

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Economics

2019

Essays in Time Series Econometrics

Dou, Liyu

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01bn9999657>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

ESSAYS IN TIME SERIES ECONOMETRICS

LIYU DOU

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF ECONOMICS
ADVISER: ULRICH K. MÜLLER

SEPTEMBER 2019

© Copyright by Liyu Dou, 2019.

All rights reserved.

Abstract

This collection of essays investigates robust inference and modelling in time series econometrics. Chapter 1 considers the problem of deriving heteroskedasticity and autocorrelation robust (HAR) inference about a scalar parameter of interest. The main finding implies that, for a given sample size, one can only be confident about the efficiency of a valid HAR test if we are willing to make *a priori* assumptions about the persistence properties of the data. This chapter demonstrates that it is advantageous to allow for bias in long-run variance estimation and adjust the critical value to explicitly account for the maximum bias. Chapter 2, jointly with Ulrich Müller, proposes a flexible asymptotic framework for the modelling of persistent time series, by generalizing the popular local-to-unity model. We establish the richness of the class of this generalized local-to-unity model, GLTU(p), in the sense that their limiting processes can well approximate a large class of stationary Gaussian processes in the total variation norm. This chapter also suggests a straightforward approximation to the limited-information asymptotic likelihood of the GLTU(p) model. Chapter 3 applies the econometric framework developed in Chapter 2 to examine and document the persistence properties of 9 macroeconomic time series over 17 advanced economies, based on the Jordá-Schularick-Taylor Macrohistory Database. It is found that allowing for the generality in modelling long-range dependence can substantially alter quantitative statements about the persistence of macroeconomic time series. Based on empirical evidence, this chapter recommends using an appropriately defined measure of the half-life in the GLTU(p) model to gauge the persistence of macroeconomic time series.

Acknowledgements

I am deeply grateful to my adviser, Ulrich K. Müller, for his excellent guidance, continuous support, and constant encouragement. His invaluable advice influences my thoughts on econometric research. He taught me what constitutes an insightful and impactful research article.

I am grateful to Bo E. Honoré, Michal Kolesár, Mikkel Plagborg-Møller, and especially Mark W. Watson for stimulating discussions and insightful suggestions. I also would like to thank many wonderful friends at Princeton, especially to Paul Ho, Mingyu Chen, Yulong Wang, Donghwa Shin, Ioannis Branikas, Michael Dobrew, and Federico Huneus. These great friends made my years at Princeton more enjoyable.

I will be eternally grateful to my wife, Shanshan Yang, for her continuous support, encouragement, and understanding, through all this journey. She, together with our son Isaac, inspires me to be better person. I would never have completed this work without them.

As always, I want to thank my parents in China, for their unconditional love, support, and encouragement throughout my life. Nothing of what I have achieved could have been possible without them.

Chapter 2 of this dissertation is based on a joint paper with my adviser Ulrich K. Müller. We thank David Papell and Mark W. Watson for helpful comments and suggestions. Müller gratefully acknowledges financial support from the National Science Foundation through grant SES-16276.

To Shanshan, Isaac, and my parents, Ruqiang and Yayun.

Contents

Abstract	iii
Acknowledgements	iv
List of Tables	ix
List of Figures	xi
1 Optimal HAR Inference	1
1.1 Introduction	1
1.2 Model and Preliminaries	8
1.3 Optimal Inference in the Diagonal Model	12
1.3.1 Optimal test	13
1.3.2 The optimal EWC test	17
1.3.3 A Rule of thumb	23
1.4 Nearly Optimal Inference in the Exact Model	24
1.4.1 The optimal EWC test	26
1.5 Monte Carlo Simulations	33
1.6 Conclusion	35
2 Generalized Local-to-Unity Models	39
2.1 Introduction	39
2.2 The GLTU(p) Model	44
2.2.1 Set-up	44

2.2.2	Limit Theory	45
2.3	Richness of the GLTU(p) Model Class	47
2.4	A Limited-Information Likelihood Framework	49
2.5	Applications	52
2.5.1	Predictive Regression with a Persistent Predictor	52
2.5.2	Persistence of Deviations from Purchasing Power Parity	55
2.6	Conclusion	59
3	Measuring Persistence of Macroeconomic Time Series: Evidence from the Jordá-Schularick-Taylor Macrohstory Database	60
3.1	Introduction	60
3.2	Models and Measures of Persistence	62
3.3	Econometric Method	66
3.4	Data Description	71
3.5	Empirical Results	74
3.6	Concluding Remarks	76
A	Appendices for Chapter 1	84
A.1	Proofs in Section 1.3	84
A.1.1	Auxiliary lemmas	85
A.1.2	Proof of Theorem 1.3.2	94
A.2	Computational Details in Section 1.3	96
A.2.1	An algorithm to compute q^* in Theorem 1.3.2	96
A.2.2	Computational details for cv_{q^*} in Theorem 1.3.2	97
A.3	More Tables in the Diagonal Model	98
A.4	Computational details in Section 1.4	98
B	Appendices for Chapter 2	103
B.1	Proof of Theorem 1	103

B.2	Proof of Theorem 2	107
C	Appendix for Chapter 3	119
C.1	More Tables of α_T and τ_T under GLTU(p)	119
	References	124

List of Tables

1.1	Optimal q and adjustment factor of the Student- t critical value of level α EWC test.	7
1.2	Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.	7
1.3	Size of the 5% level EWC test using Student- t critical values under selected q	20
1.4	Adjustment factor of the Student- t critical value and weighted average power (WAP) of 5% level EWC test under selected q	20
1.5	Weighted average power (WAP) of the optimal test and the optimal EWC test.	21
1.6	Weighted average power (WAP) of the optimal test and the optimal EWC test.	21
1.7	Diagonal model based cv_q^a , exact model based $cv_q^{a,e}$, and weighted average power (WAP) of 5% level EWC test using $cv_q^{a,e}$	29
1.8	A bound on weighted average power (WAP) and the WAP of the optimal EWC test.	30
1.9	A bound on weighted average power (WAP) and the WAP of the optimal EWC test.	30
1.10	Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test in the exact model.	32

1.11	Small sample performance for inference about population mean . . .	36
1.12	Small sample performance for inference about regression coefficient .	37
2.1	Four GLTU(2) Parameters and Resulting Null Rejection Probability of Campbell-Yogo (2006) Test	54
2.2	Bayesian Limited-Information Analysis of US/UK Real Exchange Rates	58
3.1	Available samples per variable and per country	72
3.2	The 5th, 50th and 95th percentiles of α_T for $p = 1$	77
3.3	The 5th, 50th and 95th percentiles of τ_T for $p = 1$	78
3.4	The 5th, 50th and 95th percentiles of α_T for $p = 2$	79
3.5	The 5th, 50th and 95th percentiles of τ_T for $p = 2$	80
3.6	The 5th, 50th and 95th percentiles of α_T for $p = 5$	81
3.7	The 5th, 50th and 95th percentiles of τ_T for $p = 5$	82
A.1	Optimal q and adjustment factor of the Student- t critical value of level α EWC test.	99
A.2	Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.	99
A.3	Optimal q and adjustment factor of the Student- t critical values of level α EWC test.	100
A.4	Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.	100
C.1	The 5th, 50th and 95th percentiles of α_T for $p = 3$	120
C.2	The 5th, 50th and 95th percentiles of τ_T for $p = 3$	121
C.3	The 5th, 50th and 95th percentiles of α_T for $p = 4$	122
C.4	The 5th, 50th and 95th percentiles of τ_T for $p = 4$	123

List of Figures

1.1	Power function plot of a weighted average power (WAP) bound induced test, optimal EWC test and size-adjusted EWC test using $q = 3$	5
1.2	Illustration of the least favorable distribution of H_0^d against H_{1,f_1}^d as a point mass on f^*	15
1.3	Power function plot of the test φ^* , the optimal EWC test, and the size-adjusted EWC test using $q = 3$	23
1.4	Power function plot of the optimal EWC test, the size-adjusted EWC test using $q = 3$, and the weighted average power bound induced test $\varphi_{\Lambda^\dagger, f_1}$	31
2.1	CRSP Price Dividend Ratio and Empirically Plausible Limiting Log-Spectra	53
2.2	Bayesian Limited-Information Analysis of US/UK Real Exchange Rates	55

Chapter 1

Optimal HAR Inference

1.1 Introduction

This chapter considers the problem of deriving appropriate corrections to standard errors when conducting inference with autocorrelated data. The resulting heteroskedasticity and autocorrelation robust (HAR) inference has applications in OLS and GMM settings.¹ Computing HAR standard errors involves estimating the “long-run variance ” (LRV) in econometric jargon. Classical references on HAR inference in econometrics include Newey and West (1987) and Andrews (1991), among many others. The Newey-West/Andrews approach is to use t - and F -tests based on consistent LRV estimators and to employ the critical values derived from the normal and chi-squared distributions. The resulting HAR standard errors are asymptotically justified in a large variety of circumstances.

Small sample simulations,² however, show that the Newey-West/Andrews approach can lead to tests that incorrectly reject the null far too often. A large sub-

¹For instance, OLS/GMM with HAR inference has been used in many econometric applications, such as testing long-horizon return predictability in finance (see, e.g., Kojien and Van Nieuwerburgh (2011) and Rapach and Zhou (2013)) and estimating impulse response functions by local projections in macroeconomics (see, e.g., Jordà (2005)).

²See, e.g., den Haan and Levin (1994, 1997) for early Monte Carlo evidence of the large size distortions of HAR tests computed using the Newey-West/Andrews approach.

sequent literature (surveyed in Müller (2014)) has proposed many alternative procedures. These procedures employ inconsistent LRV estimators and demonstrate better performance for controlling the null rejection rate. To implement these procedures in practice, however, the user must choose a tuning parameter. One example is the choice of b in the fixed- b scheme,³ in which a fixed- b fraction of the sample size is used as the bandwidth in kernel LRV estimators. Another example is the choice of q in orthonormal series HAR tests,⁴ in which the LRV is estimated by projections onto q mean-zero low-frequency functions of a set of orthonormal functions. The choice of the tuning parameter embeds a tradeoff between bias and variability of the LRV estimator. It subsequently leads to a size-power tradeoff in the resulting HAR inference. Previous studies address this tradeoff by restricting attention to HAR tests that are based on kernel and orthonormal series LRV estimators. They derive the optimal tuning parameter based on second-order asymptotics and under criteria that average the functions of type I and type II errors with different weights.⁵ It is not clear, however, whether the resulting HAR tests would remain optimal in finite samples if those restrictions were not imposed.

The purpose of this chapter is to provide formal results of finite-sample optimal HAR inference about a scalar parameter of interest, without restricting the class of tests and with no ad hoc loss functions. In particular, I derive finite-sample optimal HAR tests in the Gaussian location model, under nonparametric assumptions on the underlying spectral density. I find that with an appropriate adjustment to the critical value, it is nearly optimal to use the so-called equal-weighted cosine (EWC) test (cf.

³See pioneering papers by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002, 2005). Also see Jansson (2004); Müller (2004, 2007); Phillips (2005); Phillips, Sun, and Jin (2006, 2007); Sun, Phillips, and Jin (2008); Atchadé and Cattaneo (2011); Gonçalves and Vogelsang (2011); Sun and Kaplan (2012); and Sun (2014), among many others.

⁴See, e.g., Müller (2004, 2007); Phillips (2005); Ibragimov and Müller (2010); and Sun (2013), among many others.

⁵See, e.g., Sun, Phillips, and Jin (2008) and Lazarus, Lewis, and Stock (2017).

Müller (2004, 2007); Lazarus, Lewis, Stock, and Watson (2018)), where the LRV is estimated by projections onto q type II cosines.

The main assumption in this chapter is that the underlying normalized spectral density is known to lie in a function class $\mathcal{F}$, which possesses a “uniformly minimal” function. By normalized spectral density, I mean its value at the origin is normalized to unity. By “uniformly minimal” function of $\mathcal{F}$, I mean there exists a function $\underline{f}$ in $\mathcal{F}$ such that $\underline{f}(\phi) \leq f(\phi)$, $\phi \in [-\pi, \pi]$ for all f in $\mathcal{F}$. An explicit stand on possible shapes of the spectrum is necessary, because otherwise there does not exist a nontrivial HAR test (cf. Ptscher (2002)). The notion of “uniformly minimal” function further characterizes the minimal assumption on the spectrum, such that a nontrivial HAR test exists. I stress that the class $\mathcal{F}$ is of a nonparametric nature, as opposed to possibly strong parametric classes.⁶ It may contain smoothness restrictions (e.g., bounds on derivatives) and/or shape restrictions (e.g., monotonicity).

This chapter makes three main contributions. First, I establish a finite-sample theory of optimal HAR inference in the Gaussian location model. To do so, I follow Müller (2014) and recast HAR inference as a problem of inference about the covariance matrix of a Gaussian vector. The spectrum, as an infinite-dimensional nuisance parameter, complicates solution of the problem. To overcome this obstacle, I use insights from the so-called least favorable approach and identify the “least favorable distribution” over the class $\mathcal{F}$. The resulting optimal test trades off bias and variability, and requires an adjustment of the critical value to account for the maximum bias of the implied LRV estimator. Both the optimal tradeoff and the adjusted critical value are functions of $\mathcal{F}$.

⁶For examples of parametric classes, Robinson (2005) assumes that the underlying persistence is of the “fractional” type and derives consistent LRV estimators under that class of DGP’s; Müller (2014) assumes that the underlying long-run property can be approximated by a stationary Gaussian AR(1) model, with coefficient arbitrarily close to one and derives uniformly valid inference methods that maximize weighted average power.

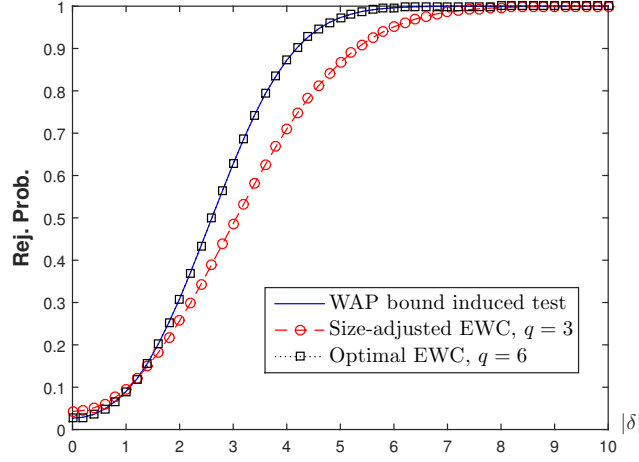
Second, I find that nearly optimal HAR inference can be obtained by using the EWC test, but only after an adjustment to the Student- t critical value. The practical implications are an explicit link between the choice of q and assumptions on the underlying spectrum, as well as a corresponding adjustment to the Student- t critical value. In detail, consider a second-order stationary scalar time series y_t . The spectral density of y_t scaled by 2π is given by the function $f : [-\pi, \pi] \mapsto [0, \infty)$. To test $H_0 : E[y_t] = 0$ against $H_1 : E[y_t] \neq 0$, the EWC test uses a t -statistic

$$t_{EWC}^q = \frac{Y_0}{\sqrt{\sum_{j=1}^q Y_j^2/q}}, \quad (1.1)$$

where Y_0 is the sample mean of y_t and Y_j , $j = 1, 2, \dots, q$ are q weighted averages of y_t as $Y_j = T^{-1}\sqrt{2}\sum_{t=1}^T \cos(\pi j(t-1/2)/T)y_t$. These weighted averages can be approximately thought of as independently distributed, each with variance $T^{-1}f(\pi j/T)$. As mentioned earlier, the choice of q embeds a bias and variance tradeoff of the LRV estimator $\sum_{j=1}^q Y_j^2/q$. The conventional wisdom is to choose q sufficiently small such that $\{Y_j\}_{j=1}^q$ can be treated as independent normal with equal variance. By doing so, one avoid possibly large bias in estimating the LRV, and the resulting EWC test has less size distortions when the Student- t critical value is employed. In contrast, the new EWC test suggests using a larger q and an appropriately enlarged critical value for more powerful inference. Both the choice of q and the critical value adjustment depend on the class $\mathcal{F}$.

Figure 1.1 illustrates this second contribution in the problem of testing $E[y_t] = 0$, $f \in \mathcal{F}$ against the local alternative $E[y_t] = \delta T^{-1/2}$ for $T = 100$, where y_t follows a Gaussian white noise and the “uniformly minimal” function of $\mathcal{F}$ is the normalized spectrum of an AR(1) model with coefficient 0.8. In this context, to avoid size distortions larger than 0.01, one needs to choose $q = 3$ when the Student- t critical value is employed. The new nearly optimal EWC test, however, has $q = 6$ and inflates

Figure 1.1: Power function plot of a weighted average power (WAP) bound induced test, optimal EWC test and size-adjusted EWC test using $q = 3$.



Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}$ and y_t follows a Gaussian white noise. Under the null, the “uniformly minimal” function of the class $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

the Student- t critical value by a factor of 1.15. This new EWC test nearly achieves a weighted average power bound for all size-controlling scale invariant tests. It has a 38.1% efficiency gain over the size-adjusted EWC test using $q = 3$, in order to achieve the same power of 0.5.⁷

Third, I propose a simple first-order adjustment to the critical value of the EWC test. The adjusted critical value is computed easily, by inverting a one-dimensional numerical integral. For practical convenience, I offer a rule of thumb to adjust the Student- t critical value of the EWC test in Table 1.2, as follows. Under a series of classes $\mathcal{F}$ in which the “uniformly minimal” function is the normalized spectrum of an AR(1) with coefficient ρ , Table 1.1 lists the optimal choice of q and the adjustment factor of the Student- t critical value for each combination of ρ and nominal level α .

⁷By efficiency gain, I mean the increase of δ^2 in percent for the size-adjusted EWC test using $q = 3$ in order to achieve the same power of the new EWC test. I note that one cannot directly appeal to Pitman efficiency measure (the increase of the number of observations required to achieve the same power) in the context of Figure 1.1, since the sample size T is fixed at 100. A different calculation, however, shows that for $T = 50$ the size-adjusted EWC test using $q = 6$ has power of 0.5, under the same δ such that the EWC test using $q = 3$ yields power of 0.5 for $T = 100$.

Table 1.2 collapses Table 1.1 into (α, q) pairs. As a practical matter, if researchers pick a value of q by some other means, then I suggest adjusting the corresponding Student- t critical value according to Table 1.2.

This chapter relates to a large literature. First, unlike the vast majority of the HAR literature, I consider optimal HAR inferences without restricting the class of tests. Second, I do not appeal to ad hoc loss functions in addressing the size-power tradeoff in HAR inference. Rather, I consider that tradeoff in the natural setting of hypothesis testing, where uniform size control is imposed and the most powerful inference is desired. Third, the majority of the literature addresses the sampling variability of LRV estimators via the so-called fixed- b asymptotics, and further accounts for bias by higher-order adjustment to the fixed- b critical value.⁸ In contrast, I concurrently tackle bias and variance in estimating the LRV by a first-order adjustment. Even so, the resulting adjusted critical value is easily computed without simulations.

The suggestion of using a larger q and enlarged critical values for the EWC test mirrors recent recommendations for nonparametric inference, such as those of Armstrong and Kolesár (2018a,b). In different contexts, Armstrong and Kolesár and I both stress the advantage of accepting bias in estimating a nonparametric function and of then using a suitably adjusted critical value to account for the maximum bias. Our frameworks are, however, different. I consider a Gaussian experiment in which the heteroskedasticity is governed by an unknown nonparametric function, while the main focus in Armstrong and Kolesár (2018a) is an unknown regression function in the mean of a homoskedastic Gaussian experiment.

The remainder of the chapter is organized as follows. Section 1.2 sets up the model and discusses preliminaries. Section 1.3 derives optimal HAR inference under an essential simplification. Section 1.4 relaxes the simplification and discusses nearly

⁸See, e.g., Velasco and Robinson (2001); Sun, Phillips, and Jin (2008); Sun (2011, 2013, 2014); and Lazarus, Lewis, and Stock (2017).

Table 1.1: Optimal q and adjustment factor of the Student- t critical value of level α EWC test.

ρ	0.5	0.7	0.8	0.9	0.95	0.98
$\alpha = 0.01$	(15, 1.07)	(10, 1.12)	(8, 1.19)	(5, 1.34)	(3, 1.57)	(2, 2.45)
$\alpha = 0.05$	(12, 1.05)	(8, 1.09)	(6, 1.13)	(4, 1.25)	(3, 1.55)	(1, 1.85)
$\alpha = 0.10$	(11, 1.04)	(7, 1.07)	(5, 1.10)	(4, 1.25)	(2, 1.36)	(1, 1.85)

Notes: Based on a series of classes $\mathcal{F}$, in which the “uniformly minimal” function is the normalized spectrum of an AR(1) with coefficient ρ . Sample size T is 100.

Table 1.2: Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	16	20
$\alpha = 0.01$	1.46 (0.93)	1.32 (0.88)	1.19 (0.80)	1.15 (0.75)	1.12 (0.70)	1.10 (0.65)	1.09 (0.60)	1.08 (0.56)	1.06 (0.45)	1.05 (0.33)
$\alpha = 0.05$	1.25 (0.90)	1.15 (0.82)	1.09 (0.70)	1.07 (0.65)	1.06 (0.60)	1.06 (0.55)	1.05 (0.50)	1.04 (0.45)	1.04 (0.35)	1.03 (0.23)
$\alpha = 0.10$	1.25 (0.90)	1.10 (0.78)	1.06 (0.65)	1.05 (0.60)	1.05 (0.55)	1.04 (0.50)	1.04 (0.45)	1.03 (0.40)	1.02 (0.28)	1.02 (0.18)

Notes: Each q is justified as the optimal choice of level α EWC test, under some class $\mathcal{F}$ and for sample size $T = 100$. An example of the corresponding class $\mathcal{F}$ is the one in which the “uniformly minimal” function is the normalized spectrum of an AR(1) model. Numbers in parentheses are the corresponding AR(1) coefficients.

optimal HAR inference. Section 1.5 contains simulation results, and Section 1.6 concludes. Proofs and computational details are provided in the appendices.

1.2 Model and Preliminaries

Throughout the chapter, I mainly focus on inference about μ in the location model,

$$y_t = \mu + u_t, \quad t = 1, 2, \dots, T, \quad (1.2)$$

where μ is the population mean of y_t and u_t is a mean-zero stationary Gaussian process with absolutely summable autocovariances $\gamma(j) = E[u_t u_{t-j}]$. The spectrum of y_t scaled by 2π is given by the even function $f : [-\pi, \pi] \mapsto [0, \infty)$ defined via $f(\lambda) = \sum_{j=-\infty}^{\infty} \cos(j\lambda) \gamma(j)$. With $y = (y_1, y_2, \dots, y_T)'$ and $e = (1, 1, \dots, 1)'$,

$$y \sim \mathcal{N}(\mu e, \Sigma(f)), \quad (1.3)$$

where $\Sigma(f)$ has elements $\Sigma(f)_{j,k} = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\lambda) e^{-i(j-k)\lambda} d\lambda$ with $i = \sqrt{-1}$.

The HAR inference problem concerns testing $H_0 : \mu = 0$ (otherwise, subtract the hypothesized mean from y_t) against $H_1 : \mu \neq 0$ based on the observation y . The derivation of powerful tests in this problem is complicated by the fact that the alternative is composite (μ is not specified under H_1), and the presence of the infinite-dimensional nuisance parameter f . I follow standard approaches to dealing with μ and mainly focus on tackling the nuisance parameter f in this chapter.

It is useful to take a spectral transformation of the model (1.3). In particular, as introduced in the introduction, consider the one-to-one transformation from $\{y_t\}_{t=1}^T$ into the sample mean $Y_0 = T^{-1} \sum_{t=1}^T y_t$ and the $T - 1$ weighted averages:

$$Y_j = T^{-1} \sqrt{2} \sum_{t=1}^T \cos(\pi j(t - 1/2)/T) y_t, \quad j = 1, 2, \dots, T - 1. \quad (1.4)$$

Define Φ as the $T \times T$ matrix with first column equal to $T^{-1}e$, and $(j+1)$ th column with elements $T^{-1}\sqrt{2}\cos(\pi j(t-1/2)/T)$, $t = 1, \dots, T$, and ι_1 as the first column of I_T . Then

$$Y = (Y_0, Y_1, \dots, Y_T)' = \Phi' y \sim \mathcal{N}(\mu \iota_1, \Omega_0(f)) \quad (1.5)$$

where $\Omega_0(f) = \Phi' \Sigma(f) \Phi$. The HAR testing problem now becomes $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ based on the observation Y .

A common device for dealing with the composite alternative in the nature of μ is to search for tests that maximize weighted average power over μ . For computational convenience, I follow Müller (2014) to consider a Gaussian weighting function for μ with mean zero and variance η^2 . The scalar η^2 governs whether closer or distant alternatives are emphasized by the weighting function. For a given f , the choice $\eta^2 = (\kappa-1)\Omega_0(f)_{1,1}$ (for analytical simplifications later) effectively changes the testing problem to $H'_0 : Y \sim \mathcal{N}(0, \Omega_0(f))$ against $H'_1 : Y \sim \mathcal{N}(0, \Omega_1(f))$, where

$$\Omega_1(f) = \Omega_0(f) + (\kappa-1)\iota_1 \iota_1' \Omega_0(f)_{1,1}. \quad (1.6)$$

This transforms the problem into one of inference about covariance matrices. The hyperparameter κ specifies a weighted average power criterion. As argued by King (1987), it makes sense to choose κ in a way such that good tests have approximately 50% weighted average power. The choice of $\kappa = 11$ would induce the resulting best 5% level (infeasible) test (reject if $Y_0^2 > 3.84\Omega_0(f)_{1,1}$) to have power of approximately $P(\chi_1^2 > 3.84/11) \approx 56\%$. I thus use $\kappa = 11$ throughout the implementations.

In most applications, it is reasonable to impose that if the null hypothesis is rejected for some observation Y , then it should also be rejected for the observation aY , for any $a > 0$. Due to this scale invariance restriction, it is without loss of generality to normalize all f such that $f(0) \equiv 1$. Furthermore, by standard testing

theory (see, e.g., Chapter 6 in Lehmann and Romano (2005)), any test satisfying this scale invariance property can be written as a function of $Y^s = Y/\sqrt{Y'Y}$. The density of Y^s under H'_i , $i = 0, 1$ is equal to (see Kariya (1980) and King (1980))

$$h_{i,f}(y^s) = C|\Omega_i(f)|^{-1/2} (y^{s'}\Omega_i(f)^{-1}y^s)^{-T/2} \quad (1.7)$$

for some constant C .

By restricting to scale invariant tests, the HAR testing problem has been further transformed into H''_0 : “ Y^s has density $h_{0,f}$ ” against H''_1 : “ Y^s has density $h_{1,f}$.” The problem remains nonstandard due to the presence of nuisance parameter f . For simplicity, I direct power at flat spectrum $f_1 = 1$. The alternative H''_1 then becomes a single hypothesis H''_{1,f_1} : “ Y^s has density h_{1,f_1} ,” where $\Omega_1(f_1) = \kappa T^{-1} \text{diag}(1, \kappa^{-1}, \dots, \kappa^{-1})$. Moreover, under the null I assume f belongs to an explicit function class $\mathcal{F}$ and seek scale invariant tests that uniformly control size over $\mathcal{F}$.

The main concern of this chapter is to test the composite null H''_0 against H''_{1,f_1} . In this context, a well-known general solution to this type of problem proceeds as follows (cf. Lehmann and Romano (2005)). Suppose Λ is some probability distribution over $\mathcal{F}$, and the composite null H''_0 is replaced by the single hypothesis $H''_{0,\Lambda}$: “ Y^s has density $\int h_{0,f}d\Lambda(f)$.” Any ad hoc test φ_{ah} that is known to be of level α under H''_0 also controls size under $H''_{0,\Lambda}$, because $\int \varphi_{\text{ah}}(y^s) \int h_{0,f}d\Lambda(f)dy^s = \int \int \varphi_{\text{ah}}(y^s)h_{0,f}dy^sd\Lambda(f) \leq \alpha$. As a result, by Neyman-Pearson lemma, the likelihood ratio test of $H''_{0,\Lambda}$ against H''_{1,f_1} , denoted by φ_{Λ,f_1} , yields a bound on the power of φ_{ah} . Furthermore, if φ_{Λ,f_1} also controls size under H''_0 , then it must be the best test of H''_0 against H''_{1,f_1} and the resulting power bound is the lowest possible power bound. In the jargon of statistical testing, the distribution that yields the best test

(should it exist) is called the “least favorable distribution,” and I denote it by Λ^* throughout the chapter.

Unfortunately, there is no systematic way of deriving the least favorable distribution. To make progress, I proceed in the following two steps. First, I consider an approximate “diagonal” model, in which for a given f the joint distribution of Y under the null is

$$Y \sim \mathcal{N}(0, T^{-1} \text{diag}(f(0), f(\pi/T), \dots, f(\pi(T-1)/T))). \quad (1.8)$$

In model (1.8), I analytically derive the least favorable distribution of H_0'' against H_{1,f_1}'' , under mild assumptions on the class $\mathcal{F}$. I also find that the “optimal” EWC test is nearly as powerful as the derived optimal test. By optimal EWC test, I mean the EWC test under an optimal choice of q and with an optimal adjustment to the critical value. Second, despite the analytical intractability of the least favorable distribution without approximation (1.8), it is still feasible to obtain upper bounds on the power of size-controlling tests of H_0'' against H_{1,f_1}'' . In particular, I use insights on optimal tests in the diagonal model (1.8) to establish tight power bounds for all valid tests in the exact model (1.5). It turns out that the optimal EWC test comes close to achieving this power bound. In light of Lemma 1 in Elliott, Müller, and Watson (2015), this implies that the resulting new EWC test is nearly optimal for HAR inference, and the proposed power bound is essentially the least upper bound. I elaborate on the above analyses in Sections 1.3 and 1.4.

Model (1.8) is in general an approximation of the exact model (1.5) by ignoring off-diagonal elements and simplifying the diagonal elements in $\Omega_0(f)$. It is motivated by the fact that (1.8) holds exactly when time series y_t follows a Gaussian white noise or a Gaussian random walk process. For stationary y_t with f falling into other parametric classes, Müller and Watson (2008) find that the covariance matrix of

$(Y_0, Y_1, \dots, Y_q)'$ is nearly diagonal for fixed q and large T . For stationary Gaussian y_t with f being in nonparametric classes, the aforementioned optimality results suggest that (1.8) is a useful simplification of (1.5) for HAR inference. I will refer to (1.8) as the diagonal model and (1.5) as the exact model hereafter.

1.3 Optimal Inference in the Diagonal Model

In this section, I derive powerful HAR tests in diagonal model (1.8). As explained in Section 1.2, I restrict attention to scale invariant tests that maximize weighted average power over μ and direct power at the flat spectrum f_1 . Under the weighted average power criterion, as specified by a given κ , I seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$H_0^d : Y \sim \mathcal{N}(0, T^{-1} \text{diag}(1, f(\pi/T), \dots, f(\pi(T-1)/T)), f \in \mathcal{F} \quad (1.9)$$

against $H_{1,f_1}^d : Y \sim \mathcal{N}(0, \kappa T^{-1} \text{diag}(1, \kappa^{-1}, \dots, \kappa^{-1}))$,

where the superscript d in H_0^d and H_{1,f_1}^d denotes the diagonal model.

The following assumptions are imposed on $\mathcal{F}$ throughout this section.

Assumption 1.3.1

- (a) *There exists a $\underline{f} \in \mathcal{F}$ such that $\underline{f}(\phi) \leq f(\phi)$, $\phi \in [-\pi, \pi]$ for all $f \in \mathcal{F}$.*
- (b) *$\underline{f}(\pi j/T) \geq \underline{f}(\pi(j+1)/T)$, $j = 0, 1, \dots, T-2$.*
- (c) *The class $\mathcal{F}$ contains all kinked functions defined by $f_a(\phi) = \max\{\underline{f}(\phi), a\}$, for $a \in [0, 1]$.*

Assumption 3.1(a) states the existence of a “uniformly minimal” function in $\mathcal{F}$, which I will use $\underline{f}$ to denote throughout the chapter, while Assumption 3.1(b) requires $\underline{f}$ to be non-increasing at $\lambda = \pi j/T$, $j = 0, 1, \dots, T-1$. Assumption 3.1(c) is needed

to ensure the existence of a point mass least favorable distribution. It is worth noting that only the evaluations of f at frequencies $\pi j/T$, $j = 0, 1, \dots, T-1$ matter in (1.9). Assumption 3.1(a) is thus sufficient but not necessary.

Assumption 1.3.1 is more general than what is assumed in the majority of HAR studies. For example, if $\underline{f}$ is the normalized spectrum of an AR(1) model with coefficient 0.8 and sample size $T = 100$, one is not committing to any parametric classes such as the AR(1) model. Rather, various kinds of parametric classes are covered, as long as the underlying normalized spectrum lies above $\underline{f}$. It includes, but is not limited to, AR(1) models with coefficient less than 0.8 and all MA(1) models and ARMA models whose spectra may oscillate but are above $\underline{f}$. Furthermore, Assumption 1.3.1 is satisfied by most function classes assumed in the nonparametric inference literature. For example, when $\mathcal{F}$ is the class in which the first derivative of the log spectrum is bounded by a constant C , the corresponding “uniformly minimal” function emerges as $\underline{f}(\phi) = \exp(-C\phi)$.

1.3.1 Optimal test

The optimal HAR test in the diagonal model is stated in the following theorem.

Theorem 1.3.2 *Let $\mathcal{F}$ be a set of f satisfying Assumption 1.3.1 with the “uniformly minimal” function $\underline{f}$, and for a given κ that specifies a weighted average power criterion,*

1. *If $\underline{f}(\pi/T) \leq \kappa^{-1}$, then the best weighted average power maximizing scale invariant test of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ is a randomized test.*
2. *If $\underline{f}(\pi/T) > \kappa^{-1}$, then the best level α weighted average power maximizing scale invariant test φ^* of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ rejects for large values of*

$$\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2 / \underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} \quad (1.10)$$

for a unique $1 \leq q^* \leq T - 1$, and with the critical value cv_{q^*} such that the test is of level α under $f = \underline{f}$.

The proof of part 1 of Theorem 1.3.2 is simple. Notice that for a given κ , if $\underline{f}(\pi/T) \leq \kappa^{-1}$, then the alternative H''_{1,f_1} is included in the null H_0^d . As a result, any nontrivial size-controlling test cannot be more powerful than a randomized test.

The idea of the proof for part 2 of Theorem 1.3.2 is to conjecture and verify that the least favorable distribution Λ^* puts a point mass on a function in $\mathcal{F}$. The logic is as follows. Suppose the conjecture is true and Λ^* concentrates on the function f^* . By the Neyman-Pearson lemma, the optimal test of H''_{0,Λ^*} against H''_{1,f_1} in the diagonal model is

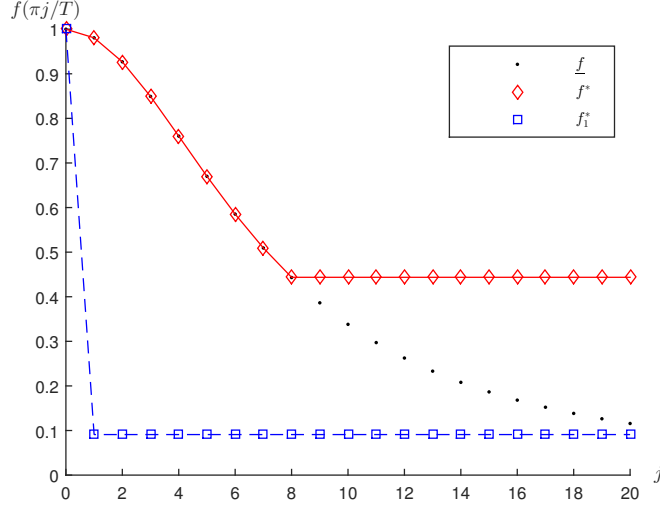
$$\varphi_{\Lambda^*,f_1} = 1 \left[\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2 / f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \text{cv} \right],$$

for some $\text{cv} \geq 0$. On the other hand, as discussed in Section 1.2, for Λ^* to be the least favorable distribution, one needs φ_{Λ^*,f_1} to uniformly control size under H_0'' . Intuitively, this requires H''_{0,Λ^*} to be as indistinguishable as possible from H''_{1,f_1} . This somewhat implies that the function f^* must mimic the discontinuous function of ϕ as $f_1^*(\phi) = \kappa^{-1} \mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$. As illustrated by Figure 1.2, the function f^* must then be kink-shaped, given the presence of $\underline{f}$. I further show that the optimal location of the kink in f^* in conjunction with the resulting cv is equivalent to ignoring Y_j with index $j > q^*$. This then gives rise to the optimal test statistic (1.10). The formal proof of Theorem 1.3.2 is given in Appendix A.1.

Discussion

Comment 1. Part 1 of Theorem 1.3.2 provides a sharper result than Ptscher (2002). In particular, it characterizes the concrete minimal smoothness assumption on the spectrum such that a nontrivial valid HAR test exists. This is beyond Pötscher's

Figure 1.2: Illustration of the least favorable distribution of H_0^d against H_{1,f_1}^d as a point mass on f^* .



Notes: The “uniformly minimal” function $\underline{f}$ is the normalized spectrum of an AR(1) model with coefficient 0.8. I use $\kappa = 11$ for f_1^* . Sample size T is 100.

negative result, which only shows that the HAR testing problem is ill-posed if no priori assumptions are imposed on the set of data generating processes.

Comment 2. The optimal test (1.10) with the resulting critical value cv_{q^*} can be rewritten as

$$\left| \frac{Y_0}{\sqrt{\sum_{j=1}^{q^*} w_j Y_j^2}} \right| > 1, \quad (1.11)$$

where the weight w_j depends on κ , $\underline{f}(\pi j/T)$ and cv_{q^*} . As can be seen, the implied (inconsistent) LRV estimator $\sum_{j=1}^{q^*} w_j Y_j^2$ does not necessarily fall into the popular kernel and orthonormal series families. This is because the optimal test is not constructed by exploiting flatness of the spectrum close to the origin. Rather, I take an explicit stand on possible shapes of the spectrum and endogenously account for the maximum bias via the weights w_j ’s.

Comment 3. A possibly small q^* may emerge in Theorem 1.3.2 for some $\mathcal{F}$. It is, however, worth noting that such q^* is an implication, not an assumption. In

particular, I do not start by restricting attention to the class of tests as functions of $(Y_1^s, \dots, Y_{q^*}^s)'$. In contrast, the approach taken by Müller (2014) assumes some fixed q as the starting point.

Comment 4. With appropriate modifications of Assumption 1.3.1, Theorem 1.3.2 can be adapted to the problem of H_0^d against $H_{1, \tilde{f}_1}^d$ for other fixed alternative $\tilde{f}_1$. In that case, the resulting q^* is also $\tilde{f}_1$ -dependent. Furthermore, Theorem 1.3.2 can be generalized to a minimax result, in which f belongs to a nonparametric class $\mathcal{G} \subset \mathcal{F}$ under H_1 . In that case, a “uniformly maximal” function in $\mathcal{F}$ must be properly defined as $\underline{f}$ in Assumption 1.3.1.

Computational considerations

The existence and uniqueness result of q^* in Theorem 1.3.2 naturally brings computational convenience in practice. For example, for a given $\mathcal{F}$ satisfying Assumption 1.3.1, one can appeal to the bisection method to locate q^* . In my implementations, this takes little computing time by using the simple algorithm in Appendix A.2.1. Moreover, for a given $\mathcal{F}$ and the resulting q^* , the critical value cv_{q^*} can easily be determined due to the following formula of Bakirov and Székely (2006):

$$P\left(Z_0^2 \geq \sum_{j=1}^n \zeta_j Z_j^2\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} du}{\sqrt{\prod_{j=1}^n (1-u^2 + \zeta_j)}}, \quad (1.12)$$

where $\{Z_j\}_{j=0}^n$ are $n+1$ i.i.d. standard normal random variables and $\zeta_j \geq 0$, $j = 1, \dots, n$. By the t -statistic expression (1.11), part 2 of Theorem 1.3.2, and (1.12), the level α constraint for the optimal test becomes

$$P\left(\frac{Z_0^2}{\sum_{j=1}^{q^*} w_j f(\pi j/T) Z_j^2} > 1\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q^*-1)/2} du}{\sqrt{\prod_{j=1}^{q^*} (1-u^2 + w_j \underline{f}(\pi j/T))}} = \alpha, \quad (1.13)$$

where $w_j = [\kappa \text{cv}_{q^*} - 1/\underline{f}(\pi j/T)] (1 - \text{cv}_{q^*})^{-1}$ is strictly monotone in cv_{q^*} under Assumption 1.3.1. The critical value cv_{q^*} is then readily determined by solving equation (1.13). Computational details are provided in Appendix A.2.2.

1.3.2 The optimal EWC test

By using higher-order expansions, Lazarus, Lewis, and Stock (2017) derive a size-power frontier for kernel and orthonormal series HAR tests under an asymptotic framework. The EWC test is shown to achieve that frontier in their context. It is, however, not clear how the EWC test performs in finite-sample contexts and in the unrestricted class of tests. The optimal HAR test derived in the last section provides a natural benchmark to gauge the performance of an ad hoc test. In this section I take up the EWC test as the ad hoc test and discuss its properties.

I have three related goals. The first is to study the (weighted average) power properties of the EWC test relative to the optimal test in Theorem 1.3.2. As it turns out, the EWC test is close to optimal, under an appropriate choice of q and with the adjusted critical value. Given the efficiency property of this new EWC test, the second goal is to develop simple procedures to implement the test. I discuss the two goals in reverse order, first elaborating on critical value adjustment and optimal choice of q for the EWC test, and then studying the power of the resulting test. My last goal is to compare the practical implications of the new EWC test with the conventional wisdom, that is, to choose a sufficiently small q and use the Student- t critical value. The general takeaway from the comparison is: One should use the EWC test with a larger q and appropriately enlarged critical values for more powerful HAR inference.

To clarify ideas and illustrate points in a consistent manner, I use the following running example throughout this section: Under the null, the “uniformly minimal” function $\underline{f}$ of the class $\mathcal{F}$ is AR(1) with coefficient 0.8; sample size T is fixed to be 100. In addition, I will frequently refer to the following two types of classes. For the first

type, the “uniformly minimal” function of $\mathcal{F}$ is the normalized spectrum of an AR(1) with coefficient ρ . For the second type, all spectra in $\mathcal{F}$ satisfy a global smoothness assumption, that is, the first derivative of the log-spectrum $\log(f)$ is bounded by a constant C .

Critical value adjustment and choice of q

The diagonal model (1.8) makes it easy to adjust the critical value for a given class $\mathcal{F}$. In particular, for a given $f \in \mathcal{F}$, the null rejection probability of the EWC test (1.1) using the critical value cv is

$$P\left(\left|\frac{Y_0}{\sqrt{\sum_{j=1}^q Y_j^2/q}}\right| \geq \text{cv}\right) = P\left(\frac{Z_0^2}{q^{-1} \text{cv}^2 \sum_{j=1}^q f(\pi j/T) Z_j^2} \geq 1\right), \quad (1.14)$$

where $\{Z_j\}_{j=0}^q$ are $q+1$ i.i.d. standard normal random variables. Under Assumption 1.3.1(a), it is not hard to see that (1.14) as a functional of f is maximized at $\underline{f}$, regardless of the choice of q and the critical value cv . Two implications are immediate. First, for the testing problem (1.9) under a given $\mathcal{F}$, it is easy to gauge the size performance of any ad hoc EWC test. In the context of the running example, Table 1.3 shows the size of the 5% EWC test using the Student- t critical value under selected choices of q . As can be seen, for size distortions less than 0.01, one needs to use $q = 3$ in the usual EWC test.

Second, by Bakirov and Szekely’s (2006) formula (1.12), it is easy to adjust the critical value of the EWC test under any ad hoc q . Specifically, as in solving for the critical value of the optimal test in Section 1.3.1, the adjusted critical value cv_q^a of the level α EWC test under given q is obtained by inverting the following level constraint:

$$\alpha = P\left(\frac{Z_0^2}{q^{-1}(\text{cv}_q^a)^2 \sum_{j=1}^q f(\pi j/T) Z_j^2} \geq 1\right)$$

$$= \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q-1)/2} du}{\sqrt{\prod_{j=1}^q (1-u^2 + q^{-1}(\text{cv}_q^a)^2 \underline{f}(\pi j/T))}}. \quad (1.15)$$

In the context of the running example, the first row in Table 1.4 summarizes the adjustment factor of the resulting adjusted critical value relative to the Student- t critical value under various q . As can be seen, in order to explicitly account for the resulting downward bias of the LRV estimator $\sum_{j=1}^q Y_j^2/q$, one must inflate the usual Student- t critical value by a factor larger than 1.

Now consider the choice of q in the EWC test. Under given q and using the adjusted critical value cv_q^a , the weighted average power of the resulting EWC test is

$$P\left(\frac{Z_0^2}{q^{-1}\kappa(\text{cv}_q^a)^2 \sum_{j=1}^q Z_j^2} \geq 1\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q-1)/2} du}{\sqrt{\prod_{j=1}^q (1-u^2 + q^{-1}\kappa(\text{cv}_q^a)^2)}}. \quad (1.16)$$

The weighted average power (1.16) can easily be computed for every q . Under a given $\mathcal{F}$ and nominal level α , the optimal choice of q for the EWC test is then defined as the one such that the resulting EWC test has the largest weighted average power. I refer to the EWC test under the optimal choice of q and with the adjusted critical value as the optimal EWC test. I stress that the notion of “optimality” for this new EWC test is with respect to the assumptions on the underlying spectrum, that is, the class $\mathcal{F}$.

Power of the optimal EWC test

Tables 1.5 and 1.6 summarize the weighted average power of the optimal EWC test and the corresponding optimal test under the aforementioned two types of classes $\mathcal{F}$, respectively. As can be seen, the optimal EWC test is nearly as powerful as the optimal test, regardless of the underlying $\mathcal{F}$ within the two types of classes. In unreported numerical results, under various $\mathcal{F}$ of other smoothness types, the near optimality property of the optimal EWC test continues to hold.

Table 1.3: Size of the 5% level EWC test using Student- t critical values under selected q .

q	3	4	6	8	10
Size	0.056	0.061	0.073	0.089	0.107

Notes: The “uniformly minimal” function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

Table 1.4: Adjustment factor of the Student- t critical value and weighted average power (WAP) of 5% level EWC test under selected q .

q	3	4	5	6	7	8	9	10
Adj. factor	1.044	1.068	1.096	1.126	1.158	1.191	1.225	1.259
WAP	0.390	0.422	0.434	0.438	0.436	0.431	0.425	0.417

Notes: The “uniformly minimal” function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

Table 1.5: Weighted average power (WAP) of the optimal test and the optimal EWC test.

ρ	0.50	0.60	0.70	0.80	0.90	0.95	0.98	0.99
WAP								
Optimal test	0.506	0.493	0.475	0.441	0.357	0.236	0.089	0.051
Optimal EWC	0.504	0.491	0.472	0.438	0.353	0.233	0.089	0.051

Notes: The “uniformly minimal” function of $\mathcal{F}$ corresponds to an $\text{AR}(1)$ with coefficient ρ . Nominal level is 5%. Sample size T is 100.

Table 1.6: Weighted average power (WAP) of the optimal test and the optimal EWC test.

C	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
WAP								
Optimal test	0.365	0.419	0.458	0.485	0.504	0.517	0.527	0.534
Optimal EWC	0.361	0.415	0.454	0.482	0.501	0.515	0.526	0.533

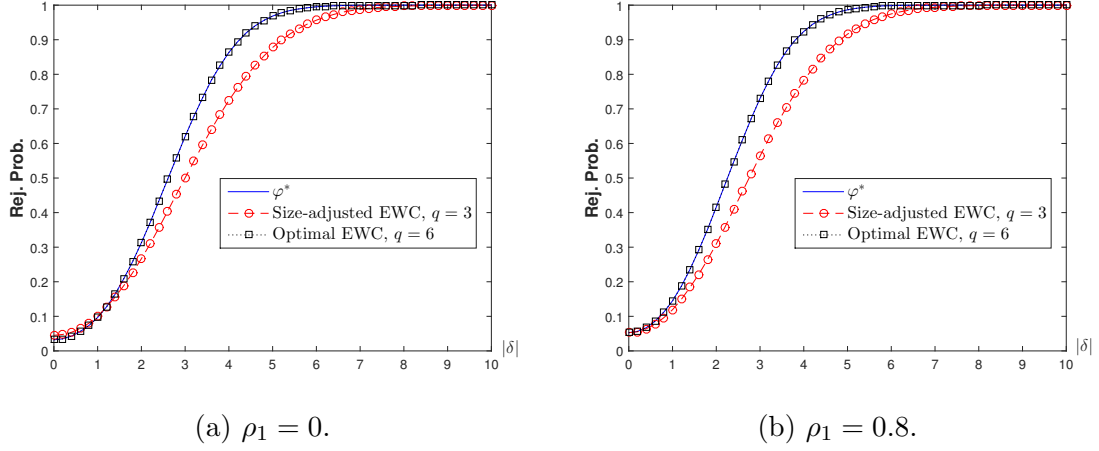
Notes: The “uniformly minimal” function of $\mathcal{F}$ is $f(\phi) = \exp(-C\phi)$. Nominal level is 5%. Sample size T is 100.

Practical implications

Recall that the conventional wisdom is to use a sufficiently small q and to employ the Student- t critical value. I find, however, that it is optimal to use a larger q and to employ an enlarged critical value. Take the running example as an illustration. As explained earlier, for size distortions less than 0.01, one needs to use $q = 3$ in the usual EWC test in which the Student- t critical value is employed. However, as highlighted in Table 1.4, the optimal EWC test has a larger $q = 6$, and the corresponding Student- t critical value must be inflated by a factor of 1.13 for exact size control. To ensure an apples-to-apples comparison, I compute the size-adjusted weighted average power of the usual EWC test using $q = 3$. In contrast to the optimal EWC test, this EWC test has about 11% weighted average power loss.

The superior power property of the optimal EWC test is further evident when local alternatives are considered. In particular, in the context of the running example, I consider $\mu = \delta T^{-1/2}(1 - \rho_1)^{-1}$ under the alternative. Panels (a) and (b) of Figure 1.3 plot the power of the test φ^* , the optimal EWC test, and the size-adjusted EWC test using $q = 3$ for various δ under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively. As can be seen in panel (a), even though the optimal EWC test underrejects under the null, it is more powerful than the EWC test using $q = 3$ in detecting local deviations from the null. Specifically, by using the optimal EWC test, a 32.0% efficiency improvement is obtained in order to achieve the same power of 0.5. In the case in which $\rho_1 = 0.8$, the efficiency gain is larger (48.7%), since the optimal EWC test then exactly controls size by construction. Furthermore, given that the optimal EWC test is numerically found to be nearly as powerful as the overall optimal test φ^* in terms of weighted average power under the white noise alternative, it is not surprising to see that the power functions of these two tests are almost identical.

Figure 1.3: Power function plot of the test φ^* , the optimal EWC test, and the size-adjusted EWC test using $q = 3$.



Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}(1 - \rho_1)^{-1}$ and y_t follows a Gaussian AR(1) with coefficient ρ_1 . Under the null, the $\underline{f}$ of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size T is 100.

1.3.3 A Rule of thumb

As a practical matter, one might like to estimate the smoothness class $\mathcal{F}$ from data. Unfortunately, the attempt is not useful. This is because the (nearly) optimal test depends on $\mathcal{F}$, and a “larger” $\mathcal{F}$ leads to a lower power. As a result, one cannot estimate $\mathcal{F}$ and still control size. In implementations of the EWC test, if q is chosen by some other approach, it still makes sense to adjust the Student- t critical value given the previous analysis of the optimal EWC test. As a rule of thumb, I suggest that practitioners implement the EWC test and adjust the Student- t critical value according to Table 1.2. In detail, the suggested test about the population mean $H_0 : \mu = \mu_0$ of an observed scalar time series $\{y_t\}_{t=1}^T$ is computed as follows.

1. Compute the T cosine weighted averages of $\{y_t\}_{t=1}^T$: $Y_0 = T^{-1} \sum_{t=1}^T (y_t - \mu_0)$ and $Y_j = T^{-1} \sqrt{2} \sum_{t=1}^T \cos(\pi j(t - 1/2)/T) y_t$, $j = 1, 2, \dots, T - 1$.
2. For the researcher’s choice of q , compute the t -statistic $t_{Y,q} = Y_0 / \sqrt{q^{-1} \sum_{j=1}^q Y_j^2}$.

3. Reject the null hypothesis at level α if $|t_{Y,q}| > B_{\alpha,q} \text{cv}_q^{na}(\alpha)$, where $\text{cv}_q^{na}(\alpha)$ is the Student- t_q critical value and $B_{\alpha,q}$ is the unparenthesized number in the (α, q) -th entry of Table 1.2.

As explained in the introduction, the adjustment factors in Table 1.2 are calibrated based on a series of classes $\mathcal{F}$ in which the “uniformly minimal” function is the normalized spectrum of an AR(1) with coefficient ρ . I make two additional remarks. First, there may be multiple ρ such that the same optimal q emerges, under the respective $\mathcal{F}$. Second, the adjustment factor in each (α, q) -th entry does not change substantially under other types of smoothness classes. Sets of tables similar to Tables 1.1 and 1.2 are provided in Appendix A.3, in which the class $\mathcal{F}$ imposes some global smoothness assumption on the spectrum. For these reasons, one should take Table 1.2 as rule of thumb. The parenthesized ρ values in Table 1.2 only serve as a reference for the underlying smoothness class.

1.4 Nearly Optimal Inference in the Exact Model

The discussions in Section 1.3 are entirely based on the diagonal model (1.8). For both theoretical interest and practical relevance, it is natural to ask whether the insights on optimal HAR inference in the diagonal model continue to hold in the exact model (1.5). This section is devoted to addressing that problem. In particular, I continue restricting attention to scale invariant tests of $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ that maximize weighted average power over μ and direct power at the flat spectrum f_1 . Under the weighted average power criterion, as specified by a given κ , the goal is to seek powerful tests as functions of $Y^s = Y/\sqrt{Y'Y}$ in the problem of

$$H_0^e : Y \sim \mathcal{N}(0, \Omega_0(f)), \quad f \in \mathcal{F} \tag{1.17}$$

$$\text{against } H_{1,f_1}^e : Y \sim \mathcal{N}(0, \kappa T^{-1} \text{diag}(1, \kappa^{-1}, \dots, \kappa^{-1})),$$

where the superscript e denotes the exact model.

First of all, I note that it is in general difficult to derive the optimal test of (1.17) under Assumption 1.3.1. This is mainly due to the complicated manner by which f enters $\Omega_0(f)$. In detail, a direct calculation shows that for a given f and $j, k = 0, 1, \dots, T-1$,

$$\begin{aligned}\Omega_0(f)_{j,k} &= \int_{-\pi T}^{\pi T} f\left(\frac{\lambda}{T}\right) w_{j,k}(\lambda) d\lambda \quad \text{with} \\ w_{j,k}(\lambda) &= \left(T^{-1} \sum_{s=1}^T \varphi_j\left(\frac{s-1/2}{T}\right) e^{-\frac{is\lambda}{T}} \right) \left(T^{-1} \sum_{t=1}^T \varphi_k\left(\frac{t-1/2}{T}\right) e^{\frac{it\lambda}{T}} \right),\end{aligned}\tag{1.18}$$

where $\varphi_j(\phi) = (\sqrt{2})^{1_{[j \neq 0]}} \cos(\pi j s)$, $0 \leq \phi \leq 1$. In this case, even if it is true that the least favorable distribution of (1.17) puts a point mass on some function $f^* \in \mathcal{F}$, the determination of f^* seems very difficult. Alternatively, one may want to impose additional assumptions on $\mathcal{F}$ such that $\Omega_0(f) = T^{-1} \text{diag}(f(0), \dots, f(\pi(T-1)/T))$ holds uniformly in $f \in \mathcal{F}$. The task is also hard, since one must then solve $(T^2 + T)/2$ functional constraints, that is, $\Omega_0(f)_{j,k} = 0$ for every $j > k$ and $\Omega_0(f)_{i,i} = f(\pi i/T)$ for every i .

Despite the difficulty in analytically deriving the exact optimal test of (1.17), one still can obtain bounds on the power of any size-controlling test by using the bounding approach of Elliott, Müller, and Watson (2015). Recall from Section 1.2 that for any probability distribution Λ over $\mathcal{F}$, the likelihood ratio test of $H''_{0,\Lambda}$ against H''_{1,f_1} yields such a power bound. Suppose there exists an ad hoc test φ_{ah} that is known to control size. If the power of φ_{ah} is close to the power bound for some Λ , then φ_{ah} is known to be close to optimal, as no substantially more powerful test exists. It turns out that the insights from the diagonal model are useful in guessing a good Λ and in justifying the near optimality of the EWC test in the exact model. In particular, for a given a in $[0, 1]$, let Λ_a be a point mass distribution on the kinked function $f_a(\phi)$, as was defined in Assumption 1.3.1. It is already known that for every a , the likelihood ratio

test of H''_{0,Λ_a} against H''_{1,f_1} yields a power bound. I then numerically search for a such that the resulting power bound is minimized. Denote this a by $a^\dagger$ and the resulting Λ by $\Lambda^\dagger$. The power bound I employ to gauge potential efficiency of the EWC test is then the power of

$$\varphi_{\Lambda^\dagger, f_1} = 1 \left[(Y' \Omega_0(f_{a^\dagger}) Y)^{-1} (Y' \Omega_1(f_1) Y) > \text{cv} \right], \quad (1.19)$$

for some cv such that $E[\varphi_{\Lambda^\dagger, f_1}] = \alpha$ under H''_{0,Λ_a} . In the following subsection, I show that the EWC test essentially achieves this bound, after optimal choice of q and critical value adjustment.

To clarify ideas and illustrate points in a consistent manner, I continue using the running example introduced in Section 1.3.2. I also use the two types of smoothness classes introduced there, except that for the first type I additionally assume every $f \in \mathcal{F}$ to be non-increasing over $[0, \pi]$.

1.4.1 The optimal EWC test

I discuss the EWC test in the exact model in the following steps. First, given the aforementioned efficiency property of the EWC test, I elaborate on how to make the critical value adjustment and choose the q for the EWC test in the exact model. Second, I use numerical exercises to study the power of the resulting EWC test. Third, as was done in Section 1.3, I compare practical implications of the new EWC test with the conventional wisdom. The general takeaway remains: One should use the EWC test with a larger q and appropriately enlarged critical values for more powerful HAR inference. Lastly, as a practical matter, I examine the robustness of the rule of thumb suggested in Section 1.3.3. I find that there is no substantial change in the adjustment factor of the Student- t critical value, even if the adjustment is made in the exact model.

Critical value adjustment and choice of q

I first note that, unlike in the diagonal case, there is no analytical expression to adjust the critical value for the EWC test in the exact model. To be precise, at given $f \in \mathcal{F}$, the null rejection probability of the EWC test under given q and with the critical value cv is

$$P \left(\left| \frac{Y_0}{\sqrt{\sum_{j=1}^q Y_j^2 / q}} \right| \geq \text{cv} \right) = P \left(\frac{Z_0^2}{q^{-1} \text{cv}^2 \sum_{j=1}^q \lambda_j(f) Z_j^2} \geq 1 \right), \quad (1.20)$$

where $\{Z_j\}_{j=0}^q$ are i.i.d. standard normal. The positive eigenvalues (normalized by the absolute value of the only negative eigenvalue) of $\Omega_{0,q}(f)^{1/2} M(\text{cv}, q) \Omega_{0,q}(f)^{1/2}$ are $\lambda_j(f), j = 1, \dots, q$, where $\Omega_{0,q}(f)$ is the upper left $(q+1) \times (q+1)$ block matrix of $\Omega_0(f)$ and $M(\text{cv}, q) = \text{diag}(-1, \text{cv}^2/q^2, \text{cv}^2/q^2, \dots, \text{cv}^2/q^2)$. It is known from Section 1.3 that (1.20) as a functional of f is maximized when all $\lambda_j(f)$'s are jointly minimized. The opaque mapping from $\lambda_j(f)$ back to f , however, prevents us from identifying the null rejection probability maximizer(s) like $\underline{f}$ in the diagonal model.

A natural reaction to this obstacle is to search for the null rejection probability maximizer numerically. To render this feasible, I approximate $f \in \mathcal{F}$ as a linear combination of basis functions. The original task is then transformed into a high-dimensional optimization problem. To be more precise, let the $n+1$ node points $\{x_i\}_{i=0}^n$ define a partition of the interval $I = [0, \pi]$ into n subintervals $I_i = [x_{i-1}, x_i], i = 1, 2, \dots, n$, each of length $h_i = x_i - x_{i-1}$, and $x_0 = 0, x_n = \pi$. Let $\mathcal{C}^0(I)$ denote the space of continuous functions on I , and $\mathcal{P}_1(I_i)$ denote the space of linear functions on I_i . Let $\{\varsigma_i\}_{i=0}^n$ be a set of basis functions for the space $\mathcal{F}_h$ of continuous piecewise linear functions defined by $\mathcal{F}_h = \{f : f \in \mathcal{C}^0(I), f|_{I_i} \in \mathcal{P}_1(I_i)\}$. The basis functions $\{\varsigma_i\}_{i=0}^n$ are normalized such that $\varsigma_j(x_i) = \mathbf{1}[i = j], i, j = 0, 1, \dots, n$. By approximating f via $\hat{f} = \sum_{i=0}^n f(x_i) \varsigma_i$ and (1.12), I approximate the rejection

probability (1.20) by

$$P\left(\frac{Z_0^2}{q^{-1} \text{cv}^2 \sum_{j=1}^q \lambda_j(\hat{f}) Z_j^2} \geq 1\right) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q-1)/2} du}{\sqrt{\prod_{j=1}^q (1-u^2 + q^{-1} \text{cv}^2 \lambda_j(\hat{f}))}}, \quad (1.21)$$

which is a function of the n -dimensional vector $(f(x_1), f(x_2), \dots, f(x_n))'$. (By normalization, $f(x_0) = 1$.) With pre-computed $\{\Omega_0(\varsigma_i)\}_{i=0}^n$ based on (1.18), the computation of (1.21) takes very little computing time for each $\hat{f}$, and it is feasible to obtain a global maximizer of (1.21) subject to implied constraints on $(f(x_1), f(x_2), \dots, f(x_n))'$ from a given $\mathcal{F}$. Denote the λ_j 's at one of those global maximizers by $\{\lambda_j^*\}_{j=1}^q$. The adjusted critical value $\text{cv}_q^{a,e}$ is then readily determined by inverting

$$\frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(q-1)/2} du}{\sqrt{\prod_{j=1}^q (1-u^2 + q^{-1} (\text{cv}_q^{a,e})^2 \lambda_j^*)}} = \alpha,$$

just like solving (1.15) in the diagonal model. I provide more computational details on numerically locating the null rejection probability maximizer in Appendix A.4.

In the context of the running example, I additionally assume that the underlying spectrum is non-increasing over $[0, \pi]$. Table 1.7 lists the resulting $\text{cv}_q^{a,e}$ and cv_q^a under selected q . As can be seen, the difference between these two adjusted critical values is slight. What's more, it is observed that neither $\text{cv}_q^{a,e}$ nor cv_q^a uniformly dominates each other as a function of q . All of these suggest that even if the exact critical value adjustment of the EWC test is complex, the simple rule proposed in Section 1.3.2 is not only practically convenient, but also without loss of generality.

Now consider the choice of q in the EWC test. I note that since the alternative hypothesis of (1.17) is identical to that of (1.9), one can proceed as in Section 1.3.2 to choose the optimal q such that the resulting EWC test has the largest weighted average power. The only difference is that one must replace the adjusted critical value

Table 1.7: Diagonal model based cv_q^a , exact model based $\text{cv}_q^{a,e}$, and weighted average power (WAP) of 5% level EWC test using $\text{cv}_q^{a,e}$.

q	3	4	5	6	7	8	9	10
cv_q^a	3.322	2.966	2.817	2.756	2.739	2.747	2.772	2.806
$\text{cv}_q^{a,e}$	3.392	3.022	2.868	2.800	2.780	2.783	2.804	2.835
WAP	0.382	0.414	0.427	0.431	0.430	0.426	0.420	0.413

Notes: The “uniformly minimal” function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. All f in $\mathcal{F}$ are non-increasing over $[0, \pi]$. Sample size T is 100.

cv_q^a by $\text{cv}_q^{a,e}$ in (1.16). I refer to the EWC test under the optimal choice of q and with the adjusted critical value $\text{cv}_q^{a,e}$ as the optimal EWC test for the rest of this section.

Power of the optimal EWC test

Table 1.8 and 1.9 summarize the weighted average power of the optimal EWC test and the weighted average power bound induced by (1.19), under the two types of classes described in the beginning of Section 1.4, respectively. As can be seen, for most $\mathcal{F}$ under consideration, the optimal EWC test essentially achieves the corresponding weighted average power bound. In the spirit of Lemma 1 in Elliott, Müller, and Watson (2015), the optimal EWC test is then known to be nearly optimal. The numerical findings also imply that the insights from the diagonal model continue to be useful in the exact model, even if the analysis of the overall optimal test is hard. I note that the relatively larger difference between the weighted average power of the optimal EWC test and the corresponding bound (e.g., under large ρ in Table 1.8 and under large C in Table 1.9) is not informative about the efficiency of the optimal EWC test, since it can arise either because the bound is far from the least upper bound, or because φ_{ah} is inefficient.

More on the practical implications and the rule of thumb

The practical implication on using the EWC test from the diagonal model continue to hold. In the context of the running example, for size distortions less than 0.01 in

Table 1.8: A bound on weighted average power (WAP) and the WAP of the optimal EWC test.

ρ	0.50	0.60	0.70	0.80	0.90	0.95	0.98	0.99
WAP of optimal EWC	0.502	0.488	0.467	0.431	0.344	0.231	0.096	0.071
WAP bound	0.505	0.492	0.473	0.438	0.361	0.257	0.132	0.088

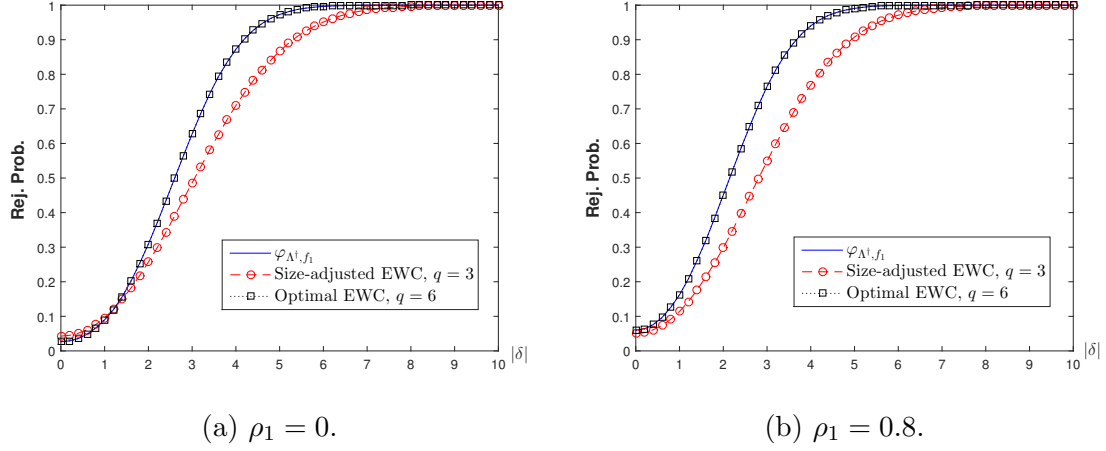
Notes: The “uniformly minimal” function of $\mathcal{F}$ corresponds to an AR(1) with coefficient ρ . All f in $\mathcal{F}$ are non-increasing over $[0, \pi]$. Nominal level is 5%. Sample size T is 100.

Table 1.9: A bound on weighted average power (WAP) and the WAP of the optimal EWC test.

C	10.0	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
WAP of optimal EWC	0.307	0.372	0.422	0.458	0.484	0.503	0.517	0.527	0.534
WAP bound	0.323	0.382	0.428	0.463	0.488	0.506	0.518	0.528	0.534

Notes: The “uniformly minimal” function of $\mathcal{F}$ is $\underline{f}(\phi) = \exp(-C\phi)$. Nominal level is 5%. Sample size $T = 100$.

Figure 1.4: Power function plot of the optimal EWC test, the size-adjusted EWC test using $q = 3$, and the weighted average power bound induced test $\varphi_{\Lambda^\dagger, f_1}$.



Notes: Under the alternative, the mean of y_t is $\delta T^{-1/2}(1 - \rho_1)^{-1}$ and y_t follows a Gaussian AR(1) with coefficient ρ_1 . Under the null, The “uniformly minimal” function of $\mathcal{F}$ corresponds to an AR(1) with coefficient 0.8. Sample size $T = 100$.

the exact model, one must use $q = 3$ in the usual EWC test. As highlighted in Table 1.7, the optimal EWC test has a larger $q = 6$. Moreover, one must enlarge the corresponding Student- t critical value by a factor of 1.15 for exact size control. In terms of weighted average power, there is a 13% gain by using the optimal EWC test. This efficiency advantage is further evident when the local alternative $\mu = \delta T^{-1/2}(1 - \rho_1)^{-1}$ is considered. Panels (a) and (b) of Figure 1.4 plot the power of $\varphi_{\Lambda^\dagger, f_1}$ as in (1.19), the optimal EWC test, and the size-adjusted EWC test using $q = 3$ for various δ under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively. To achieve the same power of 0.5, there is a 38.1% and 71.1% efficiency gain by using the optimal EWC test under $\rho_1 = 0$ and $\rho_1 = 0.8$, respectively.

In Table 1.10, I recompute the adjustment factor of the Student- t critical value for each (α, q) pair, but use the adjusted critical value $\text{cv}_q^{a,e}$ in the exact model. The calibrations are based on the type of smoothness classes $\mathcal{F}$ in which the “uniformly minimal” function is the normalized spectrum of an AR(1) with coefficient ρ and $f \in$

Table 1.10: Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test in the exact model.

q	4	6	8	9	10	11	12	13	14	15	16	18	20
$\alpha = 0.01$	1.47 (0.93)	1.34 (0.88)	1.21 (0.80)	1.16 (0.75)	1.13 (0.70)	1.11 (0.65)	1.09 (0.60)	1.09 (0.56)	1.08 (0.52)	1.08 (0.50)	1.07 (0.45)	1.06 (0.40)	1.06 (0.35)
$\alpha = 0.05$	1.28 (0.90)	1.17 (0.82)	1.10 (0.70)	1.08 (0.65)	1.07 (0.60)	1.06 (0.55)	1.05 (0.50)	1.05 (0.45)	1.04 (0.40)	1.04 (0.38)	1.04 (0.35)	1.03 (0.29)	1.03 (0.23)
$\alpha = 0.10$	1.28 (0.90)	1.12 (0.78)	1.07 (0.65)	1.06 (0.60)	1.05 (0.55)	1.05 (0.50)	1.04 (0.45)	1.04 (0.40)	1.03 (0.35)	1.03 (0.30)	1.03 (0.28)	1.02 (0.23)	1.02 (0.19)

Notes: Each q is justified as the optimal choice of level α EWC test, under some class $\mathcal{F}$ and for sample size $T = 100$. An example of the corresponding class $\mathcal{F}$ is the one in which the “uniformly minimal” function is the normalized spectrum of an AR(1) model. Numbers in parentheses are the corresponding AR(1) coefficients.

$\mathcal{F}$ is non-increasing over $[0, \pi]$, and are under sample size $T = 100$. Not surprisingly, there is no substantial change in the adjustment factors from Table 1.2 to Table 1.10. As a rule of thumb, I thus recommend that practitioners practice the EWC test in HAR inference, by following the three simple steps in Section 1.3.3 and adjusting the Student- t critical value according to Table 1.2.

1.5 Monte Carlo Simulations

The purpose of this section is twofold. First, I assess size and power performance of the suggested optimal EWC test relative to other approaches to HAR inference in the Gaussian location model. Second, I investigate the extent to which the theory derived in the Gaussian location model generalizes to inference about a scalar parameter of interest in a regression context.

I compare 18 tests in total. For the EWC test using the Student- t critical value, I consider $q = 4, 8, 12, 24$. For the optimal EWC test, labeled OEWC, I consider $q = 3, 6, 8, 10, 20$. According to Table 1.2, these q 's are the optimal choices for the EWC test, under the class $\mathcal{F}$ in which the “uniformly minimal” function is the normalized spectrum of an AR(1) with coefficient 0.95, 0.82, 0.70, 0.60, and 0.23, respectively for $T = 100$. In addition, I consider Müller’s (2014) S_q test with $q = 12, 24, 48$; Ibragimov and Müller’s (2010) test with 8 and 16 groups, IM_8 and IM_{16} ; the classical approach based on two consistent LRV estimators: Andrews’s (1991) LRV estimator $\hat{\omega}_{\text{A91}}^2$ with a quadratic spectral kernel and bandwidth selection using an AR(1) model, and Andrews and Monahan’s (1992) LRV estimator $\hat{\omega}_{\text{AM}}^2$, in which an AR(1) model is used in prewhitening; and two HAR tests based on inconsistent LRV estimators: Kiefer, Vogelsang, and Bunzel’s (2000) Bartlett kernel estimator $\hat{\omega}_{\text{KVB}}^2$ with bandwidth equal to the sample size, and Sun, Phillips, and Jin’s (2008) quadratic spectral estimator $\hat{\omega}_{\text{SPJ}}^2$ with a bandwidth that trades off asymptotic type I and type II errors

in rejection probabilities, in which the shape of the spectrum is approximated by an AR(1) model and the weight parameter is chosen to be 30.

In all simulations, the sample size is $T = 100$. The first set of simulations concerns inference about the mean of a scalar time series. In the “Gaussian AR(1)” design, the data are generated from a stationary Gaussian AR(1) model with coefficient ρ and unit innovation variance. The second set of simulations concerns inference about the coefficient on a scalar nonconstant regressor. In the “scalar nonconstant regressor” design, the regressions

$$R_t = \beta_0 + x_t\beta_1 + u_t, \quad E[u_t|x_{t-1}, x_{t-2}, \dots] = 0, \quad t = 1, \dots, T$$

contain a constant β_0 , and the nonconstant regressor x_t and the regression disturbances u_t are independent zero mean Gaussian AR(1) processes with common coefficient ρ and unit innovation variance. Under the null, the coefficient β_1 is hypothesized to be zero.

Except for the three S_q tests, I compute the test statistics based on

$$\hat{y}_t = b'\hat{\Sigma}_X^{-1}X_t\hat{u}_t, \tag{1.22}$$

where $\hat{\Sigma}_X = T^{-1} \sum_{t=1}^T X_t X_t'$ with $X_t = (1, x_t)'$, $b = (0, 1)'$, and $\hat{u}_t = R_t - \hat{\beta}_0 - x_t\hat{\beta}_1$ with $(\hat{\beta}_0, \hat{\beta}_1)'$ being the ordinary least squares (OLS) estimator for (β_0, β_1) . For the three S_q tests, I follow Section 5 in Müller (2014) to use

$$\tilde{y}_t = b'\hat{\Sigma}_X^{-1}X_t\hat{u}_t + \frac{b'\hat{\Sigma}_X^{-1}X_tX_t'\hat{\Sigma}_X^{-1}b}{b'\hat{\Sigma}_X^{-1}b}\hat{\beta}_1,$$

where b , X_t , $\hat{\Sigma}_X$, and $\hat{\beta}_1$ are the same as in (1.22).

Table 1.11 reports size and size-adjusted power of the 18 tests in the “Gaussian AR(1)” design. The size adjustment is performed on the ratio of the test statistic

and the critical value to ensure that data-dependent critical values are appropriately subsumed in the effective test. Not surprisingly, the optimal EWC test almost exactly controls size in the data generating process (DGP) that coincides with The “uniformly minimal” function of the underlying smoothness class. This can be seen in the cases of OEWC_3 under $\rho = 0.95$ and OEWC_8 under $\rho = 0.7$. Moreover, though the class of OEWC_q tests is known to essentially maximize weighted average power over μ under white noise, they also have better power performance relative to other tests when the underlying persistence is not negligible. For example, both the OEWC_3 test and the S_{24} control size under $\rho = 0.95$, but the size-adjusted power of OEWC_3 is 140% larger; the OEWC_6 test and the KVB test have roughly the same size distortions under $\rho = 0.95$, but the size-adjusted power of OEWC_6 is 36% larger.

In the “scalar nonconstant regressor” design, let $y_t = b' \Sigma_X^{-1} X_t u_t$, where Σ_X is the probability limit of $\hat{\Sigma}_X$ under suitable regularity conditions. The time series y_t is not Gaussian. On the other hand, the optimal EWC tests are based on the observable series $\hat{y}_t$ which, as argued by Müller (2014), behaves like $y_t - T^{-1} \sum_{s=1}^T y_s$ asymptotically. Despite the non-Gaussianity of the underlying time series, the optimal EWC tests, as reported in Table 1.12, continue to control size well and have better power performance relative to most alternative approaches. I note that the exceptional size and power performance of the IM_q test in Table 1.12 is specific to the design and explained by Müller.

1.6 Conclusion

This chapter considers optimal HAR inference in finite-sample contexts. The driving assumption is that the normalized spectrum of the underlying time series lies in a smoothness class, which possesses a “uniformly minimal” function. Under this assumption, I establish a finite-sample optimal theory of HAR inference in the

Table 1.11: Small sample performance for inference about population mean

ρ	EW ₄	EW ₈	EW ₁₂	EW ₂₄	OEWC ₃	OEWC ₆	OEWC ₈	OEWC ₁₀	OEWC ₂₀
Panel A: Size under Gaussian AR(1)									
0.0	5.0	5.0	5.0	5.0	1.6	3.3	3.6	4.0	4.4
0.7	5.7	6.8	8.7	15.4	1.8	4.2	5.1	6.3	12.1
0.9	9.4	17.7	25.1	40.1	2.7	9.9	14.6	19.3	34.7
0.95	17.3	32.4	42.0	57.0	4.9	20.8	28.6	35.0	52.1
0.98	36.3	54.4	62.6	73.5	13.4	42.4	50.9	57.0	70.1
0.999	83.3	89.7	91.9	94.5	68.5	85.9	88.7	90.5	93.7
Panel B: Size-adjusted power under Gaussian AR(1)									
0.0	34.0	41.9	45.0	47.9	29.0	38.6	41.9	43.9	47.3
0.7	34.6	42.9	45.4	48.8	29.5	39.5	42.9	44.8	47.9
0.9	36.6	44.5	46.9	48.9	31.4	41.7	44.5	45.8	48.5
0.95	39.5	47.1	49.2	51.0	34.3	44.9	47.1	48.5	50.8
0.98	49.1	56.5	58.6	60.3	43.2	54.3	56.5	57.8	60.0
0.999	99.7	99.9	99.9	100.0	99.3	99.9	99.9	99.9	100.0
ρ	S_{12}	S_{24}	S_{48}	$\hat{\omega}_{A91}^2$	$\hat{\omega}_{AM}^2$	$\hat{\omega}_{KVB}^2$	$\hat{\omega}_{SPJ}^2$	IM ₈	IM ₁₆
Panel A: Size under Gaussian AR(1)									
0.0	5.0	4.8	5.0	5.9	6.0	5.1	5.0	4.9	5.0
0.7	5.1	5.0	5.3	13.0	8.6	7.4	6.1	7.6	12.2
0.9	5.1	5.3	5.9	24.9	15.1	12.8	8.5	17.6	31.5
0.95	5.0	5.0	5.7	38.3	22.7	19.8	11.6	31.1	48.3
0.98	4.7	4.7	5.3	59.6	37.8	34.2	18.6	52.3	67.0
0.999	4.7	4.9	5.4	91.6	84.1	79.4	53.8	89.0	93.0
Panel B: Size-adjusted power under Gaussian AR(1)									
0.0	34.9	43.3	47.4	49.5	49.0	36.7	47.3	40.5	46.2
0.7	31.9	37.9	41.3	44.6	44.4	35.2	34.5	41.3	46.9
0.9	20.7	22.5	24.5	39.8	38.1	33.3	27.4	43.5	47.2
0.95	13.4	14.3	14.4	100.0	36.6	33.1	26.0	45.6	49.4
0.98	8.3	8.8	8.5	100.0	40.6	38.4	28.9	54.4	58.5
0.999	5.4	5.4	5.4	100.0	100.0	93.8	65.2	99.9	99.9

Note: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population mean differs from the hypothesized mean by $2T^{-1/2}(1 - \rho)^{-1}$. Based on 100000 simulations.

Table 1.12: Small sample performance for inference about regression coefficient

ρ	EW ₄	EW ₈	EW ₁₂	EW ₂₄	OEWC ₃	OEWC ₆	OEWC ₈	OEWC ₁₀	OEWC ₂₀
Panel A: Size under scalar nonconstant regressor									
0.0	5.4	5.5	5.4	5.5	1.7	3.5	4.0	4.3	4.8
0.7	6.3	6.9	7.2	8.0	2.0	4.5	5.1	5.8	6.9
0.9	9.0	10.9	12.3	17.1	3.0	7.2	8.6	9.9	14.4
0.95	12.4	16.0	19.2	28.0	4.2	10.9	13.2	15.5	23.9
0.98	18.3	25.0	30.7	42.8	7.0	17.6	21.8	25.5	38.1
0.999	28.9	39.9	46.9	58.6	12.2	30.3	36.4	41.4	54.5
Panel B: Size-adjusted power under scalar nonconstant regressor									
0.0	46.7	57.7	61.6	65.4	39.4	54.4	57.7	60.1	65.0
0.7	35.3	43.8	46.2	49.3	30.7	40.6	43.8	45.0	48.5
0.9	31.0	37.6	39.7	41.5	27.0	35.3	37.6	38.7	41.2
0.95	31.6	37.9	39.6	42.1	28.1	35.9	37.9	38.9	41.3
0.98	37.7	43.5	45.7	48.4	33.6	41.5	43.5	44.8	47.7
0.999	91.0	94.4	95.2	95.9	87.4	93.4	94.4	94.9	95.8
ρ	S_{12}	S_{24}	S_{48}	$\hat{\omega}_{A91}^2$	$\hat{\omega}_{AM}^2$	$\hat{\omega}_{KVB}^2$	$\hat{\omega}_{SPJ}^2$	IM ₈	IM ₁₆
Panel A: Size under scalar nonconstant regressor									
0.0	5.0	4.9	5.1	6.0	6.0	5.4	5.5	5.0	4.9
0.7	5.2	5.0	5.0	10.2	8.0	7.1	7.1	5.5	5.7
0.9	5.3	4.8	4.4	17.7	13.0	11.0	10.2	5.8	5.6
0.95	5.0	4.1	4.1	25.0	18.0	15.4	13.7	5.6	5.3
0.98	4.1	3.5	3.5	36.0	25.6	22.4	19.0	5.0	5.1
0.999	2.8	2.5	2.5	51.3	36.9	33.4	26.1	4.7	4.9
Panel B: Size-adjusted power under scalar nonconstant regressor									
0.0	46.5	58.4	64.2	68.2	68.0	50.8	66.6	53.1	54.2
0.7	35.1	44.0	48.4	49.7	48.8	37.9	44.9	44.4	52.1
0.9	29.7	35.1	37.9	100.0	37.6	32.7	33.9	52.9	73.9
0.95	27.3	31.5	32.8	100.0	35.3	32.5	31.1	70.3	91.6
0.98	25.5	29.2	29.6	100.0	38.8	36.7	33.7	93.5	99.7
0.999	32.0	37.0	37.0	100.0	89.5	87.8	80.3	100.0	100.0

Note: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population regression coefficient differs from the hypothesized mean by $2.5T^{-1/2}(1 - \rho^2)^{-1/2}$. Based on 100000 simulations.

Gaussian location model. The optimal test trades off bias and variability, and requires an adjustment of the critical value to account for the maximum bias of the implied long-run variance estimator. I find that the EWC test is close to optimal, but one must make an adjustment to the usual Student- t critical value and choose an optimal q accordingly. Both the critical value adjustment and the choice of q depend on assumptions on the spectrum. The main implication of my findings is that when the goal is powerful HAR inference, it is advantageous to allow for bias in LRV estimation and adjust the critical value to explicitly account for the maximum bias.

Chapter 2

Generalized Local-to-Unity Models

2.1 Introduction

This chapter proposes a flexible asymptotic framework for the modelling of persistent time series. Our starting point is an empirical observation: For many macroeconomic time series, such as the unemployment rate, interest rates, labor's share of national income, real exchange rates, price earnings ratios, etc., tests for an autoregressive unit root are often inconclusive, or rejections are not exceedingly significant. As such, the unit root model is a natural benchmark for empirically plausible persistence modelling. At the same time, most economic models assume that these time series are stationary. What is more, econometric techniques based on an assumption of an exact unit root can yield highly misleading inference under moderate deviations of the unit root model, as demonstrated by Elliott (1998).

These concerns have generated a large literature on econometric modelling and inference with the local-to-unity model: See, for instance, Bobkoski (1983), Chan and Wei (1987), Phillips (1987), Stock (1991), Elliott and Stock (1994), Cavanagh, Elliott, and Stock (1995), Moon and Phillips (2000), Rossi (2005), Campbell and Yogo (2006), Jansson and Moreira (2006), Mikusheva (2007, 2012), Müller (2014) and Moon and

Velasco (2014), among many others. Specifically, a stationary local-to-unity (LTU) model of the scalar time series $x_{T,t}$ is of the form

$$(1 - \rho_T L)(x_{T,t} - \mu) = u_t, \quad t = 1, \dots, T \quad (2.1)$$

where L is the lag operator, $\rho_T = 1 - c/T$ for some fixed $c > 0$ and u_t is a mean-zero $I(0)$ disturbance satisfying a functional central limit theorem $T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow W(\cdot)$ with W a Wiener process with variance equal to the long-run variance of u_t . In this model

$$T^{-1/2}(x_{T, \lfloor \cdot T \rfloor} - x_{T,1}) \Rightarrow J_1(\cdot) - J_1(0) \quad (2.2)$$

where J_1 is a stationary Ornstein-Uhlenbeck process with parameter c , the continuous time analogue of an AR(1) process. The process (2.1) is the local asymptotic alternative of an autoregressive unit root (cf. Elliott, Rothenberg, and Stock (1996), Elliott (1999)). As such, it is impossible to perfectly discriminate between a LTU process and a unit root process, even as $T \rightarrow \infty$. Correspondingly, for any finite c , the measure of $J_1(\cdot) - J_1(0)$ is mutually absolutely continuous with respect to the measure of W , the continuous time analogue of a unit root process. LTU model asymptotics thus properly reflect the empirical ambivalence of unit root tests noted above.

But the LTU model is clearly not the only persistence model with this feature, even with attention restricted to stationary models. After all, the properties of the limiting process J_1 are governed by a single parameter c , and the long-range dependence of J_1 are those of a continuous time AR(1). It is not clear why this particular form of long-range dependence should adequately model the persistence properties of macroeconomic time series.

This chapter proposes a more flexible asymptotic framework by allowing for p autoregressive roots and $p - 1$ moving-average roots local-to-unity for $p \geq 1$, so that

$$(1 - \rho_{T,1}L)(1 - \rho_{T,2}L) \cdots (1 - \rho_{T,p}L)(x_{T,t} - \mu) = (1 - \gamma_{T,1}L) \cdots (1 - \gamma_{T,(p-1)}L)u_t,$$

where $\rho_{T,j} = 1 - c_j/T$ and $\gamma_{T,j} = 1 - g_j/T$ for fixed $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ (with some conditions on these parameters as specified in Section 2 below). With $p = 1$, this “generalized local-to-unity” model GLTU(p) nests the familiar LTU model (2.1). A first result of this chapter is the convergence of the GLTU(p) model, that is, in analogy to (2.2),

$$T^{-1/2}(x_{T,[\cdot T]} - x_{T,1}) \Rightarrow J_p(\cdot) - J_p(0) \quad (2.3)$$

where J_p is a stationary continuous time Gaussian ARMA($p, p - 1$) process with parameters $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$.

The GLTU(p) model sets the difference between the number of local-to-unity autoregressive and moving-average parameters to exactly one. This ensures that the limit process J_p still resembles a Wiener process: For instance, if instead $(1 - \rho_{T,1}L)(1 - \rho_{T,2}L)(x_{T,t} - \mu) = u_t$, the large sample properties of $x_{T,t}$ would be more akin to an I(2) process, with the suitably scaled limit of $x_{T,[\cdot T]} - x_{T,1}$ converging to a limit process that is absolutely continuous with respect to the measure of an *integrated* Wiener process $\int_0^\cdot W(r)dr$. In contrast, the measure of $J_p(\cdot) - J_p(0)$ is mutually absolutely continuous with respect to the measure of W , so just as for the LTU model, the GLTU(p) model cannot be perfectly discriminated from the unit root model, even asymptotically.

While clearly more general than the standard local-to-unity model (2.1), one might still worry about the appropriateness of the GLTU(p) model for generic persistence modelling of macroeconomic time series. Our main theoretical result addresses this concern by establishing the richness of the GLTU(p) model class. Recall that the

total variation distance between two probability measures is the difference in the probability they assign to an event, maximized over all events. We show in Section 3 below that for any given stationary Gaussian limiting process G whose measure of $G(\cdot) - G(0)$ is mutually absolutely continuous with respect to the measure of W , and a mild regularity constraint on the spectral density of G , for any $\varepsilon > 0$ there exists a finite p_ε and GLTU(p_ε) model such that the measure of the induced limiting process J_{p_ε} is within ε of the measure of G in total variation norm. In other words, for small ε , the stochastic properties of J_{p_ε} and G are nearly identical. Thus, positing a GLTU model is nearly without loss of generality for the large sample modelling of persistent stationary processes that cannot be distinguished from a unit root process with certainty even in the limit.

As noted above, since the LTU model cannot be perfectly discriminated from the unit root model, the parameter c in (2.1) cannot be consistently estimated. By the same logic, the $2p - 1$ parameters $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ of the GLTU(p) model also cannot be consistently estimated. This impossibility is simply the flip-side of the arguably attractive property of GLTU asymptotics to appropriately capture the empirical ambivalence of unit root tests. It does, however, complicate the derivation of valid inference, especially for large values of p .

With that in mind, we suggest a limited-information framework for likelihood based inference with the GLTU(p) model. Note that (2.3) implies, for any fixed integer N

$$\{T^{-1/2}(x_{T, \lfloor jT/N \rfloor} - x_{T,1})\}_{j=1}^N \Rightarrow \{J_p(j/N) - J_p(0)\}_{j=1}^N. \quad (2.4)$$

Thus, with attention restricted to the N observations on the left-hand side of (2.4), large-sample inference about the GLTU parameters is equivalent to inference given N discretely sampled observations from a continuous time Gaussian ARMA($p, p - 1$) process. But this latter problem is well-studied (cf. Phillips (1959), Jones (1981), Bergstrom (1985), Jones and Ackerson (1990), for example), and we show how to

obtain a numerically accurate approximation to the likelihood by a straightforward Kalman filter.

We use this framework for two conceptually distinct empirical exercises. First, we show that inference methods derived to be valid in the LTU model can be highly misleading under an empirically plausible GLTU(2) model. In particular, we consider Campbell and Yogo's 2006 popular test for stock return predictability. By construction, this test controls size in the LTU model. But we find that in the GLTU(2) model, it exhibits severe size distortions, even if the GLTU(2) parameters are restricted to be within a two log-points neighborhood of the peak of the limited-information likelihood for the price-dividends ratio. In other words, unless one has good reasons to impose that the long-range persistence patterns of potential stock price predictors are of the AR(1) type, the Campbell and Yogo (2006) test is not a reliable test of the absence of predictability.

Second, and more constructively, we conduct limited-information Bayesian inference about the half-life of the US/UK real exchange rate deviations, using the long-span data from Lothian and Taylor (1996). We find that the GLTU model with $p \geq 3$ has much better fit than those with $p = 1$ or 2 , while at the same time leading to much larger half-life estimates. This illustrates that allowing the generality of the GLTU model can substantially alter conclusions about economic quantities of interest.

This chapter contributes to a large literature on alternative models of persistence, such as the fractional model (see Robinson (2002) for an overview), or more recently, the three parameter generalization of Müller and Watson (2016). A relatively close analogue to the GLTU model is the VAR(1) LTU model considered by Phillips (1988), Stock and Watson (1996), Stock (1996) or Phillips (1998): The marginal process for a scalar time series of a VAR(1) LTU model is in the GLTU class, since sums of

finite order AR processes are finite order ARMA processes with particular coefficient restrictions.

Our main theoretical result on the approximability of continuous time Gaussian processes is related in spirit to the approximability of the second order properties of discrete time stationary processes by the finite order ARMA class—see, for instance, Theorem 4.4.3 of Brockwell and Davis (1991) for a textbook exposition. The continuous time case is subtly different, though, since spectral densities are then functions on the entire real line (and not confined to the interval $[-\pi, \pi]$). What is more, we obtain approximability in total variation distance, and not just for a metric on second order properties. We are not aware of any closely related results in the literature.

The remainder of the chapter is organized as follows. Section 2.2 introduces the GLTU(p) model in detail and formally establishes its limiting properties. Section 2.3 studies the richness of the GLTU(p) model class and contains the main theoretical result. Section 2.4 develops a straightforward Kalman filter to evaluate the limited-information likelihood. Section 2.5 contains the two empirical illustrations, and is followed by a concluding Section 2.6. Proofs and computational details are collected in Appendix B.

2.2 The GLTU(p) Model

2.2.1 Set-up

We make the following assumptions about the building blocks of the GLTU(p) model

$$(1 - \rho_{T,1}L)(1 - \rho_{T,2}L) \cdots (1 - \rho_{T,p}L)(x_{T,t} - \mu) = (1 - \gamma_{T,1}L) \cdots (1 - \gamma_{T,(p-1)}L)u_t \quad (2.5)$$

where $\rho_{T,j} = 1 - c_j/T$ and $\gamma_{T,j} = 1 - g_j/T$.

Condition 2.2.1 (i) The innovations $\{u_t\}_{t=-\infty}^{\infty}$ are mean-zero covariance stationary with absolutely summable autocovariances and satisfy $T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow W(\cdot)$, where $W(\cdot)$ is a Wiener process of variance ω^2 .

(ii) The parameters $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ do not depend on T and have positive real parts. They can be complex valued, but if they are, then they appear in conjugate pairs, so that the polynomials $a(z) = \prod_{j=1}^p (c_j + z) = z^p + \sum_{j=1}^p a_j z^{p-j}$ and $b(z) = \prod_{j=1}^{p-1} (g_j + z) = z^{p-1} + \sum_{j=0}^{p-2} b_j z^j$ have real coefficients.

(iii) For all T , the process $\{x_{T,t}\}_{t=-\infty}^{\infty}$ is covariance stationary.

The high-level Condition 2.2.1 (i) allows for flexible weak dependence in the innovations u_t . Part (ii) ensures that a covariance stationary distribution of $x_{T,t}$ exists, and that the limiting continuous time Gaussian ARMA process J_p is stationary and minimum phase. Part (iii) ensures that $\{x_{T,t}\}_{t=1}^T$ is covariance stationary, implicitly restricting the initial condition $(x_{T,0}, \dots, x_{T,-p+1})$.

As noted in the introduction, the GLTU(p) model obviously nests the familiar LTU model in (2.1) as a special case with $p = 1$. Maybe more interestingly, the sum of two independent LTU processes $x_{T,t}^A$ and $x_{T,t}^B$ with parameters c_A and c_B and long-run variances $\omega^2/2$ and $\omega^2/2$ is recognized as a GLTU(2) model with parameters $c_1 = c_A$, $c_2 = c_B$ and $g_1 = (c_A^2 + c_B^2)/(2(c_A + c_B))$. The GLTU(p) model thus also encompasses aggregations of p independent LTU models. It is more general than that, though, since any aggregation of independent LTU models yields a monotone spectral density function for the limiting process, while the spectral density function of J_p has no such constraint.

2.2.2 Limit Theory

Following Brockwell (2001), a mean-zero stationary continuous time Gaussian ARMA($p, p-1$) process J_p with parameters $\{c_j\}_{j=1}^p$, $\{g_j\}_{j=1}^{p-1}$ and ω^2 of Condition 1 (ii), denoted CARMA($p, p-1$) process in the following, can be written as a scalar

observation

$$J_p(s) = \mathbf{b}'\mathbf{X}(s) \quad (2.6)$$

of the $p \times 1$ state process $\mathbf{X}$ with

$$\mathbf{X}(s) = e^{\mathbf{A}s}\mathbf{X}(0) + \int_0^s e^{\mathbf{A}(s-r)}\mathbf{e}dW(r) \quad (2.7)$$

where $\mathbf{X}(0) \sim \mathcal{N}(0, \Sigma)$ is independent of the scalar Wiener process W of variance ω^2 ,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{pmatrix}, \mathbf{e} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ 1 \end{pmatrix}$$

and the coefficients a_j and b_j are defined in Condition 2.2.1 (ii). The covariance matrix of $\mathbf{X}(0)$, and hence $\mathbf{X}(s)$, is given by

$$\Sigma = E[\mathbf{X}(0)\mathbf{X}(0)'] = \omega^2 \int_{-\infty}^0 e^{-\mathbf{A}r}\mathbf{e}\mathbf{e}'e^{-\mathbf{A}'r}dr, \quad (2.8)$$

the autocovariance function of J_p is $\gamma_p(r) = E[J_p(s)J_p(s+r)] = \mathbf{b}'e^{\mathbf{A}|r|}\Sigma\mathbf{b}$, and, with $i = \sqrt{-1}$, the spectral density $f_p : \mathbb{R} \mapsto \mathbb{R}$ of J_p satisfies

$$f_{J_p}(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda r} \gamma_p(r) dr = \frac{\omega^2 |b(i\lambda)|^2}{2\pi |a(i\lambda)|^2}. \quad (2.9)$$

It is not hard to see that under Condition 2.2.1, $f_{J_p}(\lambda)$ is a ratio of polynomials in λ^2 with real coefficients of order p and $p-1$, respectively. Also, Theorem III.17 of Ibragimov and Rozanov (1978) implies that the measure of J_p is mutually absolutely continuous with the measure of J_1 for any fixed $c = c_1 > 0$. But the measure of

$J_1(\cdot) - J(0)$ is mutually absolutely continuous with the measure of W , so the same holds true for $J_p(\cdot) - J_p(0)$.

The state space representation (2.6) and (2.7) is not obviously related to the usual state space representation of the discrete time ARMA process (2.5). But it turns out that one can rewrite the latter in the form

$$x_{T,t} = \mathbf{b}'\mathbf{Z}_{T,t} + \mu \quad (2.10)$$

$$\mathbf{Z}_{T,t} = (\mathbf{I} + \mathbf{A}/T)\mathbf{Z}_{T,t-1} + \mathbf{e}u_t \quad (2.11)$$

where $\mathbf{Z}_{T,t} \in \mathbb{R}^p$, mimicking (2.6) and (2.7). This is the key step in the proof of the following theorem, which establishes the large sample relationship between the GLTU(p) model and the corresponding CARMA($p, p-1$) model J_p .

Theorem 2.2.2 *Under Condition 1, the GLTU(p) model satisfies $T^{-1/2}(x_{T, \lfloor \cdot T \rfloor} - \mu) \Rightarrow J_p(\cdot)$.*

2.3 Richness of the GLTU(p) Model Class

In this section we explore the range of large sample persistence patterns that GLTU(p) models can induce. By Theorem 2.2.2, this amounts to studying the richness of the CARMA($p, p-1$) processes on the unit interval. As discussed in the introduction, we focus on processes that are stationary, and that cannot be distinguished from a unit root process with certainty, even asymptotically. Thus the question becomes: Does there exist a stationary Gaussian process G on the unit interval such that the measure of $G(\cdot) - G(0)$ is absolutely continuous with respect to the measure of W , yet G has a substantially different distribution than any CARMA($p, p-1$), for all finite p ?

The following theorem shows that the answer is no, at least under an additional technical assumption about the spectral density of G .

Theorem 2.3.1 *Let G be a mean-zero continuous time stationary Gaussian process on the unit interval satisfying*

- (i) $G(\cdot) - G(0)$ *is absolutely continuous with respect to the measure of W , and*
- (ii) G *has a spectral density $f_G : \mathbb{R} \rightarrow [0, \infty)$ satisfying $\sup_\lambda (1 + \lambda^2) f_G(\lambda) < \infty$ and $\inf_\lambda (1 + \lambda^2) f_G(\lambda) > 0$.*

Then for any $\varepsilon > 0$, there exists a CARMA($p_\varepsilon, p_\varepsilon - 1$) process J_{p_ε} such that the total variation distance between the measures of G and J_{p_ε} is smaller than ε .

Theorem 2.3.1 asserts that a very large class of stationary Gaussian processes on the unit interval can be arbitrarily well approximated by a CARMA($p, p - 1$) process, at least with sufficiently large p and suitably chosen parameters. In conjunction with Theorem 2.2.2, this implies that the GLTU(p) class is a nearly unrestricted starting point for approximating stationary forms of persistence that are not entirely distinct from the unit root model in large samples. To be precise, suppose the large sample properties of $x_{T,t}$ are characterized by the convergence $\hat{G}_T(s) = T^{-1/2}(x_{T,[sT]} - \mu) \Rightarrow G(s)$. Then there exists a GLTU(p_ε) model with large sample properties characterized by J_{p_ε} , and for any function $\psi(\hat{G}_T)$ that is sufficiently continuous for the continuous mapping theorem to hold, the large sample stochastic properties of $\psi(\hat{G}_T)$ are nearly indistinguishable from those of $\psi(J_{p_\varepsilon})$, for all such ψ .

Note that the spectral density of an Ornstein-Uhlenbeck process J_1 with mean reverting parameter $c = 1$ is given by $f_{J_1}(\lambda) = (2\pi)^{-1}\omega^2/(\lambda^2 + 1)$, so that $\lim_{\lambda \rightarrow \infty} (1 + \lambda^2)f_{J_1}(\lambda) = \omega^2/(2\pi)$. In fact, it follows from (2.9) that $\lim_{\lambda \rightarrow \infty} (1 + \lambda^2)f_{J_p}(\lambda) = \omega^2/(2\pi)$ for any CARMA($p, p - 1$) process. Yet the regularity assumption in part (ii) of Theorem 2 only requires that $(1 + \lambda^2)f_G(\lambda)$ is bounded away from zero and infinity uniformly in λ , but not that it converges as $\lambda \rightarrow \infty$ (so Theorem 2 covers cases where $\sup_\lambda (1 + \lambda^2)|f_G(\lambda) - f_{J_{p_\varepsilon}}(\lambda)|$ is large, even for small ε). It also immediately

follows from Theorem III.17 of Ibragimov and Rozanov (1978) that if $(1 + \lambda^q)f_G(\lambda)$ is bounded away from zero and infinity uniformly in λ for some $q > 1$, then for any $q \neq 2$, the measure of G is orthogonal to the measure of J_1 , and hence the first assumption in Theorem 2.3.1 is violated. We thus consider assumption (ii) a fairly mild regularity condition.

The proof of Theorem 2 is involved. We leverage classic results on the mutual absolute continuity (but not approximability) of Gaussian measures by Ibragimov and Rozanov (1978) to obtain a bound on the entropy norm between the measures of a countable set of characterizing random variables $\{\psi_j(G)\}_{j=1}^\infty$ and those of potential approximating process in terms of their spectral densities, and then apply a locally compact version of the Stone-Weierstrass theorem to uniformly approximate f_G by some $f_{J_{p_\varepsilon}}$. See the appendix for details.

2.4 A Limited-Information Likelihood Framework

In this section we suggest a framework for conducting large sample inference with the GLTU model. A natural place to start would be the likelihood of J_p . Pham-Dinh (1977) derives the likelihood but notes that it is “too complicated for practical use” (page 390). What is more, it wouldn’t be appropriate to treat $T^{-1/2}(x_{T,[T]} - \mu)$ as a realization of $J_p(\cdot)$ directly, since Theorem 2.2.2 only establishes weak convergence. To make further progress, we restrict attention to inference that is a function only of the N random variables $\{x_{T,[jT/N]}\}_{j=1}^N$, for some given finite integer N .

The following result is immediate from Theorem 2.2.2 and the continuous mapping theorem.

Corollary 2.4.1 *Under Condition 1, for any fixed integer $N \geq 1$,*

$$\{T^{-1/2}(x_{T,[jT/N]} - \mu)\}_{j=1}^N \Rightarrow \{J_p(j/N)\}_{j=1}^N. \quad (2.12)$$

An asymptotically justified limited-information likelihood of the GLTU(p) model is thus given by the likelihood of a discretely sampled CARMA($p, p-1$) process. The number N determines the resolution of the limited-information “lens” through which we view the original data $\{x_{T,t}\}_{t=1}^T$. The convergence in Theorem 2.2.2, and thus in (2.12), are approximations that show that under a wide range of weak dependence of u_t , Central Limit Theorem type effects yield large sample Gaussianity and a dependence structure that is completely dominated by the long-run dependence properties of the GLTU(p) model. In finite samples, a large N takes these approximations seriously even on a relatively fine grid, so in general, a large N reduces the robustness of the resulting inference. At the same time, a small N leads to a fairly uninformative limited-information likelihood. The choice of N thus amounts to a classic efficiency vs. robustness trade-off. In our applications, we set $N = 50$.

As noted in the introduction, there are a number of suggestions in the literature on how to obtain the likelihood of a discretely sampled CARMA($p, p-1$) process. One potential difficulty is the computation of covariance matrices involving matrix exponentials (cf. (2.8)). If the local-to-unity AR roots are distinct, then the companion matrix $\mathbf{A}$ is diagonalizable, so one can rotate the system by the matrix of eigenvectors to avoid this difficulty. But in general, this yields a complex valued system, which requires additional care. What is more, one might not want to rule out a pair of identical local-to-unity AR roots a priori.

We now develop an alternative approach for the computation of the likelihood of $\{J_p(j/N)\}_{j=1}^N$ that avoids these difficulties. To this end, consider the discrete time Gaussian ARMA($p, p-1$) process

$$(1 - \rho_{T_0,1}L) \cdots (1 - \rho_{T_0,p}L)(x_{T_0,t}^0 - \mu) = (1 - \gamma_{T_0,1}L) \cdots (1 - \gamma_{T_0,(p-1)}L)u_t^0 \quad (2.13)$$

for $t = 1, \dots, T_0$, where T_0 is large, $u_t^0 \sim iid\mathcal{N}(0, \omega^2)$ and $\rho_{T_0,j}$, $\gamma_{T_0,j}$ are defined below (2.5). As in (2.10) and (2.11), $x_{T_0,t}^0$ has the state space representation

$$x_{T_0,t}^0 = \mathbf{b}'\mathbf{Z}_{T_0,t}^0 + \mu \quad (2.14)$$

$$\mathbf{Z}_{T_0,t}^0 = \mathbf{\Phi}_{T_0}\mathbf{Z}_{T_0,t-1}^0 + \mathbf{e}u_t^0 \quad (2.15)$$

where $\mathbf{\Phi}_{T_0} = \mathbf{I}_p + \mathbf{A}/T_0$ and $\mathbf{\Omega}_{T_0}^0 = E[\mathbf{Z}_{T_0,0}^0(\mathbf{Z}_{T_0,0}^0)']$ satisfies $\text{vec } \mathbf{\Omega}_{T_0}^0 = \omega^2(\mathbf{I}_{p^2} - \mathbf{\Phi}_{T_0} \otimes \mathbf{\Phi}_{T_0})^{-1} \text{vec}(\mathbf{e}\mathbf{e}')$.

With $T = T_0$ and $u_t = u_t^0$, the model (2.13) clearly satisfies Condition 2.2.1, so by Corollary 2.4.1, $\{T_0^{-1/2}(x_{T_0,\lfloor jT_0/N \rfloor}^0 - \mu)\}_{j=1}^N \Rightarrow \{J_p(j/N)\}_{j=1}^N$ as $T_0 \rightarrow \infty$. Furthermore, since $\{x_{T_0,t}^0\}_{t=1}^{T_0}$ is a Gaussian process, this further implies convergence of the corresponding first two moments. Thus, the Gaussian likelihood of $\{T_0^{-1/2}(x_{T_0,\lfloor jT_0/N \rfloor}^0 - \mu)\}_{j=1}^N$ approximates the likelihood of $\{J_p(j/N)\}_{j=1}^N$ arbitrarily well as $T_0 \rightarrow \infty$. An accurate approximation to the asymptotically justified limited-information likelihood for the $2p + 1$ parameters μ , ω^2 , $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ of the GLTU(p) model can therefore be obtained from a straightforward application of the Kalman filter with state (2.15) and observations $x_{T_0,\lfloor jT_0/N \rfloor}^0 = x_{T,\lfloor jT/N \rfloor}$, $j = 1, \dots, N$, with all other observations of $x_{T_0,t}^0$ treated as missing. In our applications, we found that setting $T_0 = 1000$ leads to results that remain numerically stable also for larger values of T_0 .

A remaining difficulty is the restriction on the parameters $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ of Condition 1 (ii). Here we follow Jones (1981), who noted that under Condition 1 (ii), one can rewrite $a(z)$ and $b(z)$ as a product of quadratic factors (and a linear factor if p is odd), where each quadratic factor collapses a potentially conjugate pair of roots into a quadratic polynomial with positive coefficients. For instance, if $c_1 = c_1^r + c_1^i i$ and $c_2 = c_1^r - c_1^i i$ with $c_1^r > 0$ and $c_1^i \in \mathbb{R}$, then $(z + c_1)(z + c_2) = (c_1^i)^2 + (c_1^r)^2 + 2c_1^r z + z^2$, and if c_1 and c_2 are real and positive, $(z + c_1)(z + c_2) = c_1 c_2 + (c_1 + c_2)z + z^2$. Either

way, the resulting quadratic polynomial is of the form $h_1^2 + 2h_2z + z^2$, with $h_1, h_2 > 0$, and in this parameterization $c_{1,2} = h_2 \pm \sqrt{h_2^2 - h_1^2}$. The same argument applies to the MA polynomial, so in such a reparameterization, the limited-information likelihood evaluation as described above involves no complex numbers at all.

2.5 Applications

This section describes two applications of the GLTU(p) model and the limited-information framework of the last section. The first application considers frequentist tests of the null hypothesis of no predictability in a regression framework where the predictor is highly persistent. A large literature has considered this problem in a framework where the predictor is assumed to follow the LTU model: see, for instance, Elliott and Stock (1994), Cavanagh, Elliott, and Stock (1995), Campbell and Yogo (2006), Jansson and Moreira (2006) and Elliott, Müller, and Watson (2015). We consider the size properties of the popular test by Campbell and Yogo (2006) when the predictor is in fact generated by the GLTU(2) model. Using the CRSP data set for the price dividend ratio, we find that empirically plausible values of the GLTU(2) parameters induce large size distortions.

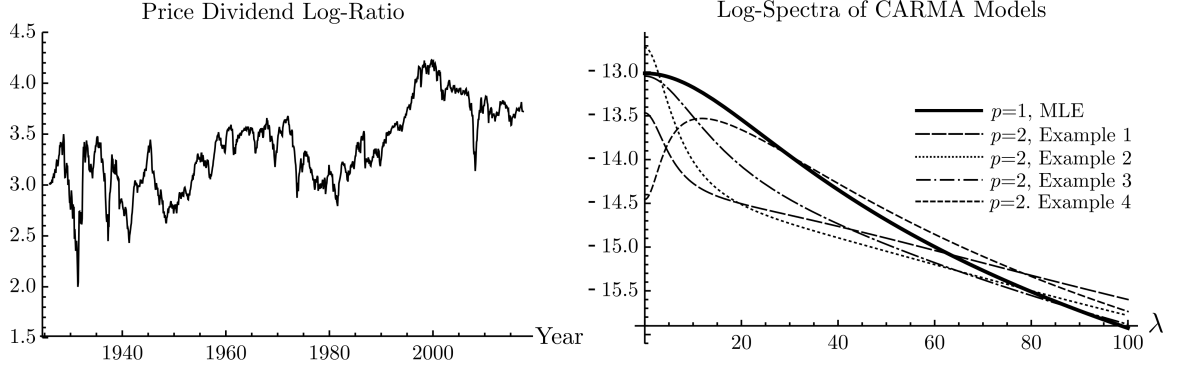
Our second application concerns the quantification of mean reversion in real exchange rates predicted by the theory of purchasing power parity, applied to the long-span data assembled by Lothian and Taylor (1996).

2.5.1 Predictive Regression with a Persistent Predictor

Let $y_{T,t}$ denote the excess stock return in period t , and let $x_{T,t-1}$ denote a potential predictor variable observed at $t-1$. Campbell and Yogo (2006) consider the regression

$$y_{T,t} = \mu_y + \beta x_{T,t-1} + e_t, \quad (2.16)$$

Figure 2.1: CRSP Price Dividend Ratio and Empirically Plausible Limiting Log-Spectra



$$(1 - \rho_T L)(x_{T,t} - \mu) = u_t \quad (2.17)$$

where $\rho_T = 1 - c/T$ for fixed $c > 0$, and the mean-zero disturbances (e_t, u_t) are weakly dependent in the sense that $T^{-1/2} \sum_{t=1}^{\lfloor T \rfloor} (e_t, u_t)' \Rightarrow (W_e, W)' = \mathbf{W}(\cdot)'$ with $\mathbf{W}$ a bivariate Wiener process with correlation r_{eu} (see Appendix A of Campbell and Yogo (2006) for a precise statement of the conditions). By construction, Campbell and Yogo's (2006) test of the null hypothesis of no predictability $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ is asymptotically valid under this LTU assumption for the predictor, that is it rejects a true null hypothesis at most 10% of the times in repeated samples.¹

As an empirical illustration, we follow Campbell and Yogo (2006) and consider the monthly excess return on the NYSE/AMSE value-weighted monthly index for $y_{T,t}$ and the corresponding price-dividend log-ratio, averaged over the preceding 12 months, for $x_{T,t}$, constructed from the database of the Center for Research in Security Prices (CRSP). We update the Campbell and Yogo (2006) data set to 1098 monthly observations from 1926:1-2018:6. The left panel in Figure 1 plots $x_{T,t}$.

¹Technically, Campbell and Yogo (2006) assume a non-stationary LTU model with zero initial condition. We find in unreported results that in analogy to the results below, an empirically plausible GLTU(2) with zero initial condition still induces large size distortions.

Table 2.1: Four GLTU(2) Parameters and Resulting Null Rejection Probability of Campbell-Yogo (2006) Test

Example No.	1	2	3	4
Value of c_1	70.9	70.0	60.6	29.8
Value of c_2	4.4	4.4	11.1	6.0
Value of g_1	7.3	11.7	24.3	3.3
Null rejection probability	49.8%	46.6%	41.2%	40.3%

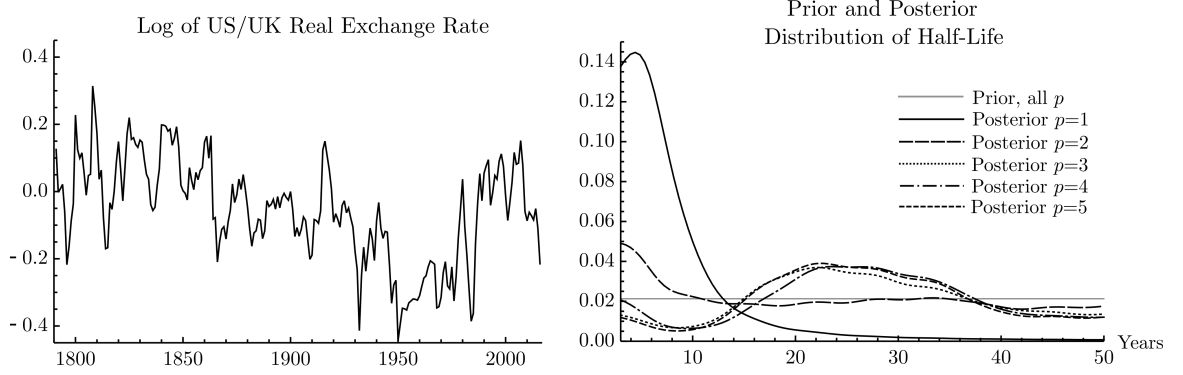
Now suppose that in contrast to Campbell and Yogo’s assumption, $x_{T,t}$ follows a GLTU(2) model

$$(1 - \rho_{T,1}L)(1 - \rho_{T,2}L)(x_{T,t} - \mu) = (1 - \gamma_{T,1}L) u_t$$

and Condition 1 holds. To obtain plausible parameters of the GLTU(2) model, we first maximize the limited-information likelihood as described in Section 2.4 in the $p = 1$ LTU model, yielding a value of c equal to 24.0. Call values of $\{c_1, c_2, g_1\}$ “empirically plausible” for the GLTU(2) model if the profiled value over μ and ω^2 of the limited-information likelihood is within two log-points of the LTU maximum likelihood. For such values, consider a data generating process with $T = 1098$, $\beta = 0$, $(e_t, u_t)'$ i.i.d. mean-zero normal and correlation equal to $r_{eu} = -0.951$, which is the value of r_{eu} estimated by Campbell and Yogo’s procedure under the LTU model assumption.

Under this GLTU(2) data generating process, we compute the rejection probability of Campbell and Yogo’s (2006) nominal 10% level two-sided test of no predictability (the test is invariant to translation shifts and scale transformations of $y_{T,t}$ and $x_{T,t}$, so the variances of e_t and u_t , as well as the means μ and μ_y are immaterial). Since $\beta = 0$ in the data generating process, numbers larger than 10% indicate size distortions. In Table 1, we report the parameter values for four fairly distinct empirically plausible GLTU(2) parameters that induce severe size distortions. The right panel in Figure 1 plots the corresponding log spectral densities of the limiting CARMA(2,1) model,

Figure 2.2: Bayesian Limited-Information Analysis of US/UK Real Exchange Rates



along with the limiting Ornstein-Uhlenbeck process with $c = 24.0$, that is at the limited-information MLE.

We conclude from this exercise that the LTU model is not well suited for generic persistence modelling: Empirically plausible deviations from the LTU model grossly invalidate inference.

2.5.2 Persistence of Deviations from Purchasing Power Parity

Lothian and Taylor (1996) assembled long-term data on the log US/UK real exchange rate from 1791 to 1990 and estimated half-life deviations of approximately 6 years based on an AR(1) specification. We consider the same data extended² through 2016, $x_{T,t}$, and plotted in the left panel of Figure 2. We are interested in quantifying for how long deviations from purchasing power parity persist assuming that the exchange rate $x_{T,t}$ follows a GLTU(p) model.

The traditional definition of the half-life is based on the impulse response of the Wold innovation to $x_{T,t}$, which in general depends not only on the GLTU(p) param-

²The extension is based on the FRED series DEXUSUK, SWPPPPI and WPSFD49207 for recent values of the exchange rate, and UK and US producer price indices.

eters $\{c_j\}$ and $\{g_j\}$, but also on the short-run dynamics of u_t . See, for instance, Andrews and Chen (1994), Murray and Papell (2002) or Rossi (2005). At the same time, as discussed in Taylor’s (2003) survey, the literature on real exchange rates emphasizes mean reversion *in the long run*, and often applies corresponding augmented Dickey-Fuller regressions, which in the context of the LTU model amount to inference about c (also see Murray and Papell (2005) and Stock (1991)).

Impulse responses are most meaningful in the context of a structural model, where innovations are given an explicit interpretation. But the structural interpretation of Wold innovations to the real exchange rates is not obvious. We therefore define the half-life in terms of the following thought experiment: Given the model parameters, suppose we learn that the value of the stationary process $x_{T,t}$ at the time $t = 0$ is one unconditional standard deviation above its mean, but we don’t observe any other values of $x_{T,t}$. What is the smallest horizon τ such that the best linear predictor of $x_{T,t}$ given $x_{T,0}$ is within $1/2$ unconditional standard deviations of its mean, for all $t \geq \tau$?

The best linear predictor of $x_{T,t}$ given $x_{T,0}$ is proportional to the correlation between $x_{T,0}$ and $x_{T,t}$. Assuming that u_t has more than two moments, Theorem 1 implies that

$$T^{-1}E[x_{T,0}x_{T,[sT]}] \rightarrow E[J_p(0)J_p(s)] = \mathbf{b}'e^{\mathbf{A}|s|}\Sigma\mathbf{b}$$

so that we obtain the large sample approximation

$$\tau \approx T \inf_r \left\{ r : \frac{\mathbf{b}'e^{\mathbf{A}|s|}\Sigma\mathbf{b}}{\mathbf{b}'\Sigma\mathbf{b}} \leq 1/2 \text{ for all } s \geq r \right\}. \quad (2.18)$$

For $p = 1$, that is in the LTU model, this definition of a half-life is equivalent to the half-life of the impulse response relative to the “long-run” shock u_t , which in large samples becomes the impulse response function of J_1 . But for $p > 2$, this equivalence breaks down, since the impulse response function of J_p is equal to $\mathbf{1}[s \geq 0]\mathbf{b}'e^{\mathbf{A}s}\mathbf{e}$ (cf.

(2.6) and (2.7)), while the autocovariance function is $\mathbf{b}'e^{\mathbf{A}|s|}\mathbf{\Sigma}\mathbf{b}$. The definition (2.18) also differs from measuring the decay of deviations by the the sum of autoregressive coefficients in the $\text{AR}(\infty)$ representation, which in a GLTU model equals to $1 - d_T/T$ with $d_T \rightarrow d = a(0)/b(0) = \prod_{j=1}^p c_j / \prod_{j=1}^{p-1} g_j$ (as implied by (2.9)).

In order to avoid evaluating the matrix exponential in (2.18), note that $\mathbf{b}'e^{\mathbf{A}|s|}\mathbf{\Sigma}\mathbf{b}$ can be arbitrarily well approximated by the autocovariance function of the discrete stationary state space system (2.14) and (2.15) as $T_0 \rightarrow \infty$, so that

$$\tau \approx T \inf_r \left\{ r : \frac{\mathbf{b}'\Phi_{T_0}^{\lfloor sT_0 \rfloor} \mathbf{\Omega}_{T_0}^0 \mathbf{b}}{\mathbf{b}'\mathbf{\Omega}_{T_0}^0 \mathbf{b}} \leq 1/2 \text{ for all } s \geq r \right\}. \quad (2.19)$$

We again find that choosing $T_0 = 1000$ generates numerically stable results.

We consider the GLTU model with $p = 1, 2, \dots, 5$, and conduct inference based on the limited-information likelihood for $N = 50$ as introduced in Section 2.4. We employ the h parameterization of $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ as discussed there, collected in the vector $\mathbf{h} = (h_1^c, \dots, h_p^c, h_1^g, \dots, h_{p-1}^g)' \in \mathbb{R}^p \times \mathbb{R}^{p-1}$, where for p odd, $c_p = 2h_p^c$, and for p even, $g_{p-1} = 2h_{p-1}^g$. We restrict each element in $\mathbf{h}$ to be in the interval $(0, 40]$, so that the real components of all c_j and g_j are smaller than 80 (a root with $c = 80$ implies very fast mean reversion, with a half-life of $(\ln 2)T/80 \approx 1.96$ years). We choose the usual improper uninformative priors for the location and scale parameters μ and ω^2 . Let $\tau(\mathbf{h})$ be the half-life in (2.18) implied by a given value of $\mathbf{h}$. Then we let the prior on $\mathbf{h}$ be proportional to $g(\tau(\mathbf{h}))$, where the function $g : \mathbb{R} \mapsto [0, \infty)$ is such that the implied prior distribution of $\tau(\mathbf{h})$ is uniform on the interval $[3, 50]$. In this way, the prior on the object of interest $\tau(\mathbf{h})$ is controlled and independent of p . (In practice, g is computed by generating many draws from a prior with $g = g_0$ for some initial guess g_0 , and g is then given by $\mathbf{1}[3 \leq \tau(\mathbf{h}) \leq 50]g_0(\tau(\mathbf{h}))$ divided by a kernel estimate of the resulting density of $\tau(\mathbf{h})$.)

Table 2.2: Bayesian Limited-Information Analysis of US/UK Real Exchange Rates

$p =$	1	2	3	4	5
Maximal log-likelihood	28.4	30.7	33.1	36.2	37.5
Post. median half-life	6.3	23.7	26.4	27.1	26.4
90% posterior interval	(3.6; 21.2)	(4.1; 47.0)	(8.1; 46.3)	(5.4; 45.9)	(9.5; 45.7)

The posterior is obtained by a standard Gibbs sampler with a random walk Metropolis-Hastings step for $\mathbf{h}$. With the Kalman filter approximation to the limited-information likelihood of Section 2.4 and the corresponding half-life approximation (2.19), evaluation is very fast and it only takes minutes to generate 100,000 draws, even for $p = 5$.

The last two rows of Table 2 provide summary statistics for the posterior half-life for $1 \leq p \leq 5$, and the right panel of Figure 2 plots the posterior densities. For $p = 1$, the posterior for the half-life is unimodal with a mode of around 4.5 years and a median of 6.3, more or less in line with the original results of Lothian and Taylor (1996). But letting $p > 1$ leads to posteriors with much more mass at substantially longer half-lives. This accords qualitatively with Murray and Papell's (2005) finding of longer half-life point estimates when allowing for many lags in the autoregression, although their half-lives are computed from impulse responses or sums of autoregressive coefficients, and are thus not directly comparable.

Remarkably, the posterior densities in Figure 2 for $p = 3, 4, 5$ are very similar to each other. It seems that once the model is flexible enough, the implications settle, with a posterior mode of the half-life at around 25 years. The maximized values of the log-likelihood is more than 4.7 log-points larger for $p \geq 3$ compared to $p = 1$, suggesting that the additional flexibility provided by the GLTU model is preferred by the data.³ Overall, these results suggest that deviations from purchasing power

³More formally, the low-frequency stationarity test LFST of Müller and Watson (2008) rejects the null hypothesis of stationarity of the LTU innovations $y_t - (1 - \hat{c}_{MLE}/T)y_{t-1}$ at the 5% level for $10 \leq q \leq 50$ ($q = 10$ corresponds to testing stationarity for frequencies below 45 year cycles, and $q = 50$ corresponds to frequencies below 9 year cycles).

parity are considerably more persistent than an analysis based on the LTU model suggests.

2.6 Conclusion

This chapter introduces the $\text{GLTU}(p)$ model as a natural generalization of the popular local-to-unity approach to modelling stationary time series persistence. The main theoretical result concerns the richness of this model class: The asymptotic properties of a very large class of persistent processes can be well approximated by some $\text{GLTU}(p)$ model. What is more, we suggest a straightforward approximation to the limited-information asymptotic likelihood of the $\text{GLTU}(p)$ model. The $\text{GLTU}(p)$ model thus seems a convenient starting point for the modelling of persistent time series in macroeconomics and finance.

Chapter 3

Measuring Persistence of Macroeconomic Time Series: Evidence from the Jordá-Schularick-Taylor Macrohistory Database

3.1 Introduction

This chapter empirically investigates a basic and important feature of business cycle fluctuations: The persistence of macroeconomic time series. By “persistent” I mean the low-frequency variability is a large, sometimes dominant, source of the overall temporal variability in a time series. This low-frequency variability or persistence has been an important and ongoing empirical issue in macroeconomics. Nelson and Plosser (1982) sparked the debate in macroeconomics by arguing that stochastic variations due to persistent real factors were an essential element of any model of macroe-

economic fluctuations. The interest in persistence of deviations from Purchasing Power Parity (PPP) arose naturally since Rogoff (1996). Moreover, some important relations among macroeconomic variables are expected to hold over long horizons, such as the so-called balanced growth relations, which refer to consumption-income and investment-income ratios. A common examination of these relations focuses on the persistence of their residuals.

Standard measures of persistence include the magnitude of the largest autoregressive roots in the individual series and in the residual from the long-run relation (see, for instance, Stock and Watson (1999)), and the so-called half-life under various definitions (see, for instance, Andrews and Chen (1994), Murray and Papell (2002, 2005) or Rossi (2005)). These two types of persistence measures essentially impose the assumption that the long-range persistence patterns of the underlying time series are of the AR(1) type. Moreover, methods for estimation of and inference about these persistence measures rely on the so-called local-to-unity (LTU) asymptotic framework (cf. Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987)). Partly, this is because of the following two observations: First, for many macroeconomic time series tests for an autoregressive unit root are often inconclusive, or rejections are not exceedingly significant; Second, it is impossible to perfectly discriminate between a LTU process and a unit root process, even in large samples.

But it is not clear that the particular AR(1) form of long-range dependence should adequately model the persistence properties of macroeconomic time series. In fact, the LTU model is not the only persistence model with the feature as mentioned above, even with attention restricted to stationary models. On the other hand, Dou and Müller (2019) recently propose a more flexible asymptotic framework called the generalized local-to-unity model, GLTU(p), for the modelling of persistent time series. They show that the GLTU(p) class is a nearly unrestricted starting point for approximating stationary forms of persistence that are not entirely distinct from the

unit root model in large samples. They also suggest a limited-information framework for likelihood based inference with the GLTU(p) model.

The purpose of this chapter is to examine and document the persistence of macroeconomic time series from a compiled long-run cross-country dataset, the Jordà-Schularick-Taylor Macrohistory Database (cf. Jordà, Schularick, and Taylor (2017)), in the GLTU(p) model. To that end, I apply Dou and Müller’s limited-information likelihood framework to conduct Bayesian inference about appropriately defined persistence measures, and compare them with those standard measures as mentioned above. As a byproduct, this chapter also provides empirical researchers a user’s guide to the econometric framework of Dou and Müller (2019), and examines the numerical stability of their framework with real data.

It is found that allowing for the generality in modelling long-range dependence can substantially alter quantitative statements about the persistence of macroeconomic time series. As a guide for practitioners, I recommend using a measure of half-life, as proposed by Dou and Müller (2019), in conjunction with the GLTU(p) model to gauge the persistence of macroeconomic time series. The suggested procedure is numerically stable throughout exercises conducted in this chapter.

The remainder of the chapter is organized as follows: Section 3.2 introduces the data-generating process (DGP) considered and the measures of persistence used. Section 3.3 describes the method to construct point estimates and credible intervals for the persistence measures. Section 3.4 explains the data used and necessary transformations taken. Section 3.5 discusses the empirical results, and Section 3.6 concludes.

3.2 Models and Measures of Persistence

Let the DGP for the observed time series, y_t , be

$$y_t = d_t + u_t, \quad t = 1, \dots, T \tag{3.1}$$

where d_t is a deterministic component that can be modelled as a constant $d_t = \mu$ or as a constant plus linear trend $d_t = \mu_0 + \mu_1 t$. This chapter focuses on measuring the persistence of the stochastic component u_t , which is modelled as a triangular array process in the GLTU(p) class

$$(1 - \rho_{T,1}L)(1 - \rho_{T,2}L) \cdots (1 - \rho_{T,p}L)u_{T,t} = (1 - \gamma_{T,1}L) \cdots (1 - \gamma_{T,(p-1)}L)\epsilon_t, \quad (3.2)$$

where $\rho_{T,j} = 1 - c_j/T$ and $\gamma_{T,j} = 1 - g_j/T$. I follow Dou and Müller (2019) to make the following assumptions about the building blocks of $u_{T,t}$.

Condition 3.2.1 (i) *The innovations $\{\epsilon_t\}_{t=-\infty}^{\infty}$ are mean-zero covariance stationary with absolutely summable autocovariances and satisfy $T^{-1/2} \sum_{t=1}^{\lfloor T \rfloor} \epsilon_t \Rightarrow W(\cdot)$, where $W(\cdot)$ is a Wiener process of variance ω^2 .*

(ii) *The parameters $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ do not depend on T and have positive real parts. They can be complex valued, but if they are, then they appear in conjugate pairs, so that the polynomials $a(z) = \prod_{j=1}^p (c_j + z) = z^p + \sum_{j=1}^p a_j z^{p-j}$ and $b(z) = \prod_{j=1}^{p-1} (g_j + z) = z^{p-1} + \sum_{j=0}^{p-2} b_j z^j$ have real coefficients.*

(iii) *For all T , the process $\{u_{T,t}\}_{t=-\infty}^{\infty}$ is covariance stationary.*

It is shown by Dou and Müller (2019) that under Condition 3.2.1, the GLTU(p) model satisfies

$$T^{-1/2}u_{T, \lfloor T \rfloor} \Rightarrow J_p(\cdot), \quad (3.3)$$

where J_p is a mean-zero stationary continuous time Gaussian ARMA($p, p-1$) process with parameters $\{c_j\}_{j=1}^p$, $\{g_j\}_{j=1}^{p-1}$ and ω^2 of Condition 3.2.1 (ii) (denoted by CARMA($p, p-1$) process in the following). Following Brockwell (2001), the process J_p can be written as a scalar *observation*

$$J_p(s) = \mathbf{b}'\mathbf{X}(s) \quad (3.4)$$

of the $p \times 1$ *state* process $\mathbf{X}$ with

$$\mathbf{X}(s) = e^{\mathbf{A}s} \mathbf{X}(0) + \int_0^s e^{\mathbf{A}(s-r)} \mathbf{e} dW(r) \quad (3.5)$$

where $\mathbf{X}(0) \sim \mathcal{N}(0, \Sigma)$ is independent of the scalar Wiener process W of variance ω^2 ,

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \cdots & -a_1 \end{pmatrix}, \mathbf{e} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-2} \\ 1 \end{pmatrix}$$

and the coefficients a_j and b_j are defined in Condition 3.2.1 (ii). The covariance matrix of $\mathbf{X}(0)$, and hence $\mathbf{X}(s)$, is given by

$$\Sigma = E[\mathbf{X}(0)\mathbf{X}(0)'] = \omega^2 \int_{-\infty}^0 e^{-\mathbf{A}r} \mathbf{e} \mathbf{e}' e^{-\mathbf{A}'r} dr, \quad (3.6)$$

the autocovariance function of J_p is $\gamma_p(r) = E[J_p(s)J_p(s+r)] = \mathbf{b}' e^{\mathbf{A}|r|} \Sigma \mathbf{b}$.

Now I discuss two measures of persistence as proposed by Dou and Müller (2019). The first measure is the sum of autoregressive coefficients in the AR(∞) representation, which in the GLTU(p) model (3.2) equals to

$$\alpha_T = 1 - d_T/T \approx 1 - T^{-1}a(0)/b(0) = 1 - T^{-1} \prod_{j=1}^p c_j / \prod_{j=1}^{p-1} g_j \quad (3.7)$$

for large T . This can be seen directly from the augmented Dickey-Fuller representation of (3.2), assuming uncorrelated innovations $\{\epsilon_t\}_{t=-\infty}^{\infty}$. Note that for $p = 1$, this measure reduces to the largest autoregressive root in the LTU model, so it is a generalization of that standard measure by accounting for more flexible patterns of low-frequency variability. It is also noted that short-run dynamics are not taken into

account in this persistence measure (and are whitened asymptotically, as implied by (3.7)), so α_T captures only long-run persistence in large samples.

The second measure used in this chapter is the half-life defined in terms of the following thought experiment: Given the model parameters, suppose we learn that the value of the stationary process $u_{T,t}$ at the time $t = 0$ is one unconditional standard deviation above its mean, but we don't observe any other values of $u_{T,t}$. What is the smallest horizon τ such that the best linear predictor of $u_{T,t}$ given $u_{T,0}$ is within 1/2 unconditional standard deviations of its mean, for all $t \geq \tau$? Specifically, assuming that ϵ_t has more than two moments, (3.3) implies that

$$T^{-1}E[u_{T,0}u_{T,[sT]}] \rightarrow E[J_p(0)J_p(s)] = \mathbf{b}'e^{\mathbf{A}|s|}\mathbf{\Sigma}\mathbf{b}.$$

Because the best linear predictor of $u_{T,t}$ given $u_{T,0}$ is proportional to the correlation between $u_{T,t}$ and $u_{T,0}$, the large sample approximation to τ is

$$\tau_T \approx T \inf_r \left\{ r : \frac{\mathbf{b}'e^{\mathbf{A}|s|}\mathbf{\Sigma}\mathbf{b}}{\mathbf{b}'\mathbf{\Sigma}\mathbf{b}} \leq 1/2 \text{ for all } s \geq r \right\}. \quad (3.8)$$

Note that this definition differs from the traditional ones that are based on the impulse response of the Wold innovation to $u_{T,t}$. As mentioned by Dou and Müller (2019), it is often not obvious what the structural interpretation of Wold innovations to $u_{T,t}$ would be in applications, for example, when y_t is the log of real exchange rate and is modelled as a GLTU(p) process. As a result, impulse responses as well as their induced measures of the half-life in those contexts are not meaningful. The definition (3.8) of the half-life, however, does not rely on any structural models. It is also noted that, for $p = 1$, the definition (3.8) is equivalent to the (asymptotic) half-life of the impulse response by ignoring short-run dynamics in ϵ_t , $(\ln 2)T/c_1$, where c_1 is the LTU parameter. But these two definitions differ for $p > 2$: The impulse

response function of J_p is equal to $\mathbf{1}[s \geq 0]\mathbf{b}'e^{\mathbf{A}s}\mathbf{e}$ (cf. (3.4) and (3.5)), while the autocovariance function is $\mathbf{b}'e^{\mathbf{A}|s|}\boldsymbol{\Sigma}\mathbf{b}$.

In Section 3.5, I report persistence measures using the above two definitions for $p = 1, 2, 3, 4, 5$, which nest the two standard measures defined in the LTU model. I then compare them and illustrate empirical relevance of the GLTU(p) class beyond the LTU model.

3.3 Econometric Method

After introducing the model and discussing two measures of persistence, I address the issue of how to conduct inference. A natural place to start would be the methods that are known to produce valid confidence intervals for standard measures of persistence in the LTU model. Given that all of the aforementioned standard measures are monotone functions of the parameter c_1 in the LTU model, those methods amount to constructing valid confidence intervals for c_1 which cannot be consistently estimated (as implied by the fact that the LTU model cannot be perfectly discriminated from the unit root model).¹

The extensions of those methods are, however, not obvious in the GLTU(p) model, especially for constructing valid confidence intervals about the two persistence measures introduced in Section 3.2. After all, for $p > 1$, the parameters $\{c_j\}_{j=1}^p, \{g_j\}_{j=1}^{p-1}$ are allowed to take complex values as in Condition 3.2.1 (ii), so no natural ordering exists between two different sets of parameters. Moreover, the two persistence measures are not necessarily monotone functions of $\{c_j\}_{j=1}^p, \{g_j\}_{j=1}^{p-1}$ for $p > 1$. It is thus not clear what the end points of the confidence sets for $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$, should they be achievable, imply for confidence intervals of α_T and τ_T .

To make progress, I borrow from Dou and Müller (2019) the limited-information likelihood framework and conduct Bayesian inference about the two persistence mea-

¹See, for instance, Mikusheva (2007) for an overview of uniform inference in the LTU model.

sures α_T and τ_T . First of all, by the continuous mapping theorem, an immediate implication of (3.2) and (3.3) is that for any fixed integer $N \geq 1$,

$$\{T^{-1/2}(y_{T, \lfloor jT/N \rfloor} - d_{\lfloor jT/N \rfloor})\}_{j=1}^N \Rightarrow \{J_p(j/N)\}_{j=1}^N. \quad (3.9)$$

It follows that the likelihood of the discretely sampled CARMA($p, p-1$) process J_p provides an asymptotically justified limited-information likelihood of the GLTU(p) model. The choice of N specifies the extent of “limited-information” and reflects a classic efficiency vs. robustness trade-off: A small N utilizes less observations but renders more flexibility in modelling weak dependence structure of $\epsilon_{T,t}$, as long as Condition 3.2.1 holds and asymptotics kick in; A large N takes (3.9) more seriously in finite samples but reduces the robustness of the resulting inference. For the empirical examinations in this chapter, I follow Dou and Müller (2019) to set N to 50.

The evaluation of the likelihood of a discretely sampled CARMA($p, p-1$) process is, however, not straightforward. Partly, this is because that for $p > 1$, the parameters $\{c_j\}_{j=1}^p, \{g_j\}_{j=1}^{p-1}$ are allowed to take complex values, so is the companion matrix $\mathbf{A}$, and more importantly the *state* process $\mathbf{X}$ in (3.5). What is more, the evaluations of the stationary covariance matrix Σ in (3.6) are not straightforward with matrix exponentials involved. Even for cases when the parameters $\{c_j\}_{j=1}^p$ are distinct, more careful treatments need to be taken, needless to mention that there is no good reason to rule out potentially identical parameters a priori.²

To avoid all these difficulties, Dou and Müller (2019) suggests to leverage on the limit result (3.9), together with a novel state space representation of the original

²See Jones (1981) for cases where $\{c_j\}_{j=1}^p$ are distinct and a rotation strategy can be taken to avoid the complication. Also, see Dini and Mandic (2012) for a general discussion of a robust scheme called *widely linear complex Kalman filter* for signal extraction in systems with complex valued states.

GLTU(p) formulation

$$y_{T,t} = \mathbf{b}'\mathbf{Z}_{T,t} + d_t \quad (3.10)$$

$$\mathbf{Z}_{T,t} = (\mathbf{I} + \mathbf{A}/T)\mathbf{Z}_{T,t-1} + \mathbf{e}\epsilon_t \quad (3.11)$$

where $\mathbf{Z}_{T,t} \in \mathbb{R}^p$. In particular, for fixed $\{c_j\}_{j=1}^p$, $\{g_j\}_{j=1}^{p-1}$ and thus $\rho_{T_0,j} = 1 - c_j/T_0$, $j = 1, \dots, p$ and $\gamma_{T_0,j} = 1 - g_j/T_0$, $k = 1, \dots, p-1$, consider the discrete time *Gaussian* ARMA($p, p-1$) process

$$(1 - \rho_{T_0,1}L) \cdots (1 - \rho_{T_0,p}L)(y_{T_0,t}^0 - d_t) = (1 - \gamma_{T_0,1}L) \cdots (1 - \gamma_{T_0,(p-1)}L)\epsilon_t^0 \quad (3.12)$$

for $t = 1, \dots, T_0$, where T_0 is sufficiently large and $\epsilon_t^0 \sim iid\mathcal{N}(0, \omega^2)$. Two implications are immediate. First, as Condition 3.2.1 holds for (3.12), similar to (3.9),

$$\{T_0^{-1/2}(y_{T_0, \lfloor jT_0/N \rfloor}^0 - d_{\lfloor jT_0/N \rfloor})\}_{j=1}^N \Rightarrow \{J_p(j/N)\}_{j=1}^N. \quad (3.13)$$

As a result, the likelihood of $\{T_0^{-1/2}(y_{T_0, \lfloor jT_0/N \rfloor}^0 - d_{\lfloor jT_0/N \rfloor})\}_{j=1}^N$ approximates the likelihood of $\{J_p(j/N)\}_{j=1}^N$ arbitrarily well (pointwisely) as $T_0 \rightarrow \infty$. (This is mainly due to the joint Gaussianity of $\{y_{T_0,t}^0\}_{t=1}^{T_0}$, as well as the convergence of the first two moments implied by (3.13).) Second, by (3.10) and (3.11), $y_{T_0,t}^0$ has the state space representation

$$y_{T_0,t}^0 = \mathbf{b}'\mathbf{Z}_{T_0,t}^0 + d_t \quad (3.14)$$

$$\mathbf{Z}_{T_0,t}^0 = \mathbf{\Phi}_{T_0}\mathbf{Z}_{T_0,t-1}^0 + \mathbf{e}\epsilon_t^0 \quad (3.15)$$

where $\mathbf{\Phi}_{T_0} = \mathbf{I}_p + \mathbf{A}/T_0$ and $\mathbf{\Omega}_{T_0}^0 = E[\mathbf{Z}_{T_0,0}^0(\mathbf{Z}_{T_0,0}^0)']$ satisfies $\text{vec } \mathbf{\Omega}_{T_0}^0 = \omega^2(\mathbf{I}_{p^2} - \mathbf{\Phi}_{T_0} \otimes \mathbf{\Phi}_{T_0})^{-1} \text{vec}(\mathbf{e}\mathbf{e}')$. In contrast to the state space representation (3.4) and (3.5) of J_p , the state variable $Z_{T_0,t}^0$ is real valued and the computation of the covariance matrix

$\Omega_{T_0}^0$ is straightforward.³ A direct application of the conventional Kalman filter gives rise to an asymptotically justified approximation to the limited-information likelihood for the parameters ω^2 , $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ of the GLTU(p), and those in d_t . I follow Dou and Müller (2019) to use $T_0 = 1000$ throughout this chapter.

In actual implementations of the above scheme, two extra things are noteworthy. First, to avoid dealing with parameter spaces of complex numbers, it is helpful to follow Jones (1981) to reparametrize the parameters $\{c_j\}_{j=1}^p$, $\{g_j\}_{j=1}^{p-1}$ of Condition 3.2.1 (ii) such that $a(z)$ and $b(z)$ are products of quadratic factors (and a linear factor if p is odd), where each quadratic factor collapses a potentially conjugate pair of roots into a quadratic polynomial with positive coefficients. Specifically, for each conjugate pair c_j and c_k (or a selected pair of real parameters without replacement if no conjugate pairs appear), consider h_j and h_k such that $h_j = \sqrt{c_j c_k}$ and $h_k = (c_j + c_k)/2$. In this way, $(z + c_j)(z + c_k) = z^2 + 2h_k z + h_j^2$, and Condition 3.2.1 (ii) ensures strict positivity of the real valued parameters h_j and h_k without further constraints. Second, we are applying Kalman filter to a discrete system with state (3.15) and observations $y_{T_0, \lfloor jT_0/N \rfloor}^0 = y_{T, \lfloor jT/N \rfloor}$, $j = 1, \dots, N$, and with all other observations of $y_{T_0, t}^0$ treated as missing. This also works even if the original data say $y_{T, \lfloor jT/N \rfloor}$ is unavailable for some j , because the Kalman filter can easily incorporate missing observations.

Now I begin describing the Bayesian inference procedure for the two persistence measures α_T and τ_T , based on the above limited-information likelihood framework. First of all, note that a useful byproduct of the above framework is an asymptotically

³In fact, to obtain more precise computations, consider the eigenvalue decomposition of $\Phi_{T_0} = \mathbf{Q}_{T_0} \Lambda_{T_0} \mathbf{Q}_{T_0}^{-1}$, then $(\mathbf{I}_{p^2} - \Phi_{T_0} \otimes \Phi_{T_0})^{-1} \text{vec}(\mathbf{e}\mathbf{e}') = (\mathbf{Q}_{T_0} \otimes \mathbf{Q}_{T_0}) (\mathbf{I}_p \otimes \mathbf{I}_p - \Lambda_{T_0} \otimes \Lambda_{T_0})^{-1} (\mathbf{Q}_{T_0}^{-1} \otimes \mathbf{Q}_{T_0}^{-1}) \text{vec}(\mathbf{e}\mathbf{e}') = (\mathbf{Q}_{T_0} \otimes \mathbf{Q}_{T_0}) (\mathbf{I}_p \otimes \mathbf{I}_p - \Lambda_{T_0} \otimes \Lambda_{T_0})^{-1} \text{vec}(\mathbf{Q}_{T_0}^{-1} \mathbf{e}\mathbf{e}' \mathbf{Q}_{T_0}^{-1})$. I am grateful to Ulrich Müller for teaching me this trick.

justified and computationally easy approximation to the persistence measure

$$\tau_T \approx T \inf_r \left\{ r : \frac{\mathbf{b}' \Phi_{T'_0}^{\lfloor sT'_0 \rfloor} \Omega_{T'_0}^0 \mathbf{b}}{\mathbf{b}' \Omega_{T'_0}^0 \mathbf{b}} \leq 1/2 \text{ for all } s \geq r \right\}, \quad (3.16)$$

where T'_0 is sufficiently large and can be chosen differently from T_0 as in the limited-information likelihood approximation. Dou and Müller (2019) claims that $T'_0 = 1000$ generates numerically stable results so I will stick to that in this chapter.

Second, for a formal (limited-information) Bayesian inference, one has to specify prior distributions for the parameters in d_t , ω^2 , and the collection of reparametrized $\{c_j\}_{j=1}^p$ and $\{g_j\}_{j=1}^{p-1}$ as the vector $\mathbf{h} = (h_1^c, \dots, h_p^c, h_1^g, \dots, h_{p-1}^g)' \in \mathbb{R}^p \times \mathbb{R}^{p-1}$, where for p odd, $c_p = 2h_p^c$, and for p even, $g_{p-1} = 2h_{p-1}^g$. Standard improper prior is used for ω^2 such that the resulting conditional posterior distribution is inverse-Gamma. Throughout empirical exercises in this chapter, d_t is assumed to be a constant μ with a standard improper prior imposed such that the resulting conditional posterior distribution is normal.⁴ There is nearly no reference for the prior on $\mathbf{h}$ in the literature so I follow Dou and Müller (2019) to impose an implicit prior (independent of p) on $\mathbf{h}$. Specifically, the prior on $\mathbf{h}$ is assumed to be proportional to $g(\tau_T(\mathbf{h}))$, where the function $g : \mathbb{R} \mapsto [0, \infty)$ is such that the implied prior distribution of $\tau_T(\mathbf{h})$ is uniform on the interval $[3, 50]$. (In practice, g is computed by generating many draws from a prior with $g = g_0$ for some initial guess g_0 , and g is then given by $\mathbf{1}[3 \leq \tau_T(\mathbf{h}) \leq 50]g_0(\tau_T(\mathbf{h}))$ divided by a kernel estimate of the resulting density of $\tau_T(\mathbf{h})$.)

Lastly, in actual implementations, each component of $\mathbf{h}$ is restricted to be in the interval $(0, 40]$, where 40 corresponds to a very fast mean reversion in the LTU model with parameter 80. With the above specified prior distributions, I obtain the posterior using a Gibbs sampler: The μ and ω^2 blocks are sampled with above known

⁴In the case that $d_t = \mu_0 + \mu_1 t$, standard improper priors on $(\mu_0, \mu_1)'$ lead to a bivariate normal conditional posterior distribution. But normality breaks down when d_t is nonlinear in parameters.

distributions, respectively;⁵ The $\mathbf{h}$ block is sampled with a random walk Metropolis-Hastings algorithm by making use of (independent) mean-zero normal distributions as jump proposals for each element of $\mathbf{h}$. The jump size is chosen differently for each empirical examination. (For convergence of the resulting MCMC chain, it requires trials-and-errors to obtain a suitable jump size. In practice, I run multiple chains of 8000 draws with different jump sizes, discard their first 4000 as burn-in draws, and then choose the smallest jump size such that the induced acceptance rate in the Metropolis-Hastings step is between 0.2 and 0.5. After that, I generate 100000 draws using the selected jump size and discard the first 50000 draws.)

3.4 Data Description

The data used in this chapter are from the Release 4 of the Jordá-Schularick-Taylor Macrohistory Database in May 2019. It is a compiled long-run macro-financial dataset that covers 17 advanced economies since 1870 on an annual basis. In this chapter, I examined 9 macroeconomic time series for all the 17 economies: Real GDP, aggregate inflation, spreads between long-term and short-term interest rates, real exchange rate (local currency/USD), real housing prices, consumption-income ratio, investment-income ratio, public debt-to-GDP ratio and the total loan-to-GDP ratio (loan to non-financial private sector). Table 3.1 gives a detailed per variable and per country overview of the coverage for the 13 original data series as needed in constructing the above 9 variables.⁶ Precise descriptions of and sources of each data series per country appear in the JST Dataset R4 documentation.

⁵For given ω^2 and GLTU parameters, analytic expressions for $E[\mu|\omega^2, \mathbf{h}, \{y_{T,t}\}_{t=1}^T]$ and $\text{Var}[\mu|\omega^2, \mathbf{h}, \{y_{T,t}\}_{t=1}^T]$ can be obtained. In the case that d_t is a time trend, element by element, one can numerically identify the mean vector and covariance matrix of $(\mu_0, \mu_1)'$. Details are omitted for brevity. A even cleaner approach is to augment the state variables in (3.15) by including μ (constant case) or $(\mu_0, \mu_1)'$ (trend case). The outputs in the last step of that Kalman filter are the objects of interest. I thank Mikkel Plagborg-Møller for this observation.

⁶Table 3.1 updates part of Table 1 in Jordà, Schularick, and Taylor (2017) with an extended sample and with corrections for previous reporting errors.

Table 3.1: Available samples per variable and per country

Country	R. GDP	N.GDP	PPP GDP	Population	CPI	L.t. rate	S.t. rate	Ex. rate	H. Prices	Pub. debt	Bnk lend.	Con.	Inv.
Australia	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1944 1948-2016	1870-2016	1870-2016	1870-2016	1870-1945 1948-2016	1870-2016	1870-1946 1949-2016
Belgium	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1912 1920-2016	1870-1914 1920-2016	1870-2016	1919-2016	1870-1913 1920-1939 1946-1979 1982-2016 1982-2016	1885-1913 1920-1940 1950-2016	1913-2016	1900-1913 1920-1939 1941 1943 1946-2016
Canada	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1934-2016	1870-2016	1921-1949 1956-2016	1870-2016	1870-2016	1871-2016	1870-2016
Switzerland	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1880-2016	1870-2016	1870-2016	1901-2016	1880-2016	1870-2016	1870-2016	1870-1913 1948-2016
Germany	1870-2016 1870-1944 1946-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1921 1924-1943 1948-2016	1870-1922 1924-1944 1950-2016	1870-1944 1946-2016	1870-1922 1924-1938 1962-2016	1871-1913 1927-1943 1950-2016	1870-1920 1924-1940 1946-2016	1870-2016	1870-1913 1920-1939 1948-2016
Denmark	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1875-2016	1870-2016	1875-2016	1880-1946 1953-1956 1960-1996 1998-2016	1870-2016	1870-2016	1870-1914 1922-2016
Spain	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1936 1940-2016	1870-2016	1870-2016	1971-2016	1880-1935 1940-2016	1900-1935 1946-2016	1870-2016	1870-2016

Continued on next page

Table 1 – Continued from previous page

Country	R. GDP	N.GDP	PPP GDP	Population	CPI	L.t. rate	S.t. rate	Ex. rate	H. Prices	Pub. debt	Bnk lend.	Con.	Inv.
Finland	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1938 1948-2016	1870-2016	1870-2016	1905-2016	1914-2016	1870-2016	1870-2016	1870-2016
France	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1914 1922-2016	1870-2016	1870-2016	1880-1913 1920-1938 1949-2016	1900-1938 1946-2016	1870-2016	1870-1918 1820-1944 1946-2016
UK	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1899-1938 1946-2016	1870-2016	1880-2016	1870-2016	1870-2016
Italy	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1871 1885-1914 1922-2016	1870-2016	1970-2016	1870-2016	1870-2016	1870-2016	1870-2016
Japan	1870-2016	1875-1944 1946-2016	1870-2016	1870-2016	1870-2016	1870-2016	1879-1938 1957-2016	1870 1873-2016	1913-1930 1936-2016	1875-1944 1946-2016	1874-2016	1874-2016	1885-1944 1946-2016
Netherlands	1870-2016	1870-1913 1921-1939 1945-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1939 1946-2016	1900-2016	1870-2016	1870-1913 1921-1939 1948-2016
Norway	1870-2016	1870-1939 1946-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-1965 1967-2016	1870-2016	1870-2016	1880-1939 1947-2016	1870-2016	1870-2016	1870-1939 1946-2016
Portugal	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1880-2016	1870-2016	1988-2016	1870-2016	1870-1903 1920-2016	1910-2016	1953-2016
Sweden	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1875-2016	1870-2016	1871-2016	1870-2016	1870-2016
USA	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1870-2016	1890-2016	1870-2016	1880-2016	1870-2016	1870-2016

The 9 series were subject to a preliminary screen for outliers and then transformed. Specifically, aggregate inflation rates during the hyperinflation period of 1915 to 1924 were excluded for Germany; aggregate inflation rates during the (post) World War II period of 1945 to 1948 were excluded for Japan; aggregate inflation rates during the World War I period of 1917 to 1918 were excluded for Finland. After these, I follow Section 4 in Müller and Watson (2008) to transform the 9 time series and focus on measuring the persistence of the remaining stationary component. The transformations used depends on the series: Linearly detrended logarithms of real GDP and real housing prices are used,⁷ while logarithms of the remaining variables or indicated ratios are used.⁸ Moreover, since the econometric method elaborated in Section 3.3 can incorporate missing observations, I do not apply special treatments to the unavailable data points as summarized in Table 3.1. Also, for asymptotic arguments outlined in Section 3.3 to be sensible, I use $T = 147$, the full time span of the Jordá-Schularick-Taylor Macrohistory dataset, for all series under consideration.

3.5 Empirical Results

Table 3.2 reports the 5th, 50th and 95th percentiles of α_T for $p = 1$, which as mentioned earlier, corresponds to percentiles of the largest autoregressive roots in the LTU model. In general, the magnitudes of the reported α_T for the U.S. turn out to be large, even relative to those computed using U.S. quarterly data (cf. Stock and Watson (1999)). This can be attributed to the fact that a uniform prior distribution for τ_T over $[3, 50]$ implicitly induces a lower bound of approximately 0.77 for α_T under $T = 147$. It would also be interesting to see how the computed α_T vary as

⁷For real DGP and real housing prices, I should have executed the implementations under the assumption that $d_t = \mu_0 + \mu_1 t$, and treated both μ_0 and μ_1 as unknowns. But I was not aware of footnote 5 and overclaimed the difficulty under $d_t = \mu_0 + \mu_1 t$ in an earlier version. The results following the suggestions in footnote 5 would be updated in near future.

⁸The Hodrick-Prescott filter were used to detrend logarithms of all data series in Jordà, Schularick, and Taylor (2017), so their reported persistence (first-order autocorrelation) is not directly comparable to those in this chapter.

the resolution of “limited-information” lens (controlled by N) changes. I leave this investigation and the scrutiny of prior sensitivity for future studies.

In line with the consensus of the literature, aggregate inflation has a very high degree of persistence for almost all economies under consideration, except for Japan. However, in contrast to what was documented in earlier literature (see, for instance, Stock and Watson (1999)), the long-short interest rate spread as well displays high persistence, especially for France, Finland, Denmark, Belgium and Canada. What’s more, high persistence remains for putative cointegration error correction terms involving consumption, income, and investment, for almost all economies in the sample. This does not line up perfectly with the simple balanced growth predictions. For the U.S., the large increase in the consumption-income ratio over the 1985-2004 period were pointed out as an apparent explanation for the high persistence of the consumption–income relationship (see, for instance, Section 4.3 in Müller and Watson (2008)).

Tables 3.3 and 3.6 report the 5th, 50th and 95th percentiles of α_T , for $p = 2$ and 5 respectively. Similar tables of α_T for $p = 3, 4$ are collected in Appendix C. Comparing these tables with Table 3.2, two things are noteworthy: First, the 90% posterior intervals for α_T get wider in general as more generality of modelling is allowed for each series (p gets larger). This is not surprising given a fixed sample size $T = 147$ and the non-negligible number of missing data points for several series, as documented in Table 3.1. Second, the 5th percentiles of α_T are uninformative (downcoded by 0) for many series under consideration for $p = 5$. This reflects a numerical drawback of using α_T as a persistence measure in the GLTU(p) model for a relatively large p : For a given sample size T , α_T as in (3.7) would fall below 0 as long as the term $\prod_{j=1}^{p-1} g_j$ is an order of magnitude smaller than $\prod_{j=1}^p c_j$. But with no further restrictions on the parameter space, this cannot be avoided numerically.

Table 3.4 shows the 5th, 50th and 95th percentiles of τ_T for $p = 1$, which is equivalent to standard measures of the half-life in the LTU model without considering short-run dynamics. The persistence of (log) real exchange rate is of particular interest given a large literature on examining deviations from PPP. For $p = 1$, the reported τ_T for that variable is more or less in line with the exiting literature (see, for instance, Rossi (2005)), with few exceptions as Sweden, Japan, Finland and Belgium. But, as can be seen from Table 3.5, much more mass has been put in posteriors on longer half-lives for almost all countries in the GLTU(2) model. This accords Dou and Müller’s (2019) finding, which is based on a longer sample of the US/UK real exchange rate only.

Surprisingly, as shown in Tables 3.5, C.2, C.4 and 3.7, for most of the macroeconomic series under consideration, percentiles of τ_T remain numerically similar for $p = 2, 3, 4, 5$. But they are substantially different from when $p = 1$. This empirical observation further substantiates the motivation of this chapter and Dou and Müller (2019): AR(1) form of long-range dependence may not adequately model the persistence properties of a generic macroeconomic time series. That being said, the GLTU(p) model with a relatively large p may also lead to numerical instabilities, like the 5th percentiles of α_T for $p = 5$ as in Table 3.6. But our empirical results suggest that τ_T is reasonably stable, even for $p = 5$.

3.6 Concluding Remarks

This chapter examines the persistence of macroeconomic time series based on the Jordá-Schularick-Taylor Macrohistory Database. It is found that quantitative statements about the persistence of macroeconomic time series can be substantially altered, when more general long-range dependence modelling other than the usual

Table 3.2: The 5th, 50th and 95th percentiles of α_T for $p = 1$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	0.77	0.93	0.90	0.94	0.97	0.97	0.96	0.93	0.94	0.96	0.80	0.95	0.91	0.94	0.84	0.95	0.91
	0.83	0.97	0.96	0.98	0.98	0.98	0.98	0.97	0.97	0.98	0.90	0.98	0.97	0.98	0.93	0.98	0.96
	0.94	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.98
Int. rate spr.	0.77	0.77	0.82	0.79	0.83	0.98	0.77	0.77	0.85	0.82	0.82	0.98	0.98	0.98	0.98	0.98	0.83
	0.81	0.83	0.87	0.89	0.84	0.98	0.81	0.81	0.89	0.87	0.87	0.98	0.98	0.98	0.98	0.98	0.88
	0.92	0.93	0.89	0.98	0.88	0.98	0.90	0.91	0.92	0.89	0.89	0.98	0.98	0.98	0.98	0.98	0.90
Inflation	0.96	0.96	0.93	0.98	0.96	0.98	0.95	0.95	0.87	0.98	0.94	0.98	0.97	0.93	0.95	0.91	0.96
	0.98	0.97	0.94	0.98	0.98	0.98	0.96	0.96	0.88	0.98	0.96	0.98	0.98	0.96	0.96	0.92	0.98
	0.98	0.98	0.95	0.98	0.98	0.98	0.97	0.97	0.93	0.98	0.96	0.98	0.98	0.97	0.98	0.93	0.98
Real ex. rate	N.A.	0.82	0.89	0.98	0.83	0.84	0.84	0.84	0.94	0.98	0.82	0.82	0.98	0.86	0.83	0.98	0.84
	N.A.	0.87	0.92	0.98	0.88	0.88	0.88	0.88	0.98	0.98	0.87	0.87	0.98	0.90	0.88	0.98	0.88
	N.A.	0.89	0.93	0.98	0.90	0.90	0.90	0.90	0.98	0.98	0.89	0.89	0.98	0.92	0.89	0.98	0.90
Real housing	0.91	0.93	0.89	0.89	0.87	0.86	0.91	0.89	0.89	0.98	0.84	0.94	0.89	0.94	0.91	0.87	0.90
	0.93	0.95	0.92	0.92	0.91	0.89	0.93	0.92	0.92	0.98	0.88	0.95	0.92	0.96	0.93	0.91	0.93
	0.94	0.96	0.93	0.93	0.92	0.92	0.94	0.93	0.93	0.98	0.90	0.96	0.93	0.96	0.94	0.92	0.94
Debt-GDP	0.96	0.94	0.92	0.95	0.91	0.97	0.85	0.86	0.89	0.95	0.86	0.93	0.87	0.89	0.95	0.92	0.94
	0.98	0.97	0.94	0.98	0.93	0.98	0.89	0.90	0.92	0.98	0.90	0.95	0.91	0.92	0.98	0.94	0.97
	0.99	0.99	0.95	0.99	0.95	0.99	0.92	0.92	0.93	0.99	0.92	0.96	0.92	0.93	0.99	0.95	0.99
Real I-GDP	0.78	0.93	0.90	0.90	0.94	0.95	0.88	0.93	0.87	0.93	0.89	0.86	0.90	0.86	0.80	0.87	0.84
	0.84	0.97	0.93	0.96	0.96	0.98	0.91	0.95	0.91	0.97	0.92	0.90	0.97	0.89	0.89	0.91	0.93
	0.95	0.99	0.94	0.99	0.96	0.98	0.92	0.96	0.92	0.99	0.93	0.92	1.00	0.92	0.98	0.92	0.98
Loan-GDP	0.89	0.94	0.94	0.97	0.91	0.96	0.90	0.92	0.98	0.97	0.84	0.89	0.98	0.98	0.96	0.89	0.94
	0.92	0.96	0.95	0.98	0.93	0.98	0.93	0.94	0.98	0.98	0.88	0.92	0.98	0.98	0.98	0.92	0.95
	0.93	0.96	0.96	0.99	0.94	0.98	0.94	0.95	0.98	0.99	0.90	0.93	0.98	0.98	0.99	0.93	0.96
Real C-I	0.90	0.84	0.84	0.97	0.87	0.87	0.98	0.98	0.90	0.95	0.98	0.85	0.89	0.90	0.83	0.89	0.90
	0.93	0.88	0.88	0.98	0.91	0.91	0.98	0.98	0.93	0.98	0.98	0.89	0.91	0.93	0.93	0.92	0.93
	0.94	0.90	0.90	0.99	0.92	0.92	0.98	0.98	0.94	0.99	0.98	0.92	0.93	0.94	0.98	0.93	0.94

Note: The column header is 3-letter ISO country code.

Table 3.3: The 5th, 50th and 95th percentiles of τ_T for $p = 1$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	3.1	10.0	7.3	12.0	20.6	22.7	19.2	9.6	10.9	18.3	3.5	12.8	7.7	11.1	4.3	13.8	7.5
	4.1	25.4	18.8	28.6	37.8	39.2	37.2	23.7	26.6	36.3	6.7	29.3	20.2	28.4	9.9	30.9	18.1
	10.8	47.2	46.1	47.4	48.8	48.8	48.7	46.8	47.3	48.5	29.0	47.5	46.5	47.2	39.7	47.8	45.8
Int. rate spr.	3.0	3.1	3.9	3.4	4.1	33.3	3.0	3.0	4.7	3.9	3.9	33.3	33.3	33.3	33.3	33.3	4.2
	3.7	4.0	5.2	6.1	4.4	40.6	3.7	3.7	6.4	5.2	5.2	40.6	33.3	40.6	40.6	40.6	5.6
	8.6	10.1	6.4	27.9	5.7	40.7	7.2	8.0	8.2	6.4	6.4	40.7	33.4	40.7	40.7	40.7	6.6
Inflation	17.3	15.4	9.6	33.3	19.6	33.3	13.5	14.2	5.2	33.3	12.0	33.3	24.7	10.0	12.7	7.6	16.5
	28.3	20.3	10.9	33.3	33.3	33.3	18.7	16.6	5.7	33.3	15.5	33.3	33.3	15.8	17.4	9.0	33.3
	33.4	33.4	13.3	33.4	33.4	33.4	24.9	19.8	9.4	33.4	17.9	33.4	33.4	23.1	31.2	10.2	33.4
Real ex. rate	N.A.	3.9	6.2	33.3	4.2	4.2	4.2	4.2	12.4	33.3	3.9	3.9	33.3	4.8	4.1	33.3	4.4
	N.A.	5.2	8.2	40.6	5.7	5.7	5.7	5.7	33.3	40.6	5.2	5.2	40.6	6.6	5.6	40.6	5.8
	N.A.	6.4	9.4	40.7	6.6	6.9	6.9	6.9	33.4	40.7	6.4	6.4	40.7	8.2	6.6	40.7	7.1
Real housing	7.6	10.1	6.0	6.3	5.4	4.8	7.6	6.2	6.3	33.3	4.2	10.8	6.1	11.2	7.4	5.4	7.1
	9.5	13.8	8.2	8.4	7.4	6.6	9.5	8.2	8.3	40.6	5.7	14.4	8.2	15.8	9.4	7.5	9.3
	11.7	17.3	9.3	9.8	9.0	8.2	12.0	9.3	9.7	40.7	6.9	16.7	9.3	18.4	11.7	9.1	11.7
Debt-GDP	16.2	12.2	8.7	13.4	7.6	22.5	4.8	5.0	6.3	12.8	4.8	10.0	5.4	6.3	13.0	8.6	10.7
	34.0	27.4	11.3	31.1	9.8	38.9	6.5	6.8	8.4	29.3	6.6	13.6	7.5	8.3	30.4	11.7	26.1
	48.4	47.4	13.8	48.1	13.1	49.0	8.2	8.3	9.7	47.8	8.2	15.8	9.0	9.6	48.3	13.7	47.0
Real I-GDP	3.1	10.3	7.1	7.1	12.5	14.2	5.6	10.5	5.4	9.8	6.2	4.8	7.1	4.8	3.5	5.3	4.2
	4.4	24.8	9.4	18.3	15.7	33.3	7.5	14.8	7.4	24.3	8.3	6.6	20.8	6.6	6.5	7.5	9.6
	13.7	47.1	11.7	46.6	19.1	33.4	9.0	17.3	9.0	47.1	9.7	8.2	$+\infty$	8.2	28.4	9.0	37.9
Loan-GDP	6.0	11.1	10.8	21.3	7.4	15.6	7.1	8.6	33.3	22.5	4.4	6.2	33.3	33.3	18.4	6.0	10.8
	8.2	15.6	14.9	39.6	9.4	33.3	9.4	11.3	40.6	38.9	5.8	8.2	33.3	33.3	37.0	8.2	14.2
	9.3	18.7	17.8	48.9	11.7	33.4	11.7	13.7	40.7	48.8	7.1	9.6	33.4	33.4	48.7	9.3	16.5
Real C-I	7.1	4.2	4.2	20.3	5.4	5.4	33.3	33.3	7.1	15.3	33.3	4.8	6.0	7.1	4.1	6.1	7.1
	9.3	5.7	5.7	37.5	7.5	7.5	33.3	40.6	9.3	33.6	40.6	6.5	8.1	9.4	9.6	8.2	9.3
	11.7	6.9	6.9	48.8	9.0	9.0	33.4	40.7	11.7	48.3	40.7	8.2	9.3	11.7	39.3	9.3	11.7

Note: The column header is 3-letter ISO country code.

Table 3.4: The 5th, 50th and 95th percentiles of α_T for $p = 2$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	0.08	0.93	0.10	0.90	0.93	0.95	0.96	0.91	0.88	0.94	0.80	0.94	0.89	0.85	0.64	0.94	0.90
	0.86	0.97	0.94	0.97	0.98	0.99	0.98	0.96	0.97	0.98	0.92	0.98	0.96	0.96	0.92	0.98	0.96
	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
Int. rate spr.	0.87	0.78	0.83	0.88	0.89	0.79	0.83	0.77	0.88	0.85	0.83	0.83	0.83	0.86	0.86	0.85	0.84
	0.96	0.94	0.95	0.95	0.96	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.95	0.95
	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97
Inflation	0.93	0.94	0.93	0.94	0.94	0.95	0.94	0.92	0.93	0.96	0.89	0.94	0.95	0.93	0.93	0.94	0.93
	0.96	0.96	0.96	0.96	0.97	0.97	0.96	0.96	0.96	0.98	0.96	0.97	0.96	0.96	0.96	0.96	0.96
	0.97	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Real ex. rate	N.A.	0.71	0.93	0.78	0.81	0.87	0.82	0.85	0.96	0.81	0.72	0.80	0.80	0.88	0.82	0.82	0.86
	N.A.	0.91	0.97	0.93	0.94	0.95	0.93	0.95	0.98	0.95	0.91	0.94	0.94	0.95	0.95	0.95	0.95
	N.A.	0.97	0.98	0.97	0.97	0.97	0.98	0.97	0.99	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.98
Real housing	0.94	0.95	0.92	0.94	0.84	0.00	0.94	0.91	0.92	0.78	0.85	0.96	0.91	0.96	0.94	0.92	0.94
	0.97	0.98	0.97	0.97	0.94	0.12	0.98	0.97	0.97	0.91	0.95	0.98	0.96	0.98	0.97	0.96	0.97
	0.99	0.99	0.98	0.99	0.98	0.72	0.99	0.98	0.99	0.97	0.97	0.99	0.99	0.99	0.99	0.98	0.98
Debt-GDP	0.96	0.87	0.95	0.94	0.94	0.96	0.87	0.86	0.93	0.93	0.86	0.95	0.91	0.92	0.94	0.95	0.92
	0.98	0.97	0.98	0.98	0.98	0.99	0.95	0.94	0.97	0.97	0.95	0.98	0.96	0.96	0.98	0.98	0.97
	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99
Real I-GDP	0.83	0.93	0.94	0.93	0.96	0.96	0.91	0.95	0.92	0.93	0.92	0.90	0.92	0.87	0.81	0.92	0.86
	0.95	0.97	0.97	0.96	0.97	0.98	0.96	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.93	0.96	0.95
	0.97	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98
Loan-GDP	0.91	0.96	0.96	0.97	0.94	0.97	0.94	0.94	0.87	0.97	0.84	0.93	0.97	0.98	0.96	0.91	0.96
	0.96	0.98	0.99	0.99	0.97	0.98	0.98	0.97	0.95	0.99	0.94	0.97	0.99	0.99	0.98	0.96	0.98
	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.98	0.99	1.00	0.99	0.98	0.99
Real C-I	0.94	0.86	0.87	0.96	0.93	0.89	0.98	0.87	0.93	0.95	0.76	0.92	0.91	0.94	0.81	0.92	0.94
	0.97	0.95	0.95	0.98	0.96	0.96	0.99	0.95	0.97	0.98	0.93	0.96	0.96	0.97	0.92	0.96	0.97
	0.98	0.97	0.97	0.99	0.97	0.98	0.99	0.97	0.99	0.99	0.97	0.97	0.98	0.98	0.98	0.98	0.99

Note: The column header is 3-letter ISO country code.

Table 3.5: The 5th, 50th and 95th percentiles of τ_T for $p = 2$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	3.1	9.3	5.9	8.4	15.3	15.4	14.7	7.5	8.4	11.8	3.8	9.8	6.9	7.1	4.3	10.6	7.1
	4.9	26.5	11.8	18.1	29.3	28.2	30.9	16.5	18.2	21.9	8.4	22.9	19.1	14.8	8.1	23.5	18.4
	40.7	47.3	42.6	44.8	47.9	47.5	48.2	44.4	44.7	45.6	42.0	46.7	46.0	44.1	39.4	46.3	45.3
Int. rate spr.	3.8	3.2	3.5	4.9	6.2	3.2	3.5	3.2	5.3	3.8	3.5	3.4	4.3	3.7	3.5	3.8	4.0
	29.0	12.5	20.4	25.6	33.5	25.3	20.9	10.1	23.1	21.6	21.0	23.5	20.4	26.8	25.1	24.5	19.4
	48.7	46.2	48.1	47.9	48.8	48.2	47.9	46.6	47.8	47.8	47.5	48.1	47.5	48.7	48.5	48.4	47.0
Inflation	8.5	11.3	6.3	12.5	13.3	21.3	11.2	6.5	6.5	25.6	4.4	12.8	15.0	7.1	6.5	5.9	11.6
	34.3	35.7	33.8	38.8	36.9	40.1	37.0	33.7	33.7	43.7	33.1	36.2	37.6	34.1	33.5	35.9	36.3
	49.0	49.0	48.8	49.2	49.2	49.2	49.1	49.1	49.0	49.5	49.0	49.1	49.2	49.0	49.1	49.1	49.1
Real ex. rate	N.A.	3.2	11.2	3.4	3.7	4.9	4.0	4.4	20.3	3.4	3.1	3.4	3.4	5.6	3.7	3.5	4.9
	N.A.	7.2	31.3	11.8	13.7	22.8	12.2	18.5	38.5	18.2	6.2	14.7	16.3	18.5	17.8	19.0	18.2
	N.A.	45.1	48.1	46.7	45.3	47.0	44.5	46.7	49.2	47.9	44.1	46.6	47.2	46.5	47.0	47.2	46.0
Real housing	11.8	14.0	8.7	9.4	5.7	4.7	10.1	8.4	8.5	3.2	4.4	10.0	6.6	16.9	10.7	9.4	13.1
	27.8	30.4	24.8	25.1	11.5	30.6	25.6	22.6	20.0	7.8	18.1	19.1	14.6	34.3	25.9	29.4	32.3
	48.2	48.5	47.9	47.9	41.5	40.0	47.3	46.6	46.0	44.8	46.3	44.7	44.2	48.8	47.5	48.7	48.5
Debt-GDP	12.5	9.0	11.2	9.8	10.4	15.3	5.0	5.0	9.3	10.1	5.6	12.2	7.6	6.9	9.1	12.3	7.9
	26.9	17.8	24.5	24.1	24.1	29.7	15.7	10.6	23.8	23.4	15.4	25.9	23.1	13.1	21.3	29.7	17.8
	47.2	44.1	47.0	46.7	47.3	47.9	45.6	41.7	47.6	46.9	45.4	47.8	47.2	41.6	45.7	48.5	44.7
Real I-GDP	3.8	10.9	12.6	11.3	20.1	20.0	8.4	16.2	9.4	10.0	9.4	6.6	10.0	5.4	3.8	8.7	4.7
	20.4	29.3	30.4	32.5	38.2	37.8	26.3	34.8	29.0	28.2	25.9	24.0	31.8	19.4	9.8	30.9	16.6
	46.9	47.8	48.7	48.2	49.0	49.1	48.1	49.0	48.4	48.1	47.9	47.6	48.5	47.2	44.2	48.7	45.9
Loan-GDP	7.2	16.0	13.7	14.0	10.4	17.6	11.8	10.6	5.1	16.0	4.4	9.7	17.8	15.4	15.1	7.5	16.9
	17.8	33.2	31.6	30.9	25.1	35.3	30.6	24.4	25.0	34.3	13.1	26.6	35.6	31.9	32.5	19.6	34.8
	45.9	48.8	48.2	48.1	47.5	49.0	48.5	47.2	48.1	48.4	45.3	48.1	49.0	48.2	48.1	46.2	49.0
Real C-I	15.0	4.6	4.9	20.7	13.8	6.2	21.8	4.3	9.0	14.0	3.2	8.8	7.3	16.5	4.1	8.7	9.8
	34.0	18.4	23.4	38.4	35.6	15.3	37.2	26.5	20.1	32.8	8.8	32.8	17.9	35.6	8.2	23.1	25.4
	49.0	47.0	47.9	49.0	49.1	45.0	49.1	48.5	45.9	48.4	47.8	48.8	45.3	49.0	42.0	46.9	47.8

Note: The column header is 3-letter ISO country code.

Table 3.6: The 5th, 50th and 95th percentiles of α_T for $p = 5$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00	0.96	0.00	0.85	0.97	0.98	0.97	0.87	0.74	0.92	0.69	0.92	0.82	0.53	0.24	0.94	0.67
	0.96	0.98	0.95	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.97	0.99	0.98	0.98	0.95	0.99	0.98
Int. rate spr.	0.88	0.00	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.00	0.31	0.79	0.00	0.62	0.54	0.56	0.00
	0.94	0.92	0.93	0.93	0.96	0.93	0.91	0.93	0.94	0.91	0.93	0.92	0.86	0.93	0.92	0.93	0.91
	0.96	0.97	0.97	0.97	0.98	0.98	0.96	0.98	0.98	0.97	0.97	0.96	0.98	0.97	0.97	0.96	0.97
Inflation	0.90	0.88	0.91	0.90	0.88	0.67	0.92	0.87	0.91	0.91	0.88	0.87	0.62	0.91	0.89	0.90	0.89
	0.96	0.94	0.95	0.95	0.96	0.97	0.95	0.95	0.95	0.97	0.95	0.96	0.95	0.95	0.94	0.94	0.96
	0.98	0.98	0.99	0.97	0.98	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.97	0.99	0.97	0.98
Real ex. rate	N.A.	0.00	0.47	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.66	0.00
	N.A.	0.74	0.96	0.90	0.88	0.92	0.89	0.91	0.97	0.93	0.89	0.92	0.90	0.93	0.93	0.94	0.91
	N.A.	0.97	0.98	0.97	0.97	0.97	0.97	0.98	0.99	0.97	0.98	0.97	0.97	0.98	0.98	0.97	0.97
Real housing	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.97	0.97	0.96	0.96	0.73	0.00	0.95	0.94	0.96	0.89	0.91	0.96	0.96	0.97	0.96	0.95	0.96
	0.99	0.99	0.98	0.99	0.97	0.48	0.99	0.98	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.98	0.98
Debt-GDP	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.00
	0.97	0.70	0.96	0.97	0.96	0.97	0.95	0.87	0.95	0.93	0.90	0.96	0.96	0.96	0.97	0.97	0.92
	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99
Real I-GDP	0.00	0.00	0.57	0.00	0.00	0.93	0.00	0.77	0.56	0.00	0.00	0.00	0.66	0.00	0.00	0.89	0.00
	0.91	0.96	0.96	0.95	0.97	0.98	0.95	0.97	0.96	0.96	0.96	0.95	0.94	0.94	0.86	0.96	0.93
	0.97	0.98	0.99	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.98
Loan-GDP	0.00	0.61	0.00	0.00	0.00	0.61	0.00	0.00	0.60	0.41	0.00	0.00	0.26	0.00	0.00	0.00	0.54
	0.95	0.98	0.98	0.98	0.96	0.98	0.97	0.96	0.94	0.98	0.85	0.95	0.99	0.99	0.98	0.96	0.97
	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.98	0.99	1.00	0.99	0.98	0.99
Real C-I	0.00	0.44	0.00	0.00	0.89	0.00	0.00	0.64	0.00	0.44	0.00	0.79	0.00	0.66	0.00	0.00	0.00
	0.96	0.94	0.93	0.97	0.95	0.93	0.98	0.93	0.96	0.97	0.93	0.93	0.92	0.96	0.85	0.95	0.97
	0.98	0.97	0.97	0.99	0.97	0.99	0.99	0.97	0.99	0.99	0.98	0.96	0.98	0.98	0.98	0.98	0.99

Notes: The column header is 3-letter ISO country code. α_T is downcoded by 0.

Table 3.7: The 5th, 50th and 95th percentiles of τ_T for $p = 5$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	3.1	8.4	5.0	7.3	13.7	14.6	14.7	6.6	7.1	9.7	3.5	8.2	5.6	6.0	3.7	8.7	5.6
	3.8	24.7	7.6	12.1	24.4	26.5	27.8	10.4	11.3	15.3	6.0	15.3	11.0	8.5	6.0	16.6	10.0
	26.9	46.3	17.6	36.8	45.9	46.0	46.6	36.5	35.6	39.0	34.3	42.0	40.6	29.7	17.3	41.5	42.3
Int. rate spr.	6.6	3.2	3.5	4.4	8.5	3.2	4.0	3.2	5.3	4.1	3.7	6.9	3.8	4.7	5.0	6.3	3.8
	33.7	8.7	19.4	21.9	35.3	26.9	20.1	19.0	20.6	17.1	25.1	26.3	8.8	27.8	23.2	25.9	18.8
	48.8	45.1	46.2	47.0	48.8	47.5	46.2	46.7	46.3	44.5	47.0	47.0	44.4	47.8	46.9	47.5	44.2
Inflation	4.7	15.4	5.9	17.8	15.6	23.4	20.6	18.1	5.1	21.8	8.7	10.0	11.0	14.7	12.3	18.4	8.8
	35.3	37.8	34.7	39.0	37.8	40.9	39.4	36.9	34.1	40.9	34.7	35.9	36.0	36.3	36.0	38.4	34.0
	49.0	49.1	48.8	49.4	49.1	49.4	49.2	49.0	48.7	49.5	48.8	49.0	49.1	49.1	48.8	49.1	48.7
Real ex. rate	N.A.	3.1	12.3	3.2	3.4	4.9	3.7	3.8	29.7	3.4	3.1	3.5	3.4	5.1	3.5	3.5	4.4
	N.A.	4.9	27.5	8.8	10.7	19.7	9.0	14.8	41.7	21.3	5.1	20.6	15.7	16.2	18.4	30.1	14.8
	N.A.	37.6	47.0	44.2	42.5	46.3	40.4	44.2	49.4	46.7	43.2	45.0	42.8	44.0	45.7	47.6	42.5
Real housing	11.3	14.6	8.2	8.1	5.1	4.0	8.5	7.2	7.8	3.2	4.0	9.0	6.6	16.2	10.3	10.0	12.9
	27.2	29.3	27.2	21.0	8.4	30.9	18.4	16.3	17.5	7.6	12.3	15.3	21.5	31.3	22.8	28.4	25.7
	47.0	47.6	46.9	46.8	24.7	39.0	44.7	43.5	44.0	42.2	43.5	41.3	46.0	48.2	45.9	47.8	47.5
Debt-GDP	10.9	6.8	10.1	9.0	9.4	13.7	4.6	4.9	8.2	8.8	5.1	11.3	7.1	6.9	8.4	10.6	6.9
	22.8	10.9	21.6	18.8	19.0	24.8	13.7	8.5	16.5	16.2	10.4	19.7	21.5	12.2	17.1	26.2	12.6
	45.1	30.1	44.5	45.7	44.5	45.9	45.7	38.1	42.2	40.9	40.3	43.7	45.9	44.2	44.2	46.7	40.9
Real I-GDP	3.7	10.0	13.8	12.2	18.8	19.4	7.5	16.2	9.6	9.1	8.7	5.9	13.8	5.0	3.7	7.4	4.3
	17.9	26.6	30.0	28.8	36.9	37.0	23.7	32.2	28.1	28.4	23.2	26.8	30.1	16.8	7.5	34.0	12.6
	45.6	46.6	47.6	47.3	48.5	49.0	46.7	48.4	47.3	46.7	46.3	47.5	47.6	45.6	38.5	48.4	44.8
Loan-GDP	6.9	15.1	12.2	12.8	10.0	15.7	10.7	9.7	11.6	14.1	3.8	10.0	17.3	14.7	14.0	6.8	16.3
	13.4	30.7	29.8	27.2	21.6	34.0	26.3	19.6	30.9	33.4	8.5	26.2	34.0	29.7	26.9	20.1	30.0
	43.4	47.6	48.1	46.9	44.8	48.4	46.5	43.8	48.1	48.7	39.5	46.5	48.4	48.1	46.6	46.6	47.3
Real C-I	17.8	4.3	4.4	21.2	20.9	5.9	22.8	8.7	7.9	16.8	3.2	15.7	6.5	20.0	3.8	8.1	9.0
	34.7	25.6	18.8	36.9	38.1	11.8	34.1	27.2	16.6	33.5	9.7	35.0	13.4	37.0	6.6	19.0	31.0
	48.5	46.7	46.0	48.5	49.0	40.6	48.1	47.3	44.7	48.4	45.3	48.8	40.1	49.0	39.2	44.1	47.0

Note: The column header is 3-letter ISO country code.

AR(1) form is allowed. In practice, I recommend computing the persistence measure τ_T as in (3.16) within the econometric framework of Dou and Müller (2019).

There is still much work to do on persistence measuring. First of all, as is well known, the limiting result (3.3) would continue to hold as long as the time variation in variances of the driving disturbances $\{u_{T,t}\}_{t=-\infty}^{\infty}$ is a stationary short-memory process. But this is often not the case for macroeconomic time series, that is, there seems to be much low-frequency variability in the second moment (cf. Müller and Watson (2008)). This can be arguably relevant for the empirical exercises conducted in this chapter, because the Jordá-Schularick-Taylor macrohistory sample spans over a long enough period on an annual basis. To make progress along this direction, one can consider a version of the GLTU(p) model with low-frequency heteroskedastic driving disturbances in their natural moving average representations, as Müller and Watson (2008) did for many alternative models of low-frequency variability. It seems that the half-life measure used in this chapter and the limited-information likelihood framework can as well be adapted to that new model. I leave this for future research.

Second, it is natural to ask how to choose p for a given series in the GLTU(p) class. In a Bayesian framework, this sort of model comparison can be investigated by using Bayes factors. But the computations are not straightforward given the implicit prior I used for $\mathbf{h}$. Alternatively, a formal statistical test needs to be developed in order to discriminate between different GLTU(p) models, which is nontrivial due to presence of (at least one) nuisance parameters that cannot be consistently estimated.

Appendix A

Appendices for Chapter 1

A.1 Proofs in Section 1.3

Before proving Theorem 1.3.2, I make some additional notations. Let $\{Z_j\}_{j=0}^n$ be $n+1$ i.i.d. standard normal random variables throughout this section. For $\zeta \in \mathbb{R}_+^n$ and $n \geq 1$, let $J_n(\zeta) = P\left(Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\right)$ and define $K_n(\zeta) = P\left(Z_0^2 \leq \sum_{j=1}^n Z_j^2 \zeta_j\right)$. For a given $1 \leq \tilde{q} \leq T-1$, a given $\underline{f}$ that only satisfies Assumption 1.3.1(a), a given f_1^* (specifically referred to as $\kappa^{-1}\mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$ in Section 1.3), and $0 < \alpha < 1$, define $\text{cv}_{\tilde{q}}$ such that

$$P\left((1 - \text{cv}_{\tilde{q}})Z_0^2 > \sum_{j=1}^{\tilde{q}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2\right) = \alpha, \quad (\text{A.1})$$

and a condition $\text{cond}_{\tilde{q}}$

$$\max_{j=1,2,\dots,\tilde{q}} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} < 0 \text{ or } \min_{j=1,2,\dots,\tilde{q}} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} > 0. \quad (\text{A.2})$$

Further, for $\zeta, \xi \in \mathbb{R}_+^n$, we call ξ majorizes ζ ($\xi \succ \zeta$) if $\sum_{j=1}^n \xi_j = \sum_{j=1}^n \zeta_j$ and $\sum_{j=1}^k \xi_j \geq \sum_{j=1}^k \zeta_j$ for $1 \leq k < n$.

The following lemmas establish some useful properties of $J_n(\cdot)$, $K_n(\cdot)$, $\text{cv}_{\tilde{q}}$, and $\text{cond}_{\tilde{q}}$. It is worth noting that if power is directed at a general function $\tilde{f}_1$ under the alternative, modified versions of the following auxiliary lemmas hold for $f_1^*(\phi) = \kappa^{-1}\tilde{f}_1(\phi)\mathbf{1}[\phi \neq 0] + \mathbf{1}[\phi = 0]$. In that case, Assumption 1.3.1 must be amended accordingly: (b) must be changed to $\underline{f}(\pi j/T)/\tilde{f}_1(\pi j/T) \geq \underline{f}(\pi(j+1)/T)/\tilde{f}_1(\pi(j+1)/T)$, $j = 0, 1, \dots, T-2$, and for the existence of nontrivial tests we need $\underline{f}(\pi/T) > \kappa^{-1}\tilde{f}_1(\pi)$. I omit details for brevity and state the lemmas and the proof of Theorem 1.3.2 with f_1^* defined as in the main text.

A.1.1 Auxiliary lemmas

Lemma A.1.1 *For any $\zeta^* \in \mathbb{R}_+^n$, define $\mathcal{M}^+(\zeta^*) \equiv \{\zeta \in \mathbb{R}_+^n | \zeta_j \geq \zeta_j^* \forall 1 \leq j \leq n\}$ and $\mathcal{M}^-(\zeta^*) \equiv \{\zeta \in \mathbb{R}_+^n | \zeta_j \leq \zeta_j^* \forall 1 \leq j \leq n\}$. Then $J_n(\zeta^*) \geq J_n(\zeta)$ for any $\zeta \in \mathcal{M}^+(\zeta^*)$ and $J_n(\zeta^*) \leq J_n(\zeta)$ for any $\zeta \in \mathcal{M}^-(\zeta^*)$. Both equalities hold only if $\zeta = \zeta^*$.*

Proof. For any $\zeta \in \mathcal{M}^+(\zeta^*)$, the event $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \subset \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$. Thus $J_n(\zeta^*) \geq J_n(\zeta)$. In the case $\zeta_j > \zeta_j^*$ for some $1 \leq j \leq n$, $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \subsetneq \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$ and $J_n(\zeta^*) > J_n(\zeta)$. For any $\zeta \in \mathcal{M}^-(\zeta^*)$, the event $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \supset \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$. Thus $J_n(\zeta^*) \leq J_n(\zeta)$. In the case $\zeta_j < \zeta_j^*$ for some $1 \leq j \leq n$, $\{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j\} \supsetneq \{Z_0^2 \geq \sum_{j=1}^n Z_j^2 \zeta_j^*\}$ and $J_n(\zeta^*) < J_n(\zeta)$. ■

Corollary A.1.2 *For any $\zeta^* \in \mathbb{R}_+^n$, $\mathcal{M}^+(\zeta^*)$ and $\mathcal{M}^-(\zeta^*)$ are as defined in Lemma A.1.1. Then $K_n(\zeta^*) \leq K_n(\zeta)$ for any $\zeta \in \mathcal{M}^+(\zeta^*)$ and $K_n(\zeta^*) \geq K_n(\zeta)$ for any $\zeta \in \mathcal{M}^-(\zeta^*)$. Both equalities hold only if $\zeta = \zeta^*$.*

Proof. Simply note that $K_n(\zeta) = 1 - J_n(\zeta)$. The conclusions then follow from Lemma A.1.1. ■

Remark A.1.3

- i) A corrected version of Remark 4 in Bakirov (1996) can lead to the same set of conclusions of Lemma A.1.1 with a different relationship ($\xi \succ \zeta$ or $\zeta \succ \xi$). The exact statement in that article is not quite right; i.e., the stated relationship $J_n(\xi) \geq J_n(\zeta)$ is not necessarily true under weak majorizations. A trivial counterexample is that $\xi = (2, 0, 0, \dots, 0) \succ_w \zeta = (0, 0, 0, \dots, 0)$ but $J(\lambda) < J(\mu) = 1$. The corrected statement is $J_n(\xi) \geq J_n(\zeta)$ if $\xi \succ \zeta$ and $\xi_j > 0$, $\zeta_j > 0$ for all j . This is proved by making use of the (Schur) convexity of $J_n(\cdot)$ and invoking Caramata inequality.
- ii) The discussion in (i) and Lemma A.1.1 provide somewhat complementary sufficient conditions to explore the possible monotone properties of $J_n(\cdot)$ over $\mathbb{R}_+^n$. Unfortunately, they have not fully characterized the necessary conditions of $J_n(\xi) \geq J_n(\zeta)$. Under the conditions $(\bar{\xi} - \bar{\zeta})(1, 1, \dots, 1) + \zeta \succ \xi$ and $\sum_{j=1}^n \xi_i \geq \sum_{j=1}^n \zeta_i$ where $\bar{\xi} = \sum_{j=1}^n \xi_j/n$ and $\bar{\zeta} = \sum_{j=1}^n \zeta_j/n$, it is less obvious to compare $J(\xi)$ and $J(\zeta)$ unless ζ lies at the boundary.

Lemma A.1.4 For $\zeta \in \mathbb{R}_+^n$ and $n \geq 1$, $J_n(\zeta) = \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} du}{\sqrt{\prod_{j=1}^n (1-u^2+\zeta_j)}}$ and $J_1(\zeta_1) = \frac{2}{\pi} \arcsin \frac{1}{\sqrt{1+\zeta_1}}$.

Proof.

$$\begin{aligned}
J_n(\zeta) &= \frac{1}{\pi} \int_0^\infty \frac{dt}{\sqrt{t(1+t)} \sqrt{\prod_{j=1}^n (1+(1+t)\zeta_j)}} = \frac{1}{\pi} \int_0^1 \frac{s^{(n-1)/2} ds}{\sqrt{(1-s) \prod_{j=1}^n (s+\zeta_j)}} \\
&= \frac{2}{\pi} \int_0^1 \frac{(1-u^2)^{(n-1)/2} du}{\sqrt{\prod_{j=1}^n (1-u^2+\zeta_j)}}, \tag{A.3}
\end{aligned}$$

where the first equality follows from Lemma 2 of Bakirov and Székely (2006), the second equality follows by a change of variable $s = -1/(1+t)$, and (A.3) follows by

another change of variable $u = \sqrt{1-s}$.

$$J_1(\zeta_1) = \frac{2}{\pi} \int_0^1 \frac{du}{\sqrt{(1-u^2+\zeta_1)}} = \frac{2}{\pi} \arcsin \frac{1}{\sqrt{1+\zeta_1}},$$

which follows from a change of variable $v = u/\sqrt{1+\zeta_1}$ and the fact that the antiderivative of $(1-v^2)^{-1/2}$ is $\arcsin v$. ■

Lemma A.1.5 *For $0 < \alpha < 1$,*

(a) cv_1 exists if and only if $\underline{f}(\pi/T) \neq \kappa^{-1}$. $\underline{f}(\pi/T) \leq \kappa^{-1}$ if and only if $1 \leq cv_1$.

(b) cond_1 holds if $\underline{f}(\pi/T) \neq \kappa^{-1}$.

Proof.

(a) In (A.1) at $\tilde{q} = 1$, if $\underline{f}(\pi/T) = \kappa^{-1}$, there does not exist a cv_1 and α such that (A.1) holds. On the other hand, a rearrangement of the event in (A.1) at $\tilde{q} = 1$ gives

$$P(Z_0^2 + Z_1^2 > (Z_0^2 + \kappa \underline{f}(\pi/T) Z_1^2) cv_1) = \alpha.$$

It follows that $\underline{f}(\pi/T) \leq \kappa^{-1}$ if and only if $1 \leq cv_1$. Moreover, if $\underline{f}(\pi/T) > \kappa^{-1}$, the second part of Lemma A.1.4 in conjunction with (A.1) at $\tilde{q} = 1$ gives

$$cv_1 = \frac{1}{\kappa \underline{f}(\pi/T) \sin^2(\alpha\pi/2) + \cos^2(\alpha\pi/2)},$$

which always exists for every $0 < \alpha < 1$. In a similar vein, if $\underline{f}(\pi/T) < \kappa^{-1}$, we have

$$cv_1 = \frac{1}{\kappa \underline{f}(\pi/T) \cos^2(\alpha\pi/2) + \sin^2(\alpha\pi/2)},$$

which always exists. Thus, cv_1 exists if and only if $\underline{f}(\pi/T) \neq \kappa^{-1}$.

(b) It follows from the above that cond_1 holds if $\underline{f}(\pi/T) \neq \kappa^{-1}$.

■

Lemma A.1.6 *For $\underline{f}(\pi/T) \neq \kappa^{-1}$ and $0 < \alpha < 1$, if $\text{cond}_{\tilde{q}}$ is violated for some $1 < \tilde{q} \leq T-2$, then cond_q is also violated for any $\tilde{q}+1 \leq q \leq T-1$.*

Proof. Suppose $\text{cond}_{\tilde{q}}$ is violated while $\text{cond}_{\tilde{q}+1}$ holds. We have

$$\max_{j=1,2,\dots,\tilde{q}} \{ \text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1 \} \geq 0 \text{ and } \min_{j=1,2,\dots,\tilde{q}} \{ \text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1 \} \leq 0. \quad (\text{A.4})$$

Case 1. Consider $\max_{j=1,2,\dots,\tilde{q}+1} \{ \text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1 \} < 0$. We must have $\text{cv}_{\tilde{q}+1} > 1$; otherwise, (A.1) does not hold at $\tilde{q}+1$. On the other hand, $0 > \max_{j=1,2,\dots,\tilde{q}+1} \{ \text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1 \} \geq \max_{j=1,2,\dots,\tilde{q}} \{ \text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1 \}$. This, in conjunction with the first part of (A.4), implies that $\text{cv}_{\tilde{q}} > \text{cv}_{\tilde{q}+1} > 1$, which we now show is impossible. Suppose $\text{cv}_{\tilde{q}} > \text{cv}_{\tilde{q}+1} > 1$ is true. Denote $A_{\tilde{q}}^- = \{j | 1 \leq j \leq \tilde{q}, \text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1 < 0\}$ ($A_{\tilde{q}}^- \neq \emptyset$; otherwise, (A.1) does not hold for $\tilde{q}$). Now (A.1) at $\tilde{q}$ gives

$$\begin{aligned} \alpha &= P \left((1 - \text{cv}_{\tilde{q}}) Z_0^2 > \sum_{j=1}^{\tilde{q}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\ &= P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j=1}^{\tilde{q}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\ &= P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right. \\ &\quad \left. + \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \notin A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\ &\leq P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \end{aligned} \quad (\text{A.5})$$

$$< P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \quad (\text{A.6})$$

$$\leq P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right. \\ \left. + \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \notin A_{\tilde{q}}^-} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \quad (\text{A.7})$$

$$\leq P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j=1}^{\tilde{q}+1} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) = \alpha,$$

where (A.5) is due to the fact that $P(A < C - B) \leq P(A < C)$ when A, B, C are independent random variables and $B \geq 0$ almost surely. The inequality (A.6) is due to Corollary A.1.2 and the fact that for any $j \in A_{\tilde{q}}^-$, $\frac{1}{1 - \text{cv}_{\tilde{q}}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1]$ under $\text{cv}_{\tilde{q}} > \text{cv}_{\tilde{q}+1} > 1$. The inequality (A.7) is due to Corollary A.1.2.

Case 2. Consider $\min_{j=1,2,\dots,\tilde{q}+1} \{\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1\} > 0$. We must have $\text{cv}_{\tilde{q}+1} < 1$; otherwise, (A.1) does not hold at $\tilde{q} + 1$. On the other hand, $0 < \min_{j=1,2,\dots,\tilde{q}+1} \{\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1\} \leq \min_{j=1,2,\dots,\tilde{q}} \{\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1\}$. This, in conjunction with the second part of (A.4), implies that $\text{cv}_{\tilde{q}} < \text{cv}_{\tilde{q}+1} < 1$, which we next show is impossible. Suppose $\text{cv}_{\tilde{q}} < \text{cv}_{\tilde{q}+1} < 1$ is true. Denote $A_{\tilde{q}}^+ = \{j | 1 \leq j \leq \tilde{q}, \text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1 > 0\}$ ($A_{\tilde{q}}^+ \neq \emptyset$; otherwise, (A.1) is violated for $\tilde{q}$). Now (A.1) at $\tilde{q}$ gives

$$\alpha = P \left((1 - \text{cv}_{\tilde{q}}) Z_0^2 > \sum_{j=1}^{\tilde{q}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\ = P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j=1}^{\tilde{q}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\ = P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right)$$

$$\begin{aligned}
& + \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \notin A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \Big) \\
& \geq P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{\tilde{q}}} \sum_{j \in A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \tag{A.8}
\end{aligned}$$

$$> P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \tag{A.9}$$

$$\begin{aligned}
& \geq P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \in A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right. \\
& \quad \left. + \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j \notin A_{\tilde{q}}^+} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \tag{A.10} \\
& \geq P \left(Z_0^2 < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} \sum_{j=1}^{\tilde{q}+1} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) = \alpha,
\end{aligned}$$

where (A.8) is due to the fact that $P(A \geq C + B) \geq P(A \geq C)$ when A, B, C are independent random variables and $B \leq 0$ almost surely. The inequality (A.9) is due to Lemma A.1.1 and the fact that for any $j \in A_{\tilde{q}}^+$, $\frac{1}{1 - \text{cv}_{\tilde{q}}} [\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1] < \frac{1}{1 - \text{cv}_{\tilde{q}+1}} [\text{cv}_{\tilde{q}+1} \kappa \underline{f}(\pi j/T) - 1]$ under $\text{cv}_{\tilde{q}} < \text{cv}_{\tilde{q}+1} < 1$. The inequality (A.10) is due to Lemma A.1.1.

Collecting the results from **Case 1** and **Case 2**, we know that it is impossible to have $\text{cond}_{\tilde{q}+1}$ hold, given that $\text{cond}_{\tilde{q}}$ does not. By induction, we have for $\underline{f}(\pi/T) \neq \kappa$ and $0 < \alpha < 1$, if $\text{cond}_{\tilde{q}}$ is violated for some $1 < \tilde{q} \leq T-2$, then cond_q is also violated for any $\tilde{q} + 1 \leq q \leq T-1$. ■

Corollary A.1.7 For $\underline{f}(\pi/T) \neq \kappa^{-1}$ and $0 < \alpha < 1$, if $\text{cond}_{\tilde{q}}$ holds for some $3 \leq \tilde{q} \leq T-1$, then cond_q also holds for any $2 \leq q \leq \tilde{q}-1$.

Proof. This is the contrapositive statement of Lemma A.1.6. ■

Corollary A.1.8 For $\underline{f}(\pi/T) \neq \kappa^{-1}$ and $0 < \alpha < 1$, either one of the following will hold:

(a) there exists a unique $1 \leq q^* \leq T-2$ such that cond_q is satisfied for all $1 \leq q \leq q^*$ and violated for all $q^* + 1 \leq q \leq T-1$;

(b) cond_q is satisfied for all $1 \leq q \leq T-1$. In this case, define $q^* = T-1$.

Proof. If cond_{T-1} holds, by Corollary A.1.7 (b) is true. Otherwise if cond_{T-2} holds, then (a) is true with $q^* = T-2$. Otherwise, given that cond_1 always holds by Lemma A.1.5, backward inductions lead (a) to be true for a unique $1 \leq q^* \leq T-3$. ■

Corollary A.1.9 *If $\underline{f}$ in Assumption 1.3.1(a) equals 1 and $0 < \alpha < 1$, then cond_q holds for all $1 \leq q \leq T-1$.*

Proof. $\underline{f}(\pi j/T) = 1$, $j = 1, 2, \dots, T-1$ reduces cond_{T-1} to $\kappa \text{cv}_{T-1} - 1 < 0$ or $\kappa \text{cv}_{T-1} - 1 > 0$, which must be true for (A.1) to hold at $\tilde{q} = T-1$. It follows that cond_q holds for all $1 \leq q \leq T-1$ by Corollary A.1.7. ■

Lemma A.1.10 *For $\underline{f}(\pi/T) \neq \kappa^{-1}$, $0 < \alpha < 1$, and q^* as defined in Corollary A.1.8, either one of the following will hold:*

(a) $\text{cv}_q > 1$ for all $1 \leq q \leq q^*$, and if $q^* \geq 2$, $\text{cv}_{q+1} > \text{cv}_q$, $q = 1, 2, \dots, q^* - 1$;

(b) $\text{cv}_q < 1$ for all $1 \leq q \leq q^*$, and if $q^* \geq 2$, $\text{cv}_{q+1} < \text{cv}_q$, $q = 1, 2, \dots, q^* - 1$.

Proof. Lemma A.1.5 leads to the conclusions for $q^* = 1$. We now focus on $q^* \geq 2$.

Case 1. Suppose $\kappa \underline{f}(\pi/T) < 1$, then Lemma A.1.5 implies that $\text{cv}_1 > 1$. Suppose there exists a $\tilde{q} = \min\{q | 2 \leq q \leq q^*, \text{cv}_q < 1\}$. (We note cv_q cannot be 1 for any $q \leq q^*$; otherwise, (A.1) cannot hold at the corresponding q .) Then we must have

$$\begin{aligned} \min_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}-1} \kappa \underline{f}(\pi j/T) - 1\} &> \min_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} \\ &\geq \min_{j=1,2,\dots,\tilde{q}} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} > 0. \end{aligned}$$

There is a contradiction, because $\max_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}-1} \kappa \underline{f}(\pi j/T) - 1\} < 0$. This subsequently implies that $\text{cv}_q > 1$ for all $1 \leq q \leq q^*$. Moreover, for each $j = 1, \dots, q^*$,

$R_j(x) = [\kappa \underline{f}(\pi j/T)x - 1] / (1-x)$ is monotonically decreasing in $(1, \infty)$. For (A.1) to hold sequentially, we necessarily need $\text{cv}_{q+1} > \text{cv}_q$, $q = 1, 2, \dots, q^* - 1$. (Otherwise, the LHS of (A.1) would always exceed α by Lemma A.1.2.)

Case 2. Suppose $\kappa \underline{f}(\pi/T) > 1$, then Lemma A.1.5 implies that $\text{cv}_1 < 1$. Suppose there exists a $\tilde{q} = \min\{q | 2 \leq q \leq q^*, \text{cv}_q > 1\}$. (We note cv_q cannot be 1 for any $q \leq q^*$; otherwise, (A.1) cannot hold at the corresponding q .) Then we must have

$$\begin{aligned} \max_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}-1} \kappa \underline{f}(\pi j/T) - 1\} &< \max_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} \\ &\leq \max_{j=1,2,\dots,\tilde{q}} \{\text{cv}_{\tilde{q}} \kappa \underline{f}(\pi j/T) - 1\} < 0. \end{aligned}$$

There is a contradiction, because $\min_{j=1,2,\dots,\tilde{q}-1} \{\text{cv}_{\tilde{q}-1} \kappa \underline{f}(\pi j/T) - 1\} > 0$. This subsequently implies that $\text{cv}_q < 1$ for all $1 \leq q \leq q^*$. Moreover, for each $j = 1, \dots, q^*$, $R_j(x) = [\kappa \underline{f}(\pi j/T)x - 1] / (1-x)$ is monotonically increasing in $(0, 1)$. (To see this, note that $\kappa \underline{f}(\pi j/T) - 1 > \kappa \underline{f}(\pi j/T) \text{cv}_{q^*} - 1 > 0$ for every $1 \leq j \leq q^*$.) For (A.1) to hold sequentially, we necessarily need $\text{cv}_{q+1} < \text{cv}_q$, $q = 1, 2, \dots, q^* - 1$. (Otherwise, the LHS of (A.1) would always be below α by Lemma A.1.1.) ■

Lemma A.1.11 *For $\underline{f}(\pi/T) \neq \kappa^{-1}$, $0 < \alpha < 1$, and q^* as defined in Corollary A.1.8, if additionally $\underline{f}(\pi j/T) \geq \underline{f}(\pi(j+1)/T)$, $j = 0, 1, \dots, T-2$ and $q^* < T-1$, then $\kappa^{-1} (\underline{f}(\pi j/T))^{-1} \geq \text{cv}_{q^*}$.*

Proof. By Assumption 1.3.1(b), it suffices to show $\kappa^{-1} (\underline{f}(\pi(q^*+1)/T))^{-1} \geq \text{cv}_{q^*}$. Suppose not; then $\kappa^{-1} (\underline{f}(\pi(q^*+1)/T))^{-1} < \text{cv}_{q^*}$. Define $Q(x, q) = P\left((1-x)Z_0^2 > \sum_{j=1}^q [x \kappa \underline{f}(\pi j/T) - 1] Z_j^2\right)$.

Case 1. Suppose $\kappa \underline{f}(\pi/T) < 1$, then $\text{cv}_{q^*} > 1$ by Lemma A.1.10. Moreover, by Assumption 1.3.1(b) we must have $\kappa^{-1} (\underline{f}(\pi(q^*+1)/T))^{-1} > 1$.

$$Q(\kappa^{-1} (\underline{f}(\pi(q^*+1)/T))^{-1}, q^*+1)$$

$$\begin{aligned}
&= P \left(\left[1 - \kappa^{-1} \left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1} \right] Z_0^2 > \right. \\
&\quad \left. \sum_{j=1}^{q^*+1} \left[\left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1} \underline{f}(\pi j/T) - 1 \right] Z_j^2 \right) \\
&= P \left(\left[1 - \kappa^{-1} \left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1} \right] Z_0^2 > \right. \\
&\quad \left. \sum_{j=1}^{q^*} \left[\left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1} \underline{f}(\pi j/T) - 1 \right] Z_j^2 \right) \\
&= 0 < \alpha.
\end{aligned}$$

Pick ε such that $1 < \text{cv}_{q^*} - \varepsilon < \kappa^{-1} \left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1} < \text{cv}_{q^*}$. Then

$$\begin{aligned}
&Q(\text{cv}_{q^*} - \varepsilon, q^* + 1) \\
&= P \left([1 - (\text{cv}_{q^*} - \varepsilon)] Z_0^2 > \sum_{j=1}^{q^*+1} [(\text{cv}_{q^*} - \varepsilon) \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\
&> P \left([1 - (\text{cv}_{q^*} - \varepsilon)] Z_0^2 > \sum_{j=1}^{q^*} [(\text{cv}_{q^*} - \varepsilon) \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\
&= Q(\text{cv}_{q^*} - \varepsilon, q^*) > Q(\text{cv}_{q^*}, q^*) = \alpha,
\end{aligned}$$

where the last but one inequality follows from the fact that $Q(\cdot, q^*)$ is monotonically decreasing in $(1, \text{cv}_{q^*})$ under $\kappa \underline{f}(\pi/T) < 1$.

By the continuity of $Q(\cdot, q^* + 1)$ in $(\text{cv}_{q^*} - \varepsilon, \kappa^{-1} \left(\underline{f}(\pi(q^* + 1)/T) \right)^{-1})$ and the intermediate value theorem, there must exist a number, denoted by cv_{q^*+1} , such that $Q(\text{cv}_{q^*+1}, q^* + 1) = \alpha$. There is a contradiction, because cond_{q^*+1} now holds, violating Corollary A.1.8.

Case 2. Suppose $\kappa \underline{f}(\pi/T) > 1$, then $\text{cv}_{q^*} < 1$ by Lemma A.1.10.

$$Q(\text{cv}_{q^*}, q^* + 1) = P \left([1 - \text{cv}_{q^*}] Z_0^2 > \sum_{j=1}^{q^*+1} [\text{cv}_{q^*} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right)$$

$$\begin{aligned}
&< P \left([1 - \text{cv}_{q^*}] Z_0^2 > \sum_{j=1}^{q^*} [\text{cv}_{q^*} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) \\
&= Q(\text{cv}_{q^*}, q^*) = \alpha.
\end{aligned}$$

On the other hand, $0 < \kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1} < \text{cv}_{q^*} < 1$. Then

$$\begin{aligned}
&Q(\kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1}, q^* + 1) \\
&= P \left(\left[1 - \kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1} \right] Z_0^2 > \right. \\
&\quad \left. \sum_{j=1}^{q^*+1} \left[(\underline{f}(\pi(q^* + 1)/T))^{-1} \underline{f}(\pi j/T) - 1 \right] Z_j^2 \right) \\
&= P \left(\left[1 - \kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1} \right] Z_0^2 > \right. \\
&\quad \left. \sum_{j=1}^{q^*} \left[(\underline{f}(\pi(q^* + 1)/T))^{-1} \underline{f}(\pi j/T) - 1 \right] Z_j^2 \right) \\
&= Q(\kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1}, q^*) > Q(\text{cv}_{q^*}, q^*) = \alpha,
\end{aligned}$$

where the last but one inequality follows from the fact that $Q(\cdot, q^*)$ is monotonically decreasing in $(0, \text{cv}_{q^*})$ under $\kappa \underline{f}(\pi/T) > 1$.

By the continuity of $Q(\kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1}, \text{cv}_{q^*})$ and the intermediate value theorem, there must exist a number, denoted by cv_{q^*+1} , such that $Q(\text{cv}_{q^*+1}, q^* + 1) = \alpha$. There is a contradiction, because cond_{q^*+1} now holds, violating Corollary A.1.8.

Both **Case 1** and **Case 2** imply that we must have $\kappa^{-1} (\underline{f}(\pi(q^* + 1)/T))^{-1} \geq \text{cv}_{q^*}$, and thus $\min_{q^*+1 \leq j \leq T-1} \kappa^{-1} (\underline{f}(\pi j/T))^{-1} \geq \text{cv}_{q^*}$ if $q^* < T - 1$. ■

A.1.2 Proof of Theorem 1.3.2

Part 1 of Theorem 1.3.2 is already proved in the main text. I now focus on proving part 2 of Theorem 1.3.2, in which $\underline{f}(\pi/T) > \kappa^{-1}$. Under Assumption 1.3.1(a) and

for $0 < \alpha < 1$, Corollary A.1.8 shows that there exists a unique q^* such that *either* (i) cond_q holds for $1 \leq q \leq q^*$ and is violated for $q^* + 1 \leq q \leq T - 1$, *or* (ii) cond_q holds for all $1 \leq q \leq T - 1$, where we define $q^* = T - 1$. We conjecture that the least favorable distribution Λ^* of (1.9) puts a point mass on $f^*(\phi) = \underline{f}(\phi)\mathbf{1}[0 \leq \phi \leq \pi q^*/T] + (\kappa \text{cv}_{q^*})^{-1}\mathbf{1}[\phi > \pi q^*/T]$. By Lemma A.1.11 and Assumption 1.3.1(c), the kinked function f^* is known to be in $\mathcal{F}$.

Using the same notation as in the main text, the best level α test of H_{0,Λ^*}^d against H_{1,f_1}^d is

$$\varphi_{\Lambda^*,f_1} = 1 \left[\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2 / f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \text{cv} \right],$$

for some $\text{cv} \geq 0$ such that $E_{P_{Y,f^*}}[\varphi_{\Lambda^*,f_1}(Y^s)] = \alpha$, where $P_{Y,\tilde{f}}$ denotes the joint distribution of Y at $f = \tilde{f}$ under H_0^d . It follows that

$$\begin{aligned} \alpha &= P_{Y,f^*} \left(\frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2 / f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \text{cv} \right) \\ &= P_{Y,f^*} \left(Y_0^2 + \sum_{j=1}^{T-1} Y_j^2 / f^*(\pi j/T) > \text{cv} \left(Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2 \right) \right) \\ &= P \left((1 - \text{cv}) Z_0^2 > \sum_{j=1}^{T-1} [\text{cv} \kappa f^*(\pi j/T) - 1] Z_j^2 \right) \\ &= P \left((1 - \text{cv}) Z_0^2 > \sum_{j=1}^{q^*} [\text{cv} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 + \sum_{j=q^*+1}^{T-1} [\text{cv} / \text{cv}_{q^*} - 1] Z_j^2 \right), \quad (\text{A.11}) \end{aligned}$$

where the last equality follows from the definition of f^* . Because Y is a continuous random vector, the critical value cv is unique. By matching (A.11) with (A.1) at $\tilde{q} = q^*$, we have $\text{cv} = \text{cv}_{q^*}$. Moreover, the events $\left\{ \frac{Y_0^2 + \sum_{j=1}^{T-1} Y_j^2 / f^*(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{T-1} Y_j^2} > \text{cv}_{q^*} \right\}$ and $\left\{ \frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2 / \underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} > \text{cv}_{q^*} \right\}$ are equivalent $P_{Y,\tilde{f}}$ -almost surely, and uniformly in $\tilde{f} \in \mathcal{F} \cup \{f_1^*\}$. This then leads to the optimal test statistic in (1.10).

Now we check that φ_{Λ^*, f_1} controls size under H_0^d . For a given $\tilde{f} \in \mathcal{F}$,

$$\begin{aligned}
& E_{P_{Y, \tilde{f}}} [\varphi_{\Lambda^*, f_1}(Y)] \\
&= P_{Y, \tilde{f}} \left(\frac{Y_0^2 + \sum_{j=1}^{q^*} Y_j^2 / \underline{f}(\pi j/T)}{Y_0^2 + \kappa \sum_{j=1}^{q^*} Y_j^2} > \text{cv}_{q^*} \right) \\
&= P \left([1 - \text{cv}_{q^*}] Z_0^2 > \sum_{j=1}^{q^*} [\text{cv}_{q^*} \kappa \underline{f}(\pi j/T) - 1] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2 \right) \\
&= P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{q^*}} \sum_{j=1}^{q^*} [\text{cv}_{q^*} \kappa \underline{f}(\pi j/T) - 1] \frac{\tilde{f}(\pi j/T)}{\underline{f}(\pi j/T)} Z_j^2 \right) \quad (\text{A.12}) \\
&\leq P \left(Z_0^2 > \frac{1}{1 - \text{cv}_{q^*}} \sum_{j=1}^{q^*} [\text{cv}_{q^*} \kappa \underline{f}(\pi j/T) - 1] Z_j^2 \right) = \alpha,
\end{aligned}$$

where (A.12) follows from (b) in Lemma A.1.10 under the condition $\underline{f}(\pi/T) > \kappa^{-1}$, and the inequality follows from the definition of q^* and Lemma A.1.1 under Assumption 1.3.1(a).

In sum, we have shown that Λ^* is indeed the least favorable distribution in testing H_0^d against H_{1, f_1}^d , under Assumption 1.3.1 and $\underline{f}(\pi/T) > \kappa^{-1}$. Further, the statistic (1.10) is the best test statistic, and the null rejection probability of this test is maximized at $f = \underline{f}$.

A.2 Computational Details in Section 1.3

A.2.1 An algorithm to compute q^* in Theorem 1.3.2

In Algorithm 1 on the following page, I first use the bisection method three times to zoom into a small set of consecutive integers and then locate q^* by a one-dimensional greedy search. One might be concerned about the sensitivity of the numerically determined q^* to the precision of empirically estimated critical values for different q .

As will be explained in the next subsection, this is not an issue in implementations with $B = 50,000$.

Algorithm 1: Bisection method to produce q^* in Theorem 1.3.2

Input: $\mathcal{F}$ with a well-defined $\underline{f}$, T , $\alpha = 0.05$ by default, $\kappa = 11$ by default,

$\left\{ \{z_j^b\}_{j=0}^{T-1} \right\}_{b=1}^B \leftarrow B$ realizations of i.i.d. standard normals for a sufficiently large B .

Output: q^*

Initialization: $\tilde{q} \leftarrow T - 1$,

$\hat{cv}_{\tilde{q}} \leftarrow \inf\{r \in \mathbb{R}_+ \mid \frac{1}{B} \sum_{b=1}^B \mathbf{1} \left[\frac{\sum_{j=0}^{\tilde{q}} (z_j^b)^2}{(z_0^b)^2 + \kappa \sum_{j=1}^{\tilde{q}} (z_j^b)^2 \underline{f}(\pi j/T)} > r \right] \geq \alpha\};$

if $\text{cond}_{\tilde{q}}$ *is true with* $cv_{\tilde{q}}$ *replaced by* $\hat{cv}_{\tilde{q}}$ **then**

$q^* \leftarrow \tilde{q};$

else

while $\text{cond}_{\tilde{q}}$ *is false with* $cv_{\tilde{q}}$ *replaced by* $\hat{cv}_{\tilde{q}}$ **do**

$\tilde{q} \leftarrow \left\lfloor \frac{1+\tilde{q}}{2} \right\rfloor;$

end

$\tilde{\tilde{q}} \leftarrow \left\lfloor \frac{3\tilde{q}-1}{2} \right\rfloor;$

while $\text{cond}_{\tilde{\tilde{q}}}$ *is false with* $cv_{\tilde{\tilde{q}}}$ *replaced by* $\hat{cv}_{\tilde{\tilde{q}}}$ **do**

$\tilde{\tilde{q}} \leftarrow \left\lfloor \frac{\tilde{\tilde{q}}+\tilde{q}}{2} \right\rfloor;$

end

for $q \leftarrow \tilde{\tilde{q}}$ **to** $2\tilde{\tilde{q}} - \tilde{q} + 1$ **by** 1 **do**

if cond_{q+1} *is false with* cv_{q+1} *replaced by* $\hat{cv}_{q+1}$ **then**

$q^* \leftarrow q;$

break ;

end

end

end

A.2.2 Computational details for cv_{q^*} in Theorem 1.3.2

Given the q^* delivered via Algorithm 1, I numerically invert (1.13) by the standard bisection method. Since the integrand in the last integral in (1.13) is smooth for typical values of cv_{q^*} , I use a 50-point Gaussian quadrature to compute the numerical integration. It suffices for my purpose with numerical errors of the order $1e - 6$. Moreover, to check whether the determination of q^* in Algorithm 1 is sensitive to simulation errors in estimating $cv_{\tilde{q}}$, I append a sanity check to Algorithm 1. In

particular, I substitute the more precisely determined $\text{cv}_{q^*}, \text{cv}_{q^*-1}, \text{cv}_{q^*-2}$ back into $\text{cond}_{q^*}, \text{cond}_{q^*-1}, \text{cond}_{q^*-2}$ and check whether any of them are inconsistent with the definition of q^* . It appears that having 50,000 realizations of $\{Z_j\}_{j=0}^{T-1}$ suffices for our purpose.

A.3 More Tables in the Diagonal Model

I follow the procedure of producing Tables 1.1 and 1.2 to provide more rule of thumb tables for other types of smoothness classes. See Tables A.1 and A.2 for the rule of thumb under the type of $\mathcal{F}$, in which the first derivative of $\log(f)$ is bounded by the constant C . See Tables A.3 and A.4 for the rule of thumb under the type of $\mathcal{F}$, in which the second derivative of $\log(f)$ is bounded by the constant C .

A.4 Computational details in Section 1.4

In this section, I explain in detail how to numerically identify the null rejection probability maximizer of the EWC test in testing (1.17), as introduced in Subsection 1.4.1.

To make things clearer, we repeat some notations from the main text. Let the $n+1$ node points $\{x_i\}_{i=0}^n$ define a partition of the interval $I = [0, \pi]$ into n subintervals $I_i = [x_{i-1}, x_i]$, $i = 1, 2, \dots, n$ of length $h_i = x_i - x_{i-1}$ with $x_0 = 0$, $x_n = \pi$. Let $\{\varsigma_i\}_{i=0}^n$ be a set of basis functions for the space $\mathcal{F}_h$ of continuous piecewise linear functions, which is defined in the main text.

Table A.1: Optimal q and adjustment factor of the Student- t critical value of level α EWC test.

C	10.0	5.6	3.2	1.8	1.0	0.6	0.3	0.2	0.1
$\alpha = 0.01$	(5, 1.59)	(6, 1.35)	(9, 1.27)	(11, 1.18)	(15, 1.13)	(20, 1.10)	(26, 1.07)	(34, 1.05)	(45, 1.04)
$\alpha = 0.05$	(4, 1.47)	(5, 1.30)	(7, 1.21)	(9, 1.15)	(11, 1.10)	(15, 1.07)	(20, 1.05)	(26, 1.04)	(34, 1.03)
$\alpha = 0.10$	(3, 1.36)	(4, 1.24)	(6, 1.19)	(7, 1.12)	(10, 1.09)	(12, 1.06)	(16, 1.04)	(22, 1.03)	(29, 1.02)

Notes: Based on a series of smoothness classes $\mathcal{F}$, in which the “uniformly minimal” function is $\underline{f}(\phi) = \exp(-C\phi)$. Sample size $T = 100$. C is log-spaced between 10 and 0.1.

Table A.2: Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.01$	1.53 (11.0)	1.35 (5.6)	1.26 (3.3)	1.27 (3.2)	1.19 (2.1)	1.18 (1.8)	1.16 (1.5)	1.15 (1.3)	1.14 (1.1)	1.13 (1.0)	1.12 (0.9)	1.12 (0.8)	1.11 (0.7)	1.10 (0.62)	1.10 (0.6)
$\alpha = 0.05$	1.47 (10.0)	1.30 (5.6)	1.19 (3.3)	1.21 (3.2)	1.11 (1.2)	1.10 (1.0)	1.09 (0.9)	1.09 (0.8)	1.08 (0.7)	1.07 (0.6)	1.07 (0.5)	1.06 (0.4)	1.06 (0.38)	1.05 (0.35)	1.05 (0.3)
$\alpha = 0.10$	1.24 (5.6)	1.19 (3.2)	1.11 (1.5)	1.10 (1.2)	1.09 (1.0)	1.07 (0.7)	1.06 (0.6)	1.06 (0.5)	1.05 (0.45)	1.05 (0.4)	1.04 (0.3)	1.04 (0.28)	1.04 (0.25)	1.04 (0.23)	1.03 (0.21)

Notes: Number in parentheses is the C in $\underline{f}(\phi) = \exp(-C\phi)$ that rationalizes the corresponding q as the optimal choice for the EWC test under $T = 100$.

Table A.3: Optimal q and adjustment factor of the Student- t critical values of level α EWC test.

C	100.0	56.2	31.6	17.8	10.0	5.6	3.2	1.8	1.0
$\alpha = 0.01$	(4, 1.43)	(5, 1.34)	(6, 1.25)	(7, 1.18)	(8, 1.13)	(10, 1.11)	(12, 1.09)	(14, 1.06)	(17, 1.05)
$\alpha = 0.05$	(3, 1.25)	(4, 1.22)	(5, 1.18)	(6, 1.14)	(7, 1.10)	(8, 1.07)	(10, 1.06)	(12, 1.05)	(14, 1.04)
$\alpha = 0.10$	(3, 1.25)	(3, 1.14)	(4, 1.12)	(5, 1.10)	(6, 1.08)	(7, 1.06)	(9, 1.05)	(10, 1.03)	(13, 1.03)

Notes: Based on a series of smoothness classes $\mathcal{F}$, in which the “uniformly minimal” function is $\underline{f}(\phi) = \exp(-C\phi^2)$. Sample size $T = 100$. C is log-spaced between 100 and 1.

Table A.4: Rule of thumb for adjustment factor of the Student- t critical value of level α EWC test.

q	4	6	8	9	10	11	12	13	14	15	16	17	18	19	20
$\alpha = 0.01$	1.43 (100.0)	1.24 (31.6)	1.13 (10.0)	1.13 (8.0)	1.11 (5.6)	1.09 (4.0)	1.09 (3.2)	1.07 (2.2)	1.06 (1.8)	1.06 (1.5)	1.06 (1.2)	1.05 (1.0)	1.05 (0.9)	1.04 (0.7)	1.04 (0.6)
$\alpha = 0.05$	1.22 (56.2)	1.14 (17.8)	1.07 (5.6)	1.06 (4.0)	1.06 (3.2)	1.05 (2.0)	1.05 (1.8)	1.04 (1.2)	1.04 (1.0)	1.04 (0.9)	1.03 (0.7)	1.03 (0.6)	1.03 (0.5)	1.03 (0.4)	1.02 (0.35)
$\alpha = 0.10$	1.12 (31.6)	1.08 (10.0)	1.05 (4.0)	1.05 (3.2)	1.03 (1.8)	1.03 (1.5)	1.03 (1.2)	1.03 (1.0)	1.03 (0.8)	1.02 (0.6)	1.02 (0.5)	1.02 (0.4)	1.02 (0.35)	1.02 (0.3)	1.02 (0.25)

Notes: Number in parentheses is the C in $\underline{f}(\phi) = \exp(-C\phi^2)$ that rationalizes the corresponding q as the optimal choice for the EWC test under $T = 100$.

In actual implementations, I choose $n = 50$, and $\{x_i\}_{i=0}^{50}$ are log-spaced nodes in $[0, \pi]$. The basis functions $\{\varsigma_i\}_{i=0}^n$ are chosen to be the hat functions

$$\varsigma_i(x) = \begin{cases} (x - x_{i-1})/h_i & , \text{ if } x \in I_i, \\ (x_{i+1} - x)/h_{i+1} & , \text{ if } x \in I_{i+1}, \\ 0 & , \text{ otherwise.} \end{cases} \quad (\text{A.13})$$

I pre-compute $\{\Omega_0(\varsigma_i)\}_{i=0}^n$ with a 5,000-point Gaussian quadrature for each non-zero element. Since each ς_i is compactly supported, these numerical integrations are nearly precise. For every $(f(x_1), f(x_2), \dots, f(x_n))'$, $\Omega_0(\hat{f})$ is simply a linear combination of these pre-computed covariance matrices.

Note, however, that the ultimate objective function I will optimize is (1.21), which involves $\Omega_0(f)$ implicitly through $\lambda_j(\hat{f})$. In a separate exercise in which, for a given EWC test and a parametric AR(1) class $\mathcal{F}$ with coefficient varying over a fine grid, I compare the rejection probabilities following the described approximate procedure and an “exact” procedure in which each entry of $\Omega_0(f)$ is evaluated by numerical integrations via Mathematica. The difference in the rejection probabilities are at most of the order 0.0001. I thus hold on to the above choice of n and $\{x_i\}_{i=0}^{50}$ throughout Section 1.4.

I proceed in three steps to identify the null rejection probability maximizer for a fixed EWC test of (1.17), in terms of $(f(x_1), f(x_2), \dots, f(x_n))'$.

1. Program up the null rejection probability at a given $(f(x_1), f(x_2), \dots, f(x_n))' \in \mathbb{R}_+^n$ from (1.21), where $\Omega_0(\hat{f}) = \sum_{i=0}^n f(x_i)\Omega_0(\varsigma_i)$ with pre-computed $\{\Omega_0(\varsigma_i)\}_{i=0}^n$.
2. Randomly draw 1,000 n -dimensional vectors such that each vector equals $(f(x_1), f(x_2), \dots, f(x_n))'$ for some $f \in \mathcal{F}$. This is in general a challenging task, since the number of numerical constraints to be checked increases exponentially with n for higher order smoothness constraints. For feasibility, I focus on two

types of smoothness classes: the class $\mathcal{F}$ in which $\underline{f}$ corresponds is AR(1) with coefficient ρ and $f \in \mathcal{F}$ is non-increasing over $[0, \pi]$; and the class $\mathcal{F}$ in which $f \in \mathcal{F}$ is Lipschitz continuous in logs with Lipschitz constant C . It is not hard to see that for the first type, it suffices to check the monotonicity constraint consecutively and the lower boundedness condition. For the second type, by the result of Beliakov (2006), the complexity of checking the global Lipschitz condition is reduced to consecutive checking of local Lipschitz conditions.

3. Use every n -dimensional vector drawn in Step 2 as the initial condition to optimize the null rejection probability function programmed in Step 1, subject to linear constraints induced by smoothness class $\mathcal{F}$ (as described in Step 2).

Under the above specifications, it takes 5 to 10 minutes to complete the optimization using *fmincon* in MATLAB via parallel computing in 12 cores.

Appendix B

Appendices for Chapter 2

B.1 Proof of Theorem 1

We first show that $x_{T,t}$ has representation (B.1) and (B.2). Set $\prod_{j=1}^p(z - \rho_{T,j}) = z^p + \sum_{j=1}^p \phi_{T,j} z^{p-j}$ and $\prod_{j=1}^{p-1}(z - \gamma_{T,j}) = z^{p-1} + \sum_{j=0}^{p-2} \theta_{T,j} z^j$. The usual state-space representation of the ARMA($p, p-1$) process $x_{T,t}$ with innovations u_t is

$$x_{T,t} = \theta'_T \mathbf{V}_t + \mu \quad (\text{B.1})$$

$$\mathbf{V}_{T,t} = \Phi_T \mathbf{V}_{T,t-1} + \mathbf{e} u_t \quad (\text{B.2})$$

where

$$\Phi_T = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -\phi_{T,p} & -\phi_{T,p-1} & -\phi_{T,p-2} & \cdots & -\phi_{T,1} \end{pmatrix},$$

and

$$\mathbf{V}_{T,t} = \begin{pmatrix} v_{T,t-p+1} \\ v_{T,t-p+2} \\ \vdots \\ v_{T,t-1} \\ v_{T,t} \end{pmatrix}, \quad \theta_T = \begin{pmatrix} \theta_{T,0} \\ \theta_{T,1} \\ \vdots \\ \theta_{T,p-2} \\ 1 \end{pmatrix}.$$

Let $\mathbf{c} = (c_1, \dots, c_p)'$ and $\mathbf{g} = (g_1, \dots, g_{p-1})'$ with elements ordered ascendingly by the real parts, and define the corresponding vectors $\rho_T = (\rho_{T,1}, \dots, \rho_{T,p})'$ and $\gamma_T = (\gamma_{T,1}, \dots, \gamma_{T,p-1})'$.

Since Φ_T and $\mathbf{A}$ are companion matrices, and the roots of $z^p + \sum_{j=1}^p \phi_{T,j} z^{p-j}$ and $a(z)$ are ρ_T and $-\mathbf{c}$, respectively, they allow the Jordan decomposition (cf. Brand (1964))

$$\Phi_T = \mathbf{Q}(\rho_T) \mathbf{J}(\rho_T) \mathbf{Q}(\rho_T)^{-1} \quad (\text{B.3})$$

$$\mathbf{A} = \mathbf{Q}(-\mathbf{c}) \mathbf{J}(-\mathbf{c}) \mathbf{Q}(-\mathbf{c})^{-1} \quad (\text{B.4})$$

where for any $\mathbf{a} = (a_1, \dots, a_k)' \in \mathbb{C}^k$, $k \in \{p-1, p\}$ with $\text{Re}(a_j) \leq \text{Re}(a_{j+1})$, $\mathbf{J}(\mathbf{a})$ is a Jordan matrix with Jordan blocks corresponding to common values of a_i , and $\mathbf{Q}(\mathbf{a})$ are the corresponding (generalized) eigenvectors. More specifically, the m columns of $\mathbf{Q}(\mathbf{a})$ corresponding to the value a of multiplicity m contain the values $\frac{d^l}{dz^l} z^{j-1} \big|_{z=a/l!}$, $j = 1, \dots, k$, $l = 0, \dots, m-1$.

From (B.4), we also have

$$\mathbf{I} + \mathbf{A}/T = \mathbf{Q}(-\mathbf{c})(\mathbf{I} + \mathbf{J}(-\mathbf{c})/T) \mathbf{Q}(-\mathbf{c})^{-1}. \quad (\text{B.5})$$

Let $\mathbf{F}$ be the $p \times p$ lower triangular Pascal matrix, that is, the first j entries in row j of $\mathbf{F}$ contain the j th binomial coefficients, and let $\mathbf{D}_T = \text{diag}(1, T^{-1}, \dots, T^{1-p})$. Further, let $\mathbf{D}_T^c$ be a diagonal matrix where the diagonal elements corresponding to

a Jordan block of $\mathbf{A}$ of size m are equal to $1, T, \dots, T^{m-1}$. Then, with $\mathbf{P}_T = \mathbf{F}\mathbf{D}_T$ we have from a straightforward calculation

$$\begin{aligned}\mathbf{Q}(\rho_T) &= \mathbf{P}_T \mathbf{Q}(-\mathbf{c}) \mathbf{D}_T^c \\ \mathbf{D}_T^c \mathbf{J}(\rho_T) (\mathbf{D}_T^c)^{-1} &= \mathbf{I} + \mathbf{J}(-\mathbf{c})/T\end{aligned}\tag{B.6}$$

so that from (B.3) and (B.5)

$$\Phi_T = \mathbf{P}_T (\mathbf{I} + \mathbf{A}/T) \mathbf{P}_T^{-1}.\tag{B.7}$$

Furthermore, since $z^{p-1} + \sum_{j=0}^{p-2} \theta_{T,j} z^j = \prod_{j=1}^{p-1} (z - \gamma_{T,j})$, we have $\theta'_T \mathbf{Q}(\gamma_T) = 0$, and similarly, $\mathbf{b}' \mathbf{Q}(-\mathbf{g}) = 0$. Now as in (B.6), $\mathbf{Q}(\gamma_T) = \mathbf{P}_T \mathbf{Q}(-\mathbf{g}) \mathbf{D}_T^g$ for some diagonal matrix $\mathbf{D}_T^g$ with nonzero diagonal elements, so that also $\theta'_T \mathbf{P}_T \mathbf{Q}(-\mathbf{g}) = 0$. Since $\mathbf{Q}(-\mathbf{g})$ is of full column rank (cf. Theorem 2 of Brand (1964)), we conclude that $\theta'_T \mathbf{P}_T$ is a scalar multiple of $\mathbf{b}'$. The last element of $\mathbf{b}'$ is equal to one, and the last element of $\theta'_T \mathbf{P}_T$ is equal to T^{1-p} , so that

$$\theta'_T \mathbf{P}_T = T^{1-p} \mathbf{b}'.\tag{B.8}$$

Finally, from $\mathbf{P}_T \mathbf{e} = T^{1-p} \mathbf{e}$,

$$\mathbf{P}_T^{-1} \mathbf{e} = T^{p-1} \mathbf{e}.\tag{B.9}$$

From (B.7), (B.8) and (B.9) it follows that the system (B.1) and (B.2) can equivalently be written as (2.10) and (2.11).

For an arbitrary matrix $\mathbf{B}$, let $\|\mathbf{B}\|$ its Frobenius norm. Since the Frobenius norm is submultiplicative, $\|\mathbf{I} + \mathbf{A}/T\| \leq \|\mathbf{Q}(-\mathbf{c})\| \cdot \|\mathbf{Q}(-\mathbf{c})^{-1}\| \cdot \|\mathbf{I} + \mathbf{J}(-\mathbf{c})/T\|$. By a direct calculation, $\|\mathbf{I} + \mathbf{J}(-\mathbf{c})/T\|^2 \leq (p-1)T^{-2} + \sum_{j=1}^p |\rho_{T,j}|^2$ and $|\rho_{T,j}|^2 = 1 - 2 \operatorname{Re}(c_j)/T + |c_j|^2/T^2$. Thus, for all large enough T , $|\rho_{T,j}|^2 + 1/T^2 \leq (1 - \frac{1}{2} \operatorname{Re}(c_1)/T)^2$ for all j ,

and we conclude that for some finite constant C_A

$$\|\mathbf{I} + \mathbf{A}/T\| \leq C_A(1 - \frac{1}{2} \operatorname{Re}(c_1)/T). \quad (\text{B.10})$$

Now from (2.10) and (2.11), for any fixed integer $K > 0$,

$$T^{-1/2}(x_{T, \lfloor sT \rfloor} - \mu) = R_T(s) + T^{-1/2} \mathbf{b}' \sum_{t=-KT+1}^{\lfloor sT \rfloor} (\mathbf{I} + \mathbf{A}/T)^{\lfloor sT \rfloor - t} \mathbf{e} u_t$$

where $R_T(s) = T^{-1/2} \mathbf{b}' (\mathbf{I} + \mathbf{A}/T)^{\lfloor sT \rfloor + KT} \mathbf{Z}_{-KT}$ and $\mathbf{Z}_{-KT} = \sum_{t=0}^{\infty} (\mathbf{I} + \mathbf{A}/T)^t \mathbf{e} u_{-KT-t}$, and we write $\mathbf{Z}_t$ for $\mathbf{Z}_{T,t}$ to ease notation. Since the autocovariances of u_t are absolutely summable, the spectral density of u_t exists and is bounded on $[-\pi, \pi]$. Let a bound be $\tilde{\sigma}_u^2/(2\pi)$. For any given T and $\mathbf{w} \in \mathbb{R}^p$, the variance of the time invariant linear filter $\mathbf{w}' \mathbf{Z}_{-KT}$ is thus weakly smaller than the variance of $\mathbf{w}' \tilde{\mathbf{Z}}_{-KT}$, where $\tilde{\mathbf{Z}}_{-KT} = \sum_{t=0}^{\infty} (\mathbf{I} + \mathbf{A}/T)^t \mathbf{e} \tilde{u}_{-KT-t}$ with $\tilde{u}_t \sim iid(0, \tilde{\sigma}_u^2)$. Furthermore

$$\operatorname{Var}[T^{-1/2} \tilde{\mathbf{Z}}_{-KT}] = \tilde{\sigma}_u^2 T^{-1} \sum_{t=0}^{\infty} (\mathbf{I} + \mathbf{A}/T)^t \mathbf{e} \mathbf{e}' (\mathbf{I} + \mathbf{A}'/T)^t$$

so that from (B.10)

$$\|\operatorname{Var}[T^{-1/2} \tilde{\mathbf{Z}}_{-KT}]\| \leq \tilde{\sigma}_u^2 \|\mathbf{e} \mathbf{e}'\| C_A^2 T^{-1} \sum_{t=0}^{\infty} (1 - \frac{1}{2} \operatorname{Re}(c_1)/T)^{2t} = O(1).$$

Thus, $\|T^{-1/2} \mathbf{Z}_{-KT}\| = O_p(1)$. Using again (B.10) and the fact that the Frobenius norm is submultiplicative, we obtain

$$\begin{aligned} \sup_{0 \leq s \leq 1} |R_T(s)| &\leq C_A \|\mathbf{b}\| \cdot \|T^{-1/2} \mathbf{Z}_{-KT}\| \cdot \sup_s (1 - \frac{1}{2} \operatorname{Re}(c_1)/T)^{\lfloor sT \rfloor + KT} \quad (\text{B.11}) \\ &\leq C_A \|\mathbf{b}\| \cdot \|T^{-1/2} \mathbf{Z}_{-KT}\| \cdot \exp(-\frac{1}{2} K \operatorname{Re} c_1) \end{aligned}$$

so that $R_T(\cdot)$ converges in probability to zero as $K \rightarrow \infty$.

Furthermore, under Condition 2.2.1,

$$W_T(\cdot) = T^{-1/2} \sum_{t=-\lfloor KT \rfloor}^{\lfloor sT \rfloor} u_t \Rightarrow W(\cdot) - W(-K)$$

where W is a Wiener process on the interval $[-K, 1]$ of variance ω^2 normalized to $W(0) = 0$. By summation by parts,

$$\begin{aligned} & T^{-1/2} \mathbf{b}' \sum_{t=-\lfloor KT \rfloor+1}^{\lfloor sT \rfloor} (\mathbf{I} + \mathbf{A}/T)^{\lfloor sT \rfloor - t} \mathbf{e} u_t \\ &= \mathbf{b}' \mathbf{e} W_T(s) + \mathbf{b}' \mathbf{A} T^{-1} \sum_{t=-\lfloor KT \rfloor+1}^{\lfloor sT \rfloor} (\mathbf{I} + \mathbf{A}/T)^{\lfloor sT \rfloor - t} \mathbf{e} W_T\left(\frac{t-1}{T}\right) \\ &\Rightarrow \mathbf{b}' \mathbf{e} (W(s) - W(-K)) + \mathbf{b}' \mathbf{A} \int_{-K}^s e^{\mathbf{A}(s-r)} \mathbf{e} (W(r) - W(-K)) dr \\ &= \mathbf{b}' \int_{-K}^s e^{\mathbf{A}(s-r)} \mathbf{e} dW(r) \\ &= R_0(s) + J_p(s) \end{aligned}$$

where $R_0(s) = -\mathbf{b}' e^{\mathbf{A}(s+K)} \mathbf{X}(-K)$ with $\mathbf{X}(-K) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ independent of W as in (2.7), the convergence relies on the well-known identity $e^{s\mathbf{A}} = \lim_{T \rightarrow \infty} (\mathbf{I} + \mathbf{A}/T)^{\lfloor sT \rfloor}$ for all s , and the second equality follows from the stochastic calculus version of integration by parts. Since $\sup_{0 \leq s \leq 1} \|e^{\mathbf{A}(s+K)}\| \rightarrow 0$ as $K \rightarrow \infty$, $\sup_{0 \leq s \leq 1} |R_0(s)|$ converges in probability to zero as $K \rightarrow \infty$. As noted below (B.11), the same holds for $R_T(\cdot)$. But convergence in probability implies convergence in distribution, and K was arbitrary, so the result follows.

B.2 Proof of Theorem 2

Overview

The proof of Theorem 2 relies heavily on the framework developed by Ibragimov and Rozanov (1978), denoted IR78 in the following. As discussed there, a continuous

time Gaussian process on the unit interval can be described in terms of a countably infinite sequence of random variables (cf. (B.20) and the discussion in the proof of Lemma B.2.4 below), whose distribution can be expressed in terms of the spectral density of the underlying process (cf. (B.19) and the discussion below (B.20)). The challenge in the proof of Theorem 2 is to establish that the “infinite tail” of this sequence contributes negligibly to the total variation distance. Intuitively, this must hold for some appropriate definition of tail if the two measures are equivalent, and appropriate equivalence results are obtained by IR78. But the construction of this tail must be such that its contribution is negligible *uniformly* over a sufficiently rich class of potential approximating processes. To this end, the sequence of random variables (and hence its tail) is constructed as a function of the properties of two Gaussian processes whose spectral densities form an upper and lower bound on the class of potential approximating spectral density functions (cf. (B.15), (B.16) and (B.17)), which turns out to be suitable to obtain such a uniform bound (cf. (B.22), (B.23) and (B.26)). With the contribution from the tail controlled, the approximability of the distribution of the finite dimensional non-tail part of the sequence of random variables follows with some additional work from Lemmas 1 and 2 below.

We first state Lemmas 1 and 2. We write z^* for the conjugate of the complex number z , and $\mathbf{v}^*$ for the conjugate transpose of a complex vector $\mathbf{v}$.

Lemma B.2.1 *Let $\mathcal{C}_0$ be the space of continuous real valued functions on $[0, \infty)$ which vanish at infinity. For any $\vartheta_0 \in \mathcal{C}_0$ and $\varepsilon > 0$, there exists an integer $q \geq 1$ such that $\sup_{\lambda \geq 0} |\vartheta_0(\lambda) - \vartheta(\lambda)| < \varepsilon$, where ϑ is a rational function of the form*

$$\vartheta(\lambda) = \frac{\sum_{j=0}^{q-1} e_j^n \lambda^{2j}}{\prod_{j=1}^q (\lambda^2 + e_j^d)} \quad (\text{B.12})$$

with $e_j^d > 0$ and $e_j^n \in \mathbb{R}$, $j = 0, \dots, q$.

Proof. Note that functions of the form ϑ form a vector subspace of $\mathcal{C}_0$ which is closed under multiplication of functions, that is, they form a sub-algebra on $\mathcal{C}_0$. It is easily seen that this sub-algebra separate points and vanishes nowhere. The locally compact version of the Stone-Weierstrass Theorem thus implies the result. ■

Lemma B.2.2 *Let $\xi_p(\lambda^2)$ be a polynomial with real coefficients of order $p - 1$ in λ^2 such that $\xi_p(\lambda^2) > 0$ for all $\lambda \in \mathbb{R}$, and with unit coefficient on $(\lambda^2)^{p-1}$. Then there exists polynomial b of order $p - 1$ of the form $b(z) = \prod_{i=1}^{p-1}(z + g_i)$ with g_i as described in Condition 1 (ii) such that $\xi_p(\lambda^2) = |b(i\lambda)|^2$ for all $\lambda \in \mathbb{R}$.*

Proof. By the fundamental theorem of algebra, and since $\xi_p(\lambda^2) > 0$ for all $\lambda \in \mathbb{R}$, $\xi_p(\lambda^2) = \prod_{j=1}^{p-1}(\lambda^2 + \eta_j)$, where the η_j 's are of two types: real and positive, or complex with positive real part, and in conjugate pairs. Now for $0 < \eta_j \in \mathbb{R}$, $\lambda^2 + \eta_j = |i\lambda + g_j|^2$ with $g_j = \sqrt{\eta_j}$. For $\eta_j = \eta_{j'}^* \in \mathbb{C}$ for $j \neq j'$,

$$(\lambda^2 + \eta_j)(\lambda^2 + \eta_{j'}) = \lambda^4 + 2\operatorname{Re}(\eta_j)\lambda^2 + |\eta_j|^2 = |i\lambda + g_j|^2 |i\lambda + g_j^*|^2$$

where $g_j = (\sqrt{|\eta_j| + \operatorname{Re}(\eta_j)} + \sqrt{|\eta_j| - \operatorname{Re}(\eta_j)}i)/\sqrt{2}$. ■

Without loss of generality, assume $\omega^2 = 2\pi$. In the following, we write G_1 for G , and f_1 for its spectral density. Let $f_0(\lambda) = (1 + \lambda^2)^{-1}$ be the spectral density of the Ornstein-Uhlenbeck process with mean reversion parameter equal to unity, denoted G_0 , and let

$$\delta_0 = \min(\tfrac{1}{2} \inf_{\lambda} (1 + \lambda^2) f_1(\lambda), 1). \quad (\text{B.13})$$

Let P_0 and P_1 be the measures of G_1 and G_0 , respectively. By Theorem III.17 of IR78, equivalence of P_0 and P_1 implies

$$\int (\frac{f_1(\lambda)}{f_0(\lambda)} - 1)^2 d\lambda < \infty \quad (\text{B.14})$$

and here and below, integrals are over the entire real line unless indicated otherwise.

Define

$$\overline{f}(\lambda) = \max(f_1(\lambda), f_0(\lambda)) + \frac{\delta_0}{(1 + \lambda^2)^2} \quad (\text{B.15})$$

$$\underline{f}(\lambda) = \min(f_1(\lambda), f_0(\lambda)) - \frac{\delta_0}{(1 + \lambda^2)^2} \quad (\text{B.16})$$

and let f_2 be some function satisfying

$$\underline{f}(\lambda) \leq f_2(\lambda) \leq \overline{f}(\lambda) \text{ for all } \lambda. \quad (\text{B.17})$$

From (B.14),

$$\int \left(\frac{\underline{f}(\lambda)}{f_0(\lambda)} - 1 \right)^2 d\lambda < \infty, \quad \int \left(\frac{\overline{f}(\lambda)}{f_0(\lambda)} - 1 \right)^2 d\lambda < \infty \quad (\text{B.18})$$

so that also for all f_2 satisfying (B.17), $\int (f_2(\lambda)/f_0(\lambda) - 1)^2 d\lambda < \infty$. Since $\overline{f}$, $\underline{f}$ and f_2 are nonnegative integrable real functions, there exist corresponding correlation functions that are positive definite. By the development in Section I.2 of IR78, there hence exist corresponding stationary Gaussian processes $\overline{G}$, $\underline{G}$ and G_2 with spectral densities $\overline{f}$, $\underline{f}$ and f_2 and measures $\overline{P}$, $\underline{P}$ and P_2 , respectively. Theorem III.17 of IR78 and (B.18) implies that $\overline{P}$, $\underline{P}$ and P_2 are equivalent to P_0 , and hence also to P_1 .

For $\psi, \varphi : \mathbb{R} \mapsto \mathbb{C}$ functions of the type $\psi(\lambda) = \sum_{k=1}^n c_k e^{i\lambda t_k}$ for $t_k \in [0, 1]$ and $c_k \in \mathbb{R}$, define the inner product

$$\langle \psi, \varphi \rangle_{F_1} = \int \psi(\lambda) \varphi(\lambda)^* f_1(\lambda) d\lambda. \quad (\text{B.19})$$

Let $L(F_1)$ be the corresponding Hilbert space. Analogously, define the inner products $\langle \psi, \varphi \rangle_{F_2}$, $\langle \psi, \varphi \rangle_{\overline{F}}$ and $\langle \psi, \varphi \rangle_{\underline{F}}$, and corresponding Hilbert spaces $L(F_2)$, $L(\overline{F})$ and $L(\underline{F})$. Since the measures P_1 , P_2 , $\overline{P}$ and $\underline{P}$ are equivalent, so are the Hilbert spaces, as noted on page 71 of IR78. Define the linear operator $A : L(F_1) \mapsto L(F_2)$ via $A\psi = \psi$, let A^* be its adjoint, and define the self-adjoint operator $\Delta : L(F_1) \mapsto L(F_1)$

via $\Delta\psi = \psi - A^*A\psi$, so that

$$\langle \Delta\psi, \varphi \rangle_{F_1} = \langle \psi, \varphi \rangle_{F_1} - \langle \psi, \varphi \rangle_{F_2}$$

and analogously for $\overline{\Delta}$ and $\underline{\Delta}$ (that is, $\langle \overline{\Delta}\psi, \varphi \rangle_{F_1} = \langle \psi, \varphi \rangle_{F_1} - \langle \psi, \varphi \rangle_{\overline{F}}$ and $\langle \underline{\Delta}\psi, \varphi \rangle_{F_1} = \langle \psi, \varphi \rangle_{F_1} - \langle \psi, \varphi \rangle_{\underline{F}}$). By Theorem III.4 of IR78, equivalence of the measures P_1 , P_2 , $\overline{P}$ and $\underline{P}$ implies that the operators Δ , $\overline{\Delta}$ and $\underline{\Delta}$ are Hilbert-Schmidt.

Let ψ_k be an arbitrary orthonormal sequence in $L(F_1)$, and define the $n \times 1$ vector η_n of Gaussian complex valued random variables

$$\eta(\psi_k) = \int \psi_k(\lambda) d\Phi_i(\lambda) \text{ for } k = 1, \dots, n \quad (\text{B.20})$$

where Φ_i is the stochastic spectral measure such that $G_i(s) = \int e^{i\lambda s} d\Phi_i(\lambda)$, $i = 1, 2$ (cf. chapter I.6 of IR78). Then $E[\eta(\psi_k)] = 0$ under both P_1 and P_2 , $E[\eta(\psi_j)\eta(\psi_k)] = \langle \psi_j, \psi_k \rangle_{F_1} = \mathbf{1}[i = k]$ under P_1 , and $E[\eta(\psi_j)\eta(\psi_k)] = \langle \psi_j, \psi_k \rangle_{F_2}$ under P_2 . Thus, under P_1 , $\eta_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, and under P_2 , $\eta_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_n)$, where $\mathbf{\Sigma}_n$ has elements $\langle \psi_j, \psi_k \rangle_{F_2}$. Since P_1 and P_2 are equivalent, $\mathbf{\Sigma}_n$ is positive definite for any n (cf. page 76 of IR78). Let $\mathbf{v}_{kn} \in \mathbb{C}^n$, $k = 1, \dots, n$ be a set of eigenvectors of $\mathbf{\Sigma}_n$ with associated eigenvalues $\sigma_{kn}^2 \leq \sigma_{k-1,n}^2$ for all $k = 2, \dots, n$, so that $\mathbf{v}_{kn}^* \eta_n \sim iid \mathcal{N}(0, 1)$ under P_1 , and $\mathbf{v}_{kn}^* \eta_n$ are independent $\mathcal{N}(0, \sigma_{kn}^2)$ under P_2 . Let d_n be the entropy distance between the distribution of η_n under P_1 and P_2 , that is the sum of the two corresponding Kullback-Leibler divergences. By a straightforward calculation (cf. equation (III.2.4) of IR78),

$$d_n = \frac{1}{2} \sum_{k=1}^n \left[\left(\frac{1}{\sigma_{kn}^2} - 1 \right) + (\sigma_{kn}^2 - 1) \right].$$

Define

$$D_n = \sum_{k=1}^n (1 - \sigma_{kn}^2)^2$$

and with $\lambda_k(\mathbf{B})$ denoting the k th largest eigenvalue of the Hermitian matrix $\mathbf{B}$, we have

$$\begin{aligned} D_n &= \sum_{k=1}^n (1 - \lambda_k(\Sigma_n))^2 \\ &= \sum_{k=1}^n \lambda_k((\mathbf{I}_n - \Sigma_n)^2) \\ &= \text{Tr}((\mathbf{I}_n - \Sigma_n)^2) \end{aligned}$$

so that

$$D_n = \sum_{j,k=1}^n |\langle \psi_j, \psi_k \rangle_{F_1} - \langle \psi_j, \psi_k \rangle_{F_2}|^2 = \sum_{j,k=1}^n |\langle \Delta \psi_j, \psi_k \rangle_{F_1}|^2. \quad (\text{B.21})$$

The following straightforward Lemma establishes a useful relationship between d_n and D_n .

Lemma B.2.3 *For any $0 < \delta < 1/4$, $D_n < \delta$ implies $d_n < \delta$.*

Proof. Note that $\sum_{k=1}^n (\sigma_k^2 - 1)^2 < 1/4$ implies $1/2 < \sigma_k^2 < 3/2$ for all $k = 1, \dots, n$, but for such σ_k^2 , $\frac{1}{2}[(\frac{1}{\sigma_k^2} - 1) + (\sigma_k^2 - 1)] \leq (\sigma_k^2 - 1)^2$, which implies the result. ■

Let $\bar{\Sigma}_n$ be the $n \times n$ Hermitian matrix with elements $\langle \psi_j, \psi_k \rangle_{\bar{F}}$. Then for any $\mathbf{v} = (v_1, \dots, v_n)' \in \mathbb{C}^n$, $\mathbf{v}^*(\bar{\Sigma}_n - \Sigma_n)\mathbf{v} = \sum_{j,k=1}^n v_k^* v_j \langle \varphi_{jn}, \varphi_{kn} \rangle_{(\bar{F}-F_2)} = \langle \sum_{j=1}^n v_j \varphi_{jn}, \sum_{k=1}^n v_k \varphi_{kn} \rangle_{(\bar{F}-F_2)} \geq 0$ from (B.17). Therefore, by Weyl's inequality, $\bar{\sigma}_{kn}^2 = \lambda_k(\bar{\Sigma}_n) \geq \lambda_k(\Sigma_n) + \lambda_n(\bar{\Sigma}_n - \Sigma_n) \geq \lambda_k(\Sigma_n) = \sigma_{kn}^2$ for all k , so that

$$\bar{D}_n = \sum_{j,k=1}^n |\langle \bar{\Delta} \psi_j, \psi_k \rangle_{F_1}|^2 = \sum_{k=1}^n (1 - \bar{\sigma}_{kn}^2)^2 \geq \sum_{k=1}^n \mathbf{1}[\sigma_{kn}^2 > 1](1 - \sigma_{kn}^2)^2.$$

By an analogous argument, also

$$\underline{D}_n = \sum_{j,k=1}^n |\langle \underline{\Delta} \psi_j, \psi_k \rangle_{F_1}|^2 \geq \sum_{k=1}^n \mathbf{1}[\sigma_{kn}^2 < 1](1 - \sigma_{kn}^2)^2$$

so that

$$D_n \leq \overline{D}_n + \underline{D}_n. \quad (\text{B.22})$$

Now let $\overline{\varphi}_k$ be a complete set of eigenvectors of the operator $\overline{\Delta}$, with associated eigenvalues $1 - \overline{\sigma}_k^2$, that is $\overline{\Delta}\overline{\varphi}_k = (1 - \overline{\sigma}_k^2)\overline{\varphi}_k$, and $\overline{\varphi}_k$ form an orthonormal basis in $L(\overline{F})$. Define $\underline{\varphi}_k$ and $\underline{\sigma}_k^2$ analogously relative to the operator $\underline{\Delta}$. Since $\overline{\Delta}$ and $\underline{\Delta}$ are Hilbert-Schmidt, $\sum_{k=1}^{\infty} (1 - \overline{\sigma}_k^2)^2 < \infty$ and $\sum_{k=1}^{\infty} (1 - \underline{\sigma}_k^2)^2 < \infty$, so that for any $\epsilon > 0$, there exists n_ϵ such that $\sum_{k=n_\epsilon}^{\infty} (1 - \overline{\sigma}_k^2)^2 < \epsilon/2$ and $\sum_{k=n_\epsilon}^{\infty} (1 - \underline{\sigma}_k^2)^2 < \epsilon/2$. Let $\overline{L}_\epsilon^0 \subset L(F_1)$ and $\underline{L}_\epsilon^0 \subset L(F_1)$ be the finite spaces spanned by $\overline{\varphi}_k$ and $\underline{\varphi}_k$, $k = 1, \dots, n_\epsilon$, respectively, and let L_ϵ^1 be the orthogonal complement of $L_\epsilon^0 = \overline{L}_\epsilon^0 \cup \underline{L}_\epsilon^0$ relative to $\langle \cdot, \cdot \rangle_{F_1}$, so that $L(F_1) = L_\epsilon^0 \cup L_\epsilon^1$. Note that L_ϵ^0 and L_ϵ^1 do not depend on f_2 . Let $\overline{L}_\epsilon^1$ be the space spanned by $\overline{\varphi}_k$, $k = n_\epsilon + 1, n_\epsilon + 2, \dots$. For any orthonormal sequence ψ_k in L_ϵ^1 , since $L_\epsilon^1 \subset \overline{L}_\epsilon^1$

$$\begin{aligned} \overline{D}_n &= \sum_{j,k=1}^n |\langle \overline{\Delta}\psi_j, \psi_k \rangle_{F_1}|^2 \\ &\leq \sum_{j=1}^n \|\overline{\Delta}\psi_j\|_{\overline{L}_\epsilon^1}^2 \\ &= \sum_{j=1}^n \sum_{k=n_\epsilon+1}^{\infty} |\langle \overline{\Delta}\psi_j, \overline{\varphi}_k \rangle_{F_1}|^2 \\ &= \sum_{k=n_\epsilon+1}^{\infty} \sum_{j=1}^n |\langle \overline{\Delta}\overline{\varphi}_k, \psi_j \rangle_{F_1}|^2 \\ &\leq \sum_{k=n_\epsilon+1}^{\infty} \|\overline{\Delta}\overline{\varphi}_k\|_{\overline{L}_\epsilon^1}^2 \\ &= \sum_{j,k=n_\epsilon+1}^{\infty} |\langle \overline{\Delta}\overline{\varphi}_j, \overline{\varphi}_k \rangle_{F_1}|^2 \\ &= \sum_{j,k=n_\epsilon+1}^{\infty} (1 - \overline{\sigma}_k^2)^2 |\langle \overline{\varphi}_j, \overline{\varphi}_k \rangle_{F_1}|^2 \\ &= \sum_{k=n_\epsilon+1}^{\infty} (1 - \overline{\sigma}_k^2)^2 < \epsilon/2 \end{aligned}$$

where the inequalities follow from Bessel's inequality. By the analogous argument, also $\underline{D}_n \leq \epsilon/2$. Thus, from (B.22), for any orthonormal sequence ψ_k in L_ϵ^1 ,

$$D_n \leq \epsilon. \quad (\text{B.23})$$

Now let ψ_k^ϵ , $k = 1, \dots, m_\epsilon \leq 2n_\epsilon$ be an orthonormal basis of L_ϵ^0 , and let ψ_k^ϵ , $k = m_\epsilon + 1, m_\epsilon + 2, \dots$ be an orthonormal basis of L_ϵ^1 , so that ψ_k^ϵ , $k = 1, 2, \dots$ is an orthonormal basis of $L(F_1)$. Note that the sequence ψ_k^ϵ does not depend on f_2 . Let $\mathfrak{U}_m^\epsilon$ be the σ -field generated by the Gaussian random variables $\eta(\psi_k^\epsilon)$ as defined in (B.20) for $k = 1, \dots, m$, $i = 1, 2$, and let $\mathfrak{U}^\epsilon$ be the σ -field generated by $\eta(\psi_k^\epsilon)$, $k = 1, 2, \dots$. Define

$$D_m^\epsilon = \sum_{j,k=1}^m |\langle \Delta \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2.$$

We have the following Lemma.

Lemma B.2.4 *For all $0 < \epsilon_0 < 1/2$, $\sup_m D_m^\epsilon < \epsilon_0^2$ implies that the total variation distance between P_1 and P_2 is smaller than ϵ_0 .*

Proof. As discussed on page 65 of IR78, the distribution on the σ -field $\mathfrak{U}^\epsilon$ equivalently characterizes the distribution of G_i relative to the σ -fields generated by the cylindric sets of the paths $G_i(\cdot)$ under P_i , $i = 1, 2$, so it suffices to show that

$$\sup_{\mathcal{A} \in \mathfrak{U}^\epsilon} |P_2(\mathcal{A}) - P_1(\mathcal{A})| \leq \epsilon_0. \quad (\text{B.24})$$

Let d_m^ϵ be the entropy distance between the distribution of $\eta(\psi_k^\epsilon)$, $k = 1, \dots, m$ under P_1 and P_2 . By Lemma B.2.3, $d_m^\epsilon \leq \epsilon_0^2$. Thus, by Pinsker's inequality

$$\sup_{\mathcal{A}_m \in \mathfrak{U}_m^\epsilon} |P_2(\mathcal{A}_m) - P_1(\mathcal{A}_m)| \leq \epsilon_0 \text{ for all } m. \quad (\text{B.25})$$

Now suppose (B.24) does not hold. Then there exists $\mathcal{A} \in \mathfrak{U}^\epsilon$ such that $P_2(\mathcal{A}) - P_1(\mathcal{A}) > \epsilon_0$. Construct a sequence of events $\mathcal{A}_m \in \mathfrak{U}_m^\epsilon$ such that $P_i(\mathcal{A}_m \ominus \mathcal{A}) \rightarrow 0$ for $i = 1, 2$ as on page 77 of IR78, where $\mathcal{A}_m \ominus \mathcal{A}$ is the symmetric difference $\mathcal{A}_m \ominus \mathcal{A} = (\mathcal{A}_m \cup \mathcal{A}) \setminus (\mathcal{A}_m \cap \mathcal{A})$. Then from $\mathcal{A} \subseteq \mathcal{A}_m \cup (\mathcal{A}_m \ominus \mathcal{A})$ and $\mathcal{A}_m \subseteq \mathcal{A} \cup (\mathcal{A}_m \ominus \mathcal{A})$, we have $|P_i(\mathcal{A}) - P_i(\mathcal{A}_m)| \leq P_i(\mathcal{A}_m \ominus \mathcal{A})$ for $i = 1, 2$. We thus obtain $P_2(\mathcal{A}_m) - P_1(\mathcal{A}_m) \rightarrow P_2(\mathcal{A}) - P_1(\mathcal{A}) > \epsilon_0$, contradicting (B.25), and the lemma is proved. ■

Given that the choice of $0 < \epsilon$ was arbitrary, in light of Lemma B.2.4 it suffices to show that for some CARMA implied f_2 satisfying (B.17), $\sup_m D_m^\epsilon < 2\epsilon$, say. Now for all $m > m_\epsilon$, from (B.21)

$$\begin{aligned} D_m^\epsilon &\leq \sum_{j,k=m_\epsilon+1}^m |\langle \Delta \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2 + 2 \sum_{j=1}^{m_\epsilon} \sum_{k=1}^m |\langle \Delta \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2 \\ &\leq \epsilon + 2 \sum_{j=1}^{m_\epsilon} \sum_{k=1}^m |\langle \Delta \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2 \end{aligned} \quad (\text{B.26})$$

where the second inequality follows from (B.23). Further

$$\begin{aligned} \sum_{k=1}^m |\langle \Delta \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2 &= \sum_{k=1}^m |\langle (\frac{f_2}{f_1} - 1) \psi_j^\epsilon, \psi_k^\epsilon \rangle_{F_1}|^2 \\ &\leq \langle (\frac{f_2}{f_1} - 1) \psi_j^\epsilon, (\frac{f_2}{f_1} - 1) \psi_j^\epsilon \rangle_{F_1}^2 \\ &= \int (\frac{f_2(\lambda)}{f_1(\lambda)} - 1)^2 |\psi_j^\epsilon(\lambda)|^2 f_1(\lambda) d\lambda \end{aligned}$$

where the inequality follows from Bessel's inequality by viewing $L(F_1)$ as a subspace of the Hilbert space of measurable square integrable functions with inner-product $\langle \cdot, \cdot \rangle_{F_1}$. Thus

$$\sup_m D_m^\epsilon \leq \epsilon + 2 \sum_{j=1}^{m_\epsilon} \int (\frac{f_2(\lambda)}{f_1(\lambda)} - 1)^2 |\psi_j^\epsilon(\lambda)|^2 f_1(\lambda) d\lambda. \quad (\text{B.27})$$

From equation (II.1.3) of IR78, every $\psi \in L(F_1)$ can be represented in the form

$$\psi(\lambda) = c_0 + (1 + i\lambda) \int_0^1 e^{i\lambda t} c(t) dt$$

for some real c_0 and some square integrable function $c : [0, 1] \mapsto \mathbb{R}$. Thus

$$\begin{aligned} |\psi(\lambda)| &\leq c_0 + \sqrt{1 + \lambda^2} \int_0^1 c(t) dt \\ &\leq c_0 + \sqrt{1 + \lambda^2} \sqrt{\int_0^1 c^2(t) dt} \end{aligned}$$

where the second inequality follows from Jensen's inequality, so that

$$\sup_{\lambda} |\psi(\lambda)|^2 f_1(\lambda) < \infty.$$

Thus, (B.27) implies that for some $M_\epsilon < \infty$ that does not depend on f_2 , for all f_2 satisfying (B.17),

$$\sup_m D_m^\epsilon \leq \epsilon + M_\epsilon \int \left(\frac{f_2(\lambda)}{f_1(\lambda)} - 1 \right)^2 d\lambda. \quad (\text{B.28})$$

It thus suffices to show that there exists a CARMA implied f_2 satisfying (B.17) that makes $\int_0^\infty (f_2(\lambda)/f_1(\lambda) - 1)^2 d\lambda$ arbitrarily small.

Let $h_1(\lambda) = f_1(\lambda)/f_0(\lambda) - 1$ and $h_2(\lambda) = f_2(\lambda)/f_0(\lambda) - 1$. Recalling the definition of δ_0 in (B.13), we have

$$\int_0^\infty \left(\frac{f_2(\lambda)}{f_1(\lambda)} - 1 \right)^2 d\lambda \leq \frac{1}{4} \delta_0^{-2} \int_0^\infty (h_1(\lambda) - h_2(\lambda))^2 d\lambda$$

and it suffices to show that for any $\epsilon_1 > 0$, there exists a CARMA implied h_2 such that $\int_0^\infty (h_1(\lambda) - h_2(\lambda))^2 d\lambda < 2\epsilon_1$.

For any $\tilde{h}_1(\lambda)$, from $(a - b)^2 \leq 2(a^2 + b^2)$,

$$\begin{aligned} \int_0^\infty (h_1(\lambda) - h_2(\lambda))^2 d\lambda &= \int_0^\infty \left(h_1(\lambda) - \tilde{h}_1(\lambda) - h_2(\lambda) + \tilde{h}_1(\lambda) \right)^2 d\lambda \\ &\leq 2 \int_0^\infty (h_1(\lambda) - \tilde{h}_1(\lambda))^2 d\lambda + 2 \int_0^\infty (h_2(\lambda) - \tilde{h}_1(\lambda))^2 d\lambda. \end{aligned}$$

By (B.14), $\int_0^\infty h_1(\lambda)^2 d\lambda < \infty$. Thus, there exists $K < \infty$ such that $\int_K^\infty h_1(\lambda)^2 d\lambda < \epsilon_1/2$. Let $\chi_K(\lambda) = 1$ for $\lambda \leq K$, $\chi_K(\lambda) = 0$ for $\lambda \geq K + 1$ and $\chi_K(\lambda) = K + 1 - \lambda$ otherwise, and define $\tilde{h}_1(\lambda) = \chi_K(\lambda)h_1(\lambda)$. Then $\int_0^\infty (h_1(\lambda) - \tilde{h}_1(\lambda))^2 d\lambda \leq \epsilon_1/2$, and since f_1 is continuous by standard Fourier arguments (see, for instance, Proposition 4.1 in Stein and Shakarchi (2005)), so is $\tilde{h}_1$. It thus suffices to show that there exists a CARMA implied h_2 that makes $\int_0^\infty (h_2(\lambda) - \tilde{h}_1(\lambda))^2 d\lambda$ smaller than $\epsilon_1/2$.

Now

$$\int_0^\infty (h_2(\lambda) - \tilde{h}_1(\lambda))^2 d\lambda = \int_0^\infty (1 + \lambda^2)^{-2} (\vartheta_2(\lambda) - \tilde{\vartheta}_1(\lambda))^2 d\lambda$$

with $\vartheta_2(\lambda) = (1 + \lambda^2)h_2(\lambda)$ and $\tilde{\vartheta}_1(\lambda) = (1 + \lambda^2)\tilde{h}_1(\lambda)$. Note that $\tilde{\vartheta}_1$ is continuous, and $\lim_{\lambda \rightarrow \infty} \tilde{\vartheta}_1(\lambda) = 0$. Thus, by Lemma B.2.1, for any $\delta > 0$, there exists an integer q and a rational function ϑ_2 of the form (B.12) such that

$$\sup_{\lambda} |\vartheta_2(\lambda) - \tilde{\vartheta}_1(\lambda)| < \delta. \quad (\text{B.29})$$

We have $\int_0^\infty (1 + \lambda^2)^{-2} (\vartheta_2(\lambda) - \tilde{\vartheta}_1(\lambda))^2 d\lambda \leq \delta^2 \int_0^\infty (1 + \lambda^2)^{-2} d\lambda$, which can be made arbitrarily small by choosing δ small. From the definitions of ϑ_2 and h_2 , we have

$$f_2(\lambda) = \frac{1}{1 + \lambda^2} + \frac{\vartheta_2(\lambda)}{(1 + \lambda^2)^2}$$

so the implied $f_2(\lambda)$ is a rational function in λ^2 of degree $p = q + 2$ in the denominator $p - 1$ in the numerator.

Furthermore, for all $\delta < \delta_0$, we have uniformly in λ ,

$$\begin{aligned}
f_2(\lambda) &\leq \frac{1}{1+\lambda^2} + \frac{\tilde{\vartheta}_1(\lambda) + \delta_0}{(1+\lambda^2)^2} \\
&= \frac{1}{1+\lambda^2} + \frac{\chi_K(\lambda)((1+\lambda^2)^2 f_1(\lambda) - (1+\lambda^2)) + \delta_0}{(1+\lambda^2)^2} \\
&= \chi_K(\lambda) f_1(\lambda) + (1 - \chi_K(\lambda)) f_0(\lambda) + \frac{\delta_0}{(1+\lambda^2)^2} \leq \bar{f}(\lambda)
\end{aligned}$$

and similarly,

$$\begin{aligned}
f_2(\lambda) &\geq \frac{1}{1+\lambda^2} + \frac{\tilde{\vartheta}_1(\lambda) - \delta_0}{(1+\lambda^2)^2} \\
&= \chi_K(\lambda) f_1(\lambda) + (1 - \chi_K(\lambda)) f_0(\lambda) - \frac{\delta_0}{(1+\lambda^2)^2} \geq \underline{f}(\lambda)
\end{aligned}$$

so that f_2 satisfies (B.17). In particular, since $\underline{f}(\lambda) > 0$ for all λ , the numerator of $f_2(\lambda)$ is a positive rational function, so by Lemma B.2.2, f_2 has the form of the spectral density of a CARMA($p, p-1$) process.

Appendix C

Appendix for Chapter 3

C.1 More Tables of α_T and τ_T under GLTU(p)

In this section, I list the tables for the 5th, 50th and 95th percentiles of α_T and τ_T , for $p = 3, 4$, respectively. A few findings stand out from these tables to supplement those stated in Section 3.5. First, regardless of the choice of country and how the persistence properties are modelled, aggregate inflations always display high degree of persistence. Second, as p gets larger, the persistence measure α_T becomes numerically unstable, which creates a seemingly heavy tail in the posterior distribution of α_T for some variables after downcoding. But this is not the case for τ_T , at least for the five p 's I considered. Of course, many further analysis can be done to provide more economic insights out of these tables. But I do not pursue that in this chapter.

Table C.1: The 5th, 50th and 95th percentiles of α_T for $p = 3$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	0.00	0.78	0.00	0.00	0.86	0.48	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.71	0.96	0.65	0.93	0.98	0.98	0.98	0.93	0.94	0.96	0.88	0.97	0.94	0.80	0.72	0.97	0.94
	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.99	0.97	0.99	0.98
Int. rate spr.	0.88	0.22	0.81	0.84	0.87	0.00	0.83	0.42	0.84	0.81	0.84	0.86	0.00	0.84	0.84	0.85	0.81
	0.94	0.93	0.93	0.94	0.96	0.95	0.92	0.94	0.94	0.91	0.94	0.93	0.91	0.94	0.93	0.93	0.92
	0.96	0.97	0.97	0.97	0.98	0.98	0.96	0.97	0.97	0.96	0.96	0.96	0.98	0.96	0.96	0.96	0.97
Inflation	0.92	0.92	0.93	0.93	0.95	0.96	0.92	0.92	0.93	0.94	0.90	0.93	0.92	0.92	0.90	0.91	0.92
	0.95	0.95	0.95	0.95	0.97	0.98	0.96	0.95	0.95	0.97	0.95	0.96	0.95	0.95	0.95	0.95	0.96
	0.97	0.97	0.97	0.97	0.98	0.99	0.98	0.97	0.99	0.99	0.97	0.98	0.97	0.97	0.97	0.97	0.98
Real ex. rate	N.A.	0.00	0.90	0.00	0.00	0.82	0.00	0.33	0.95	0.79	0.00	0.76	0.62	0.39	0.44	0.78	0.51
	N.A.	0.87	0.96	0.91	0.91	0.93	0.91	0.93	0.97	0.93	0.90	0.93	0.91	0.94	0.93	0.94	0.93
	N.A.	0.97	0.98	0.97	0.97	0.97	0.97	0.97	0.99	0.96	0.98	0.96	0.97	0.98	0.97	0.97	0.97
Real housing	0.48	0.91	0.70	0.55	0.00	0.00	0.00	0.00	0.39	0.20	0.60	0.00	0.34	0.80	0.70	0.88	0.87
	0.97	0.97	0.96	0.97	0.88	0.00	0.96	0.95	0.96	0.90	0.93	0.97	0.96	0.97	0.97	0.95	0.96
	0.99	0.99	0.98	0.99	0.98	0.59	0.99	0.98	0.99	0.97	0.97	0.99	0.99	0.99	0.99	0.98	0.98
Debt-GDP	0.42	0.00	0.48	0.65	0.45	0.20	0.80	0.00	0.36	0.00	0.33	0.67	0.85	0.38	0.39	0.55	0.00
	0.98	0.93	0.97	0.97	0.97	0.98	0.95	0.92	0.96	0.96	0.93	0.97	0.96	0.96	0.97	0.97	0.96
	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99
Real I-GDP	0.82	0.90	0.92	0.89	0.78	0.95	0.82	0.92	0.89	0.79	0.87	0.47	0.89	0.59	0.00	0.90	0.54
	0.93	0.96	0.97	0.95	0.97	0.98	0.95	0.97	0.96	0.96	0.96	0.95	0.95	0.94	0.89	0.96	0.94
	0.96	0.98	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.98
Loan-GDP	0.19	0.74	0.68	0.65	0.78	0.63	0.89	0.71	0.86	0.84	0.00	0.86	0.78	0.74	0.93	0.64	0.94
	0.96	0.98	0.98	0.98	0.97	0.98	0.97	0.97	0.94	0.98	0.91	0.96	0.99	0.99	0.98	0.96	0.97
	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.98	0.99	1.00	0.99	0.98	0.99
Real C-I	0.86	0.71	0.82	0.35	0.91	0.00	0.00	0.86	0.00	0.78	0.00	0.89	0.00	0.93	0.00	0.81	0.52
	0.96	0.94	0.94	0.97	0.95	0.94	0.98	0.93	0.96	0.97	0.92	0.94	0.94	0.96	0.89	0.96	0.97
	0.98	0.97	0.97	0.99	0.97	0.98	0.99	0.96	0.99	0.99	0.97	0.96	0.98	0.98	0.98	0.98	0.99

Notes: The column header is 3-letter ISO country code. α_T is downcoded by 0.

Table C.2: The 5th, 50th and 95th percentiles of τ_T for $p = 3$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	3.1	9.3	5.3	7.6	15.0	15.1	14.4	6.8	7.5	10.7	3.5	8.8	6.0	6.3	3.8	9.6	6.5
	4.0	25.3	8.4	13.2	27.5	28.5	29.4	12.1	13.2	16.9	6.8	19.4	14.3	9.4	6.2	20.0	15.1
	35.9	46.7	29.7	39.8	47.0	47.2	47.5	40.7	40.7	43.1	40.3	45.6	44.7	35.7	27.6	45.4	44.0
Int. rate spr.	4.4	3.2	3.5	5.0	6.9	3.2	4.3	3.2	5.1	4.0	3.7	5.6	3.7	4.4	4.9	5.1	4.0
	32.2	10.4	19.7	24.1	33.8	26.3	22.1	20.1	21.8	17.3	25.6	26.5	11.2	27.2	24.0	25.6	19.8
	48.2	46.5	46.6	47.3	48.7	47.6	46.5	47.3	46.7	46.0	47.3	47.6	45.7	47.6	46.9	47.2	46.0
Inflation	14.6	16.5	12.9	20.6	21.2	21.9	11.5	14.0	5.3	25.3	6.5	14.7	13.7	8.5	12.8	19.1	11.0
	36.9	37.3	37.9	40.0	39.8	40.1	37.5	37.2	37.5	43.4	34.7	37.6	37.5	37.2	35.3	38.8	35.1
	48.8	49.1	49.1	49.4	49.1	49.4	49.0	49.0	49.0	49.7	48.7	49.0	49.1	48.8	48.7	49.1	48.8
Real ex. rate	N.A.	3.1	11.6	3.2	3.5	4.9	3.7	4.0	24.7	3.5	3.1	3.4	3.4	5.3	3.5	3.4	4.6
	N.A.	5.3	29.5	9.3	13.2	20.1	9.7	15.6	39.0	21.8	5.3	21.6	18.2	17.9	16.8	25.9	16.3
	N.A.	42.6	47.6	45.0	44.5	46.2	44.2	45.7	49.1	46.5	46.2	46.3	45.6	45.7	46.2	48.2	45.7
Real housing	11.8	14.6	8.2	8.7	5.1	4.3	9.0	7.5	7.9	3.2	3.8	9.7	6.8	16.8	10.6	10.4	13.5
	27.2	30.9	25.0	23.7	8.8	31.3	20.6	17.6	17.3	7.8	13.8	16.5	15.4	33.1	24.0	29.4	28.4
	47.0	47.3	46.9	46.7	34.4	38.8	46.2	45.3	45.1	44.0	45.4	42.9	45.1	48.1	46.7	47.5	47.5
Debt-GDP	11.9	8.2	10.9	9.6	9.8	14.8	4.6	4.9	8.5	9.4	5.0	12.2	7.2	7.2	9.1	11.5	7.6
	24.5	13.5	22.5	21.9	21.0	27.3	14.3	9.4	18.8	19.1	12.2	23.1	21.3	13.5	19.7	27.5	15.1
	46.7	41.2	46.2	46.6	45.5	47.3	46.6	40.0	45.1	44.5	43.4	46.0	46.6	43.8	45.4	47.3	43.5
Real I-GDP	4.0	11.0	12.8	13.1	20.3	19.6	7.6	16.3	9.4	9.6	8.8	5.6	13.8	4.7	3.5	8.2	4.3
	20.7	28.2	30.4	30.6	37.0	37.0	24.0	33.1	28.4	27.8	23.5	23.8	31.6	16.2	7.9	33.2	14.6
	46.5	47.5	47.8	48.1	48.7	48.8	47.0	48.1	47.6	47.5	46.9	47.9	47.9	45.9	41.7	48.4	45.9
Loan-GDP	6.9	16.2	13.1	13.4	10.3	16.6	11.8	10.4	8.5	15.0	3.8	10.0	18.4	15.0	15.0	6.9	17.6
	15.7	32.6	30.3	28.7	23.4	34.1	28.2	21.6	29.1	31.9	9.6	27.2	35.0	32.2	30.6	18.4	32.2
	45.1	47.8	47.9	47.9	46.3	48.5	47.3	46.2	47.6	48.4	43.2	47.3	48.5	48.2	47.6	46.2	47.8
Real C-I	18.1	4.1	4.4	22.5	19.3	5.9	21.5	8.8	8.4	13.1	3.1	15.1	6.5	21.0	3.8	8.2	9.1
	34.5	21.6	21.5	37.8	37.2	12.9	33.8	26.5	18.1	32.2	7.9	33.7	14.6	38.4	7.1	20.6	25.7
	48.5	46.9	46.7	49.1	49.0	43.4	48.0	47.3	44.8	48.2	46.3	48.5	43.5	48.8	41.5	46.0	46.9

Note: The column header is 3-letter ISO country code.

Table C.3: The 5th, 50th and 95th percentiles of α_T for $p = 4$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	0.00	0.04	0.00	0.00	0.71	0.63	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00
	0.54	0.96	0.51	0.95	0.98	0.98	0.98	0.95	0.95	0.97	0.90	0.96	0.94	0.87	0.77	0.97	0.95
	0.97	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.97	0.99	0.98
Int. rate spr.	0.88	0.08	0.64	0.81	0.87	0.74	0.74	0.51	0.71	0.79	0.75	0.89	0.00	0.83	0.84	0.85	0.74
	0.94	0.94	0.94	0.94	0.96	0.95	0.93	0.94	0.95	0.93	0.94	0.94	0.95	0.94	0.93	0.94	0.93
	0.97	0.97	0.97	0.97	0.98	0.98	0.97	0.97	0.98	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97
Inflation	0.92	0.92	0.93	0.92	0.93	0.96	0.90	0.92	0.92	0.94	0.91	0.91	0.93	0.92	0.91	0.92	0.87
	0.96	0.95	0.97	0.96	0.97	0.98	0.96	0.95	0.95	0.98	0.96	0.96	0.96	0.95	0.95	0.95	0.96
	0.98	0.97	1.00	0.98	0.98	0.99	0.99	0.97	0.98	0.99	0.98	0.98	0.98	0.97	0.98	0.98	0.98
Real ex. rate	N.A.	0.00	0.69	0.00	0.00	0.43	0.00	0.00	0.95	0.44	0.00	0.61	0.66	0.13	0.48	0.76	0.36
	N.A.	0.90	0.96	0.92	0.93	0.94	0.93	0.94	0.98	0.93	0.93	0.93	0.93	0.95	0.94	0.95	0.93
	N.A.	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.99	0.97	0.98	0.97	0.97	0.98	0.97	0.97	0.98
Real housing	0.55	0.51	0.63	0.23	0.00	0.00	0.00	0.00	0.56	0.00	0.00	0.00	0.75	0.68	0.50	0.80	0.81
	0.97	0.97	0.97	0.97	0.90	0.00	0.97	0.96	0.97	0.92	0.94	0.97	0.97	0.98	0.97	0.96	0.97
	0.99	0.99	0.99	0.99	0.98	0.61	0.99	0.99	0.99	0.97	0.98	0.99	0.99	0.99	0.99	0.98	0.99
Debt-GDP	0.67	0.00	0.71	0.65	0.13	0.50	0.76	0.00	0.27	0.19	0.00	0.73	0.68	0.41	0.17	0.65	0.00
	0.98	0.95	0.97	0.97	0.97	0.98	0.96	0.92	0.96	0.97	0.94	0.98	0.96	0.96	0.98	0.97	0.96
	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99
Real I-GDP	0.77	0.73	0.91	0.88	0.27	0.94	0.44	0.93	0.87	0.64	0.66	0.35	0.89	0.42	0.00	0.90	0.36
	0.93	0.97	0.97	0.95	0.97	0.98	0.96	0.97	0.96	0.97	0.96	0.96	0.95	0.95	0.91	0.96	0.95
	0.97	0.98	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.98
Loan-GDP	0.21	0.78	0.74	0.90	0.38	0.71	0.61	0.26	0.85	0.88	0.00	0.83	0.88	0.78	0.82	0.63	0.95
	0.96	0.98	0.98	0.99	0.97	0.98	0.97	0.97	0.94	0.98	0.92	0.96	0.99	0.99	0.98	0.96	0.98
	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.98	0.98	0.99	1.00	0.99	0.98	0.99
Real C-I	0.93	0.01	0.61	0.90	0.91	0.00	0.73	0.86	0.23	0.92	0.03	0.89	0.00	0.93	0.00	0.23	0.48
	0.96	0.95	0.94	0.98	0.95	0.95	0.98	0.94	0.97	0.98	0.94	0.94	0.95	0.96	0.91	0.96	0.97
	0.98	0.98	0.98	0.99	0.97	0.99	0.99	0.96	0.99	0.99	0.97	0.96	0.99	0.98	0.98	0.98	0.99

Notes: The column header is 3-letter ISO country code. α_T is downcoded by 0.

Table C.4: The 5th, 50th and 95th percentiles of τ_T for $p = 4$.

Series	USA	GBR	CHE	SWE	ESP	POR	NOR	NLD	JPN	ITA	DEU	FRA	FIN	DNK	CAN	BEL	AUS
Real GDP	3.1	8.5	5.1	7.5	14.4	14.8	15.0	6.9	7.5	10.0	3.8	8.1	5.9	6.2	4.0	9.3	6.5
	4.0	25.9	7.8	13.5	27.3	27.0	30.6	12.3	14.1	17.8	7.3	17.6	14.3	9.1	6.6	20.4	14.6
	35.9	47.0	27.8	39.7	47.0	47.2	47.9	40.4	41.2	42.9	41.3	44.4	42.9	37.2	28.2	45.4	42.9
Int. rate spr.	4.9	3.2	3.5	4.9	7.6	3.4	4.0	3.2	5.4	4.3	3.7	8.4	4.0	4.3	4.6	5.0	4.1
	32.3	13.2	21.0	24.3	32.9	29.1	22.0	24.5	22.1	19.0	25.3	28.8	16.3	28.7	23.5	26.8	21.6
	48.7	46.3	47.0	47.5	48.4	48.4	46.6	47.5	46.7	46.0	47.5	47.6	46.2	47.9	46.7	47.8	45.4
Inflation	6.3	8.5	4.6	17.5	12.8	22.8	14.4	17.1	6.0	23.1	5.1	8.5	9.4	7.9	8.2	19.8	6.9
	34.7	37.0	31.0	38.7	37.9	41.0	39.1	38.8	33.7	42.3	32.9	36.2	35.3	36.2	34.8	37.3	32.8
	48.8	49.1	48.4	49.0	49.1	49.4	49.2	49.1	48.7	49.5	48.7	48.8	48.8	48.8	49.0	49.0	48.4
Real ex. rate	N.A.	3.1	11.3	3.2	3.5	5.1	4.0	4.3	23.5	3.4	3.1	3.5	3.5	5.6	3.7	3.5	4.7
	N.A.	7.1	29.4	10.6	14.7	21.3	12.5	17.1	39.1	21.9	7.1	23.4	20.3	18.8	20.9	28.1	17.1
	N.A.	44.2	48.1	44.4	44.5	46.5	42.8	45.7	49.2	47.0	46.2	46.3	45.3	45.9	46.6	47.9	45.0
Real housing	10.6	14.0	7.6	7.9	5.3	4.0	8.5	7.2	7.8	3.4	4.3	8.8	6.5	16.0	10.0	9.4	13.4
	27.0	29.5	25.4	21.3	8.8	30.3	21.2	18.5	18.7	8.7	15.0	15.7	18.8	31.8	23.8	28.5	28.1
	47.2	47.3	46.7	46.6	35.7	39.4	46.3	45.7	45.4	43.2	45.3	42.6	45.7	48.2	47.0	47.8	47.5
Debt-GDP	10.9	7.5	10.4	9.1	9.3	14.0	4.7	5.0	8.4	9.0	5.1	11.5	7.2	6.9	8.4	10.6	7.3
	24.5	12.9	23.4	21.0	20.7	28.5	14.8	9.1	19.4	19.6	12.8	23.2	22.6	12.2	18.4	27.5	14.8
	46.2	41.0	46.3	46.2	45.6	47.3	45.9	39.7	44.7	44.4	43.5	45.6	47.2	44.4	45.4	47.5	44.1
Real I-GDP	4.1	10.3	12.8	12.9	20.4	19.0	7.4	16.2	9.4	9.3	9.0	6.0	13.1	5.1	3.8	7.1	4.7
	20.4	29.7	30.3	30.6	37.3	36.6	24.3	31.6	27.6	28.5	25.0	25.3	31.0	17.8	8.4	32.0	16.5
	46.0	47.2	47.9	48.1	48.8	48.8	46.9	48.1	47.5	47.3	47.2	47.3	47.8	45.6	42.3	48.4	46.3
Loan-GDP	6.8	15.1	11.6	12.8	9.8	15.7	10.9	9.7	8.7	14.4	4.1	10.1	17.5	14.7	14.8	6.9	17.9
	14.8	31.5	29.1	29.7	23.2	33.2	28.4	22.0	28.8	33.2	10.1	27.6	34.5	30.3	30.4	20.3	32.2
	44.1	48.2	47.5	47.9	46.3	48.4	47.6	46.2	47.8	48.1	42.6	47.5	48.5	47.6	47.6	46.7	48.2
Real C-I	17.9	4.4	4.7	22.2	19.8	6.0	21.0	9.1	7.9	14.7	3.2	15.6	6.8	19.3	4.1	8.1	9.6
	35.6	25.4	22.6	37.9	38.1	12.8	34.4	27.8	18.4	33.8	19.8	33.5	15.1	37.9	7.6	21.6	32.0
	48.7	47.0	47.0	48.7	49.1	43.8	48.4	47.3	45.6	48.4	46.6	48.5	43.7	49.1	42.8	45.7	48.1

Note: The column header is 3-letter ISO country code.

References

- ANDREWS, D. W. K. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–858.
- ANDREWS, D. W. K., AND H.-Y. CHEN (1994): “Approximately Median-Unbiased Estimation of Autoregressive Models,” *Journal of Business and Economic Statistics*, 12, 187–204.
- ANDREWS, D. W. K., AND J. C. MONAHAN (1992): “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60(4), 953–966.
- ARMSTRONG, T. B., AND M. KOLESÁR (2018a): “Optimal Inference in a Class of Regression Models,” *Econometrica*, 86(2), 655–683.
- (2018b): “Simple and Honest Confidence Intervals in Nonparametric Regression,” *arXiv:1606.01200 [math, stat]*, arXiv: 1606.01200.
- ATCHADÉ, Y. F., AND M. D. CATTANEO (2011): “Limit Theorems for Quadratic Forms of Markov Chains,” *arXiv:1108.2743 [math]*, arXiv: 1108.2743.
- BAKIROV, N. K. (1996): “Comparison theorems for distribution functions of quadratic forms of gaussian vectors,” *Theory of Probability & Its Applications*, 40(2), 340–348.
- BAKIROV, N. K., AND G. J. SZKELY (2006): “Students t-test for Gaussian scale mixtures,” *Journal of Mathematical Sciences*, 139(3), 6497–6505.
- BELIAKOV, G. (2006): “Interpolation of Lipschitz functions,” *Journal of Computational and Applied Mathematics*, 196(1), 20–44.
- BERGSTROM, A. R. (1985): “The estimation of parameters in non-stationary higher-order continuous-time dynamic models,” *Econometric Theory*, 1, 369–385.
- BOBKOSKI, M. J. (1983): “Hypothesis Testing in Nonstationary Time Series,” Unpublished Ph.D. thesis, Department of Statistics, University of Wisconsin.
- BRAND, L. (1964): “The Companion Matrix and Its Properties,” *The American Mathematical Monthly*, 71, 629–634.

- BROCKWELL, P. J. (2001): “Continuous-Time ARMA Processes,” in *Stochastic Processes: Theory and Methods*, ed. by C. R. Rao, and D. N. Shanbhag, vol. 19 of *Handbook of Statistics*, pp. 249–276. Elsevier, Amsterdam.
- BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*, Springer series in Statistics. Springer, New York.
- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81(1), 27–60.
- CAVANAGH, C. L. (1985): “Roots Local to Unity,” Working Paper, Harvard University.
- CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): “Inference in Models With Nearly Integrated Regressors,” *Econometric Theory*, 11, 1131–1147.
- CHAN, N. H., AND C. Z. WEI (1987): “Asymptotic Inference for Nearly Nonstationary AR(1) Processes,” *The Annals of Statistics*, 15(3), 1050–1063.
- DEN HAAN, W. J., AND A. T. LEVIN (1994): “Vector Autoregressive Covariance Matrix Estimation,” Manuscript, Board of Governors of the Federal Reserve.
- (1997): “A Practitioner’s Guide to Robust Covariance Matrix Estimation,” in *Handbook of Statistics*, ed. by G. Maddala, and C. Rao, vol. 15, pp. 299–342. Elsevier.
- DINI, D. H., AND D. P. MANDIC (2012): “Class of Widely Linear Complex Kalman Filters,” *IEEE Transactions on Neural Networks and Learning Systems*, 23, 775–786.
- DOU, L., AND U. K. MÜLLER (2019): “Generalized Local-to-Unity Models,” Working Paper, Princeton University.
- ELLIOTT, G. (1998): “The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots,” *Econometrica*, 66, 149–158.
- ELLIOTT, G. (1999): “Efficient Tests for a Unit Root When the Initial Observation is Drawn From Its Unconditional Distribution,” *International Economic Review*, 40(3), 767–784.
- ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2015): “Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis,” *Econometrica*, 83(2), 771–811.
- ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64, 813–836.
- ELLIOTT, G., AND J. H. STOCK (1994): “Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown,” *Econometric Theory*, 10, 672–700.

- GONÇALVES, S., AND T. J. VOGELSANG (2011): “Block Bootstrap HAC Robust Tests: The Sophistication of the Naive Bootstrap,” *Econometric Theory*, 27(4), 745–791.
- IBRAGIMOV, I., AND Y. ROZANOV (1978): *Gaussian Random Processes*. Springer-Verlag, New York.
- IBRAGIMOV, R., AND U. K. MÜLLER (2010): “t -Statistic Based Correlation and Heterogeneity Robust Inference,” *Journal of Business & Economic Statistics*, 28(4), 453–468.
- JANSSON, M. (2004): “The Error in Rejection Probability of Simple Autocorrelation Robust Tests,” *Econometrica*, 72(3), 937–946.
- JANSSON, M., AND M. J. MOREIRA (2006): “Optimal Inference in Regression Models With Nearly Integrated Regressors,” *Econometrica*, 74, 681–714.
- JONES, R. H. (1981): “Fitting a Continuous Time Autoregression to Discrete Data,” in *Applied Time Series Analysis II*, ed. by D. F. Findley, pp. 651–682. Academic Press, New York.
- JONES, R. H., AND K. M. ACKERSON (1990): “Serial correlation in unequally spaced longitudinal data,” *Biometrika*, 77, 721–732.
- JORDÀ, O. (2005): “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95(1), 161–182.
- JORDÀ, O., M. SCHULARICK, AND A. M. TAYLOR (2017): “Macrofinancial History and the New Business Cycle Facts,” in *NBER Macroeconomic Annual 2016*, ed. by M. Eichenbaum, and J. A. Park, vol. 31, pp. 213–263. University of Chicago Press, Chicago.
- KARIYA, T. (1980): “Locally Robust Tests for Serial Correlation in Least Squares Regression,” *The Annals of Statistics*, 8(5), 1065–1070.
- KIEFER, N. M., AND T. J. VOGELSANG (2002): “Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation,” *Econometrica*, 70(5), 2093–2095.
- (2005): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *Econometric Theory*, 21(6), 1130–1164.
- KIEFER, N. M., T. J. VOGELSANG, AND H. BUNZEL (2000): “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68(3), 695–714.
- KING, M. L. (1980): “Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression,” *The Annals of Statistics*, 8(6), 1265–1271.
- KING, M. L. (1987): “Towards a theory of point optimal testing,” *Econometric Reviews*, 6(2), 169–218.

- KOIJEN, R. S., AND S. VAN NIEUWERBURGH (2011): “Predictability of Returns and Cash Flows,” *Annual Review of Financial Economics*, 3(1), 467–491.
- LAZARUS, E., D. J. LEWIS, AND J. H. STOCK (2017): “The Size-Power Tradeoff in HAR Inference,” Discussion paper, Harvard University Mimeo.
- LAZARUS, E., D. J. LEWIS, J. H. STOCK, AND M. W. WATSON (2018): “HAR Inference: Recommendations for Practice,” *Journal of Business & Economic Statistics*, 36(4), 541–559.
- LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Springer, New York, 3rd edn.
- LOTHIAN, J. R., AND M. K. TAYLOR (1996): “Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries,” *Journal of Political Economy*, 104, 488–509.
- MIKUSHEVA, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75, 1411–1452.
- (2012): “One-Dimensional Inference in Autoregressive Models with the Potential Presence of a Unit Root,” *Econometrica*, 80, 173–212.
- MOON, H. R., AND P. C. B. PHILLIPS (2000): “Estimation of Autoregressive Roots Near Unity Using Panel Data,” *Econometric Theory*, 16, 927–997.
- MOON, S., AND C. VELASCO (2014): “On the Properties of Regression Tests of Stock Return Predictability Using Dividend-Price Ratios,” *Journal of Financial Econometrics*, 12, 151–173.
- MÜLLER, U. K. (2004): “A Theory of Robust Long-Run Variance Estimation,” Working paper, Princeton University.
- (2007): “A Theory of Robust Long-Run Variance Estimation,” *Journal of Econometrics*, 141(2), 1331–1352.
- (2014): “HAC Corrections for Strongly Autocorrelated Time Series,” *Journal of Business and Economic Statistics*, 32, 311–322.
- MÜLLER, U. K., AND M. W. WATSON (2008): “Testing Models of Low-Frequency Variability,” *Econometrica*, 76, 979–1016.
- MÜLLER, U. K., AND M. W. WATSON (2016): “Measuring Uncertainty about Long-Run Predictions,” *The Review of Economic Studies*, 83(4), 1711–1740.
- MURRAY, C. J., AND D. H. PAPELL (2002): “The Purchasing Power Parity Persistence Paradigm,” *Journal of International Economics*, 56, 1–19.
- MURRAY, C. J., AND D. H. PAPELL (2005): “The purchasing power parity puzzle is worse than you think,” *Empirical Economics*, 30(3), 783–790.

- NELSON, C. R., AND C. R. PLOSSER (1982): "Trends and random walks in macroeconomic time series," *Journal of Monetary Economics*, 10(2), 139–162.
- NEWKEY, W. K., AND K. D. WEST (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.
- PHAM-DINH, T. (1977): "Estimation of parameters of a continuous time Gaussian stationary process with rational spectral density," *Biometrika*, 64, 385–399.
- PHILLIPS, A. W. (1959): "The estimation of parameters in systems of stochastic differential equations," *Biometrika*, 46, 67–76.
- PHILLIPS, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.
- (1988): "Regression Theory for Near-Integrated Time Series," *Econometrica*, 56, 1021–1043.
- (1998): "Impulse Responses and Forecast Error Variance Asymptotics in Nonstationary VARs," *Journal of Econometrics*, 83, 21–56.
- PHILLIPS, P. C. B. (2005): "HAC Estimation by Automated Regression," *Econometric Theory*, 21(1), 116–142.
- PHILLIPS, P. C. B., Y. SUN, AND S. JIN (2006): "Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels without Truncation," *International Economic Review*, 47(3), 837–894.
- (2007): "Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," *Journal of Statistical Planning and Inference*, 137(3), 985–1023.
- PTSCHER, B. M. (2002): "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters," *Econometrica*, 70(3), 1035–1065.
- RAPACH, D., AND G. ZHOU (2013): "Forecasting Stock Returns," in *Handbook of Economic Forecasting*, vol. 2, pp. 328–383. Elsevier.
- ROBINSON, P. (2002): "Long-Memory Time Series," in *Time Series with Long Memory*, ed. by P. Robinson, pp. 4–32. Oxford University Press, Oxford.
- ROBINSON, P. M. (2005): "Robust Covariance Matrix Estimation: HAC Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21(1), 171–180.
- ROGOFF, K. (1996): "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647–668.

- ROSSI, B. (2005): “Confidence Intervals for Half-Life Deviations from Purchasing Power Parity,” *Journal of Business and Economic Statistics*, 23, 432–442.
- STEIN, E., AND R. SHAKARCHI (2005): *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton Lectures in Analysis. Princeton University Press, Princeton.
- STOCK, J. H. (1991): “Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series,” *Journal of Monetary Economics*, 28, 453–459.
- (1996): “VAR, Error Corrections and Pretest Forecasts at Long Horizons,” *Oxford Bulletin of Economics and Statistics*, 58, 685–701.
- STOCK, J. H., AND M. W. WATSON (1996): “Confidence Sets in Regressions with Highly Serially Correlated Regressors,” Working Paper, Harvard University.
- STOCK, J. H., AND M. W. WATSON (1999): “Business cycle fluctuations in us macroeconomic time series,” in *Handbook of Macroeconomics*, vol. 1a, pp. 3–64. Elsevier.
- SUN, Y. (2011): “Robust trend inference with series variance estimator and testing-optimal smoothing parameter,” *Journal of Econometrics*, 164(2), 345–366.
- (2013): “A Heteroskedasticity and Autocorrelation robust F test Using an Orthonormal Series Variance Estimator,” *The Econometrics Journal*, 16(1), 1–26.
- (2014): “Fixed-Smoothing Asymptotics in a Two-Step Generalized Method of Moments Framework,” *Econometrica*, 82(6), 2327–2370.
- SUN, Y., AND D. M. KAPLAN (2012): “Fixed-Smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator,” Working paper, University of California, San Diego.
- SUN, Y., P. C. B. PHILLIPS, AND S. JIN (2008): “Optimal Bandwidth Selection in Heteroskedasticity Autocorrelation Robust Testing,” *Econometrica*, 76(1), 175–194.
- TAYLOR, M. P. (2003): “Purchasing power parity,” *Review of International Economics*, 11(3), 436–452.
- VELASCO, C., AND P. M. ROBINSON (2001): “Edgeworth expansions for spectral density estimates and studentized sample mean,” *Econometric Theory*, 17(3), 497–539.