $H < 1/2$ not satisfying (5.47) mentioned above, total wealth is expected to erode ultimately. In contrast, the expected value of the economy's total wealth does not depend on σ in the benchmark model, i.e., simply replacing σ with $\sigma\varepsilon^{H-1/2}$ leaves, on average, the evolution of the capital stock unchanged.[938] In the model with long memory ($H > 1/2$), the aggregate risk is reduced, but, at the same time, the economy's total wealth grows faster on average than in the benchmark model. Interestingly, this faster growing capital stock does not affect the wealth distribution. Instead, a different risk-free rate reflects the differing returns on capital. The reason for this may be found in the simple structure of the model. Since households cannot hold

the stock's drift. Nevertheless, if a corresponding condition such as (5.47) is not satisfied in such a setting, the stock's expected value is assumed to decrease over time.

938 Recall that $\mathbb{E}K_t^{\mathrm{BS}} = K_0^{\mathrm{BS}} \exp\left(\left(\Phi\left(\Psi(1/q)\right) - \delta\right) t\right)$.

capital, they must finance the expert's capital holdings by lending them at the risk-free rate. Therefore, the price of capital remains constant. In a more general setting, where capital trading between households and experts is permitted, the price of capital might also depend on $\varphi_t^{H,\varepsilon}$.[939]

As a result, the long memory model allows the evolution of aggregate wealth in the economy to be decoupled from the distribution of wealth. From an empirical point of view, this seems to be an interesting feature, since it is likely that inequality can be of the same (or similar) magnitude in both developing and developed countries. However, such a decoupling seems to contradict the structure of macro-financial models, as they typically focus on so-called Markov equilibria, where all model variables can be expressed as functions of a state variable (η).[940] These functional relationships are not preserved even in the rather simple model considered in this chapter. This can be seen, for example, by comparing the interest rates given in (5.39) and (5.36). In the former, the interest rate is a deterministic function of η, while in the latter it is not, since the drift effect $\varphi_t^{H,\varepsilon}$ prevents such a relationship.

Consequently, it seems questionable whether a Markov equilibrium in which all variables are functions of the state variable η can be established in a more general setting. Instead, the property of long memory seems to contradict Markov models.[941] Since the literature on continuous-time macro-financial models is overwhelmingly focused on Markov equilibria, further generalizations seem difficult and need to be brought out in future research. The following conclusive chapter elaborates on these arguments in a little more detail and suggests some possible avenues for future research.

---

[939] Such a model is considered for example in Brunnermeier and Sannikov (2016, Section 3.2 on pp. 1515ff.).

[940] See, e.g., Brunnermeier and Sannikov (2016, pp. 1519ff.).

[941] This is noted by Bollerslev et al. (2012, Footnote 19 on p. 47).

6

Conclusion

Die Wege zur Erkenntnis sind interessanter als die Erkenntnis selbst.
(Gottfried Wilhelm Leibniz)

This study was motivated by the following observation. On the one hand, there is empirical evidence of long memory in many economic time series. On the other hand, modern economic research builds strongly on stochastic models that involve, to a large extent, exogenous short memory stochastic processes to cover the uncertainty in the model. So this thesis aimed to integrate these two observations by allowing for long memory in the exogenous stochastic dynamics of the model. This was carried out in two representative models. The first model treated in Chapter 4 is a discrete-time real business cycle (RBC) dynamic stochastic general equilibrium (DSGE) model, and the second model, treated in Chapter 5, is a continuous-time macro-financial model. In the former, long memory was introduced by replacing the commonly used AR(1) shock to total factor productivity (TFP) with an ARFIMA$(1, d, 0)$ process. In the context of the continuous-time model, the exogenous driving process was assumed to be an approximation of Mandelbrot's fractional Brownian motion (fBm) in order to keep the model structure comparable to the benchmark model and to ensure the applicability of the Itô calculus. Following the solution of both models, the implications of long memory on the model outcomes were discussed and compared with the outcomes of the respective benchmark models.

Since both models, the discrete-time and the continuous-time model, have different implications and offer different paths for future research, this conclusion discusses the discrete-time and continuous-time framework separately, beginning with the discrete-time DSGE model.

The findings of Chapter 4 may be summarized as follows. Since autoregressive fractionally integrated moving average (ARFIMA) processes are not only long memory processes but also strongly persistent, they show an infinite cumulative impulse response (CIR), i.e., the cumulative impact of a shock is infinitely large. From a first point of view, one would have expected that the decisions of an infinitely-lived representative agent, who incorporates the underlying stochastic framework in his decision process by building his expectations rationally, are mostly affected by the presence of long memory. This reasoning is, in general, not valid. By considering a pure long memory ARFIMA$(0, d, 0)$ TFP shock, it turned out that the shapes of the model's impulse-response functions (IRFs) are quite similar to those of an AR(1) process with a small autoregressive parameter which is reasoned in the exponentially discounting behavior of the household. Since most of the IRF's "mass" is allocated at far distant periods, a substantial part of the cumulative effect is "discounted away" by the household before it emerges fully. This observation is in line with the permanent income hypothesis. It was shown that the response of the permanent income is much smaller in the presence of pure long memory than in the presence of short memory.

ARFIMA processes are characterized by a slowly decaying autocorrelation function (ACF), and the long memory parameter d controls the hyperbolic decay of both the ACF and IRF. A second critical remark was given in Lemma 2.4.3, illustrating that a long memory process might also have important short-term implications. More precisely, it was shown that the IRF of an ARFIMA$(1, d, 0)$ with parameters ϱ and d is equal to $\varrho + d$ in the first period after the shock. Therefore, the long memory parameter affects the model's response in the periods immediately after the shock in the same way as the short memory parameter ϱ. Consequently, when $\varrho + d > 1$, the TFP's IRF shows a hump shape, representing a technology shock with increasing impact in the periods following the shock.

As the household in the model expects this increasing shock impact, it adjusts its consumption and labor supply decisions accordingly. The model shows similar responses in the periods immediately after the shock, regardless of considering a TFP shock having a high ϱ and d or a permanent shock that pushes the economy on a new balanced growth path. As a consequence, it is less important for the household's initial response to know whether its income reaches a new long run equilibrium growth path or decays back to its old steady state value. In order to ensure a similar initial response, the income must increase at least for some periods.

Overall these considerations indicate that the asymptotic properties of long memory play a minor role in determining the model outcomes. Nevertheless, these results show that it is crucial to account for long memory as interesting short-term dynamics also result from long memory effects. From this perspective, it seems reasonable to put more effort into specifying the short-run dynamics, i.e., to consider higher-order autoregressive moving average (ARMA) specifications for rebuilding short-run dynamics adequately. However, as pointed out by Schorfheide (2011), including higher-order ARMA lags may increase the model fit, but this is somewhat arbitrary and raises questions regarding identification and interpretation.[942] From this perspective, fractionally integrated AR(1) processes offer the advantage of having only one additional parameter d to estimate, which further controls the long term dynamics of the process.

It may be owed to the simple model structure and the exogenous character of the long memory process that deeper insights into the functioning of the economy or policy implications cannot be drawn from the results derived in this thesis. Thus, this thesis has to be regarded as the first step to involve long memory dynamics in the context of stochastic economic models revealing various paths for future research. Two major directions that need further attention are estimation and generalization. This thesis aimed to show that one can solve a linearized DSGE model with an exogenous long memory process. Consequently, the thesis focused on how such a model can be solved. Whether such a model could be estimated was of secondary concern. Now, it is evident that the price for solving the long memory model is high. More specifically, the state space representation that is typically the starting point for estimating DSGE models no longer exists when the model is solved using the method described in Appendix B.5.[943] Consequently, the standard procedure for estimating DSGE models seems not directly applicable. Additionally, the state space representation allows for calculating the model-implied theoretical ACF of each model variable and the model variables' covariances.[944] Both the estimation of the model and the calculation of theoretical moments appear to be challenging in the presence of long memory. A possible path for accessing the moments implied by the long memory DSGE model is to run a Monte Carlo simulation of the model, i.e., one has to generate various trajectories of the model, then compute the moments of interest and average over the number of simulated trajectories. However, such a Monte Carlo analysis would introduce a kind of simulation bias.[945]

942 See Schorfheide (2011, p. 22).
943 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 633).
944 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 635).
945 See Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, pp. 642f.).

The estimation of a long memory DSGE model is of major empirical interest and can be motivated by the considerations made by Moretti and Nicoletti (2010). As outlined in Section 4.3, they show that long memory in the data-generating process can bias the estimates of a DSGE model's deep parameters.[946] Therefore, incorporating long memory into a DSGE model is likely to reduce this estimation bias. The pre-filtering of the data proposed by Moretti and Nicoletti (2010) implies that the agents in the model do not take into account the shock persistence that is removed by the filter. This is in contrast to the approach taken in this thesis. By incorporating long memory directly into the model structure, the household was able to account for the different types of persistence leading in some cases to different model responses.

A promising approach to estimating long memory, which might also be applicable in the context of a long memory DSGE model as the one discussed in Chapter 4, is due to Chan and Palma (1998). Chan and Palma (1998) show that an ARFIMA process with $d \neq 0$ has no finite-dimensional state space representation.[947] However, they show that given a finite sample of an ARFIMA(p, d, q) process, the corresponding likelihood function depends only on the first n entries of the state vector, where n refers to the sample size.[948] Since this may be computationally costly, they alternatively propose truncating the infinite moving average representation of an ARFIMA process. The resulting truncated process then has a finite-dimensional state space representation to which a maximum-likelihood estimator can be applied.[949] Grassi and Santucci de Magistris (2014) found, by carrying out a Monte Carlo analysis, an overall good performance of this estimation technique.[950] Andersson and Li (2020) extend the results of Chan and Palma (1998) and Grassi and Santucci de Magistris (2014) in order to allow the parameter d also spanning into the non-stationary region $d > 1/2$ and, additionally, they combine the state space estimator with a Bayesian approach.[951] However, all these approaches refer to the estimation of a univariate ARFIMA(p, d, q) model. To the best of the author's knowledge, whether and how these approaches can be generalized to a multivariate case such as the long memory DSGE model of Chapter 4 is an open question.

Another possible path for estimating a long memory DSGE model may be given by estimation techniques developed for vector-valued ARFIMA processes (so-called VARFIMA

[946] See Moretti and Nicoletti (2010, p. 22). Recall that Moretti and Nicoletti (2010) was already discussed in Section 4.3.
[947] See Chan and Palma (1998, Corollary 2.1. on p. 722).
[948] See Chan and Palma (1998, p. 725).
[949] See Chan and Palma (1998, p. 726).
[950] See Grassi and Santucci de Magistris (2014, p. 307).
[951] See Andersson and Li (2020, p. 539).

processes) as the one considered, e.g., by Abbritti et al. (2016, Section 2.2 and 3 on pp. 339ff.). Also, the approach proposed by Meyer-Gohde and Neuhoff (2015) seems promising. They solve a DSGE model given a general exogenous vector autoregressive moving average (VARMA) process without inflating the state space representation.[952] By considering the model in the frequency domain, they are able to calculate the model-implied correlation functions. Future research may generalize this approach to include exogenous ARFIMA or VARFIMA processes with long memory dynamics.

Additionally, this thesis paved the way for various generalizations. A first, rather technical question to be addressed in future research is the generalization to ARFIMA$(1, d, 0)$ processes with $d > 1/2$ and the incorporation of additional short memory components. In the following, some preliminary ideas and challenges appearing for $d > 1/2$ are outlined. Let $X = (X_t)_{t\in\mathbb{Z}}$ be an ARFIMA(p, d, q) process with $1/2 < d < 1$, then,

$$X_t = \sum_{k=1}^{t} Y_k, \tag{6.1}$$

where $Y = (Y_t)_{t\in\mathbb{Z}}$ is an ARFIMA$(p, d-1, q)$ process.[953] If the corresponding ARMA(p, q) process is stationary, so is Y; consequently, X is difference-stationary. The process X itself is no longer stationary and has infinite variance, but it can be shown that their corresponding IRF still converges to zero.[954] However, the solution method of Klein (2000) only requires the stability, not stationarity, of the exogenous process, i.e., the unconditional mean of the exogenous process has to be uniformly bounded.[955] That is, non-stationarity does not appear to be a problem at first glance. From (6.1), one would expect, though, that the unconditional mean of X is growing linearly over time, thereby violating Klein's stability condition.[956]

Nevertheless, by deriving the model's solution, the necessary condition to establish a unique solution was that the bubble solution vanishes, i.e., that (B.30) (restated in the following for convenience) holds[957]

$$\lim_{k\to\infty} \left(T_{22}^{-1} S_{22}\right)^k \mathbb{E}_t u_{t+k} = 0. \tag{B.30}$$

[952] See Meyer-Gohde and Neuhoff (2015, p. 91).
[953] See Hassler (2019, p. 115).
[954] See P. C. B. Phillips and Xiao (1998, p. 450).
[955] See (B.21) in Appendix B.5.
[956] An intuition for this may be seen from $\mathbb{E}|X_t| \leq \sum_{k=1}^{t} \mathbb{E}|Y_k| = t\mathbb{E}|Y_1|$, where the latter equality follows from the stationarity of Y.
[957] Recall that the vector u_t contains the auxiliary variables associated with the unstable subsystem. See Appendix B.5 for more details.

Clearly, Klein's stability condition appears to be a sufficient condition to ensure (B.30), but it doesn't appear to be necessary. More precisely, given that the eigenvalues of $T_{22}^{-1}S_{22}$ are smaller than one in modulus, the conditional expectations $\mathbb{E}_t u_{t+k}$ could grow at a polynomial rate in order to ensure (B.30) to hold. A hint that Klein's stability condition is too strong may also be deduced from the growth condition of Blanchard and Kahn (1980) given in (4.24). This condition allows the conditional expectations to grow at a polynomial rate which might be satisfied for ARFIMA processes with $d > 1/2$. In summary, to generalize the solution method proposed in this thesis to the non-stationary parameter range of $d > 1/2$, one has to check whether such an ARFIMA process satisfies Klein's stability condition. If Klein's stability condition holds for these processes, one shall be able to use the formulas derived in this thesis directly. If this condition is not satisfied, one may check if the weaker growth restriction (4.24) of Blanchard and Kahn (1980) is satisfied. If so, one has to check whether Klein's solution method leads to a well-defined solution given this weaker condition.

Obviously, to gain further economic insights, one has to consider a richer DSGE model involving more structure such as market imperfections and institutions such as a central bank controlling the nominal interest rate, etc. Such models involve various stochastic shocks, so extending the approach proposed in this thesis to models involving various long memory shocks with different long memory parameters would be interesting. Additionally, a New Keynesian setup would also allow for investigating more elaborate economic questions. For example, Andersson and Li (2020) propose the long memory parameter of the inflation rate as a measure of the flexibility of a central bank's inflation target (see Section 3.2.2.2). In a theoretical New Keynesian long memory DSGE model, one would, for example, be able to investigate the costs of disinflation in the presence of a more flexible inflation target.

Turning to the continuous-time model of Chapter 5, the implications are mixed. In order to preserve the applicability of the Itô calculus, fBm could not be used directly as the driving exogenous stochastic process in the model. Instead, an approximation of fBm was used. It turned out that the benchmark model could be solved accordingly, given the more general shock process. The reason for this was found in the structure of the exogenous shock that could be split into a drift effect and a volatility effect such that the evolution of the capital stock became

$$\frac{dK_t}{K_t} = \left(\Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon}\right)\,dt + \sigma\varepsilon^{H-1/2}\,dW_t.$$

Compared to the benchmark model, the drift effect is the additional term $\sigma\varphi_t^{H,\varepsilon}$ and the volatility effect is the multiplier $\varepsilon^{H-1/2}$ of σ. It may not appear surprising that all model variables independent of the capital stock's drift in the benchmark model were also independent of the drift effect induced by the long memory exogenous shock process. Exemplarily this can be seen, for example, in the law of motion of the state variable η that depends only on the volatility effect. In contrast, the equilibrium interest rate also depends on the drift effect. The reduced aggregate risk (σ reduces to $\sigma\varepsilon^{H-1/2}$ in the presence of long memory) arises as autocorrelations make the evolution of the risky capital more predictable and reduce thereby the risk premium experts can earn from their capital holdings. Consequently, it was shown that long memory in the growth rates of the capital stock and output induces η to reach its steady state value more slowly compared to the benchmark model. However, long memory does not change the model outcomes qualitatively, i.e., the steady state remains unchanged, and the fact that the experts' sector overwhelms the economy also remains true.

In addition, it was shown that the presence of long memory in the growth rates of the capital stock induces an, on average, faster growing total wealth in the economy than in the benchmark model. This is an interesting result as it allows to some extent, to decouple the evolution of an economy's total wealth from the wealth distribution. More precisely, both the benchmark model and the model with long memory can have the same dynamics of the wealth distribution. However, the corresponding macroeconomic evolution in terms of the evolution of the economy's total wealth may differ substantially. Such a feature may be interesting from an empirical point of view as large inequality may arise in both developing and industrial countries. This feature appears, however, to be a double-edged sword as it seems to be the core of more general models to express all variables in terms of certain state variables.[958] That is, macro-financial models focus on so-called Markov equilibria.[959] By expressing all variables as functions of a single state, the stochastic shocks are mapped into the evolution of the state variable, and then subsequently into the other variables through the posited functional relationship between the variable and the state variable.[960] Consequently, the remaining model variables depend on the exogenous stochastic shocks only through the state variable.

Even in the simple model considered in Chapter 5, such a structure cannot be obtained as the interest rate r_t depends on the additional drift component $\varphi_t^{H,\varepsilon} = \int_0^t (H - 1/2)(t - u + \varepsilon)^{H-3/2}\, dW_u$ and not solely on η as in the benchmark model. This simple model turned out

[958] See Brunnermeier and Sannikov (2016, p. 1519) or Brunnermeier and Sannikov (2014, p. 394).
[959] See Brunnermeier and Sannikov (2014, p. 394).
[960] See Brunnermeier and Sannikov (2014, p. 394).

to be solvable because the price of capital remained constant. In a more general setting, when for example, households are also allowed to hold capital, the price of the capital would not be constant any longer, and the drift of the assumed price process is likely to depend also on $\varphi_t^{H,\varepsilon}$.[961] Consequently, η_t is no longer the single channel through which the exogenous shocks can affect the model variables. At first glance, such a second channel does not seem problematic, since one could simply treat $\varphi_t^{H,\varepsilon}$ as a second state variable. Models with two state variables are not uncommon in the literature; see, e.g., Di Tella (2017). However, both state variables of Di Tella (2017) are described by Itô diffusions.[962] More specifically, recall from Appendix C.1 that an Itô process can be described by the following stochastic differential equation

$$dX_t = \mu_t^X \, dt + \sigma_t^X \, dW_t,$$

where μ_t^X and σ_t^X are stochastic processes. Then, X is called Itô diffusion if μ_t^X and σ_t^X are functions of X_t, i.e., $\mu_t^X = \mu^X(X_t)$ and $\sigma_t^X = \sigma^X(X_t)$. Additionally, both functions depend on time only through X_t.[963] This reflects the Markov structure of such models again. If one assumes instead that $\varphi_t^{H,\varepsilon}$ is a second state variable, it is not described as an Itô diffusion but rather as a time-weighted average of past Brownian shocks. Consequently, the applicability of this thesis's approach in more general models appears to be difficult as the Markov structure is not preserved given the more general shock process used in this thesis. In fact, this Markov structure seems to contradict the existence of long memory.[964]

However, an interesting part for future research may provide the aggregation results of Granger (1980) (and others) discussed in Section 3.3.1. More sophisticated continuous-time macro-financial models assume an infinite number of agents, each facing idiosyncratic risks and shocks.[965] However, when it comes to aggregation, these idiosyncratic risks are assumed to cancel out in the aggregate.[966] Given the results of Granger (1980), one would expect that aggregating over many micro units will affect the aggregate dynamics. So it may be reasonable to generate long memory endogenously in such a model by introducing a certain aggregation mechanism. At first glance, however, this seems to be at odds with

[961] A hint for this can be found in Brunnermeier and Sannikov (2014, p. 395), where the drift μ_t^q of the capital's price q_t is derived in a richer model. Since μ_t^q depends in this setting on the drift of the capital stock, likely, it will also depend on $\varphi_t^{H,\varepsilon}$ if one would have introduced long memory in the same way as carried out in Chapter 5.

[962] See Di Tella (2017, pp. 2044 and 2051) for the specification of the laws of motion of his state variables.

[963] See Øksendal (2013, Definition 7.1.1 on p. 116).

[964] Such reasoning was drawn by Bollerslev et al. (2012, p. 47) in the context of a stochastic volatility model.

[965] See, e.g., Brunnermeier and Sannikov (2014, p. 409) or Di Tella (2017, p. 2051).

[966] See, e.g., Brunnermeier and Sannikov (2016, p. 1534) or Di Tella (2017, p. 2045).

the often assumed elimination of idiosyncratic risk.

The approach followed by the author also reveals some drawbacks of using the shock-generating process $Z^{H,\varepsilon}$. For example, apparently, the parameter ε that measures how closely the shock-generating process is to a type II fBm is missing an economic interpretation. In order to mimic an fBm, small values of ε should be chosen. However, the effect of a variation in ε seems not negligible. An alternative specification may be the continuous-time processes proposed by Comte and Renault (1996), although these processes can preserve the Markov structure neither.

As this dissertation comes to an end, one has to say that the considered classes of well-established economic models seem not very receptive to stochastic long memory dynamics. The assumption of short memory processes in the DSGE literature and the assumption of independent-increment processes or Markov structures in the continuous-time macro-financial literature appears to be strong and restrictive.

A

Appendix to Chapter 2

A.1 Proofs

A.1.1 Proof of Lemma 2.3.2

Let $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$.

Part i): As $\sum_{k=0}^{n} \psi_k$ tends to infinity as $n \to \infty$, it follows from (2.26) that $\sum_{k=0}^{n} \gamma_X(k)$ also tends to infinity as $n \to \infty$. Since

$$\sum_{k=0}^{n} |\gamma_X(k)| \geq \sum_{k=0}^{n} \gamma_X(k) \to \infty, \text{ as } n \to \infty,$$

X is a long memory process.

Part ii): This is the contraposition of i). It can be deduced directly from

$$\sum_{k=0}^{\infty} \gamma_X(k) \leq \sum_{k=0}^{\infty} |\gamma_X(k)| < \infty.$$

Since $\sum_{k=0}^{\infty} \gamma_X(k) \geq 0$ (see Footnote 86) and $\sigma > 0$, it follows from (2.26) that X is either anti-persistent or moderately persistent.

Part iii): Without loss of generality, assume (2.28) to hold. Then, it follows again from (2.26) that $\sum_{k=0}^{n} \psi_k$ tends to infinity as $n \to \infty$.

Part iv): Consider the moving average coefficients $\psi_k = 1/(1+k)$. Then, it can be shown that $\sum_{k=0}^{n} \psi_k$ tends to infinity as $n \to \infty$. At the same time, the autocovariances are not summable and decay asymptotically with $\gamma_X(k) \sim \log(k)/k$ as $k \to \infty$.[967] Hence, X is strongly persistent and by i) a long memory process. Assume that there exist a constant $C > 0$ and $0 < d < 1/2$ such that $\gamma_X(k) \sim Ck^{2d-1}$ as $k \to \infty$. Then, $Ck^{2d-1}/(\log(k)/k)$ has to tend to 1 as $k \to \infty$. However, by the rule of l'Hospital, it yields

$$\lim_{k\to\infty} \frac{Ck^{2d-1}}{\log(k)/k} = \lim_{k\to\infty} \frac{Ck^{2d}}{\log(k)} = \lim_{k\to\infty} 2Cdk^{2d} = \infty.$$

Thus, (2.27) cannot hold. The autocovariances finally decay faster than Ck^{2d-1} but slow enough to meet the long memory condition.

Part v): See Hassler (2019, Example 3.3, Example 3.4, Example 3.5 on pp. 41ff.) □

A.1.2 Proof of Lemma 2.4.2

In order to calculate the impulse-response function (IRF) of Y, one has to derive its infinite moving average representation, i.e., one has to find a series $(\psi_n)_{n\in\mathbb{N}_0}$ such that

$$X_t = \sum_{n=0}^{\infty} \psi_n \varepsilon_{t-n}.$$

Out of (2.34), one can rewrite

$$X_t = \frac{1}{(1-\varrho B)}(1-B)^{-d}\varepsilon_t.$$

Regarding the autoregressive part, the time series expansion is given by

$$f(x) = \frac{1}{1-\varrho x} = \sum_{k=0}^{\infty} \varrho^k x^k,$$

if $|x| < 1$. Using the binomial series leads to the time series expansion[968]

$$g(x) = (1-x)^{-d} = \sum_{k=0}^{\infty} \binom{k+d-1}{k} x^k = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} x^k,$$

[967] See Hassler (2019, Example 3.5 on pp. 43ff.).

[968] See, e.g., Granger and Joyeux (1980/2001, p. 324).

if $|x| < 1$. Calculating the Cauchy-product of f and g leads to

$$f(x)g(x) = \frac{1}{1-\varrho x}(1-x)^{-d} = \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n}\varrho^i \frac{\Gamma(n-i+d)}{\Gamma(n-i+1)\Gamma(d)}\right)x^n. \tag{A.1}$$

This implies,

$$\begin{aligned} X_t = \frac{1}{1-\varrho B}(1-B)^{-d}\varepsilon_t &= \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n}\varrho^i \frac{\Gamma(n-i+d)}{\Gamma(n-i+1)\Gamma(d)}\right)B^n\varepsilon_t \\ &= \sum_{n=0}^{\infty}\left(\sum_{i=0}^{n}\varrho^i \frac{\Gamma(n-i+d)}{\Gamma(n-i+1)\Gamma(d)}\right)\varepsilon_{t-n} \\ &= \sum_{n=0}^{\infty}\psi_n\varepsilon_{t-n}. \end{aligned}$$

The second identity stated in the lemma follows directly by repeating the steps for $g(x)f(x)$ in (A.1) onward. □

A.1.3 Proof of Lemma 2.4.3

Part i): By Lemma 2.4.2, one knows that $\psi_0 = 1$ and

$$\psi_k = \sum_{i=0}^{k}\alpha_i\varrho^{k-i} = \alpha_k + \sum_{i=0}^{k-1}\alpha_i\varrho^{k-i} = \alpha_k + \varrho\sum_{i=0}^{k-1}\alpha_i\varrho^{k-1-i} = \alpha_k + \varrho\psi_{k-1}$$

Part ii): It follows directly from part i) that

$$\psi_1 = \varrho\psi_0 + \alpha_1 = \varrho + \frac{\Gamma(1+d)}{\Gamma(2)\Gamma(d)} = \varrho + d,$$

where the second equality uses $\psi_0 = 1$ and the third equality uses $\Gamma(1+d) = d\,\Gamma(d)$ and $\Gamma(2) = 1$.

Part iii): This follows from Hassler and Kokoszka (2010, Proposition 2.1 on p. 1857), since $\varrho^n n^{1-d} \to 0$, as $n \to \infty$.

Part iv): From Hassler and Kokoszka (2010, Equation (1.1) on p. 1856), it yields

$$\alpha_k = \frac{k-1+d}{k}\alpha_{k-1} \text{ for } k \geq 1 \text{ and } \alpha_0 = 1. \tag{A.2}$$

Thus, $\alpha_1 = d > 0$. By (A.2), it follows $\alpha_k > 0$ for all $k \geq 0$. The same holds true for ψ_k as, according to Lemma 2.4.2, it is the sum of non-negative numbers (recall that $0 < \varrho < 1$ by assumption). This proves the IRF to be positive.

If $\varrho + d > 1$, it follows from part i) and ii) that $\psi_1 > \psi_0$, i.e., the IRF is initially increasing. Further, from part iii), it is evident that $\lim_{k\to\infty} \psi_k = 0$. This ensures that the IRF falls below its initial value in the long run. In order to establish the hump shape of the IRF, possibly prevailing up and downswings have to be ruled out. Thus, it remains to show that the IRF decreases monotonically once the decline has set in. This is shown subsequently.

From $\lim_{k\to\infty} \psi_k = 0$ follows that there has to be an integer N such that $\psi_N \geq \psi_{N+1}$.[969] It remains to show that $\psi_k \geq \psi_{k+1}$ for all $k \geq N$. This is done inductively. By definition of N, it yields $\psi_N \geq \psi_{N+1}$. Assume that $\psi_{k-1} \geq \psi_k$ holds true for an arbitrary value of $k > N$. It remains to show that $\psi_k \geq \psi_{k+1}$.

Part i) implies

$$\psi_{k+1} = \varrho\psi_k + \alpha_k.$$

Subtracting ψ_k from both sides and applying part i) again leads to

$$\psi_{k+1} - \psi_k = \varrho\left(\psi_k - \psi_{k-1}\right) + \alpha_k - \alpha_{k-1}.$$

By (A.2), one has

$$\psi_{k+1} - \psi_k = \varrho\left(\psi_k - \psi_{k-1}\right) + \frac{d-1}{k}\alpha_{k-1}$$

The right-hand side of the latter equation is non-positive, i.e., it implies $\psi_{k+1} \leq \psi_k$. This is true since $\varrho > 0, d - 1 < 0$ by assumption, $\psi_k - \psi_{k-1} \leq 0$ by induction hypothesis and $\alpha_{k-1} > 0$.

Altogether, if $\varrho + d > 1$, the IRF increases at the beginning and decreases monotonically to zero after at least one period, implying a hump shape. □

A.1.4 Proof of Lemma 2.5.3

Part i): It is enough to show that (2.35) is equal to $\min\{s,t\}$ for $H = 1/2$ due to the equivalent characterization of the Brownian motion mentioned after Definition 2.5.1. With $H = 1/2$, (2.35) becomes

$$\gamma_{B^{1/2}}(s,t) = \frac{1}{2}\left(s + t - |s-t|\right). \tag{A.3}$$

[969] The value N is the period when the decline of the IRF sets in. Such a value exists as otherwise, it yields $\psi_{k+1} > \psi_k$ for all $k \in \mathbb{N}$ and thus $\psi_k > \psi_0 = 1$. Obviously, this is a contradiction to $\psi_k \to 0$, $k \to \infty$.

Let $t \leq s$, then $\gamma_{B^{1/2}}(s,t) = 1/2(s+t-s+t) = t$. For $t \geq s$, it yields $\gamma_{B^{1/2}}(s,t) = 1/2(s+t-t+s) = s$. Consequently, (A.3) becomes

$$\gamma_{B^{1/2}}(s,t) = \begin{cases} t & \text{if } t \leq s \\ s & \text{if } t \geq s \end{cases} = \min\{s,t\}.$$

Part ii): Recall the definition of γ_{X^H}

$$\gamma_{X^H}(k) = \operatorname{cov}\left(X_n^H, X_{n+k}^H\right) = \mathbb{E}\left(X_n^H X_{n+k}^H\right) - \mathbb{E}\left(X_n^H\right)\mathbb{E}\left(X_{n+k}^H\right) \tag{A.4}$$

Plugging the definition of X_n^H into (A.4) and using $\mathbb{E}\left(X_n^H\right) = 0$, leads to

$$\begin{aligned}\gamma_{X^H}(k) &= \mathbb{E}\left[\left(B_n^H - B_{n-1}^H\right)\left(B_{n+k}^H - B_{n+k-1}^H\right)\right] \\ &= \mathbb{E}\left(B_n^H B_{n+k}^H\right) - \mathbb{E}\left(B_n^H B_{n+k-1}^H\right) - \mathbb{E}\left(B_{n-1}^H B_{n+k}^H\right) + \mathbb{E}\left(B_{n-1}^H B_{n+k-1}^H\right)\end{aligned}$$

Applying (2.35) implies

$$\begin{aligned}\gamma_{X^H}(k) = &\frac{1}{2}\left(n^{2H} + (k+n)^{2H} - k^{2H}\right) - \frac{1}{2}\left(n^{2H} + (k+n-1)^{2H} - |k-1|^{2H}\right) \\ &- \frac{1}{2}\left(|n-1|^{2H} + (n+k)^{2H} - (k+1)^{2H}\right) + \frac{1}{2}\left(|n-1|^{2H} + (k+n-1)^{2H} - k^{2H}\right).\end{aligned}$$

After reshuffling, one finally arrives at

$$\gamma_{X^H}(k) = \frac{1}{2}\left(|k-1|^{2H} + (k+1)^{2H} - 2k^{2H}\right) \tag{A.5}$$

Since $\gamma_{X^H}(k)$ in (A.5) depends only on k but not on n, it is further emphasized that the increment process X^H is stationary according to Definition 2.1.2.

For $k > 1$ one can rewrite (A.5) and obtains[970]

$$\gamma_{X^H}(k) = \frac{1}{2}k^{2H}\left(\left(1-\frac{1}{k}\right)^{2H} + \left(1+\frac{1}{k}\right)^{2H} - 2\right).$$

Consider now the function $g(x) = (1-x)^{2H} + (1+x)^{2H} - 2$. Obviously, $g(1/k) = 2k^{-2H}\gamma_{X^H}(k)$. Therefore, the limiting behavior of $2k^{-2H}\gamma_{X^H}(k)$ as $k \to \infty$ is similar to the limiting behavior of $g(x)$ as $x \to 0$. By carrying out a second-order Taylor expansion

970 The following part of the proof is partly inspired by Beran (1994, p. 52).

of $g(\cdot)$ at the origin, one obtains

$$g(x) \sim 2H(2H-1)x^2, \text{ as } x \to 0.$$

Thus, it yields

$$2k^{-2H}\gamma_{X^H}(k) \sim 2H(2H-1)k^{-2}, \text{ as } k \to \infty. \tag{A.6}$$

The first part of the statement follows immediately from (A.6), i.e., as $k \to \infty$, it yields

$$\gamma_{X^H}(k) \sim H(2H-1)k^{2H-2}.$$

Since the series $\sum_{k=1}^{\infty} k^{2H-2}$ converges if and only if $2-2H > 1$ or $H < 1/2$, the increment process X^H is a short memory process if $H < 1/2$ and a long memory process if $H > 1/2$.[971]

With $H = 1/2$, the process is a short memory process as well. This follows from Part i) and Definition 2.5.1, since $B^{1/2}$ is a Brownian motion whose increments are stochastically independent. Consequently, $\gamma_{X^{1/2}}(k) = 0$ for all $k > 1$ and, thus, $\sum_{k=0}^{\infty} |\gamma_{X^{1/2}}(k)| < \infty$. □

A.2 On the Gamma and Gaussian Hypergeometric Function

This section briefly reviews the definitions of the gamma function and the Gaussian hypergeometric function. Both functions are needed, e.g., for the definition of the autoregressive fractionally integrated moving average (ARFIMA) processes in (2.29) or for expressing the key moments of an ARFIMA$(1, d, 0)$ process in Appendix A.3. Additionally, the Gaussian hypergeometric function is used for the derivation of the long memory dynamic stochastic general equilibrium (DSGE) model's solution in Appendix B.5.4.

A.2.1 The Gamma Function

The gamma function extends the well-known factorial $n!$ of a natural number n to general complex numbers. Let z be a complex number with $z \neq 0, -1, -2, \ldots$, then the gamma function $\Gamma(\cdot)$ can be defined by[972]

$$\Gamma(z) = \lim_{k\to\infty} \frac{k!k^{z-1}}{z(z+1)\cdots(z+k-1)}.$$

971 Hassler (2016, pp. 119f.) proves the convergence properties of $\sum_{k=1}^{\infty} k^{-p}$.

972 See G. E. Andrews et al. (1999, Definition 1.1.1 on p. 3).

This rather abstract formulation can be accessed better by considering the essential properties of the gamma function. The generalization of the factorial follows from the identity[973]

$$\Gamma(z+1) = z\Gamma(z) \qquad \text{(A.7)}$$
$$\Gamma(0) = 1.$$

For $z = n \in \mathbb{N}$ it follows from an iterative application of (A.7), that $\Gamma(n+1) = n!$[974] This generalization of the factorial is used in (2.30) to define the binomial coefficient for non-integer values.

A.2.2 The Gaussian Hypergeometric Function

Let $a, b, c, z \in \mathbb{C}$ with $|z| < 1$. The Gaussian hypergeometric function ${}_2F_1(a, b; c; z)$ is defined by[975]

$$\begin{aligned} {}_2F_1(a,b;c;z) &= \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+n-1)b(b+1)\cdots(b+n-1)}{c(c+1)\cdots(c+n-1)n!} z^n \\ &= 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{c(c+1)2!}z^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)3!}z^3 + \cdots . \end{aligned} \qquad \text{(A.8)}$$

Due to $a(a+1)\cdots(a+n-1) = \Gamma(a+n)/\Gamma(a)$, the Gaussian hypergeometric function can alternatively be expressed in terms of the gamma function, i.e.,[976]

$${}_2F_1(a,b;c;z) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+k)} \frac{z^k}{\Gamma(k+1)}.$$

As will turn out in the derivation of the IRFs of the long memory DSGE model, the following identity is useful to simplify some expressions.[977] Let $\lambda \in \mathbb{R}$ with $|\lambda| < 1$ and $\alpha_k = \Gamma(k+d)/(\Gamma(k+1)\Gamma(d))$. Then,

$$\sum_{k=0}^{\infty} \lambda^k \alpha_{t+k-1} = \sum_{k=0}^{\infty} \lambda^k \frac{\Gamma(t+k-1+d)}{\Gamma(t+k)\Gamma(d)}$$

[973] See G. E. Andrews et al. (1999, p. 3).
[974] See G. E. Andrews et al. (1999, p. 3).
[975] See G. E. Andrews et al. (1999, Definition 2.1.5 on p. 64).
[976] This follows from G. E. Andrews et al. (1999, p. 2) together with (A.7).
[977] See the derivation of (B.48) in Appendix B.

$$= \frac{\Gamma(t-1+d)}{\Gamma(t)\Gamma(d)} \sum_{k=0}^{\infty} \frac{\Gamma(k+1)\Gamma(t+k-1+d)\Gamma(t)}{\Gamma(1)\Gamma(t-1+d)\Gamma(t+k)} \frac{\lambda^k}{\Gamma(k+1)}$$

$$= \alpha_{t-1}\ {}_2F_1(1, t-1+d; t; \lambda), \tag{A.9}$$

where the second line uses $\Gamma(1) = 1$.

For the calculations made in this thesis, the Gaussian hypergeometric function is computed numerically by using the series expansion (A.8) instead of using the inbuilt Matlab function *hypergeom*. The Matlab function appeared to be unstable and slow for large values of t in (A.9).[978] Therefore, Pearson et al. (2017) investigate various methods to calculate the Gaussian hypergeometric function and found an overall good performance of the series approximation (A.8).[979,980]

For the purposes of this thesis, the series approximation of (A.8) mentioned in Pearson et al. (2017) seems satisfactory. In order to give an example, a value like ${}_2F_1(1, d+n; 1+n; 0.93)$ has to be calculated for $n = 1, 2, \ldots, 140$ in order to derive the IRF of the long memory DSGE model over 140 periods.[981] The inbuilt Matlab function computes these values in 2.715 seconds and the code used by the author needs just 0.0134 seconds. Furthermore, the maximum absolute deviation between both methods is of size 3.6×10^{-14} that appears to be sufficiently small. Additionally, the more periods are simulated, the more striking the computational advantages of the procedure followed by the author compared to the inbuilt Matlab function.

978 That the inbuilt Matlab function may not be reliable for all parameter values is also stated in Pearson et al. (2017, p. 824).

979 See Pearson et al. (2017, pp. 841ff.).

980 The code used for the calculation of the Gaussian hypergeometric function in this thesis is inspired by the one of Garcia (2023), URL in list of references. The code of Garcia (2023), URL in list of references, was rewritten and augmented by an additional stopping criterion proposed by Pearson et al. (2017, p. 841) that breaks the approximation if the absolute value of the next summand in (A.8) divided by the already calculated sum is smaller than 10^{-16}. Moreover, as in the original code, the user can specify a total number of summands to which the approximation should be made. For all calculations within this thesis, the maximal number of summands was chosen to be 10.000. However, this boundary was never reached for the parameter values considered in this thesis as the additionally implemented stopping criterion was binding.

981 See (B.48).

A.3 Some Moments of the ARFIMA(1,d,0) Process

This section provides the closed-form expressions of an ARFIMA$(1, d, 0)$ process's second-order moments considered in Section 2.4.2. Let $X = (X_t)_{t\in\mathbb{Z}}$ be an ARFIMA$(1, d, 0)$ process as in Section 2.4.2. Then, the autocovariance function is given by[982]

$$\gamma_X(k) = \sigma_\varepsilon^2 \frac{(-1)^k \Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)} \times \frac{{}_2F_1(1, d+k; 1-d+k; \varrho) + {}_2F_1(1, d-k; 1-d-k; \varrho) - 1}{1-\varrho^2}. \quad \text{(A.10)}$$

The corresponding autocorrelation function (ACF) is given by[983]

$$\rho_X(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} \frac{{}_2F_1(1, d+k; 1-d+k; \varrho) + {}_2F_1(1, d-k; 1-d-k; \varrho) - 1}{(1-\varrho)\, {}_2F_1(1, 1+d; 1-d; \varrho)}. \quad \text{(A.11)}$$

From these two equations, the variance of X and the first order autocorrelation of X is found to be[984,985]

$$\gamma_X(0) = \frac{\sigma_\varepsilon^2 \Gamma(1-2d)}{\Gamma(1-d)^2} \frac{{}_2F_1(1, 1+d; 1-d; \varrho)}{(1+\varrho)} \quad \text{(A.12)}$$

and

$$\rho_X(1) = \frac{(1+\varrho^2)\, {}_2F_1(1, d; 1-d; \varrho) - 1}{\varrho(2\, {}_2F_1(1, d; 1-d; \varrho) - 1)}. \quad \text{(A.13)}$$

[982] See Hosking (1981, Lemma 1 on p. 172).
[983] See Hosking (1981, Lemma 1 on p. 172).
[984] See Hosking (1981, Lemma 1 on p. 172).
[985] Note that these equations can be proven to be equivalent to the ones given in Palma (2007, p. 47) if $\varrho = 0$. These are just different representations induced by the properties of the gamma function. Note that these equations are equivalent to (A.10) and (A.11) for $k = 0$. The different forms are induced by the properties of the Gaussian hypergeometric function and the gamma function.

B

Appendix to Chapter 4

B.1 Derivation of the Model's Equilibrium

Equations (4.10) to (4.13) are derived from the household's and firm's maximization problem (4.3) and (4.2), respectively. The Lagrangian of the household's maximization problem is given by

$$\mathcal{L} = \mathbb{E}_t \sum_{s=t}^{\infty} \beta^{s-t} \left\{ U\left(C_s, H_s\right) - \lambda_s \left[C_s + K_{s+1} - (1-\delta)K_s - W_s H_s - R_s K_s \right] \right\},$$

where λ_t is the Lagrange multiplier. Then, the first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial C_t} = U_{C,t} - \lambda_t \equiv 0 \tag{B.1}$$

$$\frac{\partial \mathcal{L}}{\partial H_t} = U_{H,t} + \lambda_t W_t \equiv 0 \tag{B.2}$$

$$\frac{\partial \mathcal{L}}{\partial K_{t+1}} = -\lambda_t + \beta \mathbb{E}_t \left[\lambda_{t+1} \left(1 - \delta + R_{t+1}\right)\right] \equiv 0. \tag{B.3}$$

Now, (4.10) follows directly from (B.1) and (B.2). The Euler equation (4.11) follows directly from (B.1) and (B.3).

By maximizing the firm's objective function $\Pi_t = P_t Y_t - W_t H_t - R_t K_t$ with respect to K_t

and H_t, the first order conditions are given by

$$\frac{\partial \Pi_t}{\partial H_t} = P_t \frac{\partial Y_t}{\partial H_t} - W_t \equiv 0 \tag{B.4}$$

$$\frac{\partial \Pi_t}{\partial K_t} = P_t \frac{\partial Y_t}{\partial K_t} - R_t \equiv 0. \tag{B.5}$$

By the production function (4.1), the marginal products of labor and capital are given by

$$\frac{\partial Y_t}{\partial H_t} = (1-\alpha)\frac{Y_t}{H_t}, \quad \frac{\partial Y_t}{\partial K_t} = \alpha \frac{Y_t}{K_t},$$

respectively. Equation (4.12) and (4.13) follow directly from (B.4), (B.5) and (4.8).

B.2 Frisch Elasticity of the Additive Separable Utility Function

Here, it is derived that the Frisch elasticity of the utility function specified in (4.18) is given by $1/\varphi$. Recall the definition of the Frisch elasticity mentioned below of (4.18). The Frisch elasticity given the level $\bar{U}_C$ of the marginal utility of consumption is given by[986]

$$\frac{\partial H_t}{\partial W_t} \frac{W_t}{H_t}\Big|_{U_{C,t}=\bar{U}_C}. \tag{B.6}$$

From (4.10), one can derive labor supply in the equilibrium, i.e.,

$$C_t^{-\varsigma} W_t = \varkappa H_t^{\varphi} \text{ or } H_t = \left(\frac{C_t^{-\varsigma} W_t}{\varkappa}\right)^{1/\varphi}. \tag{B.7}$$

It follows directly from (B.7) that

$$\frac{\partial H_t}{\partial W_t}\Big|_{U_{C,t}=\bar{U}_C} = \frac{\bar{U}_C}{\varkappa\varphi}\left(\frac{\bar{U}_C W_t}{\varkappa}\right)^{-1+1/\varphi} = \frac{1}{\varphi}\frac{H_t}{W_t}. \tag{B.8}$$

Inserting (B.8) into (B.6) illustrates that the Frisch elasticity is given by $1/\varphi$.

[986] This is adapted from Furlanetto and Seneca (2014, p. 115).

B.3 Summary of the Model with Additive Utility Function

B.3.1 The Nonlinear Model Equations

The model with additive separable utility function specified in (4.18) is summarized in Table B.1. As the utility function is incompatible with a balanced growth path, it is abstracted from labor augmenting technical progress, i.e., $\bar{A}_t \equiv 1$. The exogenous process $(\varepsilon_t^A)_{t\in\mathbb{Z}}$ is a zero mean independent and identically distributed (i.i.d.) sequence.

Model equation		equation
$\varkappa C_t^{\varsigma} H_t^{\varphi} = W_t$	labor supply	(4.10)
$1 = \beta \mathbb{E}_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\varsigma} ((1-\delta) + R_{t+1}) \right]$	Euler equation	(4.11)
$K_{t+1} = (1-\delta)K_t + I_t$	capital accumulation	(4.5)
$L_t = 1 - H_t$	leisure choice	(4.6)
$Y_t = A_t K_t^{\alpha} H_t^{1-\alpha}$	production function	(4.1)
$K_t = \alpha \frac{Y_t}{R_t}$	capital demand	(4.13)
$H_t = (1-\alpha)\frac{Y_t}{W_t}$	labor demand	(4.12)
$Y_t = C_t + I_t$	goods market clearing	(4.7)
$(1-\varrho_A B)(\log(A_t) - \log(A_{ss})) = (1-B)^{-d}\varepsilon_t^A$	transitory TFP	(4.14)

Table B.1: Model equations of the nonlinear model with additive utility function. The right column shows the related equation in the text from which the model equations are derived. Some model equations are the same as their corresponding equation in the text; they are restated here to provide a complete overview of the model.

B.3.2 The Model's Steady State

Due to the simple model structure, closed-form expressions can be derived for the variables' steady state values. For the model summarized in Table B.1, they are given recursively in

Table B.2.

variable		steady state	variable		steady state
A_{ss}	$=$	exogenous parameter	I_{ss}	$=$	δK_{ss}
R_{ss}	$=$	$\frac{1}{\beta} - (1-\delta)$	H_{ss}	$=$	$(1-\alpha)\frac{Y_{ss}}{W_{ss}}$
W_{ss}	$=$	$A_{ss}^{1/(1-\alpha)} \left(\frac{\alpha}{R_{ss}}\right)^{\alpha/(1-\alpha)} (1-\alpha)$	L_{ss}	$=$	$1 - H_{ss}$
Y_{ss}	$=$	$\left[\frac{1}{\varkappa}\left(\frac{R_{ss}}{R_{ss}-\delta\alpha}\right)^{\varsigma} (1-\alpha)^{-\varphi} W_{ss}^{\varphi+1}\right]^{1/(\varphi+\varsigma)}$	C_{ss}	$=$	$Y_{ss} - I_{ss}$
K_{ss}	$=$	$\alpha\frac{Y_{ss}}{R_{ss}}$			

Table B.2: Steady state of the nonlinear model with additive utility function. The values are sorted recursively from top to bottom, starting from the left column.

Note that the steady state value of A_t is not determined by (4.14) and has to be set exogenously. To be more precise, (4.14) specifies how the deviation of the logarithm of A_t from its steady state value evolves over time without determining the steady state value itself. It is often implicitly assumed that $A_{ss} = 1$ in the literature. For example, Aguiar and Gopinath (2007) specify the transitory component of total factor productivity (TFP) according to[987]

$$A_t = \exp(z_t), \quad z_t = \varrho z_{t-1} + \varepsilon_t^A. \tag{B.9}$$

Equation (B.9) implies

$$\log(A_t) = z_t = \varrho z_{t-1} + \varepsilon_t^A. \tag{B.10}$$

In contrast to (4.14), (B.10), however, pins down the steady state value of $\log(A_{ss}) = 0$ or $A_{ss} = 1$.

B.3.3 The Linearized Model Equations

Table B.3 shows the linearized model equations. The linearization was taken around the steady state values of the model given in Table B.2.

[987] See, e.g., Aguiar and Gopinath (2007, Equations (1) and (2) on p. 78).

linearized Model equation		equation
$\varsigma\tilde{C}_t + \varphi\tilde{H}_t = \tilde{W}_t$	labor supply	(4.10)
$\frac{\varsigma}{\beta}\left(\mathbb{E}_t\tilde{C}_{t+1} - \tilde{C}_t\right) = R_{ss}\mathbb{E}_t\tilde{R}_{t+1}$	Euler equation	(4.11)
$\tilde{K}_{t+1} = (1-\delta)\tilde{K}_t + \delta\tilde{I}_t$	capital accumulation	(4.5)
$L_{ss}\tilde{L}_t = -H_{ss}\tilde{H}_t$	leisure choice	(4.6)
$\tilde{Y}_t = \tilde{A}_t + \alpha\tilde{K}_t + (1-\alpha)\tilde{H}_t$	production function	(4.1)
$\tilde{K}_t = \tilde{Y}_t - \tilde{R}_t$	capital demand	(4.13)
$\tilde{H}_t = \tilde{Y}_t - \tilde{W}_t$	labor demand	(4.12)
$Y_{ss}\tilde{Y}_t = C_{ss}\tilde{C}_t + I_{ss}\tilde{I}_t$	goods market clearing	(4.7)
$(1-\varrho_A B)\tilde{A}_t = (1-B)^{-d}\varepsilon_t^A$	transitory TFP	(4.14)

Table B.3: Linearized model equations of the model with additive utility function. The right column gives the corresponding equation in the text. The order of the equations is the same as in Table B.1. The involved steady state values are given in Table B.2.

B.3.4 Matrices of the Model's Canonical Form

Before the canonical form of the model can be given, the predetermined (or backward-looking) and non-predetermined (forward-looking) variables have to be specified. Out of Table B.3, there are two predetermined variables, namely the capital stock and the transitory TFP process, and two forward-looking variables, namely consumption and the real rental rate on capital. The other five variables are so-called static variables, which are linear combinations of predetermined and forward-looking variables. The static variables can be seen from Table B.3 as these variables occur only with time index t but without any lead or lag.[988]

Therefore, the first step is to eliminate them from the system, leaving only the predetermined and forward-looking variables. Then the model can be cast into the canonical

[988] See, for example, Adjemian et al. (2022, p. 72). Some models may also contain mixed variables that appear in the model equations at time $t, t+1$ and $t-1$. There are no such variables in the models considered in this thesis.

form needed for the solution method of Klein (2000), and the model can be solved. Afterward, the static variables are recovered from the formerly found model's solution and the postulated linear relationship between static and non-static variables.

To proceed as described, the model stated in Table B.3 can be written in the following matrix equation:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \varsigma/\beta & -R_{ss} & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbb{E}_t \begin{pmatrix} \tilde{A}_t \\ \tilde{K}_{t+1} \\ \tilde{C}_{t+1} \\ \tilde{R}_{t+1} \\ \tilde{H}_{t+1} \\ \tilde{L}_{t+1} \\ \tilde{W}_{t+1} \\ \tilde{I}_{t+1} \\ \tilde{Y}_{t+1} \end{pmatrix}$$

$$= \begin{pmatrix} \varrho_A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\delta & 0 & 0 & 0 & 0 & 0 & \delta & 0 \\ 0 & 0 & \varsigma/\beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 1-\alpha & 0 & 0 & 0 & -1 \\ 0 & 0 & \varsigma & 0 & \varphi & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & H_{ss} & L_{ss} & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & -C_{ss} & 0 & 0 & 0 & 0 & -I_{ss} & Y_{ss} \end{pmatrix} \begin{pmatrix} \tilde{A}_{t-1} \\ \tilde{K}_t \\ \tilde{C}_t \\ \tilde{R}_t \\ \tilde{H}_t \\ \tilde{L}_t \\ \tilde{W}_t \\ \tilde{I}_t \\ \tilde{Y}_t \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \nu_t^A \quad \text{(B.11)}$$

This system can be partitioned according to the number of non-static (i.e., predetermined and forward-looking) and static variables. Let the non-static variables be collected in the vector $x_{t+1} = \left(\tilde{A}_t, \tilde{K}_{t+1}, \tilde{C}_{t+1}, \tilde{R}_{t+1}\right)^T \in \mathbb{R}^{4\times 1}$ and the remaining static variables in the vector $y_{t+1} \in \mathbb{R}^{5\times 1}$. Then (B.11) can be written as

$$\begin{pmatrix} \mathbf{A} & 0_{4\times 5} \\ 0_{5\times 4} & 0_{5\times 5} \end{pmatrix} \mathbb{E}_t \begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} \mathbf{G} \\ 0_{5\times 1} \end{pmatrix} \nu_t^A, \quad \text{(B.12)}$$

where $\mathbf{A}, \mathbf{D}_{11} \in \mathbb{R}^{4\times 4}, \mathbf{D}_{12} \in \mathbb{R}^{4\times 5}, \mathbf{D}_{21} \in \mathbb{R}^{5\times 4}, \mathbf{D}_{22} \in \mathbb{R}^{5\times 5}$ and $\mathbf{G} \in \mathbb{R}^{4\times 1}$. Considering the last five rows of the system (B.12), one obtains

$$\mathbf{D}_{21} x_t + \mathbf{D}_{22} y_t = 0_{5\times 1}. \quad \text{(B.13)}$$

If $\mathbf{D}_{22}$ is invertible, (B.13) can be solved for y_t, i.e.,[989]

$$y_t = -\mathbf{D}_{22}^{-1}\mathbf{D}_{21}x_t. \tag{B.14}$$

Equation (B.14) states that the static variables stored in the vector y_t depend linearly on the predetermined and non-predetermined variables stored in the vector x_t.

What remains to do is to solve the first four rows of the system (B.12) in order to find an appropriate solution for the non-static variables. Plugging (B.14) back in (B.13) leads to

$$\mathbf{A}\mathbb{E}_t x_{t+1} = \left(\mathbf{D}_{11} - \mathbf{D}_{12}\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\right) x_t + \mathbf{G}\nu_t^A. \tag{B.15}$$

Equation (B.15) defines a system with predetermined and non-predetermined variables stored in the vector x_t. Further, this system can now be solved with the method of Klein (2000) as it fits the required canonical form stated in (B.20) with $\mathbf{B} = \left(\mathbf{D}_{11} - \mathbf{D}_{12}\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\right)$ and $z_t = \nu_t^A$.[990]

After the solution to (B.15) has been found, the solution to the whole system (B.11) is then found by solving for the static variables with (B.14).

B.3.5 On the Permanent Income

By linearizing (4.36), one obtains

$$\begin{aligned} Y_{ss}^p\tilde{Y}_t^p &= Y_{ss}\tilde{Y}_t + (1-\delta)K_{ss}\tilde{K}_t \\ &\quad + W_{ss}H_{ss}\mathbb{E}_t\sum_{j=1}^{\infty}(R_{ss} + (1-\delta))^{-j}\left(\tilde{Y}_{t+j} - \frac{R_{ss}}{R_{ss} + (1-\delta)}\sum_{i=1}^{j}(\tilde{R}_{t+i} + (1-\delta))\right) \\ &= Y_{ss}\tilde{Y}_t + (1-\delta)K_{ss}\tilde{K}_t + W_{ss}H_{ss}\mathbb{E}_t\sum_{j=1}^{\infty}\beta^j\left(\tilde{Y}_{t+j} - \beta R_{ss}\sum_{i=1}^{j}(\tilde{R}_{t+i} + (1-\delta))\right), \end{aligned} \tag{B.16}$$

where $Y_{ss}^P = Y_{ss} + (1-\delta)K_{ss} + W_{ss}H_{ss}\sum_{j=1}^{\infty}\beta^j = Y_{ss} + (1-\delta)K_{ss} + W_{ss}H_{ss}\beta/(1-\beta)$ refers to the steady state value of the permanent income around which the linearization was taken. Note that $R_{ss} + (1-\delta) = \beta^{-1}$ by the steady state relationships mentioned in

[989] As can be seen easily from (B.11), one has that $\det(\mathbf{D}_{22}) = I_{ss}L_{ss}(1+\varphi)$. Since the inverse of the Frisch elasticity φ is assumed to be positive, $\mathbf{D}_{22}$ is invertible as long as the steady state investments and the steady state time devoted to leisure are different from zero. From an economic point of view, these requirements seem not demanding.

[990] Since the first and the last row of $\mathbf{A}$ are linearly dependent (see (B.11)), $\mathbf{A}$ is not invertible and thus the method of Blanchard and Kahn (1980) is indeed not applicable to the system (B.15).

Table B.2. Further, dividing (B.16) by Y^p_{ss} results in

$$\tilde{Y}^p_t = \frac{Y_{ss}}{Y^p_{ss}}\tilde{Y}_t + \frac{(1-\delta)K_{ss}}{Y^p_{ss}}\tilde{K}_t + \frac{W_{ss}H_{ss}}{Y^p_{ss}}\mathbb{E}_t\sum_{j=1}^{\infty}\beta^j\left(\tilde{Y}_{t+j} - \beta R_{ss}\sum_{i=1}^{j}\tilde{R}_{t+i}\right). \tag{B.17}$$

From (B.17), the three contributions to the percentage deviation of the permanent income depicted in Figure 4.5 become evident. The contribution of current income is given by $Y_{ss}\tilde{Y}_t/Y^p_{ss}$, the contribution of the current capital stock is given by $(1-\delta)K_{ss}\tilde{K}_t/Y^p_{ss}$ and the remaining term in (B.17) is the contribution of the discounted future labor income.

The IRF of $\tilde{Y}^p_t$ is then calculated by replacing $\tilde{Y}_t, \tilde{K}_t, \tilde{R}_t$ by their corresponding IRFs. Furthermore, the expectation operator in (B.17) can be dropped along the IRF, as all future shocks are equal to zero, see (B.51) below. The calculation of the discounted labor income for Figure 4.5 is more demanding as for each instant of time $t = 1, \ldots, 140$ (the time range of Figure 4.5) the limit of a series has to be determined.[991] As the whole model and thus the IRFs of the other variables are solved numerically and not analytically, the series in (B.17) has to be approximated numerically as well. In order to do so, the IRFs of the whole model are calculated for $10.000 + 140$ periods. Then at each instant of time $t = 1, \ldots, 140$, the discounted future labor income is calculated along (B.17) by incorporating the periods $t+1, t+2, \ldots, t+10.000$.[992]

B.4 Summary of the Model with Cobb-Douglas Utility

B.4.1 The Stationarized (Nonlinear) Model Equations

A summary of the model with growth and Cobb-Douglas preferences can be found in Table B.4. Note that the model is expressed in the stationarized variables given in (4.20). Since the growing variables were divided by $\bar{A}_t$, $\bar{A}_t$ canceled out from the model equations. The remaining model is driven by the two exogenous processes specifying the transitory TFP component A_t and the growth rate of the labor augmenting technological progress g_t.

[991] Note that the limit of the series is well-defined due to the boundedness of the IRFs of $\tilde{Y}_t$ and $\tilde{R}_t$ and $0 < \beta < 1$.

[992] To cut off the series after 10.000 periods seems arbitrary. From the permanent income hypothesis, however, it follows that consumption is a fixed fraction of permanent income if the intertemporal elasticity of substitution is equal to one, i.e., it yields $\tilde{Y}^p_t \equiv \tilde{C}_t$ if $\varsigma = 1$, see Wen (2001, Equation (5) on p. 1226). Given the 1% transitory technology shock and using $\varsigma = 1$ instead of $\varsigma = 2$, Figure 4.5 results in an absolute deviation between $\tilde{Y}^p_t$ and $\tilde{C}_t$ of order 10^{-15}, which seems acceptable from a numerical point of view. Note that the conclusions drawn from Figure 4.5 are essentially the same for $\varsigma = 1$ as for $\varsigma = 2$.

Model equation		equation
$\frac{1-\gamma}{\gamma} w_t = \frac{c_t}{1-H_t}$	labor supply	(4.10)
$1 = \beta \mathbb{E}_t \left[b_{t+1}^{\bar{\gamma}} \left(\frac{c_{t+1}}{c_t} \right)^{\bar{\gamma}} \left(\frac{1-H_{t+1}}{1-H_t} \right)^{\gamma(1-\tau)} ((1-\delta) + R_{t+1}) \right]$	Euler equation	(4.11)
$k_{t+1} = \frac{(1-\delta)}{b_t} k_t + i_t$	capital accumulation	(4.5)
$L_t = 1 - H_t$	leisure choice	(4.6)
$y_t = A_t k_t^{\alpha} H_t^{1-\alpha} b_t^{-\alpha}$	production function	(4.1)
$k_t = \alpha b_t \frac{y_t}{R_t}$	capital demand	(4.13)
$H_t = (1-\alpha) \frac{y_t}{w_t}$	labor demand	(4.12)
$y_t = c_t + i_t$	goods market clearing	(4.7)
$(1 - \varrho_A B) (\log(A_t) - \log(A_{ss})) = (1-B)^{-d} \varepsilon_t^A$	transitory TFP	(4.14)
$b_t = (1 + g_t)$	growth factor	(4.16)
$(1 - \varrho_g B) (\log(b_t) - \log(b_{ss})) = \varepsilon_t^g$	evolution of the growth	(4.17)

Table B.4: Model equations of the stationarized nonlinear model with Cobb Douglas utility function. The right column shows the related equation in the text from which the model equations are derived. Some model equations are the same as their corresponding equation in the text; they are restated here to provide a complete overview of the model. Note that $\bar{\gamma} = (1-\gamma)(1-\tau) - 1$.

B.4.2 The Model's Steady State

As in the case of the model with additive utility function, closed-form expressions can be derived for the variables' steady state values. For the model summarized in Table B.4, they are given recursively in Table B.5. The composite parameter $\bar{\gamma} = (1-\gamma)(1-\tau) - 1$ is as in Appendix B.4.1.

variable		steady state	variable		steady state
A_{ss}	$=$	exogenous parameter	i_{ss}	$=$	$\left(1 - \frac{1-\delta}{b_{ss}}\right) k_{ss}$
$b_{ss} = (1+g_{ss})$	$=$	exogenous parameter	H_{ss}	$=$	$(1-\alpha)\frac{y_{ss}}{w_{ss}}$
R_{ss}	$=$	$\frac{1}{\beta} b_{ss}^{-\bar{\gamma}} - (1-\delta)$	L_{ss}	$=$	$1 - H_{ss}$
w_{ss}	$=$	$A_{ss}^{1/(1-\alpha)} \left(\frac{\alpha}{R_{ss}}\right)^{\alpha/(1-\alpha)} (1-\alpha)$	c_{ss}	$=$	$y_{ss} - i_{ss}$
y_{ss}	$=$	$\frac{1-\gamma}{\gamma} w_{ss} \left[1 + \frac{(1-\gamma)(1-\alpha)}{\gamma} - (g_{ss}+\delta)\frac{\alpha}{R_{ss}}\right]^{-1}$			
k_{ss}	$=$	$\alpha b_{ss} \frac{y_{ss}}{R_{ss}}$			

Table B.5: Steady state of the stationarized nonlinear model with Cobb-Douglas utility function. The values are sorted recursively from top to bottom, starting from the left column. Note that $\bar{\gamma} = (1-\gamma)(1-\tau) - 1$.

B.4.3 The Linearized Model Equations

Table B.6 shows the linearized model equations. The linearization was taken around the steady state values of the model given in Table B.5.

linearized Model equation		equation
$\tilde{w}_t = \tilde{c}_t - \tilde{L}_t$	labor supply	(4.10)
$\gamma(1-\tau)\tilde{L}_t + \bar{\gamma}\tilde{c}_t = \beta b_{ss}^{\bar{\gamma}} R_{ss}\mathbb{E}_t\tilde{R}_{t+1} + \gamma(1-\tau)\mathbb{E}_t\tilde{L}_{t+1} + \bar{\gamma}\mathbb{E}_t\tilde{c}_{t+1} + \bar{\gamma}\mathbb{E}_t\tilde{b}_{t+1}$	Euler equation	(4.11)
$\tilde{k}_{t+1} = \frac{(1-\delta)}{b_{ss}}\left(\tilde{k}_t - \tilde{b}_t\right) + \left(1 - \frac{1-\delta}{b_{ss}}\right)\tilde{i}_t$	capital accumulation	(4.5)
$L_{ss}\tilde{L}_t = -H_{ss}\tilde{H}_t$	leisure choice	(4.6)
$\tilde{y}_t = \tilde{A}_t + \alpha\tilde{k}_t + (1-\alpha)\tilde{H}_t - \alpha\tilde{b}_t$	production function	(4.1)
$\tilde{k}_t = \tilde{y}_t - \tilde{R}_t + \tilde{b}_t$	capital demand	(4.13)
$\tilde{H}_t = \tilde{y}_t - \tilde{w}_t$	labor demand	(4.12)
$y_{ss}\tilde{y}_t = c_{ss}\tilde{c}_t + i_{ss}\tilde{i}_t$	goods market clearing	(4.7)
$(1-\varrho_A B)\tilde{A}_t = (1-B)^{-d}\varepsilon_t^A$	transitory TFP	(4.14)
$\tilde{b}_t = \varrho_g\tilde{b}_{t-1} + \varepsilon_t^g$	growth factor	(4.16)

Table B.6: Linearized model equations of the model with Cobb-Douglas utility function. The right column gives the corresponding equation in the text. The order of the equations is the same as in Table B.1. The involved steady state values are given in Table B.5.

B.4.4 Matrices of the Model's Canonical Form

There are three predetermined (the capital stock, transitory TFP process, and the growth factor) and three forward-looking variables (consumption, the rental rate on capital and leisure) in the model given in Table B.6. The other variables are static again. The procedure is similar to the one considered in Appendix B.3.4. The model can be written in the following matrix equation

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & (1-\delta)/b_{ss} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \bar{\gamma}\varrho_g & 0 & \gamma(1-\tau) & \bar{\gamma} & \beta b_{ss}^{\bar{\gamma}} R_{ss} & 0 & 0 & 0 & 0 \\ -1 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \mathbb{E}_t \begin{pmatrix} \tilde{A}_t \\ \tilde{b}_t \\ \tilde{k}_{t+1} \\ \tilde{L}_{t+1} \\ \tilde{c}_{t+1} \\ \tilde{R}_{t+1} \\ \tilde{H}_{t+1} \\ \tilde{w}_{t+1} \\ \tilde{i}_{t+1} \\ \tilde{y}_{t+1} \end{pmatrix}$$

$$= \begin{pmatrix} \varrho_A & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varrho_g & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\delta)/b_{ss} & 0 & 0 & 0 & 0 & 0 & 1-(1-\delta)/b_{ss} & 0 \\ 0 & 0 & 0 & \gamma(1-\tau) & \bar{\gamma} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 & 1-\alpha & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & L_{ss} & 0 & 0 & H_{ss} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & c_{ss} & 0 & 0 & 0 & i_{ss} & -y_{ss} \end{pmatrix} \begin{pmatrix} \tilde{A}_{t-1} \\ \tilde{b}_{t-1} \\ \tilde{k}_t \\ \tilde{L}_t \\ \tilde{c}_t \\ \tilde{R}_t \\ \tilde{H}_t \\ \tilde{w}_t \\ \tilde{i}_t \\ \tilde{y}_t \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \nu_t^A \\ \varepsilon_t^g \end{pmatrix} \tag{B.18}$$

Similar to Appendix B.3.4, this system can be partitioned according to the number of non-static (i.e., predetermined and forward-looking) and static variables. The non-static (i.e., the predetermined and non-predetermined) variables are collected in the vector $x_{t+1} = \left(\tilde{A}_t, \tilde{b}_t, \tilde{k}_{t+1}, \tilde{L}_{t+1}, \tilde{c}_{t+1}, \tilde{R}_{t+1}\right)^T \in \mathbb{R}^{6\times 1}$ and the remaining static variables in the vector $y_{t+1} \in \mathbb{R}^{4\times 1}$. Then (B.18) can be written as

$$\begin{pmatrix} \mathbf{A} & 0_{6\times 4} \\ 0_{4\times 6} & 0_{4\times 4} \end{pmatrix} \mathbb{E}_t \begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} \mathbf{G} \\ 0_{4\times 2} \end{pmatrix} \nu_t^A, \tag{B.19}$$

where $\mathbf{A}, \mathbf{D}_{11} \in \mathbb{R}^{6\times 6}, \mathbf{D}_{12} \in \mathbb{R}^{6\times 4}, \mathbf{D}_{21} \in \mathbb{R}^{4\times 6}, \mathbf{D}_{22} \in \mathbb{R}^{4\times 4}$ and $\mathbf{G} \in \mathbb{R}^{6\times 2}$.

The steps to solve the model are essentially the same as those described after (B.13). The difference is that the matrices $\mathbf{A}$ and $\mathbf{B}$ of the model's canonical form are now of size 6×6 instead of 4×4 due to more predetermined and forward-looking variables in the model with Cobb-Douglas preferences. Furthermore, the exogenous stochastic process $z_t = \left(\nu_t^A, \varepsilon_t^g\right)^T$ turns out to be two-dimensional, in contrast to the model with additive preferences, which includes only the transitory shock component. Thus, the corresponding matrix $\mathbf{G}$ has to be of size 6×2 instead of 4×1.

Again, if $\mathbf{D}_{22}$ is invertible, (B.13) can be solved, and the static variables y_t are determined by an analogous equation to (B.14).[993,994]

The matrices of the DSGE model's canonical form (See (B.20) below) are then again given by (B.15) with the matrices $\mathbf{A}, \mathbf{D}_{11}, \mathbf{D}_{22}, \mathbf{D}_{21}$ and $\mathbf{G}$ as in (B.19).

B.5 The Method of Klein

In this thesis, the solution method proposed by Klein (2000) is generalized and applied to a model with long memory. Here, the solution procedure and how it is applied in the two models presented in Chapter 4 are described. Before the details on the solution methodology are outlined, some introducing remarks and relations to the solution methodology of Blanchard and Kahn (1980) treated in Section 4.2.1 are stated in the next section.

993 As can be seen easily from (B.18), one has that $\det(\mathbf{D}_{22}) = -H_{ss} i_{ss}$. Thus, $\mathbf{D}_{22}$ is invertible as long as the steady state investments and the steady state time devoted to work differ from zero. These conditions are quite similar to the one in the model with additive separable utility function, see Footnote 989.

994 Since, e.g., the last two columns of $\mathbf{A}$ are linearly dependent (see (B.18)), $\mathbf{A}$ is indeed not invertible and thus the method of Blanchard and Kahn (1980) is again not applicable here, see Footnote 990.

B.5.1 Introducing Remarks and Relations to Blanchard and Kahn

Klein's method requires the canonical form of the linearized DSGE model to be[995]

$$\mathbf{A}\mathbb{E}_t x_{t+1} = \mathbf{B}x_t + \mathbf{G}z_t, \tag{B.20}$$

where the vector x_t is of size $n \times 1$ and n corresponds to the number of variables in the model. The matrices $\mathbf{A}$ and $\mathbf{B}$ are of size $n \times n$ and the matrix G is of size $n \times n_z$, where n_z is the number of exogenous stochastic shocks specified in the $n_z \times 1$ vector z_t. The elements of these matrices are functions of the model parameters and variables' steady state values.[996]

It can be seen easily that (B.20) can be transformed into the form required for Blanchard and Kahn's solution method only if the matrix $\mathbf{A}$ is invertible.[997] On the other hand, (4.22) always satisfies (B.20). However, in both models (the one without growth and additive separable utility and the one with growth and Cobb-Douglas utility), matrix $\mathbf{A}$ is not invertible, and the solution method of Blanchard and Kahn (1980) is not applicable.[998]

Furthermore, the vector x_t in (B.20) contains variables of different types, namely predetermined or backward-looking and non-predetermined or forward-looking variables.[999] This division into two types of variables is common to both methods. However, the definition of a predetermined or backward-looking variable is slightly different.[1000]

A backward-looking variable in Klein's sense generalizes Blanchard and Kahn's concept of predetermined variables in the sense that there may be additional (exogenously given) prediction errors in the method of Klein (2000). To be more precise, a variable is backward-looking in the sense of Klein if $\mathbb{E}_t x^b_{t+1} = x^b_{t+1} + \xi_{t+1}$, where x^b_t refers to the part of the vector x_t containing the backward-looking variables and ξ_t is an exogenous stochastic process (prediction error) satisfying $E_t\xi_{t+1} = 0$.[1001] It is now easy to see that a predetermined

[995] See Klein (2000, Equation (2.1) on p. 1408).

[996] The derivation of (B.20) for the non-growing DSGE model with additive separable utility and the stationarized model with Cobb-Douglas preferences are given in Appendix B.3.4 and Appendix B.4.4, respectively.

[997] In this case multiply (B.20) with $\mathbf{A}^{-1}$ from left, then define $\mathbf{A}_0 = \mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{G}_0 = \mathbf{A}^{-1}\mathbf{G}$ in (4.22).

[998] See Footnote 990 and Footnote 994.

[999] Recall from the derivation of (B.20) for the two linearized DSGE models considered in this thesis, the so-called static variables that turn out be linear combinations of predetermined and non-predetermined were already eliminated from the system.

[1000] There are other solution methods for linear DSGE models that need no distinction between these two kinds of variables, see Sims (2002, pp. 1f.). Such a method is outlined in Sims (2002).

[1001] See Klein (2000, Definition 4.3 on p. 1412).

variable in the sense of Blanchard and Kahn is also a backward-looking variable in the sense of Klein.[1002] Since there are no exogenous prediction errors specified in the models stated in Section 4.1 (see, e.g., the example in Footnote 663), both concepts of predetermined and backward-looking variables coincide in the context of this thesis. Note that all variables that are not predetermined are non-predetermined or forward-looking.

A process $x = (x_t)_{t \in \mathbb{Z}}$ satisfying (B.20) is called a solution to the model. This solution is called stable if there is a constant M such that[1003]

$$\max_{1 \le i \le n} \left(\mathbb{E}|x_{i,t}|\right)^{1/2} \le M \quad \text{for all } t \ge 0, \tag{B.21}$$

where $x_{i,t}$ refers to the ith entry of the vector x_t.[1004] Equation (B.21) states that the unconditional expectations of the process $(x_t)_{t \in \mathbb{Z}}$ are uniformly bounded, and thus, the unconditional expectations cannot grow without bounds.[1005] The solution method of Klein focuses on finding a unique stable solution similar to the method of Blanchard and Kahn presented in Section 4.2. They restrict possible solutions to the set of processes that satisfy condition (4.24). However, condition (B.21) is stronger than (4.24) as the latter is implied by the former.[1006] The reverse, however, does not need to hold.[1007] A class of processes satisfying (B.21) is, for example, the set of stationary processes.[1008] It is further assumed that the exogenous process is stable as well.[1009]

Recall that the corresponding condition (4.24) of Blanchard and Kahn ensures that the bubble term of the solutions vanishes.[1010] The same holds for (B.21) in Klein's solution methodology.[1011] However, as stressed by Gourieroux, Jasiak, et al. (2020), the model becomes indeterminate, i.e., there are many solutions to the model, if one relaxes (B.21).[1012,1013] The authors show that then there is an infinite number of solutions containing a strictly stationary bubble term.[1014] More precisely, recall from the text after

[1002] Simply let $\xi_t \equiv 0$, see Klein (2000, p. 1412).
[1003] See Klein (2000, Definition 4.1 on p. 1412).
[1004] This definition of stability is in line with Binder and Pesaran (1995, p. 151 and Footnote 22 on p. 151/180).
[1005] See Klein (2000, p. 1411).
[1006] This follows from G. W. Evans and McGough (2005, p. 621) who showed that for a process with uniformly bounded expectations, the conditional expectations are uniformly bounded as well.
[1007] See G. W. Evans and McGough (2005, pp. 621f.).
[1008] See G. W. Evans and McGough (2005, p. 622).
[1009] See Klein (2000, Assumption 4.1 on p. 1412).
[1010] See Section 4.2.3 for details.
[1011] See (B.30) below.
[1012] See Gourieroux, Jasiak, et al. (2020, pp. 717 and 722f.).
[1013] Note that Gourieroux, Jasiak, et al. (2020) consider only univariate models.
[1014] See Gourieroux, Jasiak, et al. (2020, p. 722).

Definition 2.1.2 that strict stationarity means a process has time-consistent distributions. The processes considered by Gourieroux, Jasiak, et al. (2020) have no finite second moments; hence these processes cannot be covariance stationary as the first condition in Definition 2.1.2 is not satisfied by these processes nor do they satisfy (B.21).[1015] Such processes are, for example, processes with fat-tailed distributions.[1016] The overall (infinitely many) model solutions can then be described as the sum of the forward solution plus a strictly stationary bubble term.[1017]

This short discussion illustrates that condition (B.21) is rather sharp and rules out many possibly economically meaningful model solutions. However, the occurring indeterminacy when one allows for solutions not satisfying (B.21) raises the question of whether and how a specific solution should be selected. For this reason, the approach followed in this thesis is the one followed by a large strand of the DSGE literature, i.e., the focus lies on stable (in the sense of (B.21)) solutions. Nevertheless, it should be kept in mind that the procedure described in the sequel finds a unique solution only under all stable solutions. Whether there are additional unstable solutions (in a sense not satisfying (B.21)) goes beyond the capabilities of the methods discussed in this thesis.

B.5.2 The Solution Methodology

Here, the solution methodology of Klein (2000) is summarized for general exogenous processes z_t. This is done because some essential equations in Klein (2000) are only given when z_t is a VAR(1). However, the exogenous process in the long memory model cannot be described as a VAR(1) process. The subsequent sections build on this general case and derive the solutions to the models in Chapter 4.

Consider the model given in (B.20) and divide the vector $x_t = \left(x_t^b, x_t^f\right)^T$ into two sub-vectors containing the backward- and forward-looking variables, respectively. Assume the dimension of x_t^b and x_t^f to be $n_b \times 1$ and $n_f \times 1$, respectively, with $n_b + n_f = n$. Then, (B.20) can be written as

$$\mathbf{A}\mathbb{E}_t \begin{pmatrix} x_{t+1}^b \\ x_{t+1}^f \end{pmatrix} = \mathbf{B} \begin{pmatrix} x_t^b \\ x_t^f \end{pmatrix} + \mathbf{G} z_t. \tag{B.22}$$

In the following, consider the generalized Schur (or QZ) decomposition of the matrices $\mathbf{A}$

[1015] See Gourieroux, Jasiak, et al. (2020, p. 717).
[1016] See Gourieroux, Jasiak, et al. (2020, p. 722).
[1017] See Gourieroux, Jasiak, et al. (2020, p. 722).

and $\mathbf{B}$, i.e., there are $n \times n$ unitary matrices Q, Z satisfying

$$QQ' = Q'Q = I = Z'Z = Z'Z,$$

and $n \times n$ upper triangular matrices S and T such that the following equations hold

$$\begin{aligned} Q\mathbf{A}Z &= S \text{ or } \mathbf{A} = Q'SZ' \\ Q\mathbf{B}Z &= T \text{ or } \mathbf{B} = Q'TZ'. \end{aligned} \tag{B.23}$$

Note that X' is the Hermitian transpose of the matrix X.[1018]

The generalized eigenvalues of the matrices $\mathbf{A}$ and $\mathbf{B}$ can then be calculated by taking quotients of the diagonal entries of the matrices T and S.[1019] That is, the generalized eigenvalues are given by $\lambda_i = T(i,i)/S(i,i)$ for $i = 1, \dots, n$, where $T(i,i)$ is the element of matrix T contained in the i-th row and i-th column (and similar for S). An eigenvalue λ_i is called stable if $|\lambda_i| < 1$ and unstable if $|\lambda_i| > 1$. It is assumed that $|\lambda_i| \neq 1$ for all $i = 1, \dots, n$. Thus, let n_s be the number of stable and $n_u = n - n_s$ be the number of unstable eigenvalues. One can assume further that the generalized Schur decomposition is sorted in a way that $|\lambda_i| < 1$ for $i = 1, \dots, n_s$ and $|\lambda_i| > 1$ for $i = n_s + 1, \dots, n$.[1020]

Inserting (B.23) into (B.22) and multiplying with Q from left results in

$$SZ' \, \mathbb{E}_t \begin{pmatrix} x_{t+1}^b \\ x_{t+1}^f \end{pmatrix} = TZ' \begin{pmatrix} x_t^b \\ x_t^f \end{pmatrix} + Q\mathbf{G}z_t. \tag{B.24}$$

Then, decompose the matrices S, Z, Z', T, Q, and G such that all stable eigenvalues are contained in the upper n_s rows of the system. As in Klein (2000), define the two auxiliary variables s_t and u_t associated with the stable and unstable subsystem, respectively. Overall, (B.24) then becomes[1021]

$$\begin{pmatrix} S_{11} & S_{12} \\ 0_{n_u \times n_s} & S_{22} \end{pmatrix} \mathbb{E}_t \begin{pmatrix} s_{t+1} \\ u_{t+1} \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ 0_{n_u \times n_s} & T_{22} \end{pmatrix} \begin{pmatrix} s_t \\ u_t \end{pmatrix} + \begin{pmatrix} \hat{G}_1 \\ \hat{G}_2 \end{pmatrix} z_t, \tag{B.25}$$

[1018] Note that the matrices Q and Z are generally complex-valued. The Hermitian transpose of a complex-valued matrix X is built by first taking the complex conjugate of each element of X and then building the transpose of the resulting matrix.

[1019] A generalized eigenvalue λ of the matrices $\mathbf{A}$ and $\mathbf{B}$ with (generalized) eigenvector x_λ satisfies $\mathbf{B}x_\lambda = \lambda \mathbf{A} x_\lambda$, see Klein (2000, p. 1410).

[1020] See Klein (2000, Theorem 3.3. on p. 1410).

[1021] The matrix dimensions are as follows: $T_{11}, S_{11} \in \mathbb{R}^{n_s \times n_s}$, $T_{12}, S_{12} \in \mathbb{R}^{n_s \times n_u}$, $T_{22}, S_{22} \in \mathbb{R}^{n_u \times n_u}$, $\hat{G}_1 \in \mathbb{R}^{n_s \times n_z}$, $\hat{G}_2 \in \mathbb{R}^{n_u \times n_z}$, $Z'_{11}, Q_{11} \in \mathbb{R}^{n_s \times n_b}$, $Z'_{12}, Q_{12} \in \mathbb{R}^{n_s \times n_f}$, $Z'_{21}, Q_{21} \in \mathbb{R}^{n_u \times n_b}$, $Z'_{22}, Q_{22} \in \mathbb{R}^{n_u \times n_f}$, $G_1 \in \mathbb{R}^{n_b \times n_z}$ and $G_2 \in \mathbb{R}^{n_f \times n_z}$.

where[1022]

$$\begin{pmatrix} s_t \\ u_t \end{pmatrix} := Z' \begin{pmatrix} x_t^b \\ x_t^f \end{pmatrix} = \begin{pmatrix} Z'_{11} & Z'_{12} \\ Z'_{21} & Z'_{22} \end{pmatrix} \begin{pmatrix} x_t^b \\ x_t^f \end{pmatrix} \tag{B.26}$$

and

$$\begin{pmatrix} \hat{G}_1 \\ \hat{G}_2 \end{pmatrix} := Q \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{pmatrix}.$$

Until now, the initial problem stated in (B.20) has been transformed into the triangular system (B.25). This system allows to consider the subsystems associated with the stable and unstable eigenvalues of the system separately.

Consider first the second row of (B.25), i.e., the subsystem associated with the unstable eigenvalues. It follows directly that[1023]

$$u_t = T_{22}^{-1} S_{22} \mathbb{E}_t u_{t+1} - T_{22}^{-1} \hat{G}_2 z_t. \tag{B.27}$$

Equation (B.27) can be solved by forward iteration. By shifting (B.27) one period ahead and taking expectations at time t, one obtains

$$\mathbb{E}_t u_{t+1} = T_{22}^{-1} S_{22} \mathbb{E}_t \left(\mathbb{E}_{t+1} u_{t+2} \right) - T_{22}^{-1} \hat{G}_2 \mathbb{E}_t z_{t+1}. \tag{B.28}$$

By the law of iterated expectations, it yields that $\mathbb{E}_t \left(\mathbb{E}_{t+1} u_{t+2} \right) = \mathbb{E}_t u_{t+2}$.[1024] Thus, by plugging (B.28) back in (B.27), one obtains

$$u_t = \left(T_{22}^{-1} S_{22} \right)^2 \mathbb{E}_t u_{t+2} - T_{22}^{-1} S_{22} T_{22}^{-1} \hat{G}_2 \mathbb{E}_t z_{t+1} - T_{22}^{-1} \hat{G}_2 z_t.$$

By repeating these steps k-times, one finally arrives at

$$u_t = \left(T_{22}^{-1} S_{22} \right)^k \mathbb{E}_t u_{t+k} - \sum_{j=0}^{k-1} \left(T_{22}^{-1} S_{22} \right)^j T_{22}^{-1} \hat{G}_2 \mathbb{E}_t z_{t+j}, \tag{B.29}$$

where the first term on the right-hand side of (B.29) is often called bubble and the second

[1022] To avoid confusion regarding the notation, note that Z_{11} refers to the first block of the partitioned matrix Z and Z'_{11} to the first block of Z' and not to the transpose of Z_{11}. The latter would be denoted by $(Z_{11})'$.

[1023] Note, that the matrix T_{22} is invertible by the construction of the generalized Schur decomposition, see Klein (2000, p. 1415).

[1024] See, e.g., Klenke (2013, p. 178).

term fundamental solution.[1025] As one is interested in a stable solution to the model, i.e., a solution that satisfies (B.21), the unconditional expectations of u_t have to be bounded.[1026] Since u_t has to be stable, one can deduce that $\mathbb{E}_t u_{t+k}$ is uniformly bounded in k as well.[1027]

Further, note that all eigenvalues of $T_{22}^{-1}S_{22}$ are by construction less than one in modulus.[1028] Overall, it follows that

$$\lim_{k\to\infty}\left(T_{22}^{-1}S_{22}\right)^k \mathbb{E}_t u_{t+k} = 0. \tag{B.30}$$

By letting $k \to \infty$ and by plugging (B.30) back into (B.29), one obtains the unique and stable solution for u_t given by [1029,1030]

$$u_t = -\sum_{j=0}^{\infty}\left(T_{22}^{-1}S_{22}\right)^j T_{22}^{-1}\hat{G}_2\mathbb{E}_t z_{t+j}. \tag{B.31}$$

Further, one can show inductively that (B.31) can equally be written as

$$u_t = -T_{22}^{-1}\sum_{j=0}^{\infty}\left(S_{22}T_{22}^{-1}\right)^j \hat{G}_2\mathbb{E}_t z_{t+j}. \tag{B.32}$$

Now, consider the upper row of (B.25), i.e., the subsystem associated with the stable eigenvalues. It follows directly that

$$S_{11}\mathbb{E}_t s_{t+1} + S_{12}\mathbb{E}_t u_{t+1} = T_{11}s_t + T_{12}u_t + \hat{G}_1 z_t,$$

or equivalently

$$\mathbb{E}_t s_{t+1} = S_{11}^{-1}T_{11}s_t + S_{11}^{-1}T_{12}u_t - S_{11}^{-1}S_{12}\mathbb{E}_t u_{t+1} + S_{11}^{-1}\hat{G}_1 z_t. \tag{B.33}$$

Further, it follows from (B.26) that

$$x_{t+1}^b = Z_{11}s_{t+1} + Z_{12}u_{t+1} \tag{B.34}$$

[1025] See Blanchard and Fischer (1993, p. 221) and Section 4.2.3.

[1026] Recall that the vector $(s_t, u_t)^T$ is just a linear transformation of $x_t = (x_t^b, x_t^f)^T$, see (B.26). Thus, if the stability condition is imposed on x_t, the vector u_t is also stable.

[1027] See Appendix B.5.1 and especially Footnote 1006.

[1028] Since T_{22} and S_{22} are upper triangular matrices, the same holds true for $T_{22}^{-1}S_{22}$. Thus, the diagonal elements of $T_{22}^{-1}S_{22}$ are equal to the eigenvalues of $T_{22}^{-1}S_{22}$ and they are given by $S_{22}(i,i)/T_{22}(i,i)$ for $i = 1, \ldots, n_u$. Recall that the Schur decomposition was sorted in a way that $|\lambda_i| = |T(i,i)/S(i,i)| > 1$ for $i = n_s + 1, \ldots, n$ or equivalently $|T_{22}(i,i)/S_{22}(i,i)| > 1$ for $i = 1, \ldots, n_u$. Hence, the eigenvalues of $T_{22}^{-1}S_{22}$ are smaller than one in modulus.

[1029] Note that (B.31) does not correspond to Equation (5.5) in Klein (2000) due to a typo in the published version of Klein (2000). A corrected version of Klein's Equation (5.5) can be found in Klein (no date), URL in list of references. The corrected versions coincides with (B.31).

[1030] Equation (B.31) corresponds to (4.33) in Section 4.2.3.

Since all variables in x_t^b are predetermined by definition, one has that $\mathbb{E}_t x_{t+1}^b - x_{t+1}^b = 0$.[1031] Thus, (B.34) becomes

$$0 = Z_{11}\left(\mathbb{E}_t s_{t+1} - s_{t+1}\right) + Z_{12}\left(\mathbb{E}_t u_{t+1} - u_{t+1}\right),$$

from where one can derive

$$\mathbb{E}_t s_{t+1} = s_{t+1} - Z_{11}^{-1} Z_{12}\left(E_t u_{t+1} - u_{t+1}\right). \tag{B.35}$$

Plugging (B.35) back into (B.33) leads to

$$\begin{aligned} s_{t+1} = S_{11}^{-1} T_{11} s_t + S_{11}^{-1} T_{12} u_t - S_{11}^{-1} S_{12} \mathbb{E}_t u_{t+1} \\ + Z_{11}^{-1} Z_{12}\left(\mathbb{E}_t u_{t+1} - u_{t+1}\right) + S_{11}^{-1} \hat{G}_1 z_t. \end{aligned} \tag{B.36}$$

In summary, with (B.32) one can calculate the values for u_t, u_{t+1} and $\mathbb{E}_t u_{t+1}$. These can then be plugged in (B.36) that defines a recursive law on determining s_{t+1} from s_t and these values. It remains to find the initial value s_0 to start the recursion. From (B.34), it follows, however, that

$$s_0 = Z_{11}^{-1}\left(x_0^b - Z_{12} u_0\right), \tag{B.37}$$

where x_0^b contains the initial values of the predetermined variables, which are assumed to be exogenously given. Now, the solution to (B.22) is given in terms of the auxiliary variables u_t and s_t. With (B.26), the solution in terms of x_t^b and x_t^f can be recovered by multiplying $(s_t, u_t)^T$ from left with Z.

Note that (B.37) requires the matrix Z_{11} to be invertible. By Footnote 1021 the dimension of Z_{11} is $n_s \times n_b$. Thus, a necessary condition for Z_{11} to be invertible is $n_s = n_b$, i.e., the number of backward-looking variables has to equal the number of stable eigenvalues. This condition corresponds to the Blanchard-Kahn condition for the existence of a unique solution mentioned in Section 4.2.1. For the method of Klein (2000), however, the Blanchard-Kahn condition is not sufficient to pin down the unique solution.[1032] Nevertheless, common to both solution strategies is that the initial values of the stable subsystem have to be uniquely determined by the initial values of the predetermined or backward-looking variables x_0^b and by the initial values of the forward solution to the unstable part of the system u_0 in order to find a unique solution (see (B.37)).

[1031] This is precisely the definition of a predetermined variable.

[1032] See Klein (2000, pp. 1418ff.). Clearly, $n_b = n_s$ implies the matrix Z_{11} to be square but nothing can be said about its invertibility.

B.5.3 The Standard Case

Evidently, the solution to the DSGE model depends on the expectations of the exogenous process $(z_t)_{t\in\mathbb{Z}}$, see (B.32) and (B.36). Here it is illustrated shortly that in the case of a normally distributed white noise exogenous shock process $(z_t)_{t\in\mathbb{Z}}$, the model's solution is again a short memory process.[1033] To do so, observe that in this case

$$\mathbb{E}_t z_{t+j} = \begin{cases} z_t & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{B.38}$$

Plugging (B.38) back in (B.32), one obtains

$$u_t = -T_{22}^{-1}\hat{G}_2 z_t. \tag{B.39}$$

Further, it follows from (B.39), that $\mathbb{E}_t u_{t+1} = 0$. Plugging this back in (B.36), the recursive law of motion for s_t is given by

$$s_{t+1} = S_{11}^{-1}T_{11}s_t + S_{11}^{-1}T_{12}u_t - Z_{11}^{-1}Z_{12}u_{t+1} + S_{11}^{-1}\hat{G}_1 z_t \tag{B.40}$$

Now, (B.39) and (B.40) can be rewritten as

$$\begin{pmatrix} s_t \\ u_t \end{pmatrix} = \mathbf{U}\begin{pmatrix} s_{t-1} \\ u_{t-1} \end{pmatrix} + \mathbf{V}\check{z}_{t-1} + \mathbf{W}\check{z}_t, \tag{B.41}$$

with matrices

$$\mathbf{U} := \begin{pmatrix} \mathbf{P} & 0_{n_s\times n_u} \\ 0_{n_u\times n_s} & 0_{n_u\times n_u} \end{pmatrix}, \mathbf{V} := \begin{pmatrix} \mathbf{Q} & 0_{n_s\times(n-n_z)} \\ 0_{n_u\times n_z} & 0_{n_u\times(n-n_z)} \end{pmatrix}, \mathbf{W} := \begin{pmatrix} \mathbf{R} & 0_{n_s\times(n-n_z)} \\ \mathbf{S} & 0_{n_u\times(n-n_z)} \end{pmatrix}$$

and $\mathbf{P} = S_{11}^{-1}T_{11} \in \mathbb{R}^{n_s\times n_s}, \mathbf{Q} = -S_{11}^{-1}T_{12}T_{22}^{-1}\hat{G}_2 + S_{11}^{-1}\hat{G}_1 \in \mathbb{R}^{n_s\times n_z}, \mathbf{R} = Z_{11}^{-1}Z_{12}T_{22}^{-1}\hat{G}_2 \in \mathbb{R}^{n_s\times n_z}, \mathbf{S} = -T_{22}^{-1}\hat{G}_2 \in \mathbb{R}^{n_u\times n_z}$. The process $\check{z}_t$ is obtained by augmenting zeros to z_t, i.e., $\check{z}_t = \left(z_t, 0_{1\times(n-n_z)}\right)^T \in \mathbb{R}^{n\times 1}$.

By appending $\check{z}_t$ to the state vector (s_t, u_t), (B.41) can equally be written as

$$v_t = \mathbf{M}v_{t-1} + \mathbf{N}\varepsilon_t \tag{B.42}$$

[1033] Regarding the models stated in Table B.3 and Table B.6, this corresponds to the situation where $d = 0$, i.e., both the transitory productivity process and the productivity growth processes are assumed to be AR(1) processes.

with

$$v_t := \begin{pmatrix} s_t \\ u_t \\ \check{z}_t \end{pmatrix} \in \mathbb{R}^{2n\times 1},\ \mathbf{M} := \begin{pmatrix} \mathbf{U} & \mathbf{V} \\ 0_{n\times n} & 0_{n\times n} \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \text{ and}$$

$$\mathbf{N} := \begin{pmatrix} \mathbf{W} & 0_{n\times n} \\ \mathbf{I}_{n\times n} & 0_{n\times n} \end{pmatrix} \begin{pmatrix} \Sigma_z & 0_{n_z\times(2n-n_z)} \\ 0_{(2n-n_z)\times n_z} & 0_{(2n-n_z)\times(2n-n_z)} \end{pmatrix} \in \mathbb{R}^{2n\times 2n}.$$

The process ε_t is a standard normally distributed random vector of size $2n \times 1$ and $\Sigma_z \in \mathbb{R}^{n_z\times n_z}$ is the covariance matrix of the white process $(z_t)_{t\in\mathbb{Z}}$.

Equation (B.42) is exactly the already mentioned representation (4.25) and illustrates that the solution to a DSGE model in the standard case of exogenous normally distributed white noise shocks, can be represented as a multivariate AR(1) process.[1034] Further, all eigenvalues of $\mathbf{M}$ are less than one in modulus by construction.[1035]

For the multivariate case, a similar equation to (2.10) exists. To be more precise, one can show that each component process v_t^i of v_t has a geometrically bounded ACF, i.e., there are constants $0 < C$ and $0 < \beta < 1$ such that $\gamma_{v^i}(k) < C\beta^k$.[1036,1037] Hence, each component of v_t is again a short memory process.

This result illustrates that a linearized DSGE model with canonical form (B.22) cannot generate long memory endogenously if the exogenous shock process is assumed to be a normally distributed white noise process. Thus, this result strongly supports introducing long memory exogenously into the model, i.e., the approach followed in this thesis. The procedure for solving the model with exogenous long memory is outlined in the next section.

[1034] Note that (B.42) is written in terms of the auxiliary variables s_t and u_t. To recover the original variables $x_t = (x^b, x^f)^T$ from (B.42), one can use $x_t = (Z, 0_{n\times n})\, v_t$, see (B.26).

[1035] It can easily be seen from the definition of $\mathbf{M}$ that $\det(\mathbf{M} - \lambda \mathbf{I}_{n\times n}) = (-\lambda)^{n+n_u}\det(\mathbf{P} - \lambda \mathbf{I}_{n_s\times n_s})$. Hence, the eigenvalues of $\mathbf{M}$ are just zero and the ones of $\mathbf{P}$. Recall the definiton of $\mathbf{P} = S_{11}^{-1}T_{11}$. An analogous argument to Footnote 1028 establishes that all eigenvalues of $\mathbf{P}$ are smaller than one in modulus. The eigenvalues of $\mathbf{P}$ are essentially the n_s stable eigenvalues of the initial system.

[1036] See Brockwell and Davis (1987, p. 410).

[1037] Note that this result does not only hold for the autocorrelation of each component of v_t but also for the cross-correlations between the components of v_t. The condition that all eigenvalues of $\mathbf{M}$ are less than one in modulus is required for the process to be causal, which is a requirement for the geometrical boundedness of the ACF, see again Brockwell and Davis (1987, p. 410). Note that there are closed-form expressions for calculating the autocovariance matrix for the process v_t, see Fernández-Villaverde, Rubio-Ramírez, and Schorfheide (2016, p. 635) for the VAR(1) case and Brockwell and Davis (1987, Equation (11.3.15) on p. 410) for causal vector autoregressive moving average (VARMA) processes. They are not stated here since the geometrical boundedness is enough for the purposes of this section.

B.5.4 The Solution to the Model with CD Preferences

This section illustrates how to derive the solution to the model given in Table B.6. The procedure for solving the model stated in Table B.3 is quite similar to the one presented here since it just involves $z_t = \nu_t^A$ as a one-dimensional exogenous stochastic process instead of $z_t = \left(\nu_t^A, \varepsilon_t^g\right)^T$ as a two-dimensional stochastic process.

Recall that (B.22) for the model given in Table B.6 was derived in Appendix B.4.4. Here, the focus lies on (B.32) and (B.36). More precisely, the expectation operators in these equations are eliminated in the following by plugging the definition of the exogenous process $z_t = \left(\nu_t^A, \varepsilon_t^g\right)^T$ into these equations.

Recall, that ν_t^A is a fractional white noise process with infinite moving average representation[1038]

$$\nu_t^A = \sum_{i=0}^{\infty} \alpha_i \, \varepsilon_{t-i}^A, \tag{B.43}$$

where $\alpha_i = \dfrac{\Gamma(i+d)}{\Gamma(i+1)\Gamma(d)}$.

Since $(\varepsilon_t^A)_{t\in\mathbb{Z}}$ and $(\varepsilon_t^g)_{t\in\mathbb{Z}}$ are independent Gaussian white noise processes, one can further make use of the following identity

$$\mathbb{E}_t \varepsilon_{t+k}^x = \begin{cases} \varepsilon_{t+k}^x & \text{if } k \leq 0 \\ 0 & \text{if } k > 0 \end{cases} \quad \text{for } k \in \mathbb{Z}, x \in \{A, g\}. \tag{B.44}$$

Especially, by using (B.43) and (B.44), it yields

$$\mathbb{E}_t \nu_{t+k}^A = \mathbb{E}_t \left(\sum_{i=0}^{\infty} \alpha_i \varepsilon_{t+k-i}^A \right) = \sum_{i=0}^{\infty} \alpha_i \mathbb{E}_t \varepsilon_{t+k-i}^A = \sum_{i=k}^{\infty} \alpha_i \, \varepsilon_{t+k-i}^A = \sum_{i=0}^{\infty} \alpha_{i+k} \, \varepsilon_{t-i}^A. \tag{B.45}$$

By inserting this into (B.32) and defining $\Lambda \coloneqq S_{22} T_{22}^{-1}$, one further obtains

$$\begin{aligned} u_t &= -T_{22}^{-1} \sum_{k=0}^{\infty} \Lambda^k \hat{G}_2 \mathbb{E}_t \begin{pmatrix} \nu_{t+k}^A \\ \varepsilon_{t+k}^g \end{pmatrix} \\ &= -T_{22}^{-1} \hat{G}_2 \varepsilon_t^g e_2 - T_{22}^{-1} \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \Lambda^k \alpha_{i+k} \hat{G}_2 e_1 \, \varepsilon_{t-i}^A, \end{aligned} \tag{B.46}$$

where $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

[1038] See (4.15), Definition 2.4.1 and Lemma 2.4.2.

Remark B.5.1

Note that the term in (B.46) that involves the two sums can equally be written as[1039]

$$-T_{22}^{-1}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\Lambda^k\alpha_{j+k}\hat{G}_2e_1\,\varepsilon_{t-j}^A = -T_{22}^{-1}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\Lambda^k\alpha_{j+k}\hat{G}_2e_1\,\varepsilon_{t-j}^A = -T_{22}^{-1}\sum_{j=0}^{\infty}b_j\,\varepsilon_{t-j}^A \quad \text{(B.47)}$$

with $b_j := \sum_{k=0}^{\infty}\Lambda^k\alpha_{j+k}\hat{G}_2e_1 \in \mathbb{R}^{n_u\times 1}$. The right-hand side illustrates that the part of u_t associated with the long memory process can again be interpreted as an infinite moving average process with coefficients b_j. Unfortunately, each moving average coefficient occurs as the limit of a (complicated) series. This limit is well defined since from (A.2), it follows that $\alpha_k \leq 1$ for all $k \geq 0$ and all eigenvalues of Λ are smaller than one in modulus by construction. Thus, the series $\sum_{k=0}^{\infty}\Lambda^k = (\mathbf{I}_{n_u\times n_u} - \Lambda)^{-1}$ converges and the values of b_j are finite for all j.[1040] The limit of the series involved in the definition of the b_j's can be calculated by the use of the Gaussian hypergeometric function given the additional assumption that the matrix $\Lambda = S_{22}T_{22}^{-1}$ is diagonalizable, i.e., there exists $n_u \times n_u$ matrices P and D so that[1041]

$$\Lambda = PDP^{-1} \text{ where } D = \begin{pmatrix} \lambda_{n_s+1}^{-1} & & \\ & \ddots & \\ & & \lambda_n^{-1} \end{pmatrix}.$$

Note that $\Lambda^k = PD^kP^{-1}$. Using this decomposition leads to

$$\begin{aligned} b_j &= \sum_{k=0}^{\infty} PD^kP^{-1}\alpha_{k+j}\hat{G}_2e_1 \\ &= P\left(\sum_{k=0}^{\infty} D^k\alpha_{k+j}\right)P^{-1}\hat{G}_2e_1 \\ &= \alpha_j P\ {}_2F_1(1, j+d; j+1; D)P^{-1}\hat{G}_2e_1, \end{aligned} \quad \text{(B.48)}$$

1039 This follows from Palma (2007, Theorem 1.5 on p. 7).

1040 See Olver and Shakiban (2018, p. 499).

1041 The assumption that Λ is of this form is made due to simplification issues. A sufficient condition for Λ to be diagonalizable is that Λ has n_u distinct eigenvalues, see Olver and Shakiban (2018, Theorem 8.23 and Theorem 8.25). This property is equivalent to saying that all unstable generalized eigenvalues of **A** and **B** have to be different. This condition is satisfied for the models considered in this thesis and the considered parameter values. If this condition is not met, it may be useful to consider the Jordan canonical form of Λ that provides a decomposition of Λ as the sum of a diagonal matrix and a nilpotent matrix that offers another path for calculating the limit of this series. Such generalizations need, however, deeper considerations and are left for future research.

where

$$
{}_2F_1(a,b;c;D) := \begin{pmatrix} {}_2F_1(a,b;c;\lambda_{n_s+1}^{-1}) & & \\ & \ddots & \\ & & {}_2F_1(a,b;c;\lambda_n^{-1}) \end{pmatrix}. \tag{B.49}
$$

Note that the derivation of (B.48) involves the already mentioned useful identity (A.9). Equation (B.49) further illustrates the usefulness of assuming Λ to be diagonalizable. Essentially, the limit of the series can then be calculated by applying the Gaussian hypergeometric function to the eigenvalues of the matrix Λ. If Λ is not diagonalizable, an alternative effort must be made to calculate this limit.

Note that (B.49) is well-defined as $|\lambda_i| > 1$ for all $i = n_s + 1, \dots, n$, see Appendix A.2.2. [1042] ⋊

Now, (B.46) expresses the auxiliary variable u_t as function of the exogenous processes $(\varepsilon_t^g)_{t\in\mathbb{Z}}$ and $(\varepsilon_t^A)_{t\in\mathbb{Z}}$. To do the same for s_{t+1}, one has to find similar expressions for u_{t+1} and $\mathbb{E}_t u_{t+1}$ and then one has to plug these values into (B.36) to get the model's solution. The expression for u_{t+1} is obtained by shifting (B.46) one period ahead. Using the law of iterated expectations and doing a similar calculation as for u_t, one can show that

$$
\begin{aligned}
\mathbb{E}_t u_{t+1} &= -T_{22}^{-1} \sum_{k=0}^{\infty} \Lambda^k \hat{G}_2 \mathbb{E}_t \nu_{t+1+k}^A \\
&= -T_{22}^{-1} \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} \Lambda^k \alpha_{i+k+1} \hat{G}_2 e_1 \, \varepsilon_{t-i}^A \\
&= -T_{22}^{-1} \sum_{i=0}^{\infty} b_{i+1} \, \varepsilon_{t-i}^A,
\end{aligned} \tag{B.50}
$$

where the last line uses the infinite moving average coefficients introduced in Remark B.5.1. Inserting (B.46), (B.50) and $z_t = (\nu_t^A, \varepsilon_t^g)^T$ into (B.36) leads then to the desired expression for s_{t+1}.[1043] Since this expression is rather long and complicated and provides no further insights, it is not stated here.[1044]

The analytical difficulty of introducing long memory into a linearized DSGE model can be seen from a comparison between (B.39) and (B.46). In the former case, u_t depends linearly on the instantaneous shock z_t. This is also the case for the white noise trend shock ε_t^g in

[1042] To calculate the Gaussian hypergeometric function in a tractable numerical way, a Taylor series expansion is recommended and realized during in this thesis, see Appendix A.2.2 for details.
[1043] To recover the original variables from the auxiliary variables u_t and s_t one has to use (B.26) again.
[1044] The formula for the particular case of no long memory, i.e., $d = 0$, is given in (B.40).

(B.46). However, in the case of a transitory long memory TFP shock, u_t depends, on the history of all past shocks ε_t^A. This dependence comes essentially from (B.45), i.e., since the expectations on the exogenous process depend on the whole history of past shocks, so does the whole model solution.

Overall, in this section, a solution to the model with Cobb-Douglas (CD) preferences and long memory was derived along the method of Klein (2000). As outlined earlier, similar calculations can be carried out to derive a solution to the model with additive utility. The following theorem summarizes the necessary assumptions made for deriving the model's solution.

Theorem B.5.2 ——

Given the canonical form (B.22) *where the matrices* $\boldsymbol{A}, \boldsymbol{B}$ *and* $\boldsymbol{G}$ *are specified in Appendix B.4.4 and given the (stable) exogenous stochastic process* $z_t = \left(\nu_t^A, \varepsilon_t^g\right)^T$, *then under the following assumptions a unique and stable solution to* (B.22) *exists:*

- *The generalized Schur-decomposition of the matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *exists*
- *There is no complex number* λ *with* $|\lambda| = 1$ *and* $\det(\lambda \mathbf{A} - \mathbf{B}) = 0$, *i.e., there is no generalized eigenvalue with modulus 1*
- *The* $n_s \times n_b$ *matrix* Z_{11} *is invertible, i.e., necessarily* $n_s = n_b$
- *The matrix* $\Lambda = S_{22} T_{22}^{-1}$ *is diagonalizable.*

PROOF

See the derivation of the solution above. □

In the next section, a recursive law of motion for the IRFs of the model is derived from the formulas above.

B.5.5 Deriving IRFs of the Model with CD Preferences

Impulse-response functions show the effects of an initial shock in $t = 1$ on future outcomes of the model's variables. Definition 2.1.4 defines an Impulse-Response function as the sequence of coefficients of an infinite moving average process. Meyer-Gohde (2010) states that whenever the exogenous stochastic process has an infinite moving average representation, so does the solution of the linear rational expectations model (if it exists).[1045] Equations (B.47) and (B.50) state such infinite moving average representations for the auxiliary variable

[1045] See Meyer-Gohde (2010, p. 986), and also Taylor (1986, p. 2046) and Taylor (1986, Section 2.2. on pp. 2016ff.) for bivariate models.

u_t and its expectations. Instead of deriving the infinite moving average representation of the solution process directly, the infinite moving average representation of u_t and its expectations together with (B.36) is used to derive a recursion to compute the model's IRFs.

However, as outlined by Meyer-Gohde (2010), the reverse approach is also possible, i.e., one can set up a system of difference equations in the infinite moving average coefficients instead of the model's variables itself.[1046] Since Meyer-Gohde (2010) uses then again Klein (2000)'s method to solve for the infinite moving average coefficients, the approach followed in this thesis and by Meyer-Gohde (2010), appear to be quite similar.

To summarize, the model solution has an infinite moving average representation; thus, Definition 2.1.4 still holds in the context of the linearized DSGE models considered in this thesis, although the solution's moving average representation is not stated explicitly.

The following section derives the model's IRF of the analytically more demanding transitory TFP shock process ε_t^A and the second subsection outlines the procedure for the shock ε_t^g.

B.5.5.1 IRF of the Transitory Productivity Process

In order to calculate the IRF of a one standard deviation long memory shock, one has to set $\varepsilon_t^g \equiv 0$ and

$$\varepsilon_t^A = \begin{cases} \sigma_{\varepsilon^A} & \text{if } t = 1 \\ 0 & \text{else} \end{cases} \tag{B.51}$$

The IRF is calculated by assuming that the economy is in the steady state before the shock materializes. Since all variables are expressed in percentage deviations from the steady state, the initial values of the backward-looking variables are equal to zero, i.e., $x_0^b = 0$. Further, (B.46) involves for $t = 0$ only values of ε_t^A for $t < 0$. Due to (B.51), these values are also assumed to be zero. Hence, $u_0 = 0$. By (B.37), the initial value s_0 is also equal to zero. Moreover, inserting (B.51) into (B.50) leads to $\mathbb{E}_0(u_1) = 0$.

For $t > 0$, plugging (B.51) into (B.47) results in

$$u_t = -\sigma_{\varepsilon^A} T_{22}^{-1} b_{t-1} = -\sigma_{\varepsilon^A} \alpha_{t-1} T_{22}^{-1} P \; {}_2F_1(1, t-1+d; t; D) P^{-1} \hat{G}_2 e_1, \tag{B.52}$$

where the second identity follows from (B.48). Similarly, plugging (B.51) into (B.50) leads

[1046] See Meyer-Gohde (2010, pp. 986f.).

to

$$\mathbb{E}_t(u_{t+1}) = \begin{cases} 0 & \text{if } t = 0 \\ -\sigma_{\varepsilon^A}\alpha_t T_{22}^{-1} P \; {}_2F_1(1, t+d; t+1; D) P^{-1} \hat{G}_2 e_1 & \text{if } t \geq 1 \end{cases}. \tag{B.53}$$

Given the already calculated values of u_t, s_t and $\mathbb{E}_t(u_{t+1})$, the next iteration step for the auxiliary variable s_{t+1} in accordance with (B.36) is given by

$$\begin{aligned} s_{t+1} = {} & S_{11}^{-1} T_{11} s_t + S_{11}^{-1} T_{12} u_t - S_{11}^{-1} S_{12} \mathbb{E}_t(u_{t+1}) \\ & + Z_{11}^{-1} Z_{12}(\mathbb{E}_t(u_{t+1}) - u_{t+1}) + \begin{cases} 0 & \text{if } t = 0 \\ \sigma_{\varepsilon^A} S_{11}^{-1} \alpha_{t-1} \hat{G}_1 e_1 & \text{if } t \geq 1 \end{cases} \end{aligned} \tag{B.54}$$

Along the shock process, (B.51) one further has that $\mathbb{E}_t(u_{t+1}) - u_{t+1} = 0$ for $t > 0$. Thus, (B.54) simplifies to

$$s_{t+1} = \begin{cases} -Z_{11}^{-1} Z_{12} u_1 & \text{if } t = 0 \\ S_{11}^{-1} T_{11} s_t + S_{11}^{-1} T_{12} u_t - S_{11}^{-1} S_{12} \mathbb{E}_t(u_{t+1}) + \sigma_{\varepsilon^A} S_{11}^{-1} \alpha_{t-1} \hat{G}_1 e_1 & \text{if } t \geq 1 \end{cases}. \tag{B.55}$$

In summary, the recursion starts with computing u_t from (B.52) and $\mathbb{E}_t u_{t+1}$ from (B.53). Then s_{t+1} is computed from these values with (B.55). These steps are then repeated for each instant of time t. To get the IRF of the original model variables, one has to transform the auxiliary variables s_t and u_t back into the model's variables with (B.26), i.e.,

$$\begin{aligned} x_t^b &= Z_{11} s_t + Z_{12} u_t \\ x_t^f &= Z_{21} s_t + Z_{22} u_t. \end{aligned}$$

Finally, in order to get the IRFs of the static variables, one has to apply (B.14) with the matrices given in (B.19).

B.5.5.2 IRF of the Growth Process

A recursive law of motion of the growth process's IRF can be obtained similarly to the one illustrated in the previous subsection. However, these expressions are computationally simpler because the growth shock was assumed to be a white noise process. In order to calculate the IRF of a one standard deviation growth shock, one has to set $\nu_t^A \equiv 0$ and

$$\varepsilon_t^g = \begin{cases} \sigma_{\varepsilon^g} & \text{if } t = 1 \\ \\ 0 & \text{else} \end{cases} . \tag{B.56}$$

As above, the IRF is calculated assuming the economy to be in the steady state before the shock materializes. Since all variables are expressed in percentage deviations from the steady state, the initial values of the backward-looking variables are equal to zero, i.e., $x_0^b = 0$. Plugging (B.56) in (B.46) leads to $u_0 = 0$. Further, (B.50) does not depend on any value of ε_t^g, i.e., $\mathbb{E}_t(u_{t+1}) = 0$ for all $t > 0$. By (B.37), the initial value s_0 is also equal to zero.

For $t > 0$, plugging (B.56) into (B.46) results in

$$u_t = \begin{cases} -\sigma_{\varepsilon^g} T_{22}^{-1} \hat{G}_2 e_2 & \text{if } t = 1 \\ \\ 0 & \text{else} \end{cases} . \tag{B.57}$$

Given the values of u_t from (B.57), s_t and $\mathbb{E}_t(u_{t+1}) = 0$, the next iteration step for the auxiliary variable s_{t+1} in accordance with (B.36) is given by

$$s_{t+1} = \begin{cases} -Z_{11}^{-1} Z_{12} u_1 & \text{if } t = 0 \\ S_{11}^{-1} T_{11} s_t + S_{11}^{-1} T_{12} u_t + \sigma_{\varepsilon^g} S_{11}^{-1} \hat{G}_2 e_2 & \text{if } t = 1 \\ S_{11}^{-1} T_{11} s_t & \text{if } t > 1 \end{cases} .$$

To obtain the response of the original model variables to a one standard deviation shock in the growth process, one has to use (B.26) again.

B.6 Comparison of Both Models without Growth

Figure B.1 compares the responses of both models to a 1% transitory short memory technology shock. The dashed lines refer to the responses of the model with additive separable utility of Figure 4.1. As mentioned in Section 4.4.3, consumption rises initially stronger in the case of CD utility than in the case of additive separable utility. However, the negative effect on employment is smaller because the dependence of labor supply on consumption expenditures is less pronounced in the model with CD utility. For all values of ϱ_A, the initial response of hours worked is positive in the model with CD utility. Overall, the responses of factor prices are quite similar in both models. Furthermore, due to the overall higher employment in the case of CD utility, the effect on output is also more pronounced and allows for additional investment expenditures in addition to expanded consumption expenditures, which lead to an intensified capital accumulation.

The same pattern can be observed in Figure B.2, which compares the models' responses to a pure long memory shock. An increase in the long memory parameter leads to smaller differences between the two models compared to an increase in the short memory parameter ϱ_A. Interestingly, the shape of consumption's response is characterized by the initial drop of TFP in the model with CD utility but remains stronger compared to the model with additive separable utility function.

Figure B.3 finally compares the responses to an ARFIMA$(1, d, 0)$ TFP shock. The reduced negative employment effects in the model with CD utility become apparent. The initial response of employment keeps to be negative in the model with CD utility for $\varrho_A = 0.95$ and $d = 0.4$, but afterward, employment overshoots its steady state value quickly. This behavior indicates that the upward shift in the labor demand curve induced by the increasing TFP in the first periods outweighs the upward shift in the labor supply curve caused by increasing consumption expenditures. In the model with additive separable utility function, the shift in the labor supply curve is stronger, yielding an overall negative effect on employment.

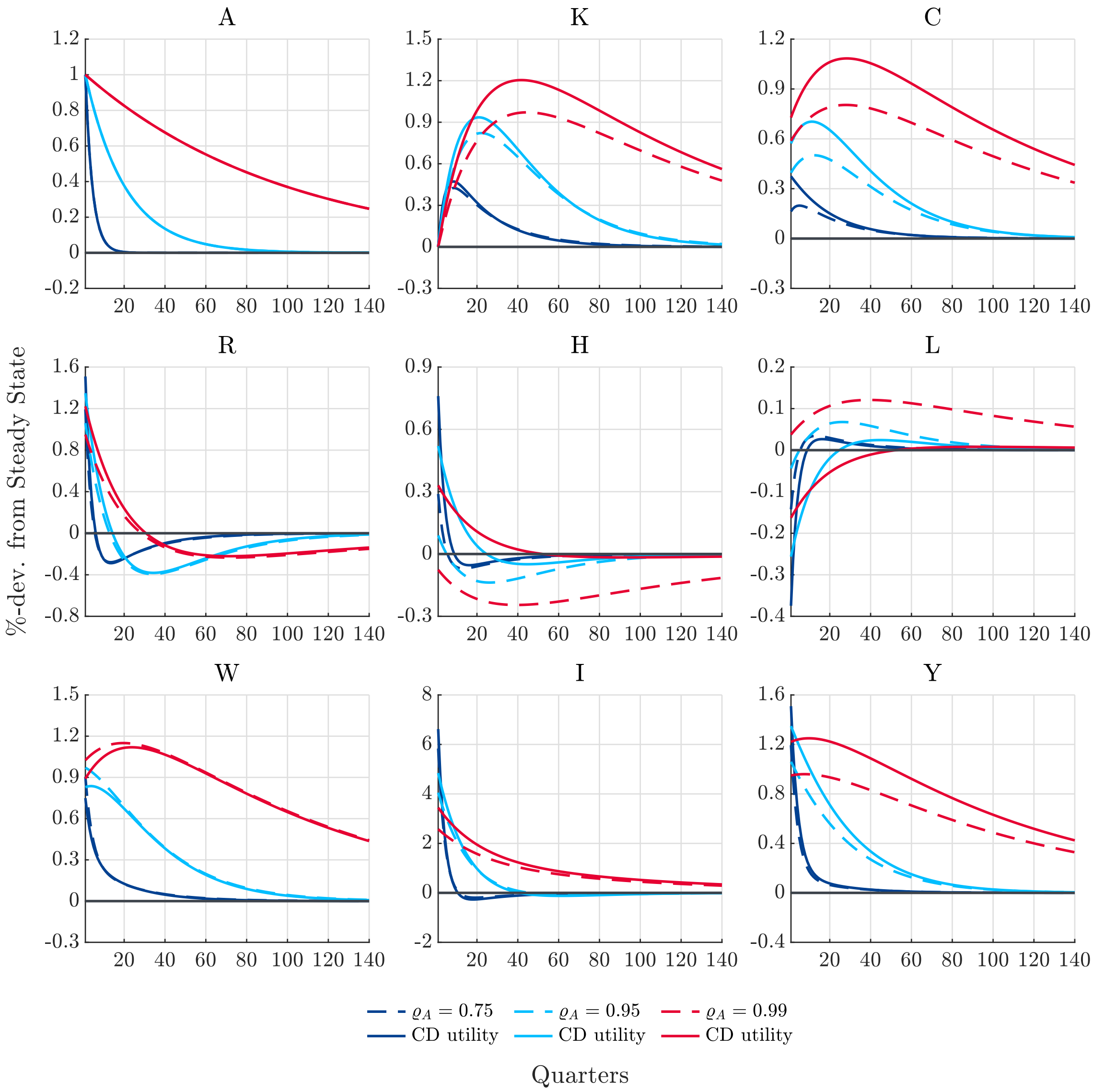


Figure B.1: Comparison of the responses of both models to a 1% transitory short memory technology shock for various values of ϱ_A. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Dashed lines belong to the model with additive separable utility function, and solid lines to the model with CD utility. Note the different axis scaling of the vertical axis.

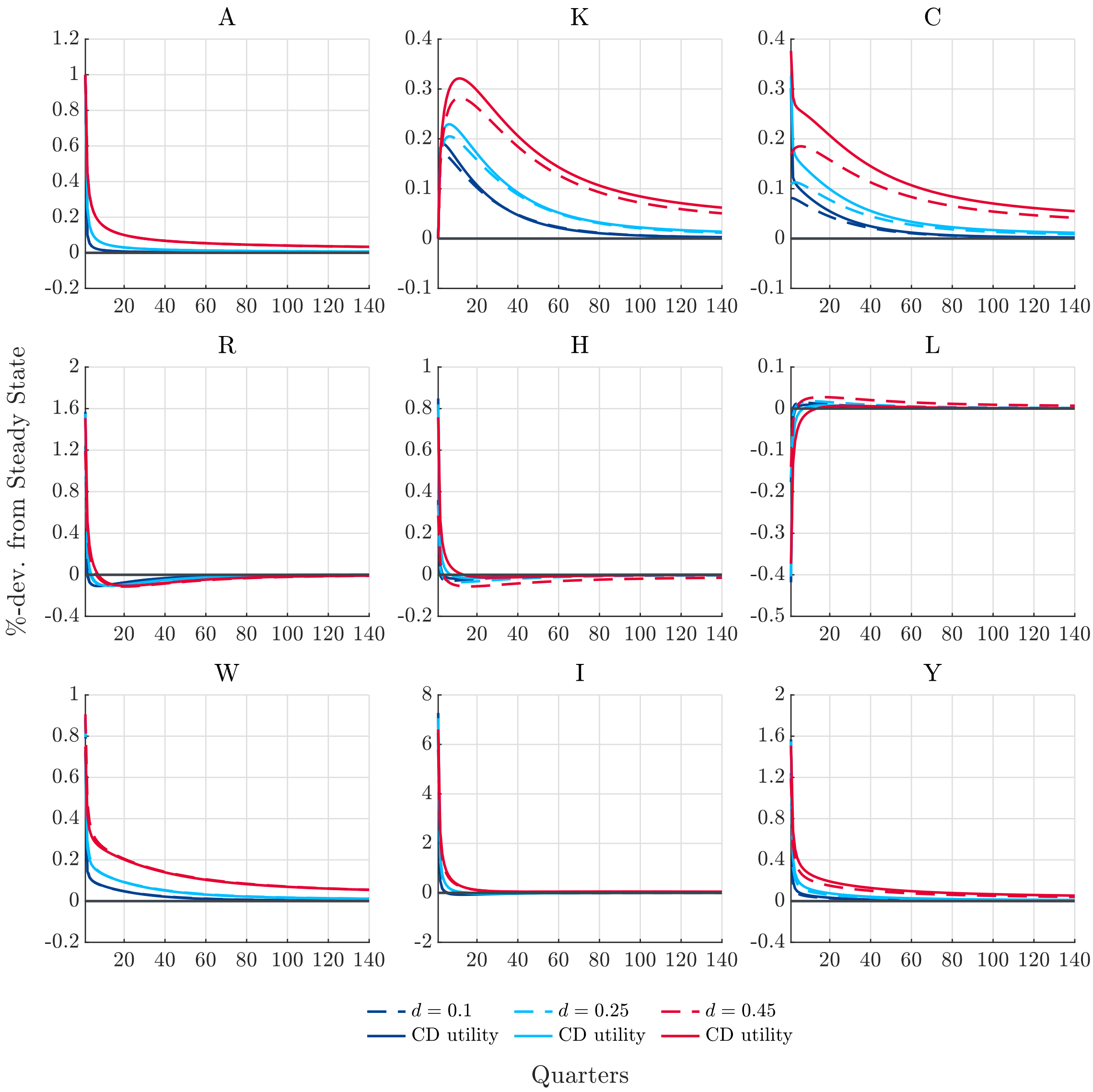


Figure B.2: Comparison of the responses of both models to a 1% transitory long memory technology shock for various values of d. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Dashed lines belong to the model with additive separable utility function, and solid lines to the model with CD utility. Note the different axis scaling of the vertical axis.

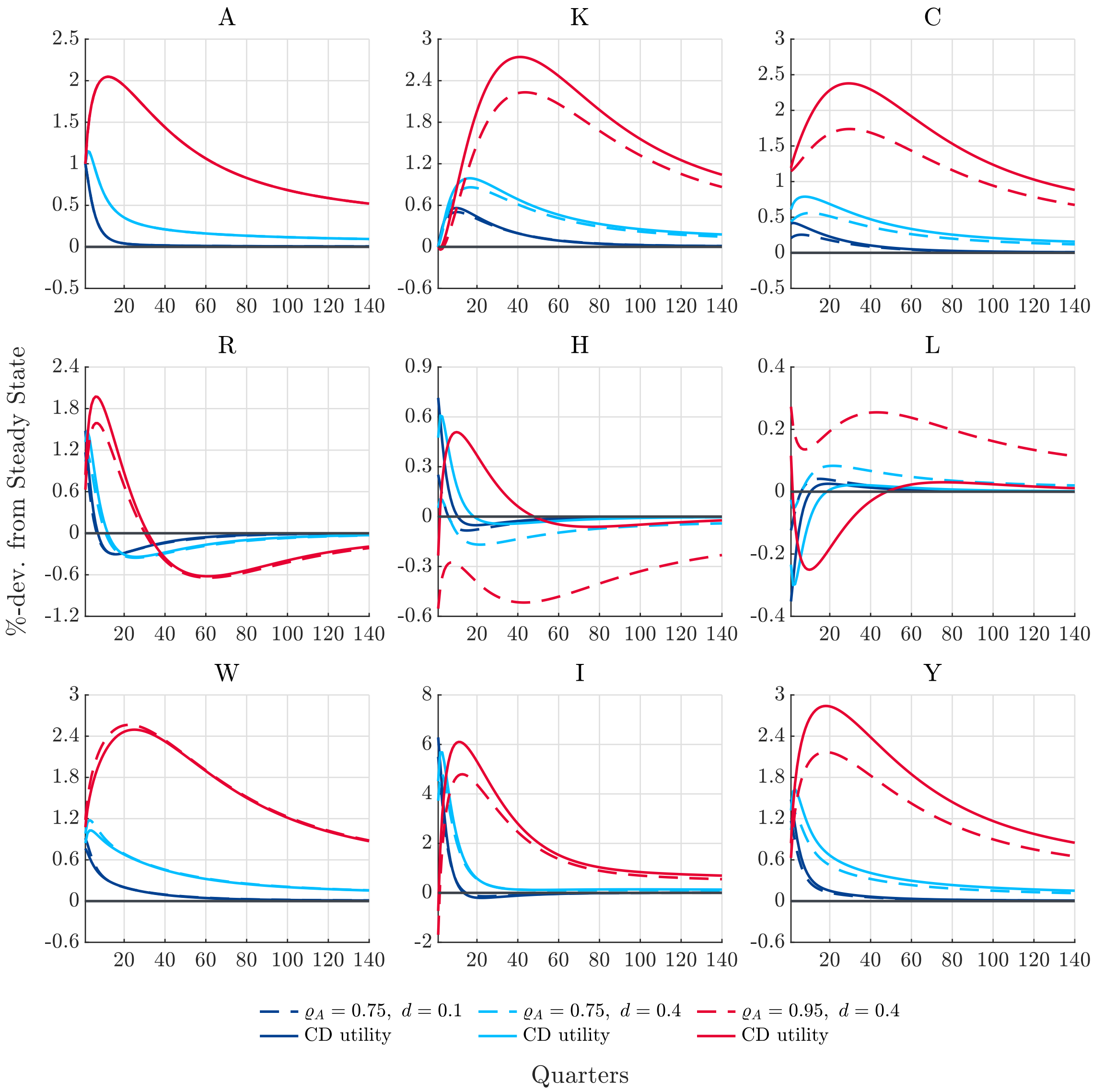


Figure B.3: Comparison of the responses of both models to a 1% transitory long memory technology shock for various values of ϱ_A and d. The vertical axes report percentage deviations from the respective steady state value. The horizontal axes report quarters. The dark gray line in each subfigure marks the zero line where the variables are at their steady state values. Dashed lines belong to the model with additive separable utility function, and solid lines to the model with CD utility. Note the different axis scaling of the vertical axis.

C

Appendix to Chapter 5

C.1 On the Itô Calculus

In this appendix, some useful formulas for handling stochastic differential equations are stated and partly derived. Some of them are used in the text; others are needed for the proofs of Chapter 5's lemmas and propositions given in Appendix C.2.

Before some forms of the Itô formula are outlined, recall the definition of a so-called Itô process. A continuous-time stochastic process $X = (X_t)_{t \geq 0}$ is called Itô process, if X can be written as

$$X_t - X_0 = \int_0^t \mu_s^X \, ds + \int_0^t \sigma_s^X \, dW_s, \tag{C.1}$$

where $(W_t)_{t \geq 0}$ is a Brownian motion as defined in Definition 2.5.1 and $(\mu_s^X)_{s \geq 0}$ and $(\sigma_s^X)_{s \geq 0}$ are certain stochastic processes.[1047] The integral with respect to the Brownian motion W, is called Itô integral. Equation (C.1) can equally be written in the shorthand differential notation, i.e.,[1048]

$$dX_t = \mu_t^X \, dt + \sigma_t^X \, dW_t.$$

The Itô calculus offers nice analytic tools to deal with Itô processes. As outlined in the following, it can be shown that stochastic integrals with respect to Itô processes can easily

[1047] See Shreve (2004, Definition 4.4.3. on p. 143). In order to be well-defined, the stochastic processes involved in (C.1) have to satisfy some regularity and integrability conditions. For example, they have to be adapted to the filtration generated by the Brownian motion $(W_t)_{t \geq 0}$. The details can be found in the mentioned definition of Shreve (2004).

[1048] See Shreve (2004, Equation (4.4.18) on p. 144).

be defined and that functions of Itô processes are again Itô processes. To see this, let $(\lambda_s)_{s\geq 0}$ be a stochastic process and X as in (C.1), then the integral $\int_0^t \lambda_s \, dX_s$ is defined by[1049]

$$\int_0^t \lambda_s \, dX_s := \int_0^t \lambda_s \mu_s^X \, ds + \int_0^t \lambda_s \sigma_s^X \, dW_s. \tag{C.2}$$

For an Itô process X of the form (C.1), let $\langle X \rangle_t := \displaystyle\int_0^t (\sigma_s^X)^2 \, ds$ be the so-called quadratic variation process of X.[1050]

The Itô formula says that if X is an Itô process as in (C.1) and if $f : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ is a time-dependent function whose first-order derivative with respect to t ($\partial f / \partial t$) and whose first-order and second-order derivatives with respect to x ($\partial f / \partial x, \partial^2 f / \partial^2 x$) exist, then the process $Y_t = f(t, X_t)$ is again an Itô process and it yields[1051]

$$f(t, X_t) - f(0, X_0) = \int_0^t \frac{\partial}{\partial t} f(s, X_s) \, ds + \int_0^t \frac{\partial}{\partial x} f(s, X_s) \, dX_s + \frac{1}{2} \int_0^t \frac{\partial^2}{\partial^2 x} f(s, X_s) \, d\langle X \rangle_s. \tag{C.3}$$

By using the definition of the quadratic variation process and (C.2) to evaluate the integral with respect to X, (C.3) becomes

$$\begin{aligned} f(t, X_t) - f(0, X_0) = \int_0^t \left(\frac{\partial}{\partial t} f(s, X_s) + \frac{\partial}{\partial x} f(s, X_s) \mu_s^X + \frac{1}{2} \frac{\partial^2}{\partial^2 x} f(s, X_s) (\sigma_s^X)^2 \right) ds \\ + \int_0^t \frac{\partial}{\partial x} f(s, X_s) \sigma_s^X \, dW_s. \end{aligned}$$

Consequently, $Y_t = f(t, X_t)$ is again an Itô process and an expression like (C.1) can be derived easily from (C.3), i.e.,

$$Y_t - Y_0 = \int_0^t \mu_s^Y \, ds + \int_0^t \sigma_s^Y \, dW_s,$$

where $\mu_s^Y = \dfrac{\partial}{\partial t} f(s, X_s) + \dfrac{\partial}{\partial x} f(s, X_s) \mu_s^X + \dfrac{1}{2} \dfrac{\partial^2}{\partial^2 x} f(s, X_s) (\sigma_s^X)^2$ and $\sigma_s^Y = \dfrac{\partial}{\partial x} f(s, X_s) \sigma_s^X$.

Note that (C.3) can also be applied if f does not depend on t. Then, one can omit the first argument of f and set $\dfrac{\partial}{\partial t} \equiv 0$.

The one-dimensional Itô formula (C.3) can be generalized to higher dimensions. For

[1049] See Shreve (2004, Definition 4.4.5. on p. 145).

[1050] This definition is made for the sake of simplicity. In general, it follows from Itô's integration calculus that the quadratic variation of X has this form, see Shreve (2004, Lemma 4.4.4. on p. 143). The quadratic variation can be defined for general time-dependent functions; see Shreve (2004, Definition 3.4.1. on p. 101). However, such generality is not needed for the context of this thesis.

[1051] See Shreve (2004, Theorem 4.4.6. on p. 146).

this thesis, a simplified two-dimensional Itô formula appears to be sufficient. Let X and Y be general Itô processes as in (C.1) with corresponding drift and volatility processes μ_t^X, μ_t^Y and σ_t^X, σ_t^X, respectively. Additionally, assume that the Brownian motion in their respective Itô integrals is the same and $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a function such that all partial derivatives up to the second order exist, then, the process $f(X_t, Y_t)$ is again an Itô process and it yields[1052]

$$f(X_t, Y_t) - f(X_0, Y_0) = \int_0^t \frac{\partial}{\partial x} f(X_s, Y_s)\, dX_s + \int_0^t \frac{\partial}{\partial y} f(X_s, Y_s)\, dY_s$$
$$+\frac{1}{2}\left(\int_0^t \frac{\partial^2}{\partial^2 x} f(X_s, Y_s)\, d\langle X\rangle_s + 2\int_0^t \frac{\partial^2}{\partial x \partial y} f(X_s, Y_s)\, d\langle X, Y\rangle_s + \int_0^t \frac{\partial^2}{\partial^2 y} f(X_s, Y_s)\, d\langle Y\rangle_s\right), \tag{C.4}$$

where $\langle X, Y\rangle_s := \int_0^t \sigma_s^X \sigma_s^Y\, ds$ is the cross-variation of X and Y. The reason why the integral associated with $\partial^2 f/\partial x \partial y$ is multiplied with two in (C.4) is the symmetry of $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$ and $\langle X, Y\rangle_t = \langle Y, X\rangle_t$.[1053]

By plugging the definitions of the quadratic variation, the cross-variation process, and (C.2) into (C.4), one can deduce that $f(X_t, Y_t)$ is again an Itô process. This step is not carried out here, but instead, (C.4) is applied to two examples frequently used in Chapter 5.

In the following, let X and Y be given by

$$\frac{dX_t}{X_t} = \bar{\mu}_t^X\, dt + \bar{\sigma}_t^X\, dW_t \text{ and } \frac{dY_t}{Y_t} = \bar{\mu}_t^Y\, dt + \bar{\sigma}_t^Y\, dW_t,$$

i.e., $\mu_t^A = A_t \bar{\mu}_t^A$ and $\sigma_t^A = A_t \bar{\sigma}_t^A$ for $A \in \{X, Y\}$. Consider the function $f(x, y) = xy$, then (C.4) becomes

$$X_t Y_t - X_0 Y_0 = \int_0^t Y_s\, dX_s + \int_0^t X_s\, dY_s + \langle X, Y\rangle_t.$$

Using (C.2) to replace the integrals with respect to X and Y and the definition of the cross-variation process leads to

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s Y_s \bar{\mu}_s^X\, ds + \int_0^t X_s Y_s \bar{\sigma}_s^X\, dW_s + \int_0^t X_s Y_s \bar{\mu}_s^Y\, ds$$
$$+ \int_0^t X_s Y_s \bar{\sigma}_s^Y\, dW_s + \int_0^t X_s Y_s \bar{\sigma}_s^X \bar{\sigma}_s^Y\, ds.$$

[1052] See Lamberton and Lapeyre (2008, Proposition 3.4.18 on pp. 71f.).
[1053] See Lamberton and Lapeyre (2008, Remark 3.4.19 on p. 72).

Rearranging implies,

$$X_tY_t - X_0Y_0 = \int_0^t X_sY_s\left(\bar{\mu}_s^X + \bar{\mu}_s^Y + \bar{\sigma}_s^X\bar{\sigma}_s^Y\right) ds + \int_0^t X_sY_s\left(\bar{\sigma}_s^X + \bar{\sigma}_s^Y\right) dW_s. \tag{C.5}$$

Rewriting (C.5) in differential notation leads to

$$d(X_tY_t) = X_tY_t\left(\bar{\mu}_t^X + \bar{\mu}_t^Y + \bar{\sigma}_t^X\bar{\sigma}_t^Y\right) dt + X_tY_t\left(\bar{\sigma}_t^X + \bar{\sigma}_t^Y\right) dW_t$$

or

$$\frac{d(X_tY_t)}{X_tY_t} = \left(\bar{\mu}_t^X + \bar{\mu}_t^Y + \bar{\sigma}_t^X\bar{\sigma}_t^Y\right) dt + \left(\bar{\sigma}_t^X + \bar{\sigma}_t^Y\right) dW_t. \tag{C.6}$$

Brunnermeier and Sannikov (2016) refer to (C.6) as "Itô's formula for products"[1054].

Now, let $f(x, y) = x/y$, then (C.4) becomes

$$\frac{X_t}{Y_t} - \frac{X_0}{Y_0} = \int_0^t \frac{1}{Y_s}\, dX_s - \int_0^t \frac{X_s}{Y_s^2}\, dY_s + \frac{1}{2}\left(-2\int_0^t \frac{1}{Y_s^2}\, d\langle X, Y\rangle_s + 2\int_0^t \frac{X_s}{Y_s^3}\, d\langle Y\rangle_s\right).$$

Again, using (C.2) to replace the integrals with respect to X and Y and the definitions of the variation processes, leads to

$$\int_0^t \frac{X_s}{Y_s}\left(\bar{\mu}_s^X - \bar{\mu}_s^Y - \bar{\sigma}_s^X\bar{\sigma}_s^Y + \left(\bar{\sigma}_s^Y\right)^2\right) ds + \int_0^t \frac{X_s}{Y_s}\left(\bar{\sigma}_s^X - \bar{\sigma}_s^Y\right) dW_s. \tag{C.7}$$

By rewriting (C.7) in differential form, one obtains "Itô's formula for ratios"[1055].

$$\frac{d(X_t/Y_t)}{X_t/Y_t} = \left(\bar{\mu}_t^X - \bar{\mu}_t^Y - \bar{\sigma}_t^X\bar{\sigma}_t^Y + \left(\bar{\sigma}_t^Y\right)^2\right) dt + \left(\bar{\sigma}_t^X - \bar{\sigma}_t^Y\right) dW_t. \tag{C.8}$$

C.2 Proofs

C.2.1 Proof of Lemma 5.2.1

Since $dK_t = \left(\Phi(\iota) - \delta + \sigma\varphi_t^{H,\varepsilon}\right) K_t\, dt + \sigma\varepsilon^{H-1/2} K_t\, dW_t$, an application of Itô's formula (see (C.3) in Appendix C.1) to $f(x) = \log(x)$ implies

$$\log(K_t) - \log(K_0) = \int_0^t \frac{1}{K_s}\, dK_s + \frac{1}{2}\int_0^t -\frac{1}{K_s^2}(\sigma\varepsilon^{H-1/2})^2 K_s^2\, ds. \tag{C.9}$$

[1054] See Brunnermeier and Sannikov (2016, p. 1506).
[1055] See Brunnermeier and Sannikov (2016, p. 1508).

Rearranging (C.9) leads to

$$\log\left(\frac{K_t}{K_0}\right) = \int_0^t \left(\Phi(\iota) - \delta + \sigma\varphi_s^{H,\varepsilon} - \frac{(\sigma\varepsilon^{H-1/2})^2}{2}\right) ds + \int_0^t (\sigma\varepsilon^{H-1/2})\, dW_s. \tag{C.10}$$

By applying the exponential on both sides of (C.10) and using $\int_0^t dW_s = W_t$, one obtains

$$K_t = K_0 \exp\left(\left(\Phi(\iota) - \delta - \frac{(\sigma\varepsilon^{H-1/2})^2}{2}\right) t + \int_0^t \sigma\varphi_s^{H,\varepsilon}\, ds + \sigma\varepsilon^{H-1/2} W_t\right). \tag{C.11}$$

The stated expression for K_t follows from Lemma 5.1.2, since $\int_0^t \sigma\varphi_s^{H,\varepsilon}\, ds + \sigma\varepsilon^{H-1/2}W_t = \sigma Z_t^{H,\varepsilon}$. The corresponding expression for K_t^{BS} follows from (C.11) by setting $H = 1/2$.
□

C.2.2 Proof of Proposition 5.3.2

To solve the utility maximization problem of the expert i, the analogy of this problem to classical problems in finance is used.[1056] Especially the finance results given in Karatzas and Shreve (1998, Chapter 1 and 3) are applied to the to model context of Chapter 5. Let initially $(\iota_t^i)_{t\geq 0}$ be a given process of re-investment rates. Then the utility maximization problem of expert i is the same as for an investor in a financial market with the following investment possibilities:

A risk-free bond S^0 with price process given by

$$dS_t^0 = S_t^0 r_t\, dt$$

and a risky stock S^1 whose price process is given by

$$dS_t^1 = \mu_t^S S_t^1 dt + \sigma_t^S S_t^1\, dW_t,$$

with

$$\mu_t^S = \left[\frac{(a - \iota_t^i)}{q_t} + \left(\Phi(\iota_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2}\sigma\sigma_t^q\right)\right] \text{ and } \sigma_t^S = (\sigma\varepsilon^{H-1/2} + \sigma_t^q).$$

Consider, in the following, a portfolio process $\pi_t = (\pi_t^0, \pi_t^1)$, where π_t^0 and π_t^1 denote the

[1056] According to Karatzas and Shreve (1998, Definition 7.2 on p. 28), certain technical integrability assumptions are required to solve the maximization problems of experts (and households). Specifically, for any finite time horizon $T > 0$, it is necessary that $\int_0^T |r_t|, dt < \infty$, $\int_0^T |\mu_t^q|, dt < \infty$, $\int_0^T \left|\iota_t^i\right|, dt < \infty$, and $\int_0^T (\sigma_t^q)^2, dt < \infty$ almost surely. These assumptions ensure that the relevant quantities are all well-behaved.

value of bonds and stocks held by the investor, respectively. Following Karatzas and Shreve (1998), the wealth process of the investor with consumption process $(c_t)_{t\geq 0}$ is given by[1057]

$$dn_t = n_t r_t \, dt + \pi_t^1(\mu_t^S - r_t) \, dt + \pi_t^1 \sigma_t^S \, dW_t - c_t.$$

Let $x_t = \pi_t^1 / n_t$ be the share of wealth invested into the stock. Then, one has

$$dn_t = n_t r_t \, dt + (\mu_t^S - r_t) x_t n_t \, dt + \sigma_t^S x_t n_t \, dW_t - c_t,$$

which corresponds exactly to (5.21). Therefore, the maximization problem of an investor in the financial market described by the bond process S_0 and the stock process S_1 is the same as for expert i in the model economy of Chapter 5.

Hence, one can apply Theorem 3.9.11 of Karatzas and Shreve (1998) to obtain experts i's optimal consumption process $(\hat{c}_t^i)_{t\geq 0}$ and optimal wealth process $(\hat{n}_t^i)_{t\geq 0}$. Recall that $n_0^i = q_0 \bar{k}_0^i$ is expert i's initial wealth, then

$$\hat{c}_t^i = \rho e^{-\rho t} n_0^i \frac{1}{\xi_t^i} \tag{C.12}$$

and[1058]

$$\hat{n}_t^i = \frac{1}{\xi_t^i} \mathbb{E}\left[\int_t^\infty \xi_u^i \hat{c}_u^i \, du \middle| \mathcal{F}_t \right]. \tag{C.13}$$

Plugging the expression of optimal consumption (C.12) into the expression of optimal wealth (C.13) results in

$$\hat{n}_t^i = \frac{1}{\xi_t^i} \int_t^\infty n_0^i \rho e^{-\rho u} \, du = e^{-\rho t} n_0^i \frac{1}{\xi_t^i}. \tag{C.14}$$

Comparing (C.14) and (C.12), one can see that it is optimal for expert i to consume the fraction ρ of his wealth, i.e.,

$$\hat{c}_t^i = \rho \hat{n}_t^i. \tag{C.15}$$

Applying Itô's formula (see (C.3) in Appendix C.1) to (C.14) and using (5.29) leads to

[1057] This follows from Karatzas and Shreve (1998, Equation (3.3) on p. 11) if one replaces their $d\Gamma(t)$ with $-c_t \, dt$. This replacement is carried out by Karatzas and Shreve (1998, p. 137) as well.

[1058] Here, $\mathcal{F}_t$ again refers to a sigma algebra over Ω. More precisely, the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ is generated by the Brownian motion $(W_t)_{t\geq 0}$, augmented by the null-sets of the probability measure $\mathbb{P}$. For technical details, see Karatzas and Shreve (1998, p. 2).

the dynamics of optimal wealth[1059]

$$\begin{aligned} d\hat{n}_t^i &= -\rho \hat{n}_t^i \, dt - e^{-\rho t} n_0^i \left(\frac{1}{\xi_t^i}\right)^2 d\xi_t^i + \frac{1}{2} e^{-\rho t} n_0^i \frac{2}{(\xi_t^i)^3} \, d\langle \xi^i \rangle_t \\ &= -\rho \hat{n}_t^i \, dt - e^{-\rho t} n_0^i \left(\frac{1}{\xi_t^i}\right)^2 \left[-r_t \xi_t^i \, dt - \vartheta_t^i \xi_t^i \, dW_t\right] + e^{-\rho t} n_0^i \frac{1}{(\xi_t^i)^3} (\vartheta_t^i)^2 (\xi_t^i)^2 \, dt \\ &= \left[(\vartheta_t^i)^2 + r_t - \rho\right] \hat{n}_t^i \, dt + \vartheta_t^i \hat{n}_t^i \, dW_t. \end{aligned} \tag{C.16}$$

All expressions above yield for arbitrary re-investment rate processes $(\iota_t^i)_{t\geq 0}$. Expert i chooses the optimal re-investment rate process $(\hat{\iota}_t^i)_{t\geq 0}$ that maximizes

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t} \log\left(\hat{c}_t^i\right) dt\right]$$

under the conditions stated in (5.22). Since ρ does not depend on ι and consumption is proportional to wealth, see (C.15), this is equivalent to maximize

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t} \log\left(\hat{n}_t^i\right) dt\right]. \tag{C.17}$$

Using (C.16) and applying Itô's formula (see (C.3) in Appendix C.1) to $\log(\hat{n}_t^i)$, one obtains

$$\log\left(\hat{n}_t^i\right) = \log\left(\hat{n}_0^i\right) + \int_0^t \left(\frac{1}{2}(\vartheta_s^i)^2 + r_s - \rho\right) ds + \int_0^t \vartheta_s^i \, dW_s.$$

Inserting this back in (C.17), neglecting some constants and applying Fubini's Theorem to switch expectation and integration, one obtains[1060]

$$\int_0^\infty e^{-\rho t} \int_0^t \mathbb{E}\left[\frac{1}{2}(\vartheta_s^i)^2 + r_s - \rho\right] ds \, dt.$$

Since all experts are assumed to be price-takers, the interest rate r_t is independent of ι_t^i. Therefore, one has to maximize

$$\mathbb{E}\left[\frac{1}{2}(\vartheta_t^i)^2\right].$$

An ω-wise maximization leads to the first-order condition

$$\vartheta_t^i \left(\Phi'(\hat{\iota}_t^i) - \frac{1}{q_t}\right) = 0.$$

[1059] More specifically, Itô's formula is applied to the function $f(t,x) = n_0^i e^{-\rho t}/x$.
[1060] See Hassler (2016, Proposition 8.2 on p. 182).

By the concavity of Φ, it can easily be verified that the maximum is attained if

$$\Phi'(\hat{\iota}_t^i) = \frac{1}{q_t} \text{ or } \hat{\iota}_t^i = \Psi\left(\frac{1}{q_t}\right) \text{ as long as } \hat{\vartheta}_t^i > 0 \tag{C.18}$$

where Ψ is the inverse function of Φ'. Inserting the optimal re-investment rate process in (C.16) and (C.12) leads to the optimal wealth and consumption process of expert i. It remains to determine the optimal fraction of wealth invested into capital. On the one hand, the dynamics of the optimal wealth process is given by (C.16) if one replaces ϑ_t^i with $\hat{\vartheta}_t^i$. On the other hand, inserting (C.15) and (C.18) into the dynamics of the wealth process given by (5.21), one obtains

$$\begin{aligned} dn_t^i = &\left[\frac{(a-\hat{\iota}_t^i)}{q_t} + \left(\Phi(\hat{\iota}_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2}\sigma\sigma_t^q\right)\right] x_t^i n_t^i \, dt \\ &+ (1-x_t^i) r_t n_t^i \, dt - \rho n_t^i \, dt + (\sigma\varepsilon^{H-1/2} + \sigma_t^q) x_t^i n_t^i \, dW_t. \end{aligned} \tag{C.19}$$

A comparison of (C.16) with (C.19) leads to the equations

$$\begin{aligned} \left[(\hat{\vartheta}_t^i)^2 + r_t - \rho\right] n_t^i = &\left[\frac{(a-\hat{\iota}_t^i)}{q_t} + \left(\Phi(\hat{\iota}_t^i) - \delta + \sigma\varphi_t^{H,\varepsilon} + \mu_t^q + \varepsilon^{H-1/2}\sigma\sigma_t^q\right)\right] x_t^i n_t^i \\ &+ (1-x_t^i) r_t n_t^i - \rho n_t^i \end{aligned}$$

and $\hat{\vartheta}_t^i n_t^i = (\sigma\varepsilon^{H-1/2} + \sigma_t^q) x_t^i n_t^i$. Solving both equations for x_t^i leads to the optimal fraction of wealth to be invested in capital

$$x_t^i = \frac{\hat{\vartheta}_t^i}{(\sigma\varepsilon^{H-1/2} + \sigma_t^q)} \text{ if } \hat{\vartheta}_t^i > 0 \text{ and } x_t^i = 0 \text{ otherwise.} \tag{C.20}$$

□

C.2.3 Proof of Proposition 5.3.3

Recall that the wealth process follows (5.25). Let the discounted wealth process be defined by

$$\tilde{\tilde{n}}_t^j := \exp\left(-\int_0^t r_s \, ds\right) \tilde{n}_t^j. \tag{C.21}$$

By applying Itô's formula (see (C.3) in Appendix C.1) to $f(t,x) = \exp\left(-\int_0^t r_s \, ds\right) x$, one obtains

$$d\tilde{\tilde{n}}_t^j = -r_t \tilde{\tilde{n}}_t^j \, dt + \exp\left(-\int_0^t r_s \, ds\right)(\tilde{n}_t^j r_t - \tilde{c}_t^j) \, dt = -\exp\left(-\int_0^t r_s \, ds\right) \tilde{c}_t^j \, dt. \tag{C.22}$$

This implies

$$\bar{\tilde{n}}_0^j = \mathbb{E}\left[\tilde{\tilde{n}}_t^j + \int_0^t \exp\left(-\int_0^s r_u\,du\right)\tilde{c}_s^j\,ds\right] \geq \mathbb{E}\left[\int_0^t \exp\left(-\int_0^s r_u\,du\right)\tilde{c}_s^j\,ds\right],$$

since $\tilde{\tilde{n}}_0^j = \tilde{n}_0^j = \bar{\tilde{n}}_0^j$ and $\tilde{\tilde{n}}_t^j \geq 0$. Applying the monotone convergence theorem, one obtains[1061]

$$\mathbb{E}\left[\int_0^\infty \exp\left(-\int_0^s r_s\,du\right)\tilde{c}_s^j\,ds\right] \leq \bar{\tilde{n}}_0^j.$$

This equation can be interpreted as a budget constraint in the sense that total expected discounted future consumption expenditures cannot be larger than the initial wealth. It is clear that for the optimal consumption process, this constraint is binding. One can now set up a Lagrangian function with corresponding multiplier λ to solve the maximization problem:

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t}\log\!\left(\tilde{c}_t^j\right)dt\right] - \lambda\left(\bar{\tilde{n}}_0^j - \mathbb{E}\left[\int_0^\infty \exp\left(-\int_0^t r_s\,ds\right)\tilde{c}_t^j\,dt\right]\right).$$

Again, an ω-wise maximization leads to the first-order condition

$$e^{-\rho t}\frac{1}{\hat{\tilde{c}}_t^j} + \lambda\exp\left(-\int_0^t r_s\,ds\right) = 0$$

and thus

$$\hat{\tilde{c}}_t^j = -\frac{1}{\lambda}e^{-\rho t}\exp\left(\int_0^t r_s\,ds\right) = \bar{\tilde{n}}_0^j\rho e^{-\rho t}\exp\left(\int_0^t r_s\,ds\right), \tag{C.23}$$

where the last line uses the budget constraint to solve for $\lambda = -1/(\bar{\tilde{n}}_0^j\rho)$. Inserting (C.23) in (C.22), leads to

$$\tilde{\tilde{n}}_t^j = \tilde{\tilde{n}}_0^j - \int_0^t \exp\left(-\int_0^s r_u\,du\right)\hat{\tilde{c}}_s^j\,ds = \tilde{\tilde{n}}_0^j - \int_0^t \bar{\tilde{n}}_0^j\rho e^{-\rho s}\,ds = \bar{\tilde{n}}_0^j e^{-\rho t},$$

and finally with (C.21), one obtains

$$\hat{\tilde{n}}_t^j = \bar{\tilde{n}}_0^j e^{-\rho t}\exp\left(\int_0^t r_s\,ds\right).$$

The latter equation implies that $\hat{\tilde{c}}_t^j = \rho\hat{\tilde{n}}_t^j$. □

[1061] See, e.g., Klenke (2013, Satz 5.3 on p. 104).

C.2.4 Proof of Lemma 5.3.4

The optimal aggregate wealth of experts follows the dynamics

$$\begin{aligned} dN_t &= \left[\hat{\vartheta}_t^2 + r_t - \rho\right] N_t\, dt + \hat{\vartheta}_t N_t\, dW_t \\ &= \left[\left(\frac{\varepsilon^{H-1/2}\sigma}{\eta_t}\right)^2 + r_t - \rho\right] N_t\, dt + \frac{\varepsilon^{H-1/2}\sigma}{\eta_t} N_t\, dW_t, \end{aligned}$$

see (C.16). It follows from (5.31) that the value of the capital stock qK_t evolves according to

$$\begin{aligned} d(qK_t) &= \left(\Phi(\Psi(1/q)) - \delta + \sigma\varphi_t^{H,\varepsilon}\right) qK_t\, dt + \sigma\varepsilon^{H-1/2} qK_t\, dW_t \\ &= \left(r_t - \rho + \frac{(\varepsilon^{H-1/2}\sigma)^2}{\eta_t}\right) qK_t\, dt + \sigma\varepsilon^{H-1/2} qK_t\, dW_t. \end{aligned}$$

Applying Itô's formula for ratios (see (C.8) in Appendix C.1) leads to

$$\begin{aligned} \frac{d\eta_t}{\eta_t} = &\left[\left(\frac{\varepsilon^{H-1/2}\sigma}{\eta_t}\right)^2 + r_t - \rho - \left(r_t - \rho + \frac{(\varepsilon^{H-1/2}\sigma)^2}{\eta_t}\right) + (\varepsilon^{H-1/2}\sigma)^2 - \frac{(\varepsilon^{H-1/2}\sigma)^2}{\eta_t}\right] dt \\ &+ \left[\frac{\varepsilon^{H-1/2}\sigma}{\eta_t} - \varepsilon^{H-1/2}\sigma\right] dW_t. \end{aligned}$$

Rearranging terms proves the statement. □

C.2.5 Proof of Lemma 5.4.1

It follows directly from (5.44), that

$$\mathbb{E}K_t = K_0 \exp\left(\left(\Phi\left(\Psi(1/q)\right) - \delta - \frac{(\varepsilon^{H-1/2}\sigma)^2}{2}\right) t\right) \mathbb{E}\left(\exp\left(\sigma Z_t^{H,\varepsilon}\right)\right), \qquad \text{(C.24)}$$

i.e., it suffice to calculate $\mathbb{E}\left(\exp\left(\sigma Z_t^{H,\varepsilon}\right)\right)$.

Recall the definition of $Z^{H,\varepsilon}$ from (5.6), i.e., $Z_t^{H,\varepsilon} = \int_0^t (t - s + \varepsilon)^{H-1/2}\, dW_s$. It follows from the properties of the Itô integral that $Z^{H,\varepsilon}$ is normally distributed with mean zero and variance $\int_0^t (t-s+\varepsilon)^{2H-1}\, ds$.[1062] Thus, it yields

$$\operatorname{var}(Z_t^{H,\varepsilon}) = \int_0^t (t - s + \varepsilon)^{2H-1}\, ds = \frac{1}{2H}\left((t+\varepsilon)^{2H} - \varepsilon^{2H}\right).$$

[1062] See Shreve (2004, p. 173).

Consequently, the variable $\exp\left(\sigma Z_t^{H,\varepsilon}\right)$ is log-normally distributed with mean[1063]

$$\exp\left(\frac{\sigma^2}{4H}\left((t+\varepsilon)^{2H}-\varepsilon^{2H}\right)\right).$$

Plugging this back into (C.24) proofs the stated result. □

C.3 A Note on the Equilibrium Price

This appendix briefly outlines why (5.32) determines the model's equilibrium price uniquely. To see this, let within this section, $f(x) = a - \Psi(1/x)$ and $g(x) = \rho x$, i.e., $f(q_t)$ and $g(q_t)$ correspond to the right-hand and left-hand side of (5.32), respectively. The equilibrium price q_t is the solution to $f(q_t) = g(q_t)$.

Before the solution is established, one can see from the monotonicity of both functions f and g that it has to be unique if a solution exists. More specifically, assume that there are two solutions x_1 and x_2 to $f(x) = g(x)$ with $x_1 \neq x_2$. Without loss of generality, assume that $x_1 < x_2$, then

$$g(x_1) = f(x_1) > f(x_2) = g(x_2),$$

since f is monotonically decreasing. By the definition of g, however, the latter equation corresponds to $\rho x_1 > \rho x_2$, which contradicts the assumption of $x_1 < x_2$.

The existence of a solution is given in the following. Since Ψ is the inverse function of Φ' and $\Phi'' < 0$ by assumption, one knows that $\Psi' < 0$.[1064] Consequently, $f' < 0$ and, since $\rho > 0$, $g' > 0$. Recall, from Section 5.2.1 that $\Phi'(0) = 1$. Therefore, one knows that $\Psi(1) = 0$ and, thus, $f(1) = a$ and $g(1) = \rho$.

Obviously, if $a = \rho$, a solution is given by $q_t \equiv 1$. If $a < \rho$, then $f(1) = a < \rho = g(1)$. By the monotonicity of f, one thus knows that $f(x) > a$ for all $x < 1$. One has to find a value of $x = x^* < 1$, such that $g(x^*) > f(x^*)$. By assumption $x^* = a/\rho < 1$ and by the definition of g, $g(x^*) = a$. Since $f(x) > a$ for all $x < 1$, it follows that $g(x^*) < f(x^*)$. In summary, one has that

$$f(1) - g(1) < 0 \text{ and } f(x^*) - g(x^*) > 0.$$

Since both functions f and g are continuous in x, the intermediate value theorem implies

[1063] See Hassler (2016, p. 166).

[1064] Essentially, this follows from the fact that the inverse function shares the same monotonicity properties as the original function, see Ovchinnikov (2021, p. 81).

that there has to be $\hat{x} \in [x^*, 1]$ with $f(\hat{x}) - g(\hat{x}) = 0$, i.e., $\hat{x}$ is the unique solution and the corresponding equilibrium price satisfies $q_t \equiv \hat{x}$.[1065] An analogous argument can be established in the case of $a > \rho$.

Moreover, if $a < \rho$, the equilibrium price is smaller than 1; on the contrary, if $a > \rho$, the equilibrium price is larger than one.

[1065] See Ovchinnikov (2021, Theorem 3.18 on p. 77) for a statement of the intermediate value theorem.

References

Abadir, Karim M., Walter Distaso, and Liudas Giraitis (2007). "Nonstationarity-extended local Whittle estimation". In: *Journal of Econometrics* 141.2, pp. 1353–1384.

Abadir, Karim M. and Gabriel Talmain (2002). "Aggregation, Persistence and Volatility in a Macro Model". In: *The Review of Economic Studies* 69.4, pp. 749–779.

Abbritti, Mirko, Luis A. Gil-Alaña, Yuliya Lovcha, and Antonio Moreno (2016). "Term Structure Persistence". In: *Journal of Financial Econometrics* 14.2, pp. 331–352.

Adelman, Irma (1965). "Long cycles, fact or antifact?" In: *The American Economic Review* 55.3, pp. 444–463.

Adjemian, Stéphane, Houtan Bastani, Michel Juillard, Fréderic Karamé, Ferhat Mihoubi, Willi Mutschler, Johannes Pfeifer, Marco Ratto, Normann Rion, and Sébastien Villemot (2022). *Dynare: Reference Manual Version 5: Dynare Working Papers*. 72. CEPREMAP.

Adrian, Tobias and Nina Boyarchenko (2012). *Intermediary Leverage Cycles and Financial Stability*. Federal Reserve Bank of New York Staff Report No. 567.

Aguiar, Mark and Gita Gopinath (2007). "Emerging Market Business Cycles: The Cycle Is the Trend". In: *Journal of Political Economy* 115.1, pp. 69–102.

Aguiar, Mark and Gita Gopinath (2004). *Emerging Market Business Cycles: The Cycle is the Trend*. NBER Working Paper 10734. Cambridge, MA.

Andersson, Fredrik N. G. and Yushu Li (2020). "Are Central Bankers Inflation Nutters? An MCMC Estimator of the Long-Memory Parameter in a State Space Model". In: *Computational Economics* 55.2, pp. 529–549.

Andrews, Donald W. K. and Hong-Yuan Chen (1994). "Approximately Median-Unbiased Estimation of Autoregressive Models". In: *Journal of Business & Economic Statistics* 12.2, pp. 187–204.

Andrews, Donald W. K. and Patrik Guggenberger (2003). “A Bias-Reduced Log-Periodogram Regression Estimator for the Long-Memory Parameter”. In: *Econometrica* 71.2, pp. 675–712.

Andrews, George E., Richard Askey, and Ranjan Roy (1999). *Special functions*. Vol. 71. Encyclopedia of mathematics and its applications. Cambridge: Cambridge Univ. Press.

Ashley, Richard A. and Douglas M. Patterson (2010). “Apparent Long Memory in Time Series as an Artifact of a Time-Varying Mean: Considering Alternatives to the Fractionally Integrated Model”. In: *Macroeconomic Dynamics* 14.S1, pp. 59–87.

Assaf, Ata, Luis A. Gil-Alaña, and Khaled Mokni (2022). “True or spurious long memory in the cryptocurrency markets: evidence from a multivariate test and other Whittle estimation methods”. In: *Empirical economics* 63.3, pp. 1543–1570.

Baillie, Richard T. (1996). “Long memory processes and fractional integration in econometrics”. In: *Journal of Econometrics* 73.1, pp. 5–59.

Baillie, Richard T. and George Kapetanios (2016). “On the estimation of short memory components in long memory time series models”. In: *Studies in Nonlinear Dynamics & Econometrics* 20.4, pp. 365–375.

Bårdsen, Gunnar and Luca Fanelli (2015). “Frequentist Evaluation of Small DSGE Models”. In: *Journal of Business & Economic Statistics* 33.3, pp. 307–322.

Basak, Suleyman and Domenico Cuoco (1998). “An Equilibrium Model with Restricted Stock Market Participation”. In: *Review of Financial Studies* 11.2, pp. 309–341.

Bechlioulis, Alexandros P. and Sophocles N. Brissimis (2021). “Identifying key aspects of household behavior in a representative agent framework”. In: *Economic Modelling* 97, pp. 105–117.

Beran, Jan (1994). *Statistics for long-memory processes*. Vol. 61. Monographs on statistics and applied probability. New York: Chapman & Hall.

Beran, Jan, Yuanhua Feng, Sucharita Ghosh, and Rafal Kulik (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Berlin: Springer.

Biagini, Francesca, Yaozhong Hu, Bernt K. Øksendal, and Tusheng Zhang (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. London: Springer.

Biblica (2021). *Genesis 41 - New International Version (NIV): On the pages of Biblica*. URL: `https://www.biblica.com/bible/niv/genesis/41/` (visited on May 16, 2023).

Binder, Michael and M. Hashem Pesaran (1995). “Multivariate Rational Expectations Models and Macroeconometric Modeling: A Review and Some New Results”. In: *Handbook of Applied Econometrics. Volume I: Macroeconomics*. Ed. by M. Hashem Pesaran and Michael R. Wickens. Oxford, UK: Blackwell Publishing Ltd, pp. 139–187.

Bisaglia, Luisa (2002). "Model selection for long-memory models". In: *Quaderni di Statistica* 4, pp. 33–49.

Björk, Tomas and Henrik Hult (2005). "A note on Wick products and the fractional Black-Scholes model". In: *Finance and Stochastics* 9.2, pp. 197–209.

Black, Fischer and Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81.3, pp. 637–654.

Blanchard, Olivier Jean (2009). "The State of Macro". In: *Annual Review of Economics* 1.1, pp. 209–228.

Blanchard, Olivier Jean and Stanley Fischer (1993). *Lectures on macroeconomics*. 6th ed. Cambridge, Mass.: MIT Press.

Blanchard, Olivier Jean and Charles M. Kahn (1980). "The Solution of Linear Difference Models under Rational Expectations". In: *Econometrica* 48.5, pp. 1305–1311.

Bollerslev, Tim, Natalia Sizova, and George Tauchen (2012). "Volatility in Equilibrium: Asymmetries and Dynamic Dependencies". In: *Review of Finance* 16.1, pp. 31–80.

Boubaker, Heni, Giorgio Canarella, Rangan Gupta, and Stephen M. Miller (2021). "Long-memory modeling and forecasting: evidence from the U.S. historical series of inflation". In: *Studies in Nonlinear Dynamics & Econometrics* 25.5, pp. 289–310.

Brockwell, Peter J. and Richard A. Davis (2016). *Introduction to Time Series and Forecasting*. Third edition. Springer Texts in Statistics. Cham: Springer.

Brockwell, Peter J. and Richard A. Davis (1987). *Time series: theory and methods*. Springer series in statistics. New York: Springer.

Brunnermeier, Markus K. and Yuliy Sannikov (2014). "A Macroeconomic Model with a Financial Sector". In: *American Economic Review* 104.2, pp. 379–421.

Brunnermeier, Markus K. and Yuliy Sannikov (2016). "Macro, Money, and Finance". In: *Handbook of macroeconomics*. Ed. by John B. Taylor and Harald Uhlig. Vol. 2. Amsterdam: Elsevier, pp. 1497–1545.

Brzoza-Brzezina, Michał, Marcin Kolasa, and Krzysztof Makarski (2013). "The anatomy of standard DSGE models with financial frictions". In: *Journal of Economic Dynamics and Control* 37.1, pp. 32–51.

Brzoza-Brzezina, Michał and Jacek Suda (2021). "Are DSGE models irreparably flawed?" In: *Bank i Kredyt* 52.3, pp. 227–252.

Buiter, Willem H. (1984). "Saddlepoint Problems in Continuous Time Rational Expectations Models: A General Method and Some Macroeconomic Examples". In: *Econometrica* 52.3, pp. 665–680.

Bullard, James and Stefano Eusepi (2014). "When Does Determinacy Imply Expectational Stability?" In: *International Economic Review* 55.1, pp. 1–22.

Burmeister, Edwin (1980). "On Some Conceptual Issues in Rational Expectations Modeling". In: *Journal of Money Credit Banking* 12.4, pp. 800–816.

Cajueiro, Daniel O. and Benjamin M. Tabak (2005). "Ranking efficiency for emerging equity markets II". In: *Chaos, Solitons & Fractals* 23.2, pp. 671–675.

Cajueiro, Daniel O. and Benjamin M. Tabak (2004a). "Ranking efficiency for emerging markets". In: *Chaos, Solitons & Fractals* 22.2, pp. 349–352.

Cajueiro, Daniel O. and Benjamin M. Tabak (2004b). "The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient". In: *Physica A: Statistical Mechanics and its Applications* 336.3-4, pp. 521–537.

Canarella, Giorgio and Stephen M. Miller (2017). "Inflation Persistence Before and After Inflation Targeting: A Fractional Integration Approach". In: *Eastern Economic Journal* 43.1, pp. 78–103.

Candelon, Bertrand and Luis A. Gil-Alaña (2004). "Fractional integration and business cycle features". In: *Empirical economics* 29, pp. 343–359.

Cantore, Cristiano, Vasco J. Gabriel, Paul Levine, Joseph Pearlman, and Bo Yang (2013a). "The science and art of DSGE modelling: I – construction and Bayesian estimation". In: *Handbook of research methods and applications in empirical macroeconomics*. Ed. by Nigar Hashimzade and Michael A. Thornton. Handbooks of research methods and applications series. Cheltenham, U.K and Northampton, MA, USA: Edward Elgar, pp. 411–440.

Cantore, Cristiano, Vasco J. Gabriel, Paul Levine, Joseph Pearlman, and Bo Yang (2013b). "The science and art of DSGE modelling: II – model comparisons, model validation, policy analysis and general discussion". In: *Handbook of research methods and applications in empirical macroeconomics*. Ed. by Nigar Hashimzade and Michael A. Thornton. Handbooks of research methods and applications series. Cheltenham, U.K and Northampton, MA, USA: Edward Elgar, pp. 441–463.

Cantore, Cristiano, Paul Levine, Joseph Pearlman, and Bo Yang (2015). "CES technology and business cycle fluctuations". In: *Journal of Economic Dynamics and Control* 61, pp. 133–151.

Caporale, Guglielmo Maria and Luis A. Gil-Alaña (2013). "Long memory in US real output per capita". In: *Empirical economics* 44.2, pp. 591–611.

Caporale, Guglielmo Maria, Luis A. Gil-Alaña, and Yuliya Lovcha (2016). "Testing Unemployment Theories: A Multivariate Long Memory Approach". In: *Journal of Applied Economics* 19.1, pp. 95–112.

Caporale, Guglielmo Maria, Luis A. Gil-Alaña, and Manuel Monge (2020). "GDP per Capita in Europe: Time Trends and Persistence". In: *Journal of Business Management & Economics* 08.09, pp. 1–10.

Caporale, Guglielmo Maria, Luis A. Gil-Alaña, and Alex Plastun (2019). "Long memory and data frequency in financial markets". In: *Journal of Statistical Computation and Simulation* 89.10, pp. 1763–1779.

Caporale, Guglielmo Maria and Marinko Škare (2014). *Long memory in UK real GDP, 1851 – 2013: an ARFIMA-FIGARCH analysis*. DIW Discussion Paper 1395. Berlin.

Chambers, Marcus J. (1998). "Long Memory and Aggregation in Macroeconomic Time Series". In: *International Economic Review* 39.4, pp. 1053–1072.

Chan, Ngai Hang and Wilfredo Palma (1998). "State space modeling of long-memory processes". In: *The Annals of Statistics* 26.2, pp. 719–740.

Chen, Guanrong (2005). "Stability of Nonlinear Systems". In: *Encyclopedia of RF and Microwave Engineering*. Ed. by Kai Chang. Hoboken, NJ, USA: John Wiley & Sons, Inc, pp. 4881–4896.

Cheridito, Patrick (2001). "Mixed Fractional Brownian Motion". In: *Bernoulli* 7.6, pp. 913–934.

Chetty, Raj (2006). "A New Method of Estimating Risk Aversion". In: *American Economic Review* 96.5, pp. 1821–1834.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber (2011). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins". In: *American Economic Review* 101.3, pp. 471–475.

Cheung, Ying Lun (2020). "Nonstationarity-extended Whittle estimation with discontinuity: A correction". In: *Economics Letters* 187, pp. 1–5.

Cheung, Ying Lun and Uwe Hassler (2020). "Whittle-type estimation under long memory and nonstationarity". In: *AStA Advances in Statistical Analysis* 104.3, pp. 363–383.

Chevillon, Guillaume and Sophocles Mavroeidis (2017). "Learning can generate long memory". In: *Journal of Econometrics* 198.1, pp. 1–9.

Chib, Siddhartha and Srikanth Ramamurthy (2014). "DSGE Models with Student-t Errors". In: *Econometric Reviews* 33.1-4, pp. 152–171.

Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin (2010). "DSGE Models for Monetary Policy Analysis". In: *Handbook of monetary economics*. Ed. by Benjamin M. Friedman and Michael Woodford. Vol. 3. San Diego, CA and Amsterdam: North-Holland/Elsevier, pp. 285–367.

Christoffel, Kai and Keith Kuester (2008). “Resuscitating the wage channel in models with unemployment fluctuations”. In: *Journal of Monetary Economics* 55.5, pp. 865–887.

Clarida, Richard, Jordi Galí, and Mark Gertler (2002). “A simple framework for international monetary policy analysis”. In: *Journal of Monetary Economics* 49.5, pp. 879–904.

Coeurdacier, Nicolas, Hélène Rey, and Pablo Winant (2011). “The Risky Steady State”. In: *American Economic Review* 101.3, pp. 398–401.

Cogley, Timothy and James Michael Nason (1995). “Output Dynamics in Real-Business-Cycle Models”. In: *The American Economic Review* 85.3, pp. 492–511.

Comte, F. and E. Renault (1996). “Long memory continuous time models”. In: *Journal of Econometrics* 73.1, pp. 101–149.

Cont, Rama (2001). “Empirical properties of asset returns: stylized facts and statistical issues”. In: *Quantitative Finance* 1.2, pp. 223–236.

Cooley, Thomas F. and Edward C. Prescott (1995). “Economic Growth and Business Cycles”. In: *Frontiers of Business Cycle Research*. Ed. by Thomas F. Cooley. Princeton University Press, pp. 1–38.

Costa Junior, Celso Jose (2016). *Understanding DSGE models*. Wilmington: Vernon Press.

Cúrdia, Vasco and Ricardo Reis (2010). *Correlated Disturbances and U.S. Business Cycles*. NBER Working Paper 15774. Cambridge, MA.

Cutland, Nigel J., P. Ekkehard Kopp, and Walter Willinger (1995). “Stock Price Returns and the Joseph Effect: A Fractional Version of the Black-Scholes Model”. In: *Seminar on Stochastic Analysis, Random Fields and Applications*. Ed. by Erwin Bolthausen, Marco Dozzi, and Francesco Russo. Basel: Birkhäuser Basel, pp. 327–351.

Czado, Claudia and Thorsten Schmidt (2011). *Mathematische Statistik*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Dahlhaus, Rainer (2006). “Correction: Efficient Parameter Estimation for Self-Similar Processes”. In: *The Annals of Statistics* 34.2, pp. 1045–1047.

Dahlhaus, Rainer (1989). “Efficient Parameter Estimation for Self-Similar Processes”. In: *The Annals of Statistics* 17.4, pp. 1749–1766.

Davidson, James and Philipp Sibbertsen (2005). “Generating schemes for long memory processes: regimes, aggregation and linearity”. In: *Journal of Econometrics* 128.2, pp. 253–282.

Davies, Robert B. and David S. Harte (1987). “Tests for Hurst effect”. In: *Biometrika* 74.1, pp. 95–101.

DeJong, David N. and Charles H. Whiteman (1991). "Reconsidering 'trends and random walks in macroeconomic time series'". In: *Journal of Monetary Economics* 28.2, pp. 221–254.

Di Tella, Sebastian (2017). "Uncertainty Shocks and Balance Sheet Recessions". In: *Journal of Political Economy* 125.6, pp. 2038–2081.

Dias, Daniel A. and Carlos Robalo Marques (2010). "Using mean reversion as a measure of persistence". In: *Economic Modelling* 27.1, pp. 262–273.

Diebold, Francis X. and Atsushi Inoue (2001). "Long memory and regime switching". In: *Journal of Econometrics* 105.1, pp. 131–159.

Diebold, Francis X. and Glenn D. Rudebusch (1989). "Long memory and persistence in aggregate output". In: *Journal of Monetary Economics* 24.2, pp. 189–209.

Diebolt, Claude and Vivien Guiraud (2005). "A Note On Long Memory Time Series". In: *Quality & Quantity* 39.6, pp. 827–836.

Diebolt, Claude and Vivien Guiraud (2000). "Long memory time series and fractional integration. A cliometric contribution to French and German economic and social history". In: *Historical Social Research* 25.3/4, pp. 4–22.

Domeij, David and Martin Flodén (2006). "The labor-supply elasticity and borrowing constraints: Why estimates are biased". In: *Review of Economic Dynamics* 9.2, pp. 242–262.

Doob, Joseph L. (1953). *Stochastic processes*. Wiley publications in statistics. New York and London: John Wiley & Sons Inc and Chapman & Hall Limited.

Dung, Nguyen Tien (2013). "Fractional stochastic differential equations with applications to finance". In: *Journal of Mathematical Analysis and Applications* 397.1, pp. 334–348.

Dung, Nguyen Tien (2011). "Semimartingale approximation of fractional Brownian motion and its applications". In: *Computers & Mathematics with Applications* 61.7, pp. 1844–1854.

Elaydi, Saber N. (2005). *An introduction to difference equations*. 3rd ed. Undergraduate texts in mathematics. New York, NY: Springer.

Enriquez, Nathanaël (2004). "A simple construction of the fractional Brownian motion". In: *Stochastic Processes and their Applications* 109.2, pp. 203–223.

Evans, George W. (1985). "Expectational Stability and the Multiple Equilibria Problem in Linear Rational Expectations Models". In: *The Quarterly Journal of Economics* 100.4, pp. 1217–1233.

Evans, George W. and Seppo Honkapohja (2001). *Learning and expectations in macroeconomics*. Princeton, New Jersey and Oxford: Princeton University Press.

Evans, George W. and Bruce McGough (2005). “Stable sunspot solutions in models with predetermined variables”. In: *Journal of Economic Dynamics and Control* 29.4, pp. 601–625.

Evans, Richard W. and Kerk L. Phillips (2018). “Advantages of an Ellipse when Modeling Leisure Utility”. In: *Computational Economics* 51.3, pp. 513–533.

Fernández-Villaverde, Jesús, Olaf Posch, and Juan F. Rubio-Ramírez (2012). *Solving the new Keynesian model in continuous time*. Unpublished Manuscript.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson (2007). “ABCs (and Ds) of Understanding VARs”. In: *American Economic Review* 97.3, pp. 1021–1026.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, and Frank Schorfheide (2016). “Solution and Estimation Methods for DSGE Models”. In: *Handbook of macroeconomics*. Ed. by John B. Taylor and Harald Uhlig. Vol. 2. Amsterdam: Elsevier, pp. 527–724.

Fiorito, Riccardo and Giulio Zanella (2012). “The anatomy of the aggregate labor supply elasticity”. In: *Review of Economic Dynamics* 15.2, pp. 171–187.

Fuhrer, Jeffrey C. (2010). “Inflation Persistence”. In: *Handbook of monetary economics*. Ed. by Benjamin M. Friedman and Michael Woodford. Vol. 3. San Diego, CA and Amsterdam: North-Holland/Elsevier, pp. 423–486.

Fuhrer, Jeffrey C. (1995). “The persistence of inflation and the cost of disinflation”. In: *New England economic review* 1995, pp. 3–16.

Furlanetto, Francesco and Martin Seneca (2014). “Investment shocks and consumption”. In: *European Economic Review* 66, pp. 111–126.

Galí, Jordi (2008). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework*. Princeton, NJ: Princeton Univ. Press.

Galizia, Dana (2021). “Saddle cycles: Solving rational expectations models featuring limit cycles (or chaos) using perturbation methods”. In: *Quantitative economics* 12, pp. 869–901.

Galor, Oded (2007). *Discrete dynamical systems*. Berlin and Heidelberg: Springer.

Gandolfo, Giancarlo (1997). *Economic dynamics*. Berlin and Heidelberg: Springer.

Garcia, Diego (2023). *Gaussian Hypergeometric Function: On the pages of MathWorks*. URL: `https://de.mathworks.com/matlabcentral/fileexchange/1844-gaussian-hypergeometric-function` (visited on May 16, 2023).

Geweke, John and Susan Porter-Hudak (1983). “The Estimation and Application of long memory time series models”. In: *Journal of Time Series Analysis* 4.4, pp. 221–238.

Giacomini, Raffaella (2013). “The Relationship Between DSGE and VAR Models”. In: *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*. Ed. by Thomas B. Fomby, Lutz Kilian, and Anthony Murphy. Vol. 32. Advances in Econometrics. Emerald Group Publishing Limited, pp. 1–25.

Gil-Alaña, Luis A. (2020). “The Kuznets Environmental Curve Using Fractional Integration and Long Range Dependence Techniques”. In: *International Journal of Economics and Politics* 1.1, pp. 69–94.

Gil-Alaña, Luis A. (2001). “The persistence of unemployment in the USA and Europe in terms of fractionally ARIMA models”. In: *Applied Economics* 33.10, pp. 1263–1269.

Gil-Alaña, Luis A., OlaOluwa S. Yaya, and Olanrewaju I. Shittu (2015). “GDP per capita in Africa before the global financial crisis: Persistence, mean reversion and long memory features”. In: *CBN journal of applied statistics* 06.1, pp. 219–239.

Giraitis, Liudas, Hira L. Koul, and Donatas Surgailis (2012). *Large sample inference for long memory processes*. London: Imperial College Press.

Gomme, Paul and Damba Lkhagvasuren (2013). “Calibration and simulation of DSGE models”. In: *Handbook of research methods and applications in empirical macroeconomics*. Ed. by Nigar Hashimzade and Michael A. Thornton. Handbooks of research methods and applications series. Cheltenham, U.K and Northampton, MA, USA: Edward Elgar, pp. 575–592.

Gourieroux, C., J. Jasiak, and A. Monfort (2020). “Stationary bubble equilibria in rational expectation models”. In: *Journal of Econometrics* 218.2, pp. 714–735.

Gourieroux, C., J. J. Laffont, and A. Monfort (1982). “Rational Expectations in Dynamic Linear Models: Analysis of the Solutions”. In: *Econometrica* 50.2, pp. 409–425.

Granger, Clive W. J. (1980). “Long memory relationships and the aggregation of dynamic models”. In: *Journal of Econometrics* 14.2, pp. 227–238.

Granger, Clive W. J. (1966). “The Typical Spectral Shape of an Economic Variable”. In: *Econometrica* 34.1, pp. 150–161.

Granger, Clive W. J. and Zhuanxin Ding (1996). “Varieties of long memory models”. In: *Journal of Econometrics* 73.1, pp. 61–77.

Granger, Clive W. J. and Roselyne Joyeux (1980/2001). “An introduction to long–memory time series models and fractional differencing”. In: *Journal of Time Series Analysis* 1.1, pp. 15–29. Reprinted in: Eric Ghysels, Norman R. Swanson, and Mark W. Watson (eds.). *Essays in Econometrics: Collected Papers of Clive W. J. Granger - Volume II: Causality, Integration and Cointegration, and Long Memory*. Cambridge: Cambridge Univ. Press, 2001, pp. 321–337.

Grassi, Stefano and Paolo Santucci de Magistris (2014). “When long memory meets the Kalman filter: A comparative study”. In: *Computational Statistics & Data Analysis* 76, pp. 301–319.

Graves, Timothy, Robert Gramacy, Nicholas Watkins, and Christian Franzke (2017). “A Brief History of Long Memory: Hurst, Mandelbrot and the Road to ARFIMA, 1951–1980”. In: *Entropy* 19.9, Article 437.

Greenwood, Jeremy, Zvi Hercowitz, and Gregory W. Huffman (1988). “Investment, Capacity Utilization, and the Real Business Cycle”. In: *American Economic Review* 78.3, pp. 402–417.

Guasoni, Paolo (2006). “No Arbitrage under transaction costs, with fractional Brownian motion and beyond”. In: *Mathematical Finance* 16.3, pp. 569–582.

Guégan, Dominique (2005). “How can we Define the Concept of Long Memory? An Econometric Survey”. In: *Econometric Reviews* 24.2, pp. 113–149.

Haldrup, Niels and J. Eduardo Vera-Valdés (2017). “Long memory, fractional integration, and cross-sectional aggregation”. In: *Journal of Econometrics* 199.1, pp. 1–11.

Hamilton, James D. (1994). *Time series analysis*. Princeton, NJ: Princeton Univ. Press.

Hansen, Gary D. (1985). “Indivisible labor and the business cycle”. In: *Journal of Monetary Economics* 16.3, pp. 309–327.

Hartley, James E., Kevin D. Hoover, and Salyer Kevin D. (1997). “The limits of business cycle research: assessing the real business cycle model”. In: *Oxford Review of Economic Policy* 13.3, pp. 34–54.

Hassler, Uwe (2011). “Estimation of fractional integration under temporal aggregation”. In: *Journal of Econometrics* 162.2, pp. 240–247.

Hassler, Uwe (2016). *Stochastic Processes and Calculus*. Cham: Springer International Publishing.

Hassler, Uwe (2019). *Time series analysis with long memory in view*. Wiley series in probability and statistics. Hoboken, NJ: John Wiley & Sons.

Hassler, Uwe and Piotr Kokoszka (2010). “Impusle responses of fractionally integrated processes with long memory”. In: *Econometric Theory* 26.6, pp. 1855–1861.

Hassler, Uwe and Jürgen Wolters (2009). “Hysteresis in Unemployment Rates? A Comparison between Germany and the US”. In: *Jahrbücher für Nationalökonomie und Statistik* 229.2-3, pp. 119–129.

Hassler, Uwe and Jürgen Wolters (1995). “Long Memory in Inflation Rates: International Evidence”. In: *Journal of Business & Economic Statistics* 13.1, pp. 37–45.

Haubrich, Joseph G. and Andrew W. Lo (2001). "The sources and nature of long-term memory in aggregate output". In: *Economic Review-Federal Reserve Bank of Cleveland* 37.2, pp. 15–30.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2008). "Insurance and opportunities: A welfare analysis of labor market risk". In: *Journal of Monetary Economics* 55.3, pp. 501–525.

Henry, Marc and Paolo Zaffaroni (2003). "The Long-Range Dependence Paradigm for Macroeconomics and Finance". In: *Theory and applications of long-range dependence*. Ed. by Paul Doukhan, George Oppenheim, and Murad S. Taqqu. Boston, MA, Basel, and Berlin: Birkhäuser, pp. 417–438.

Hosking, J. R. M. (1981). "Fractional Differencing". In: *Biometrika* 68.1, pp. 165–176.

Hu, Yaozhong and Bernt K. Øksendal (2003). "Fractional white noise calculus and applications to Finance". In: *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 06.01, pp. 1–32.

Hull, Matthew and Frank McGroarty (2014). "Do emerging markets become more efficient as they develop? Long memory persistence in equity indices". In: *Emerging Markets Review* 18, pp. 45–61.

Hurst, H. E. (1951). "Long-Term Storage Capacity of Reservoirs". In: *Transactions of the American Society of Civil Engineers* 116.1, pp. 770–799.

Hurst, H. E. (1956). "The Problem of Long-term Storage in Reservoirs". In: *International Association of Scientific Hydrology. Bulletin* 1.3, pp. 13–27.

Hurvich, Clifford M. and Kaizo I. Beltrao (1994). "Automatic semiparametric estimation of the memory parameter of a long memory time series". In: *Journal of Time Series Analysis* 15.3, pp. 285–302.

Hurvich, Clifford M., Rohit Deo, and Julia Brodsky (1998). "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series". In: *Journal of Time Series Analysis* 19.1, pp. 19–46.

Ireland, Peter N. (2004a). "A method for taking models to the data". In: *Journal of Economic Dynamics and Control* 28.6, pp. 1205–1226.

Ireland, Peter N. (2004b). "Technology Shocks in the New Keynesian Model". In: *Review of Economics and Statistics* 86.4, pp. 923–936.

Iskrev, Nikolay (2010). "Local identification in DSGE models". In: *Journal of Monetary Economics* 57.2, pp. 189–202.

Jensen, Mark J. and Ming Liu (2006). "Do long swings in the business cycle lead to strong persistence in output?" In: *Journal of Monetary Economics* 53.3, pp. 597–611.

Karatzas, Ioannis and Steven E. Shreve (1998). *Methods of mathematical finance*. New York, NY: Springer.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (1988). “Production, growth and business cycles”. In: *Journal of Monetary Economics* 21.2-3, pp. 195–232.

King, Robert G., Charles I. Plosser, and Sergio T. Rebelo (2002). “Production, Growth and Business Cycles: Technical Appendix”. In: *Computational Economics* 20.1, pp. 87–116.

King, Robert G. and Mark W. Watson (1998). “The Solution of Singular Linear Difference Systems under Rational Expectations”. In: *International Economic Review* 39.4, pp. 1015–1026.

Klein, Paul (no date). *Erratum to Klein (2000): On the pages of Paul Klein*. URL: http://paulklein.ca/newsite/research/erratum.pdf (visited on May 16, 2023).

Klein, Paul (2000). “Using the generalized Schur form to solve a multivariate linear rational expectations model”. In: *Journal of Economic Dynamics and Control* 24.10, pp. 1405–1423.

Klenke, Achim (2013). *Wahrscheinlichkeitstheorie*. 3., überarbeitete und ergänzte Auflage. Masterclass. Berlin and Heidelberg: Springer Spektrum.

Kohlscheen, Emanuel and Jouchi Nakajima (2021). “Steady–state growth”. In: *International Finance* 24.1, pp. 40–52.

Koop, Gary, Eduardo Ley, Jacek Osiewalski, and Mark F.J. Steel (1997). “Bayesian analysis of long memory and persistence using ARFIMA models”. In: *Journal of Econometrics* 76.1-2, pp. 149–169.

Koustas, Zisimos and William Veloce (1996). “Unemployment hysteresis in Canada: an approach based on long-memory time series models”. In: *Applied Economics* 28.7, pp. 823–831.

Kumar, Manmohan S. and Tatsuyoshi Okimoto (2007). “Dynamics of Persistence in International Inflation Rates”. In: *Journal of Money, Credit and Banking* 39.6, pp. 1457–1479.

Lamberton, Damien and Bernard Lapeyre (2008). *Introduction to stochastic calculus applied to finance*. 2. ed. Boca Raton: Chapman & Hall CRC.

Lieberman, Offer, Roy Rosemarin, and Judith Rousseau (2012). “Asymptotic theory for Maximum Likelihood Estimation of the memory parameter in stationary Gaussian processes”. In: *Econometric Theory* 28.2, pp. 457–470.

Lim, Kian-Ping and Robert Brooks (2011). “The Evolution of Stock Market Efficiency Over Time: A Survey of the Empirical Literature”. In: *Journal of Economic Surveys* 25.1, pp. 69–108.

Lindé, Jesper (2009). “The effects of permanent technology shocks on hours: Can the RBC-model fit the VAR evidence?” In: *Journal of Economic Dynamics and Control* 33.3, pp. 597–613.

Lindé, Jesper, Frank Smets, and Rafael Wouters (2016). “Challenges for Central Banks' Macro Models”. In: *Handbook of macroeconomics*. Ed. by John B. Taylor and Harald Uhlig. Vol. 2. Amsterdam: Elsevier, pp. 2185–2262.

Linden, Mikael (1999). “Time series properties of aggregated AR(1) processes with uniformly distributed coefficients”. In: *Economics Letters* 64.1, pp. 31–36.

Lo, Andrew W. (1991). “Long-Term Memory in Stock Market Prices”. In: *Econometrica* 59.5, pp. 1279–1313.

Lütkepohl, Helmut (2005). *New Introduction to Multiple Time Series Analysis*. Berlin Heidelberg: Springer-Verlag GmbH.

Maheswaran, S. and Christopher A. Sims (1993). “Empirical Implications of Arbitrage-free Asset Markets”. In: *Models, methods, and applications of econometrics*. Ed. by Peter C. B. Phillips. Cambridge, Mass.: Blackwell, pp. 301–316.

Mandelbrot, Benoit B. (2002). *Gaussian self-affinity and fractals: Globality, the earth, 1/f noise and R/S*. Vol. H. Selecta. New York, NY: Springer.

Mandelbrot, Benoit B. (1975/2002). “Limit theorems on the self-normalized range for weakly and strongly dependent processes”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 31.4, pp. 271–285. Reprinted in: *Gaussian self-affinity and fractals: Globality, the earth, 1/f noise and R/S*. Vol. H. Selecta. New York, NY: Springer, 2002, pp. 517–537.

Mandelbrot, Benoit B. (1972). “Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis”. In: *Annals of economic and social measurement: journal of computers, information retrieval, and research methodology* 1.3, pp. 259–290.

Mandelbrot, Benoit B. (1971). “When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models”. In: *Review of Economics and Statistics* 53.3, pp. 225–236.

Mandelbrot, Benoit B. and John W. van Ness (1968). “Fractional Brownian Motions, Fractional Noises and Applications”. In: *SIAM Review* 10.4, pp. 422–437.

Mandelbrot, Benoit B. and James R. Wallis (1968/2002). “Noah, Joseph, and Operational Hydrology”. In: *Water Resources Research* 4.5, pp. 909–918. Reprinted in: Benoit B. Mandelbrot. *Gaussian self-affinity and fractals: Globality, the earth, 1/f noise and R/S*. Vol. H. Selecta. New York, NY: Springer, 2002, pp. 236–251.

Mandelbrot, Benoit B. and James R. Wallis (1969/2002). “Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence”. In: *Water Resources Research* 5.5, pp. 967–988. Reprinted in: Benoit B. Mandelbrot. *Gaussian* *self-affinity and fractals: Globality, the earth, 1/f noise and R/S*. Vol. H. Selecta. New York, NY: Springer, 2002, pp. 483–516.

Mankiw, N. Gregory (1989). “Real Business Cycles: A New Keynesian Perspective”. In: *The Journal of Economic Perspectives* 3.3, pp. 79–90.

Marinucci, D. and P. M. Robinson (1999). “Alternative forms of fractional Brownian motion”. In: *Journal of Statistical Planning and Inference* 80.1-2, pp. 111–122.

Mayoral, Laura (2006). “Further evidence on the statistical properties of real GNP”. In: *Oxford bulletin of economics and statistics* 68, pp. 901–920.

McCallum, Bennett T. (1983). “On non-uniqueness in rational expectations models”. In: *Journal of Monetary Economics* 11.2, pp. 139–168.

Merton, Robert C. (1973). “Theory of Rational Option Pricing”. In: *The Bell Journal of Economics and Management Science* 4.1, pp. 141–183.

Meyer-Gohde, Alexander (2010). “Linear rational-expectations models with lagged expectations: A synthetic method”. In: *Journal of Economic Dynamics and Control* 34.5, pp. 984–1002.

Meyer-Gohde, Alexander and Daniel Neuhoff (2018). *Generalized exogenous processes in DSGE: A Bayesian approach*. IMFS Working Paper Series, No. 125.

Meyer-Gohde, Alexander and Daniel Neuhoff (2015). “Solving and estimating linearized DSGE models with VARMA shock processes and filtered data”. In: *Economics Letters* 133, pp. 89–91.

Miao, Jianjun (2020). *Economic dynamics in discrete time*. Second edition. Cambridge, Massachusetts: The MIT Press.

Michelacci, Claudio (2004). “Cross-sectional heterogeneity and the persistence of aggregate fluctuations”. In: *Journal of Monetary Economics* 51.7, pp. 1321–1352.

Michelacci, Claudio and Paolo Zaffaroni (2000). “(Fractional) beta convergence”. In: *Journal of Monetary Economics* 45, pp. 129–153.

Miller, J. Isaac and Joon Y. Park (2010). “Nonlinearity, nonstationarity, and thick tails: How they interact to generate persistence in memory”. In: *Journal of Econometrics* 155.1, pp. 83–89.

Mitra, Kaushik, George W. Evans, and Seppo Honkapohja (2013). “Policy change and learning in the RBC model”. In: *Journal of Economic Dynamics and Control* 37.10, pp. 1947–1971.

Moretti, Gianluca and Giulio Nicoletti (2010). *Estimating DSGE models with unknown data persistence*. Banca d’Italia Temi di discussione (Working Papers) Number 750.

Morris, Stephen D. (2016). “VARMA representation of DSGE models”. In: *Economics Letters* 138, pp. 30–33.

Mrázek, Milan, Jan Pospíšil, and Tomáš Sobotka (2016). “On calibration of stochastic and fractional stochastic volatility models”. In: *European Journal of Operational Research* 254.3, pp. 1036–1046.

Necula, Ciprian (2002). *Option Pricing in a Fractional Brownian Motion Environment: Manuscript. On the pages of SSRN.* URL: `https://ssrn.com/abstract=1286833` (visited on May 16, 2023).

Ohanissian, Arek, Jeffrey R. Russell, and Ruey S. Tsay (2008). “True or Spurious Long Memory? A New Test”. In: *Journal of Business & Economic Statistics* 26.2, pp. 161–175.

Øksendal, Bernt K. (2013). *Stochastic differential equations: An introduction with applications.* Sixth edition, sixth corrected printing. Universitext. Berlin et al.: Springer.

Olver, Peter J. and Chehrzad Shakiban (2018). *Applied Linear Algebra.* Cham: Springer International Publishing.

Onali, Enrico and John Goddard (2011). “Are European equity markets efficient? New evidence from fractal analysis”. In: *International Review of Financial Analysis* 20.2, pp. 59–67.

Oppenheim, Georges and Marie-Claude Viano (2004). “Aggregation of random parameters Ornstein-Uhlenbeck or AR processes: some convergence results”. In: *Journal of Time Series Analysis* 25.3, pp. 335–350.

Ovchinnikov, Sergei (2021). *Real Analysis: Foundations.* Cham: Springer International Publishing.

Palma, Wilfredo (2007). *Long-memory time series: Theory and methods.* Wiley series in probability and statistics. Hoboken, N.J: Wiley-Interscience.

Parke, William R. (1999). “What is Fractional Integration?” In: *The Review of Economics and Statistics* 81.4, pp. 632–638.

Parra-Alvarez, Juan Carlos (2018). “A Comparison of Numerical Methods for the Solution of Continuous-Time DSGE Models”. In: *Macroeconomic Dynamics* 22.6, pp. 1555–1583.

Paya, Ivan, Agustin Duarte, and Ken Holden (2007). “On the Relationship between Inflation Persistence and Temporal Aggregation”. In: *Journal of Money, Credit and Banking* 39.6, pp. 1521–1531.

Pearson, John W., Sheehan Olver, and Mason A. Porter (2017). “Numerical methods for the computation of the confluent and Gauss hypergeometric functions”. In: *Numerical Algorithms* 74.3, pp. 821–866.

Peterman, William B. (2016). “Reconciling micro and macro estimates of the frisch labor supply elasticity”. In: *Economic Inquiry* 54.1, pp. 100–120.

Phillips, Peter C. B. and Katsumi Shimotsu (2004). “Local Whittle estimation in nonstationary and unit root cases”. In: *The Annals of Statistics* 32.2, pp. 656–692.

Phillips, Peter C. B. and Zhijie Xiao (1998). “A Primer on Unit Root Testing”. In: *Journal of Economic Surveys* 12.5, pp. 423–470.

Pivetta, Frederic and Ricardo Reis (2007). “The persistence of inflation in the United States”. In: *Journal of Economic Dynamics and Control* 31.4, pp. 1326–1358.

Ravenna, Federico (2007). “Vector autoregressions and reduced form representations of DSGE models”. In: *Journal of Monetary Economics* 54.7, pp. 2048–2064.

Rebelo, Sergio T. (2005). “Real Business Cycle Models: Past, Present and Future”. In: *Scandinavian Journal of Economics* 107.2, pp. 217–238.

Robinson, P. M. (1995a). “Log-Periodogram Regression of Time Series with Long Range Dependence”. In: *The Annals of Statistics* 23.3, pp. 1048–1072.

Robinson, P. M. (1995b). “Gaussian Semiparametric Estimation of Long Range Dependence”. In: *The Annals of Statistics* 23.5, pp. 1630–1661.

Robinson, P. M. (1994). “Efficient Tests of Nonstationary Hypotheses”. In: *Journal of the American Statistical Association* 89.428, pp. 1420–1437.

Rogers, Leonard C. G. (1997). “Arbitrage with Fractional Brownian Motion”. In: *Mathematical Finance* 7.1, pp. 95–105.

Rogers, Leonard C. G. and David Williams (1987). *Diffusions, Markov Processes, and Martinagles: Itô calculus*. Vol. 2. Wiley series in probability and mathematical statistics. Chichester: Wiley & Sons.

Rostek, Stefan (2009). *Option pricing in fractional Brownian markets: Zugl.: Tübingen, Univ., Diss., 2008*. 1. Aufl. Vol. 622. Lecture notes in economics and mathematical systems. Berlin and Heidelberg: Springer.

Rostek, Stefan and R. Schöbel (2013). “A note on the use of fractional Brownian motion for financial modeling”. In: *Economic Modelling* 30, pp. 30–35.

Salopek, D. M. (1998). “Tolerance to arbitrage”. In: *Stochastic Processes and their Applications* 76.2, pp. 217–230.

Schennach, Susanne M. (2018a). “Long Memory via Networking”. In: *Econometrica* 86.6, pp. 2221–2248.

Schennach, Susanne M. (2018b). *Supplement to “Long Memory Via Networking”*. Econometrica Supplementary Materials.

Schorfheide, Frank (2011). *Estimation and Evaluation of DSGE Models: Progress and Challenges*. NBER Working Paper 16781. Cambridge, MA.

Schüler, Yves S. (2018). *Detrending and financial cycle facts across G7 countries: Mind a spurious medium term!* ECB Working Paper No. 2138. European Central Bank.

Sewell, Martin Victor (2012). "The Efficient Market Hypothesis: Empirical Evidence". In: *International Journal of Statistics and Probability* 1.2, pp. 164–178.

Shi, Wendong and Jingwei Sun (2016). "Aggregation and long-memory: An analysis based on the discrete Fourier transform". In: *Economic Modelling* 53, pp. 470–476.

Shimotsu, Katsumi (2010). "Exact local Whittle estimation of fractional integration with unknown mean and time trend". In: *Econometric Theory* 26.2, pp. 501–540.

Shimotsu, Katsumi and Peter C. B. Phillips (2005). "Exact local Whittle estimation of fractional integration". In: *The Annals of Statistics* 33.4, pp. 1890–1933.

Shreve, Steven E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Finance. New York, NY: Springer New York.

Siddiqui, M. M. (1976). "The asymptotic distribution of the range and other functions of partial sums of stationary processes". In: *Water Resources Research* 12.6, pp. 1271–1276.

Silverberg, Gerald and Bart Verspagen (2000). *A note on Michelacci and Zaffaroni, long memory, and time series of economic growth*. UNU-MERIT Research Memoranda 32.

Silverberg, Gerald and Bart Verspagen (2003). "Long Memory and Economic Growth in the World Economy Since the 19th Century". In: *Processes with long-range correlations*. Ed. by Govindan Rangarajan and Mingzhou Ding. Berlin and Heidelberg: Springer, pp. 270–285.

Silverberg, Gerald and Bart Verspagen (1999). *Long memory in time series of economic growth and convergence*. UNU-MERIT Research Memoranda 16.

Sims, Christopher A. (2002). "Solving Linear Rational Expectations Models". In: *Computational Economics* 20, pp. 1–20.

Smets, Frank and Rafael Wouters (2007). "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach". In: *American Economic Review* 97.3, pp. 586–606.

Smith, Aaron (2005). "Level Shifts and the Illusion of Long Memory in Economic Time Series". In: *Journal of Business & Economic Statistics* 23.3, pp. 321–335.

Smith, Jeremy, Nick Taylor, and Sanjay Yadav (1997). "Comparing the bias and misspecification in ARFIMA models". In: *Journal of Time Series Analysis* 18.5, pp. 507–527.

Souza, Leonardo Rocha (2005). "A Note on Chambers's "Long Memory and Aggregation in Macroeconomic Time Series". In: *International Economic Review* 46.3, pp. 1059–1062.

Souza, Leonardo Rocha (2008). “Why Aggregate Long Memory Time Series?” In: *Econometric Reviews* 27.1-3, pp. 298–316.

Sowell, Fallaw (1992a). “Maximum likelihood estimation of stationary univariate fractionally integrated time series models”. In: *Journal of Econometrics* 53.1-3, pp. 165–188.

Sowell, Fallaw (1992b). “Modeling long-run behavior with the fractional ARIMA model”. In: *Journal of Monetary Economics* 29, pp. 277–302.

Sowell, Fallaw (1990). “The Fractional Unit Root Distribution”. In: *Econometrica* 58.2, pp. 495–505.

Stiglitz, Joseph E. (2018). “Where modern macroeconomics went wrong”. In: *Oxford Review of Economic Policy* 34.1-2, pp. 70–106.

Stock, James H. (1991). “Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series”. In: *Journal of Monetary Economics* 28.3, pp. 435–459.

Struyf, Gunther (2014). *bar3c expanded to support colouring by range: On the pages of GitHub*. URL: `https://github.com/GuntherStruyf/matlab-tools/blob/master/bar3crange.m` (visited on May 16, 2023).

Swanson, Eric T. (2012). “Risk Aversion and the Labor Margin in Dynamic Equilibrium Models”. In: *American Economic Review* 102.4, pp. 1663–1691.

Taylor, John B. (1986). “New econometric approaches to stabilization policy in stochastic models of macroeconomic fluctuations”. In: *Handbook of Econometrics*. Ed. by Zvi Griliches and Michael D. Intriligator. Vol. 3. Handbook of Econometrics. Elsevier, pp. 1997–2055.

Teverovsky, Vadim, Murad S. Taqqu, and Walter Willinger (1999). “A critical look at Lo’s modified R/S statistic”. In: *Journal of Statistical Planning and Inference* 80.1-2, pp. 211–227.

Thao, Tran Hung (2003). “A Note on Fractional Brownian Motion”. In: *Vietnam Journal of Mathematics* 31.3, pp. 255–260.

Thao, Tran Hung (2006). “An approximate approach to fractional analysis for finance”. In: *Nonlinear Analysis: Real World Applications* 7.1, pp. 124–132.

Thao, Tran Hung and Christine Thomas-Agnan (2003). “Évolution des Cours Gouvernée par un Processus de Type Arima Fractionnaire”. In: *Studia Universitatis Babeş-Bolyai Mathematica* XLVIII.2, pp. 107–115.

Turnovsky, Stephen J. (2000). *Methods of macroeconomic dynamics*. 2. ed. Cambridge, Mass.: MIT Press.

Uhlig, Harald (1999). “A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily”. In: *Computational methods for the study of dynamic economies*. Ed. by Ramon Marimon and Andrew Scott. Oxford: Oxford Univ. Press, pp. 30–61.

Velasco, Carlos (1999a). “Non-stationary log-periodogram regression”. In: *Journal of Econometrics* 91.2, pp. 325–371.

Velasco, Carlos (1999b). “Gaussian Semiparametric Estimation of Non-stationary Time Series”. In: *Journal of Time Series Analysis* 20.1, pp. 87–127.

Vera-Valdés, J. Eduardo (2021). “Nonfractional Long-Range Dependence: Long Memory, Antipersistence, and Aggregation”. In: *Econometrics* 9.4, pp. 1–18.

Vera-Valdés, J. Eduardo (2020). “On long memory origins and forecast horizons”. In: *Journal of Forecasting* 39.5, pp. 811–826.

Vera-Valdés, J. Eduardo (2022). “The persistence of financial volatility after COVID-19”. In: *Finance research letters* 44, pp. 1–9.

Wang, Xiao-Tian (2010). “Scaling and long-range dependence in option pricing I: Pricing European option with transaction costs under the fractional Black–Scholes model”. In: *Physica A: Statistical Mechanics and its Applications* 389.3, pp. 438–444.

Wen, Yi (2001). “Understanding self-fulfilling rational expectations equilibria in real business cycle models”. In: *Journal of Economic Dynamics and Control* 25.8, pp. 1221–1240.

Yang, Ben-zhang, Xinjiang He, and Nan-jing Huang (2016). *Equilibrium price and optimal insider trading strategy under stochastic liquidity with long memory: Manuscript. On the pages of arxiv.org*. URL: `https://arxiv.org/pdf/1901.00345` (visited on May 16, 2023).

Zaffaroni, Paolo (2004). “Contemporaneous aggregation of linear dynamic models in large economies”. In: *Journal of Econometrics* 120.1, pp. 75–102.