ESSAYS IN HONOR OF JOON Y. PARK

Econometric Theory

Edited by Yoosoon Chang, Sokbae Lee and J. Isaac Miller

ADVANCES IN ECONOMETRICS

VOLUME 45A

ESSAYS IN HONOR OF JOON Y. PARK

ADVANCES IN ECONOMETRICS

Series editors: Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano, Eric Hillebrand, Daniel L. Millimet and Rodney Strachan

Recent Volumes:

Volume 25	Nonparametric Econometric Methods – Edited by Qi Li and Jeffrey S. Racine
Volume 26	Maximum Simulated Likelihood Methods and Applications – Edited by R. Carter Hill and William Greene
Volume 27A	Missing Data Methods: Cross-sectional Methods and Applications – Edited by David M. Drukker
Volume 27B	Missing Data Methods: Time-series Methods and Applications – Edited by David M. Drukker
Volume 28	DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments – Edited by Nathan Balke, Fabio Canova, Fabio Milani and Mark Wynne
Volume 29	Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill
Volume 30	30th Anniversary Edition – Edited by Dek Terrell and Daniel Millmet
Volume 31	Structural Econometric Models – Edited by Eugene Choo and Matthew Shum
Volume 32	VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy
Volume 33	Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park
Volume 34	Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier
Volume 35	Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman
Volume 36	Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee
Volume 37	Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage and R. Kelley Pace
Volume 38	Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano
Volume 39	The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi
Volume 40A	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part A – Edited by Ivan Jeliazkov and Justin L. Tobias
Volume 40B	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part B – Edited by Ivan Jeliazkov and Justin L. Tobias
Volume 41	Essays in Honor of Cheng Hsiao – Edited by Tong Li, M. Hashem Pesaran and Dek Terrell
Volume 42	The Econometrics of Networks – Edited by Áureo de Paula, Elie Tamer and Marcel-Cristian Voia
Volume 43A	Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling – Edited by Alexander Chudik, Cheng Hsiao and Allan Timmermann
Volume 43B	Essays in Honor of M. Hashem Pesaran: Panel Modeling, Micro Aplications, and Econometric Methodology – Edited by Alexander Chudik, Cheng Hsiao and Allan Timmermann
Volume 44A	Essays in Honour of Fabio Canova Part A – Edited by Juan J. Dolado, Luca Gambetti and Christian Matthes
Volume 44B	Essays in Honour of Fabio Canova Part B – Edited by Juan J. Dolado, Luca Gambetti and Christian Matthes

ADVANCES IN ECONOMETRICS VOLUME 45A

ESSAYS IN HONOR OF JOON Y. PARK: ECONOMETRIC THEORY

EDITED BY

YOOSOON CHANG
Indiana University, USA

SOKBAE LEE
Columbia University, USA

And

J. ISAAC MILLER
University of Missouri, USA

emerald PUBLISHING

United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2023

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83753-209-4 (Print)
ISBN: 978-1-83753-208-7 (Online)
ISBN: 978-1-83753-210-0 (Epub)

ISSN: 0731-9053 (Series)

CONTENTS

LIST OF CONTRIBUTORS

Chafik Bouhaddioui	United Arab Emirates University, UAE
Jean-Marie Dufour	McGill University, Canada
Jiti Gao	Monash University, Australia
Nikolay Gospodinov	Federal Reserve Bank of Atlanta, USA
Uwe Hassler	Goethe-Universität Frankfurt, Germany
Javier Hidalgo	London School of Economics, UK
Kohtaro Hitomi	Kyoto Institute of Technology, Japan
Mehdi Hosseinkouchack	EBS Universität, Germany
Hsein Kew	Monash University, Australia
Jiwoong Kim	University of South Florida, USA
Kun Ho Kim	Concordia University, Canada
Yun-Yeong Kim	Dankook University, South Korea
Hira L. Koul	Michigan State University, USA
Heejun Lee	Brown University, USA
Jungyoon Lee	University of London, Royal Holloway, UK
Han-Ying Liang	Tongji University, China
Yingqian Lin	Shanghai University of Finance and Economics, China
Alex Maynard	University of Guelph, Canada
Keiji Nagai	Yokohama National University, Japan
Yoshihiko Nishiyama	Kyoto University, Japan
Elena Pesavento	Emory University, USA
Peter C. B. Phillips	Yale University, USA; University of Auckland, New Zealand; Singapore Management University, Singapore; and University of Southampton, UK
Myung Hwan Seo	Seoul National University, South Korea
Yu Shen	Tongji University, China
Yixiao Sun	University of California, San Diego, USA
Masaya Takano	McGill University, Canada

Junfan Tao	Kyoto University, Japan
Yundong Tu	Peking University, China
Qiying Wang	The University of Sydney, Australia
Xiaohu Wang	Fudan University, China
Weilin Xiao	Zhejiang University, China
Jun Yu	Singapore Management University, Singapore
Ying Zhou	Monash University, Australia

INTRODUCTION

Volume 45 of *Advances in Econometrics* honors Professor Joon Y. Park, who has made numerous and substantive contributions to the field of econometrics over a career spanning four decades since the 1980s and counting. Volume 45 consists of 28 chapters and is in fact split between two volumes with the first focusing on econometric theory and the second focusing on econometric applications. These papers have been contributed by Joon's friends, colleagues, coauthors, former students, and even his dissertation advisor, Professor Peter C. B. Phillips, and the volume is edited by his wife and most frequent collaborator, Professor Yoosoon Chang, and two of his former students.

In the typical fashion of *Advances in Econometrics*, the papers were to be submitted in early 2021 after a conference in Joon's honor in April 2020, which would have nearly coincided with his 65th birthday. Of course, the COVID-19 pandemic forced much of the world into lockdown in April 2020, so plans changed. Papers were still submitted in 2021, but the conference was delayed and, as of this writing, is scheduled for September 29–30, 2023, in Bloomington, Indiana, which Joon and Yoosoon have called home for nearly 15 years.

We introduce the 13 chapters of the first volume, which are loosely grouped into three sections that are closely related to Professor Park's contribution to the theoretical analysis of time series and particularly related to the research of the first two or so decades of his career.

After graduating from Yale under the supervision of Professor Phillips, Joon's early work in the late 1980s and 1990s focused on nonstationary time series and particularly on cointegration and common stochastic trends, where some of his most highly cited contributions were made. These include foundational work on regressions with nonstationary series (Park & Phillips, 1988, 1989); the variable addition test for cointegration (Park, 1990), which remains one of the most highly cited papers in *Advances in Econometrics*; and perhaps his most well-known contribution to the field on canonical cointegrating regressions (CCR)[1] (Park, 1992).

Shifting his research, Park published a series of papers in the late 1990s and 2000s on nonlinear transformations of unit root processes, which introduce nontrivial obstacles in the form of nonstandard rates of convergence and limiting distributions. One could say that his work helped to redefine *nonstandard* in the sense that up to this point, nonstandard typically meant rate-T with limiting Dickey–Fuller type distributions. The rates of convergence in these papers generally involve powers of the sample size other than ½ or 1, and the limits usually include nonlinear functions of stochastic integrals and/or Brownian local times. Park's most well-cited contributions to the study of nonlinear transformation of nonstationary series are Park and Phillips (1999, 2001), but his work on nonlinearity has also spilled over into time-varying coefficients (Park & Hahn, 1999), instrumental variables (Chang et al., 2004), functional coefficients (Cai et al., 2009), and other areas.

Following the themes of nonstationarity and nonlinearity, the papers in this volume are grouped as follows: (I) nonstationarity, unit roots, and fractional noise; (II) nonlinearity; and (III) inference and prediction using models with trending series.

Part I: Nonstationarity, Unit Roots, and Fractional Noise

A contribution by Peter C. B. Phillips, not only Joon's dissertation advisor but also longtime editor of *Econometric Theory*, appropriately opens the volume on econometric theory and the section on nonstationarity, unit roots, and fractional noise. Specifically, his article "Discrete Fourier Transforms of Fractional Processes With Econometric Applications" presents an exact representation of the discrete Fourier transform in terms of the component data, which he finds to be particularly useful for analyzing the asymptotic behavior of the periodogram when the memory parameter exceeds the threshold for stationarity. He shows that smoothed periodogram spectral estimates remain consistent for frequencies away from the origin as long as the memory parameter is strictly less than unity.

Also studying fractional noise, Xiaohu Wang, Weilin Xiao, and Jun Yu contribute the article "Asymptotic Properties of the Least Squares Estimator in Local to Unity Processes With Fractional Gaussian Noises." They derive the asymptotic properties of the autoregressive parameter in local to unity processes with errors generated as fractional Gaussian noise with the Hurst parameter over the interval (0,1). The rates of convergence are standard rate-T over the upper half of this interval, but nonstandard and dependent of the Hurst parameter over the lower half. They derive limiting distributions over this interval that are new to the literature except at 1/2.

Critical to ascertaining stationarity or lack thereof are unit root tests. In their contribution, "Powerful Self-normalizing Tests for Stationarity Against the Alternative of a Unit Root," Uwe Hassler and Mehdi Hosseinkouchack introduce a new and powerful tool to address this well-known problem. Specifically, they propose a family of tests for stationarity against a local unit root that builds on the Karhunen–Loève expansions of the limiting CUSUM process under the null hypothesis and a local alternative. They find that the proposed tests are more powerful than the classic KPSS test.

Also on the topic of testing for unit roots, Kohtaro Hitomi, Keiji Nagai, Yoshihiko Nishiyama, and Junfan Tao contribute "A Sequential Test for a Unit Root in Monitoring a p-th Order Autoregressive Process." They study unit root tests for autoregressive processes of order p under sequential sampling schemes using stopping times based on the observed Fisher information. They derive the joint limit of the test statistics and the stopping time under the null and local alternatives, which are nonstandard.

Part II: Nonlinearity

As we mentioned, both cointegration and functional coefficients are areas in which Professor Park has made contributions to the literature. Han-Ying Liang, Yu Shen, and Qiying Wang contribute to the volume and this literature with "Functional-coefficient Cointegrating Regression With Endogeneity." As the title suggests, they explore nonparametric estimation of cointegrating regression models with functional coefficients and where the structural equation errors are serially dependent and the regressor is endogenous. In this context, they show the self-normalized local kernel and local linear estimators to be asymptotically normal.

In "A Specification Test Based on Convolution-type Distribution Function Estimates for Non-linear Autoregressive Processes," Kun Ho Kim, Hira L. Koul, and Jiwoong Kim develop a test for a parametric specification of the autoregressive function of a given stationary autoregressive time series. Their test is based on the integrated square difference between the empirical distribution function estimate and a convolution-type distribution function estimate of the stationary distribution function obtained from the autoregressive residuals.

Yingqian Lin and Yundong Tu contribute "Transformation Models With Cointegrated and Deterministically Trending Regressors," which contains important and interesting extensions of the statistical foundation for the nonlinear cointegrated models pioneered by Park and his coauthors. For a general transformation model with a time trend, stationary regressors, and unit root regressors, they estimate the transformation parameter and other model parameters by minimizing the concentrated loss function, and they obtain the asymptotic distributions of the proposed estimators.

The threshold model has been frequently used to model the nonlinearity of time series. Park and Shintani (2016) examine testing issues surrounding threshold effects and unit roots. In "Minimax Risk in Estimating Kink Threshold and Testing Continuity," Javier Hidalgo, Heejun Lee, Jungyoon Lee, and Myung Hwan Seo derive a risk lower bound in estimating the threshold parameter without knowing whether the threshold regression model is continuous or not. They show that the bound goes to zero as the sample size grows only at the cube root rate. Motivated by this finding, they develop a continuity test for the threshold regression model and a bootstrap to compute its p-values.

Part III: Inference and Prediction Using Models With Trending Series

Articles in the final section of this volume deal with models containing stochastic and/or deterministic trends, as do many of Professor Park's papers, from his earliest work on cointegration (Park & Phillips, 1988, 1989) and his widely cited CCR paper (Park, 1992) through his more recent work, such as that on estimating stochastic trends in state-space models (Chang et al., 2009).

In the first of these, "Semiparametric Independence Tests Between Two Infinite-order Cointegrated Series," Chafik Bouhaddioui, Jean-Marie Dufour, and Masaya Takano propose a semiparametric approach for testing independence between two cointegrated vector autoregressive series of infinite order. The residual-based tests allow for computational simplicity and weak assumptions on the form of the underlying process. The authors derive the asymptotic distributions of the test statistics under the null hypothesis and establish consistency of the tests against fixed alternatives of serial cross-correlation of unknown form.

Nikolay Gospodinov, Alex Maynard, and Elena Pesavento contribute "Inference in Conditional Vector Error Correction Models With a Small Signal-to-Noise Ratio," in which they study vector error correction models when the error correction term is characterized simultaneously by high persistence (near-unit-root behavior) and very small (near zero) variance. The importance of these features lies in the fact that conventional cointegration tests may fail to detect cointegration. The authors develop asymptotic theory for the parameter estimators for unconditional and conditional vector error correction models with these features.

Yixiao Sun, in his contribution entitled "Some Extensions of Asymptotic *F* and *t* Theory in Nonstationary Regressions," extends the asymptotic theory for *F*- and *t*-tests to linear regression models where the regressors could contain deterministic trends, unit-root processes, and near-unit-root processes. The tests themselves are implemented in the usual ways, but approximations to the limiting distributions are more accurate than the more commonly used chi-squared and normal approximations.

The last two contributions focus on predictive models with nonstationary series. Ying Zhou, Hsein Kew, and Jiti Gao contribute "Non-stationary Parametric Single-index Predictive Models: Simulation and Empirical Studies." Their model is designed to handle a wide variety of nonlinear relationships between the regressand and a single-index component containing either the cointegrated predictors or the non-cointegrated predictors. They introduce a new estimation procedure and investigate its finite-sample properties.

We opened the volume with a contribution from Joon's advisor, so it seems appropriate to close the volume with a contribution from one of his many students. In "Best Linear Prediction in Cointegrated Systems," Yun-Yeong Kim introduces the best linear predictor with the asymptotic minimum mean squared forecasting error among linear predictors of variables in cointegrated systems with unknown error specification. He suggests a switching predictor that automatically selects the random walk or cointegration model according to the size of the estimated autocorrelation coefficient estimated from the residuals.

We hope you enjoy reading "Essays in Honor of Joon Y. Park: Econometric Theory" and learning about the advances in econometrics made by the authors as much as we have!

NOTE

1. In case you have ever wondered… yes, Joon has always been a fan of the music of Credence Clearwater Revival, also abbreviated as CCR!

ADDITIONAL REFERENCES

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, *148*, 101–113.

Chang, Y., Miller, J. I., & Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications. *Journal of Econometrics*, *150*, 231–247.

Chang, Y., Park, J. Y., & Phillips, P. C. B. (2004). Nonlinear instrumental variable estimation of an autoregression. *Journal of Econometrics*, *118*, 219–246.

Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition. In G. F. Rhodes & T. B. Fomby (Eds.), *Advances in econometrics* (pp. 107–133). JAI Press.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica*, *60*, 199–143.

Park, J. Y., & Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, *15*, 664–703.

Park, J. Y., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*, 468–497.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*, 95–131.

Park, J. Y., & Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, *15*, 269–298.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, *69*, 117–161.

Park, J. Y., & Shintani, M. (2016). Testing for a unit root against transitional autoregressive models. *International Economic Review*, *57*(2), 635–664.

PART I

NONSTATIONARITY, UNIT ROOTS, AND FRACTIONAL NOISE

This page intentionally left blank

CHAPTER 1

DISCRETE FOURIER TRANSFORMS OF FRACTIONAL PROCESSES WITH ECONOMETRIC APPLICATIONS*

Peter C. B. Phillips[a,b,c,d]

[a] *Yale University, New Haven, CT, USA*
[b] *University of Auckland, Auckland, New Zealand*
[c] *Singapore Management University, Singapore*
[d] *University of Southampton, Southampton, UK*

ABSTRACT

The discrete Fourier transform (dft) of a fractional process is studied. An exact representation of the dft is given in terms of the component data, leading to the frequency domain form of the model for a fractional process.

* This paper has a long history. It was presented at the Cowles Foundation Conference "New Developments in Time Series Econometrics," October 23–24, 1999, and to the New Zealand Econometric Study Group, July 1999, Auckland, New Zealand. The paper originated in some notes on fractional processes in the nonstationary case that were written in May 1998 and circulated at Yale. It was completed in 1999 while the author was living on Waiheke Island and visiting the University of Auckland. That version (Phillips, 1999a) was conditionally accepted subject to revision by the *Annals of Statistics* but the revise-by-date expired and the revision was never submitted. The paper was revised and updated in 2021, leaving its main results essentially unchanged. Thanks go to the Editor and referees of the *Annals of Statistics*, Chang Sik Kim, Katsumi Shimotsu, Yixiao Sun, and Karim Abadir for comments on the original paper. Katsumi and Yixiao made further helpful comments on the revision, which are greatly appreciated. Thanks are due to the NSF for research support under Grant Nos. SBR 94-22922, 97-30295, and SES 18-50860.

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 3–71

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A001

This representation is particularly useful in analyzing the asymptotic behavior of the dft and periodogram in the nonstationary case when the memory parameter $\mathrm{d} \geq \frac{1}{2}$. *Various asymptotic approximations are established including some new hypergeometric function representations that are of independent interest. It is shown that smoothed periodogram spectral estimates remain consistent for frequencies away from the origin in the nonstationary case provided the memory parameter* $\mathrm{d} < 1$. *When* $\mathrm{d} = 1$, *the spectral estimates are inconsistent and converge weakly to random variates. Applications of the theory to log periodogram regression and local Whittle estimation of the memory parameter are discussed and some modified versions of these procedures are suggested for nonstationary cases.*

Keywords: Discrete Fourier transform; fractional Brownian motion; fractional integration; log periodogram regression; nonstationarity; operator decomposition; semiparametric estimation; Whittle likelihood;

JEL classification: C22

1. INTRODUCTION

Studies of nonstationary time series over the last four decades have produced a vast body of knowledge that has transformed the conduct of empirical research in economics. The impact of this research is now manifest in empirical work throughout the social and business sciences. A catalyst supporting these developments was the widespread recognition that real world processes in society, economics, and politics are influenced in fundamental ways by advances in technology, firm investments, and individual human decision-making. These processes are rarely, if ever, stationary. Inevitably they evolve in uncertain ways over time, reflecting the arrival of new shocks to the system, some of which have persistent effects. Recognizing this reality led to an understanding that methods of data analysis need to account for the fact that the way in which memory is carried in the data differs in a fundamental manner among stationary, near-stationary, and nonstationary processes.

Acknowledgment of the importance of this distinction is evident in early researches of statisticians and economists at the turn of the twentieth century (Hooker, 1901; Pearson & Elderton, 1923; Yule, 1926) on nonsense correlations[1] and the work of the mathematician Bachelier (1900) on speculative prices, which introduced the notion of a stochastic process. Methods began to emerge later that provided probabilistic underpinnings and foundations for statistical inference with data that demonstrated long range memory or dependence (Granger & Joyeux, 1980; Hosking, 1981; Hurst, 1951; 1956; Mandelbrot & Van Ness, 1968) and various types of random wandering behavior over time. In economics in the 1980s, advances in the use of function space limit theory were made that enabled the full trajectory features of nonstationary data to be reflected in regression asymptotics, leading to new understanding of such regressions, including both cointegrating and spurious regressions, and new methods of testing and inference for analyzing nonstationary data (Durlauf & Phillips, 1988; Phillips, 1986b; 1987; 1988; Phillips & Durlauf, 1986).

Joon Park played a big part in these developments, starting with his doctoral dissertation research and early research at Yale (Park & Phillips, 1988; 1989) and a sustained series of subsequent works that have helped to push out the envelope of econometric methodology for linear, nonlinear, and continuous time methods of analysis with nonstationary data. Many of these works have been jointly conducted with the present author in a longstanding collaboration that has been as pleasurable and special an academic fellowship as much as it has enriched this field of research.

My contribution to this symposium of works honoring Joon Park relates to his research on nonstationary processes and focuses on some of the defining properties of long range-dependent time series. The present work has a history reaching back more than two decades and it is hoped that a good part of its value is retained amidst the considerable body of work that has emerged since the original version of the paper (Phillips, 1999a) was written. The first contribution of this chapter is to provide an exact representation of the dft of a fractional process, which enables asymptotic analysis of its behavior and various functionals such as the periodogram in the nonstationary case when the memory parameter $d \geq \frac{1}{2}$. The methods reveal that smoothed periodogram spectral estimates remain consistent for frequencies away from the origin in the nonstationary case provided the memory parameter $d < 1$. When $d = 1$, the spectral estimates are inconsistent and converge weakly to random variates. Some useful applications of this theory are given for log periodogram regression and local Whittle estimation of the memory parameter in nonstationary cases. For an advanced textbook treatment of long memory processes, readers are referred to Surgailis et al. (2012).

The plan of this chapter is as follows. Various preliminaries are given in the following Section 2. Some useful new decompositions and representations in the frequency domain are developed in Section 3 that extend related decompositions in the time domain. Section 4 develops asymptotic approximations for dfts involving special functions that help to simplify representations and enable development of limit theory for dfts of fractional processes in nonstationary cases. These results extend earlier work on the limit theory of dfts of stationary processes to the fractional case. For higher levels of dependence, when $d = 1$, the leakage from the zero frequency becomes dominant and affects the limit theory at all frequencies, so that dfts are spatially correlated across frequency asymptotically, quite unlike the stationary case. Section 5 provides some applications of the results to spectral estimation and to semiparametric estimation of the memory parameter. Particular attention in the latter case is given to log periodogram regression and local Whittle estimation. Some modified versions of these procedures are suggested which conveniently extend their range of applicability to the nonstationary case. Final remarks on long memory and autoregressive approaches to nonstationarity close out Section 5. Proofs and technical results are in the Appendix in Section 6. A notational summary is given at the end of this chapter in Section 7.

A final word of introduction. While our focus is on the case where $d \in \left(\frac{1}{2}, 1\right]$, the methods introduced here are applicable when $d > 1$, and in modified form when $|d| < \frac{1}{2}$. A particularly useful approach is to combine the exact representation (3.7) that applies when $d = 1$ with results for fractional d to produce valid representations for the $d > 1$ case. The remarks and results in paragraphs 3.6–3.8 indicate some of these possibilities.

2. PRELIMINARIES

We consider the fractional process X_t generated by the model

$$(1-L)^d X_t = u_t, \quad t = 0,1,... \tag{2.1}$$

Our interest is primarily in the case where X_t is nonstationary and $d \geq \frac{1}{2}$, so in (2.1) we work from a given initial date $t = 0$, set $u_j = 0$ for all $j \leq 0$, and assume that $u_t(t > 0)$ is stationary with 0 mean and continuous spectrum $f_u(\lambda) > 0$. This formulation corresponds to a Type II fractional process (Davidson & Hashimzade, 2009; Marinucci & Robinson, 1999). Expanding the binomial in (2.1) gives the form

$$\sum_{k=0}^{t} \frac{(-d)_k}{k!} X_{t-k} = u_t, \tag{2.2}$$

where

$$(d)_k = \frac{\Gamma(d+k)}{\Gamma(d)} = (d)(d+1)\ldots(d+k-1)$$

is Pochhammer's symbol for the forward factorial function and $\Gamma(\cdot)$ is the gamma function. When d is a positive integer, the series in (2.2) terminates, giving the usual formulae for the model (2.1) in terms of differences and higher order differences of X_t. An alternate form for X_t is obtained by inversion of (2.1), giving

$$X_t = (1-L)^{-d} u_t = \sum_{k=0}^{t} \frac{(d)_k}{k!} u_{t-k}. \tag{2.3}$$

Throughout this chapter it will be convenient to assume that the stationary component u_t in (2.1) is a linear process of the form

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} j\left|c_j\right| < \infty, \quad C(1) \neq 0, \tag{2.4}$$

for all t and with $\varepsilon_t = iid\left(0,\sigma^2\right)$ with finite fourth moments. The summability condition in (2.4) is satisfied by a wide class of parametric and nonparametric models for u_t, enables the use of the techniques in Phillips and Solo (1992), and ensures that partial sums of u_t satisfy a functional central limit theorem, which will be needed later.

Under (2.4), the spectrum is $f_u(\lambda)=\frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty}c_je^{ij\lambda}\right|^2$ and $f_u(0)=\frac{\sigma^2}{2\pi}C(1)^2>0.$[2] In view of (2.1), it is natural to define

$$f_x(\lambda)=|1-e^{i\lambda}|^{-2d}f_u(\lambda). \tag{2.5}$$

The function $f_x(\lambda)$ gives the spectrum of X_t when it exists and X_t is stationary (i.e., for $|d|<\frac{1}{2}$ and under infinite past initialization of X_t in (2.3)) and is the analog of the spectrum in the nonstationary case when $d\geq\frac{1}{2}$ even though it is not integrable. In that case, Solo (1992) gave a formal justification of $f_x(\lambda)$ as a spectrum in terms of the limit of the expectation of the periodogram. Taking logarithms of (2.5) produces the equation

$$\ln(f_x(\lambda))=-2d\ln(|1-e^{i\lambda}|)+\ln(f_u(\lambda)), \tag{2.6}$$

which motivates a linear log periodogram regression for the estimation of d, in which $f_x(\lambda)$ is replaced by periodogram ordinates $I_x(\lambda)$ evaluated at the fundamental frequencies $\lambda_s=\frac{2\pi s}{n}, s=0,1,\ldots,n-1$. Here, $I_a(\lambda_s)=w_a(\lambda_s)w_a(\lambda_s)^*$, $w_a(\lambda_s)$ is the dft, $w_a(\lambda_s)=\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n}a_te^{it\lambda_s}$ of a time series a_t, and w^* is the complex conjugate of w. With this substitution (2.6) becomes

$$\ln\left(I_x(\lambda_s)\right)=-2d\ln\left|1-e^{i\lambda_s}\right|+\ln\left(f_u(\lambda_s)\right)+U(\lambda_s), \tag{2.7}$$

where $U(\lambda_s)=\ln\left[I_x(\lambda_s)/f_x(\lambda_s)\right]$. By virtue of the continuity of f_u, $f_u(\lambda_s)$ is effectively constant for frequencies in a shrinking band around the origin, suggesting a linear least squares regression of $\ln\left(I_x(\lambda_s)\right)$ on $\ln\left|1-e^{i\lambda_s}\right|$ over frequencies $s=1,\ldots,m$ (with m a truncation number). The method has undoubted appeal, is easy to perform in practice and has been commonly employed in applications. However, (2.6) is a moment condition, not a data generating mechanism, and the analysis of this regression estimator is complicated by the difficulty of characterizing the asymptotic behavior of the dft $w_x(\lambda_s)$, which is the central element in determining the properties of the regression residual $U(\lambda_s)$ in (2.7).

An important contribution by Künsch (1986) showed that, for fractional processes like (2.1), $w_x(\lambda_s)$ has quite different statistical properties from the corresponding dft, $w_u(\lambda_s)$, of the stationary process u_t for frequencies in the immediate neighborhood of the origin. In particular, for $\lambda_s=\frac{2\pi s}{n}\to 0$, with s fixed as $n\to\infty$ the dft ordinates are asymptotically correlated, not uncorrelated. Analyses by Robinson (1995b) and Hurvich et al. (1998) for Gaussian u_t have provided an asymptotic theory in the stationary case, thereby placing log periodogram regression on a rigorous footing. More recent work has dealt with nonstationary cases where $d \geq \frac{1}{2}$ (Kim & Phillips, 2006; Phillips, 2007; Velasco, 1999). Another semiparametric estimation procedure, suggested by Künsch (1987), is the Gaussian estimator which maximizes a local version of the Whittle likelihood, which is known to have a smaller variance than log periodogram regression in the stationary case (Robinson, 1995a). This estimator also relies on the behavior of $w_x(\lambda_s)$ for frequencies in the vicinity of the origin. More recent work on Whittle estimation has focused on nonstationary cases where $d \geq \frac{1}{2}$ (Abadir et al., 2007; Phillips, 2014; Phillips & Shimotsu, 2004; Shao, 2010; Shimotsu & Phillips, 2005; 2006; Velasco & Robinson, 2000) and cases of noise contaminated data (Sun & Phillips, 2003) such as in the estimation of the Fisher equation (Sun & Phillips, 2004).

This chapter provides new methods for studying the asymptotic behavior of $w_x(\lambda_s)$ for nonstationary values of d. The approach relies on an exact representation of $w_x(\lambda_s)$ in terms of the dft $w_u(\lambda_s)$ and certain residual components. This representation aids in the analysis of the properties of $w_x(\lambda_s)$ and, thereby, in the study of log periodogram regression and local Whittle estimation. The representation also provides a frequency domain version of the data generating mechanism (2.1) above. As such, it is useful in motivating some alternative approaches to inference about d that are proposed here and which have been explored in subsequent work that has appeared since the first version of this chapter circulated in 1999.

3. FREQUENCY DOMAIN DECOMPOSITIONS

It is convenient to manipulate the operator $(1-L)^d$ in (2.1), with its polynomial expansion (2.2), in a form that more readily accommodates dfts. This can be done algebraically, as in Phillips and Solo (1992), by expanding the polynomial operator about its value at the complex exponential $e^{i\lambda}$, leading to the following decomposition.

3.1. Lemma

Define the fractional operator expansion $D_n(L;d) = \sum_{k=0}^{n} \frac{(-d)_k}{k!} L^k$. *Then*

$$D_n(L;d) = D_n\left(e^{i\lambda};d\right) + \tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)\left(e^{-i\lambda}L - 1\right), \tag{3.1}$$

where $\tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right) = \sum_{p=0}^{n-1} \tilde{d}_{\lambda p} e^{-ip\lambda} L^p$ *and* $\tilde{d}_{\lambda p} = \sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda}$.

The representation (3.1) is an immediate consequence of formula (32) in Phillips and Solo (1992) and can be obtained by straightforward algebraic manipulation. No summability conditions are required here for its validity since it is a finite sum. However, the value of d does affect the order of the terms in this expansion and, consequently, the order of magnitude of these terms when $n \to \infty$, a fact that does affect subsequent theory. Additionally, when λ depends on n, the order of these terms is affected and this too needs to be accounted for in the asymptotic theory. Much of this chapter is devoted to this accounting to assist in characterizing the limit behavior of the dft $w_x(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} X_t e^{it\lambda}$.

Using the operator (3.1), we may write the model (2.1) in the following form for all $t \le n$

$$\begin{aligned} u_t &= D_n(L;d) X_t \\ &= D_n(e^{i\lambda};d) X_t + \tilde{D}_{n\lambda}(e^{-i\lambda}L;d)(e^{-i\lambda}L - 1) X_t \end{aligned} \tag{3.2}$$

Taking dfts of the left and right sides of (3.2) now yields an exact expression for $w_x(\lambda)$ in terms of $w_u(\lambda)$. The result is stated as follows.

3.2. Theorem

$$w_u(\lambda) = w_x(\lambda) D_n(e^{i\lambda};d) + \frac{1}{\sqrt{2\pi n}}\left(\tilde{X}_{\lambda 0}(d) - e^{in\lambda}\tilde{X}_{\lambda n}(d)\right) \tag{3.3}$$

where $D_n(e^{i\lambda};d) = \sum_{k=0}^{n} \frac{(-d)_k}{k!} e^{ik\lambda}$,

$$\tilde{X}_{\lambda t}(d) = \tilde{D}_{n\lambda}(e^{-i\lambda}L;d) X_t = \sum_{p=0}^{n-1} \tilde{d}_{\lambda p} e^{-ip\lambda} X_{t-p},$$

and

$$\tilde{D}_{n\lambda}(e^{-i\lambda}L;d) = \sum_{p=0}^{n-1} \tilde{d}_{\lambda p} e^{-ip\lambda} L^p, \quad \text{with } \tilde{d}_{\lambda p} = \sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda}. \tag{3.4}$$

3.3. Remark

Equation (3.3) provides an exact representation of $w_x(\lambda)$ in terms of $w_u(\lambda)$ and a residual component involving $n^{-\frac{1}{2}}\tilde{X}_{\lambda n}(d)$. Explicitly,

$$w_x(\lambda) = D_n(e^{i\lambda};d)^{-1} w_u(\lambda) - \frac{1}{\sqrt{2\pi n}} D_n(e^{i\lambda};d)^{-1}\left(\tilde{X}_{\lambda 0}(d) - e^{in\lambda}\tilde{X}_{\lambda n}(d)\right). \tag{3.5}$$

In fact, (3.3) or (3.5) may be interpreted as a frequency domain version of the original model (2.1). In terms of periodogram ordinates, we have the corresponding equation

$$\begin{aligned} I_x(\lambda_s) = |w_x(\lambda_s)|^2 &= \left| D_n(e^{i\lambda_s};d)^{-1}\left[w_u(\lambda_s) - \frac{1}{\sqrt{2\pi n}}\left(\tilde{X}_{\lambda 0}(d) - e^{in\lambda_s}\tilde{X}_{\lambda_s n}(d)\right)\right]\right|^2 \\ &= \left|D_n(e^{i\lambda_s};d)\right|^{-2}\left[I_u(\lambda_s) - 2\mathrm{Re}\left\{\frac{1}{\sqrt{2\pi n}}\left(\tilde{X}_{\lambda_s 0}(d) - \tilde{X}_{\lambda_s n}(d)\right) w_u(\lambda_s)^*\right\}\right. \\ &\quad \left. + \frac{1}{2\pi n}\left|\left(\tilde{X}_{\lambda_s 0}(d) - \tilde{X}_{\lambda_s n}(d)\right)\right|^2\right], \end{aligned} \tag{3.6}$$

which may be interpreted as the data generating mechanism for the ordinates $I_x(\lambda_s)$ that are used in a log periodogram regression. Equation (3.6) reveals the model that is implicit in (2.7). To the extent that $\left|D_n\left(e^{i\lambda_s};d\right)\right|^{-2}$ can be replaced by $\left|1-e^{i\lambda_s}\right|^{-2d}$ and the component $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s n}(d)$ is small enough to be neglected, (3.6) and (2.5) might seem to suggest that $U(\lambda_s)=\ln\left[I_x(\lambda_s)/f_x(\lambda_s)\right]$ will behave like the corresponding functional, $\log\left[I_u(\lambda_s)/f_u(\lambda_s)\right]$, of the errors in (2.1). However, as will become apparent in our analysis, the residual component $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s n}(d)$ in (3.5) and (3.6) cannot be neglected, in general, and its importance grows as d increases.

3.4. Remark

When $d=1$, the forward factorial $(-d)_k=0$ for all $k>1$, so that series involving these coefficients terminate at $k=1$. In this case $D_n\left(e^{i\lambda};1\right)=\left(1-e^{i\lambda}\right)$, $\tilde{d}_{\lambda 0}=-e^{i\lambda}$, $\tilde{X}_{\lambda 0}(1)=-e^{i\lambda}X_0$, and $\tilde{X}_{\lambda n}(1)=-e^{i\lambda}X_n$. Equation (3.3) then reduces to the simple form

$$w_u(\lambda)=\left(1-e^{i\lambda}\right)w_x(\lambda)+\frac{e^{i\lambda}}{\sqrt{2\pi n}}\left(e^{in\lambda}X_n-X_0\right), \tag{3.7}$$

an expression obtained by the author in earlier work and used in Corbae et al. (2002, Lemma B). In this case, it is apparent that $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s n}(d)=e^{i\lambda_s}n^{-\frac{1}{2}}X_n=O_p(1)$ for all λ_s. Thus, in the unit root case, the residual correction term $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s n}(d)$ definitely matters, plays a role in the asymptotic behavior of $w_x(\lambda_s)$ at all frequencies and thereby affects the asymptotic theory of estimators of d like those arising from log periodogram regression and local Whittle estimation. Indeed, in those cases the author has shown in other works (Phillips, 2007; Phillips & Shimotsu, 2004) that these estimators have limiting mixed normal distributions rather than normal distributions when $d=1$.

3.5. Remark

When $u_t=0$ for $t\leq 0$, in (2.1), it follows that $X_t=0$ for $t\leq 0$ and hence $\tilde{X}_{\lambda 0}(d)=0$. In this event, expression (3.3) becomes

$$\begin{aligned} w_u(\lambda)&=w_x(\lambda)D_n\left(e^{i\lambda};d\right)-\frac{e^{in\lambda}}{\sqrt{2\pi n}}\tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)X_n \\ &=w_x(\lambda)D_n\left(e^{i\lambda};d\right)-\frac{e^{in\lambda}}{\sqrt{2\pi n}}\tilde{X}_{\lambda n}(d), \end{aligned} \tag{3.8}$$

or, in the unit root case,

$$w_u(\lambda)=\left(1-e^{i\lambda}\right)w_x(\lambda)+\frac{e^{i\lambda}}{\sqrt{2\pi n}}e^{in\lambda}X_n, \tag{3.9}$$

in place of (3.7). Since these initial conditions are assumed in (2.1), and since the effect of relaxing them will usually be apparent, we will henceforth use (3.8) in place of (3.3).

3.6. Remark

Another useful representation for the dft of X_t can be obtained by combining the representation (3.8) with the unit root decomposition (3.9). It is especially useful when $d > 1$. Write (2.1) as

$$(1-L)X_t = (1-L)^{1-d} u_t := z_t \tag{3.10}$$

so that $X_t = \sum_{j=1}^{t} z_j + X_0$. Then, taking dfts in (3.10), we first apply (3.9) to write $w_x(\lambda_s)$ in terms of $w_z(\lambda_s)$ and then use (3.8) to reduce $w_z(\lambda_s)$ in terms of $w_u(\lambda_s)$ and a correction term. The outcome is formalized in the following theorem.

3.7. Theorem

If X_t follows (2.1), then

$$w_x(\lambda)\left(1-e^{i\lambda}\right) = w_z(\lambda) - e^{i\lambda}\frac{e^{i\lambda n}X_n}{\sqrt{2\pi n}} \tag{3.11}$$

$$= D_n\left(e^{i\lambda}; f\right) w_u(\lambda) - \frac{e^{i\lambda n}}{\sqrt{2\pi n}}\tilde{U}_{\lambda n}(f) - e^{i\lambda}\frac{e^{i\lambda n}X_n}{\sqrt{2\pi n}}, \tag{3.12}$$

where $f = 1-d$,

$$\tilde{U}_{\lambda n}(f) = \tilde{D}_{n\lambda}\left(e^{-i\lambda}L; f\right)u_n = \sum_{p=0}^{n-1}\tilde{f}_{\lambda p}e^{-ip\lambda}u_{n-p}, \quad \text{and} \quad \tilde{f}_{\lambda p} = \sum_{k=p+1}^{n}\frac{(-f)_k}{k!}e^{ik\lambda}. \tag{3.13}$$

3.8. Remark

Some further decomposition beyond (3.11) and (3.12) is possible. As in Phillips and Solo (1992), we can decompose the operator $C(L)$ that appears in $u_t = C(L)\varepsilon_t$ as

$$C(L) \;=\; C\left(e^{i\lambda}\right) + \tilde{C}_\lambda\left(e^{-i\lambda}L\right)\left(e^{-i\lambda}L-1\right), \quad \tilde{C}_\lambda(L) = \sum_{j=0}^{\infty}\tilde{c}_{j\lambda}L^j,$$

$$\tilde{c}_{j\lambda} = e^{-i\lambda j}\sum_{k=j+1}^{\infty}c_k e^{i\lambda k},$$

where $\sum_{j=0}^{\infty}|\tilde{c}_{j\lambda}| < \infty$ in view of the summability condition on c_j in (2.4). Then,

$$u_t = C(L)\varepsilon_t = C\left(e^{i\lambda}\right)\varepsilon_t + e^{-i\lambda}\varepsilon_{\lambda t-1} - \varepsilon_{\lambda t}, \tag{3.14}$$

is a valid decomposition of u_t into the *iid* component $C\left(e^{i\lambda}\right)\varepsilon_t$ and a stationary error that telescopes under the dft operation, with $\varepsilon_{\lambda t} = \tilde{C}_\lambda\left(e^{-i\lambda}L\right)\varepsilon_t = \sum_{j=0}^{\infty}\tilde{c}_{j\lambda}e^{-i\lambda j}\varepsilon_{t-j}$. In particular,

$$w_u(\lambda) = C\left(e^{i\lambda}\right) w_\varepsilon(\lambda) + \frac{1}{\sqrt{2\pi n}}\left(\varepsilon_{\lambda 0} - e^{in\lambda}\varepsilon_{\lambda n}\right) = C\left(e^{i\lambda}\right) w_\varepsilon(\lambda) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Using this representation in (3.12) we get

$$\begin{aligned} w_x(\lambda)\left(1-e^{i\lambda}\right) &= D_n\left(e^{i\lambda}; f\right) C\left(e^{i\lambda}\right) w_\varepsilon(\lambda) - \frac{e^{i\lambda n}}{\sqrt{2\pi n}}\tilde{U}_{\lambda n}(f) \\ &\quad - e^{i\lambda}\frac{e^{i\lambda n} X_n}{\sqrt{2\pi n}} + D_n\left(e^{i\lambda}; f\right) \times O_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{3.15}$$

Additionally, z_t in (3.10) can be written as

$$z_t = C\left(e^{i\lambda}\right)(1-L)^f \varepsilon_t + (1-L)^f \left(e^{-i\lambda}L - 1\right)\varepsilon_{\lambda t}. \tag{3.16}$$

Set $\eta_t = (1-L)^f \varepsilon_t$, $\eta_{\lambda t} = (1-L)^f \varepsilon_{\lambda t}$ in (3.16) and take dfts, giving

$$\begin{aligned} w_z(\lambda) &= C\left(e^{i\lambda}\right) w_\eta(\lambda) + \frac{1}{\sqrt{2\pi n}}\left(\eta_{\lambda 0} - e^{in\lambda}\eta_{\lambda n}\right) \\ &= C\left(e^{i\lambda}\right) w_\eta(\lambda) + O_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{3.17}$$

since $\eta_{\lambda t}$ is stationary with finite variance for all $d \in \left(\frac{1}{2}, \frac{3}{2}\right)$ because then $|f| < \frac{1}{2}$. (Note that $\eta_{\lambda t} = \varepsilon_{\lambda t}$ when $d = 1$). Next write

$$\eta_t = (1-L)^f \varepsilon_t = \left[D_n(L; f) + R_n(L; f)\right]\varepsilon_t \tag{3.18}$$

with

$$R_n(L; f) = \sum_{k=n+1}^{\infty} \frac{(-f)_k}{k!} L^k,$$

and note that

$$\varepsilon_{nt} := R_n(L; f)\varepsilon_t = O_p\left(\frac{1}{n^{\frac{1}{2}+f}}\right).$$

Applying (3.3) to the dft $w_\eta(\lambda)$ calculated from (3.18) we have

$$w_\eta(\lambda) = w_\varepsilon(\lambda) D_n\left(e^{i\lambda}; f\right) + \frac{1}{\sqrt{2\pi n}}\left(\tilde{\varepsilon}_{\lambda 0}(f) - e^{in\lambda}\tilde{\varepsilon}_{\lambda n}(f)\right) + w_{n\varepsilon}(\lambda), \tag{3.19}$$

with

$$\tilde{\varepsilon}_{\lambda n}(f) = \sum_{p=0}^{n-1} \tilde{f}_{\lambda p} e^{-ip\lambda} \varepsilon_{n-p}, \quad \tilde{f}_{\lambda p} = \sum_{k=p+1}^{n} \frac{(-f)_k}{k!} e^{ik\lambda}, \tag{3.20}$$

and

$$w_{n\varepsilon}(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \varepsilon_{nt} e^{i\lambda_t}.$$

Now $w_{n\varepsilon}(\lambda) = O_p\left(n^{-f}\right)$ because

$$E\left[w_{n\varepsilon}(\lambda) w_{n\varepsilon}(\lambda)^*\right] = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} e^{i\frac{2\pi}{n}(t-s)} E\left(\varepsilon_{nt}\varepsilon_{ns}\right) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} O\left(n^{-1-2f}\right) = O\left(n^{-2f}\right). \tag{3.21}$$

Using (3.19) and (3.21) in (3.17) we get

$$\begin{aligned} w_z(\lambda) = C\left(e^{i\lambda}\right) & \left[D_n\left(e^{i\lambda}; f\right) w_\varepsilon(\lambda) + \frac{1}{\sqrt{2\pi n}} \left(\tilde{\varepsilon}_{\lambda 0}(f) - e^{in\lambda} \tilde{\varepsilon}_{\lambda n}(f) \right) \right] \\ & + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{n^f}\right). \end{aligned} \tag{3.22}$$

Then, combining (3.22) with the unit root decomposition (3.11) leads to the representation

$$\begin{aligned} w_x(\lambda)\left(1 - e^{i\lambda}\right) &= C\left(e^{i\lambda}\right) D_n\left(e^{i\lambda}; f\right) w_\varepsilon(\lambda) - e^{i\lambda} \frac{X_n}{\sqrt{2\pi n}} \\ & \quad + \frac{1}{\sqrt{2\pi n}} C\left(e^{i\lambda}\right) \left(\tilde{\varepsilon}_{\lambda 0}(f) - e^{in\lambda} \tilde{\varepsilon}_{\lambda n}(f) \right) + O_p\left(\frac{1}{n^f}\right). \end{aligned} \tag{3.23}$$

This representation holds uniformly over λ and is likely to be most useful when $\lambda = \lambda_s = \frac{2\pi s}{n} \to 0$ and $s \to \infty$.

3.9. Remark

The representations (3.8), (3.11), and (3.12) hold for all fundamental frequencies $\lambda_s = \frac{2\pi s}{n}$. They are helpful in providing asymptotic representations of $w_x\left(\lambda_s\right)$. In such expansions, it is useful to allow for situations where $s \to \infty$ as well as $n \to \infty$. In some cases, as in spectral density estimation at some frequency $\phi \neq 0$, we want the expansion rate of s to be the same as n, so that we can accommodate $\lambda_s \to \phi$ as $n \to \infty$. In other cases, as in log periodogram and Gaussian semiparametric regression, interest centers on frequencies λ_s in the vicinity of the origin, so then we consider cases where s is fixed or $s \to \infty$ and $\frac{s}{n} \to 0$ as $n \to \infty$. The following section gives results that are helpful in the determination of the asymptotic form of these representations as $n \to \infty$ under these various conditions.

4. ASYMPTOTIC APPROXIMATIONS

4.1. Component Approximations

We start with the sinusoidal polynomial $D_n\left(e^{i\lambda};d\right)=\sum_{k=0}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}$ that appears in the decomposition (3.1) and Theorems 3.2 and 3.7. The series can be summed in terms of hypergeometric functions and the asymptotic form taken as $n\to\infty$ depends on λ. The behavior is described in the following lemma.

4.2. Lemma[3]

Suppose $d>0$ and is noninteger. Then

$$D_n\left(e^{i\lambda};d\right)=\left(1-e^{i\lambda}\right)^d-e^{i(n+1)\lambda}\frac{(-d)_{n+1}}{(n+1)!}\;{}_2F_1\left(n+1-d,1;n+2;e^{i\lambda}\right), \tag{4.1}$$

and, for $\cos(\lambda)<\frac{1}{2}$,

$$D_n\left(e^{i\lambda};d\right)=\left(1-e^{i\lambda}\right)^d+\frac{e^{i(n+1)\lambda}}{e^{i\lambda}-1}\frac{(-d)_{n+1}}{(n+1)!}\;{}_2F_1\left(1+d,1;n+2;\frac{e^{i\lambda}}{e^{i\lambda}-1}\right). \tag{4.2}$$

The following asymptotic representations hold:

(a) *For fixed* $\lambda\neq 0$

$$D_n\left(e^{i\lambda};d\right)=\left(1-e^{i\lambda}\right)^d-\frac{1}{\Gamma(-d)n^{1+d}}\frac{e^{in\lambda}}{1-e^{i\lambda}}\left[1+O\left(\frac{1}{n}\right)\right].$$

(b) *For* $\lambda=\lambda_s=\frac{2\pi is}{n}\to 0$ *and* $s\to\infty$ *as* $n\to\infty$

$$D_n\left(e^{i\lambda_s};d\right)=\left(1-e^{i\lambda_s}\right)^d+\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^d s}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right).$$

(c) *For* $\lambda=\lambda_s=\frac{2\pi is}{n}\to 0$ *and* s *fixed as* $n\to\infty$

$$D_n\left(e^{i\lambda_s};d\right)=\frac{1}{\Gamma(1-d)n^d}\;{}_1F_1(1,1-d;-2\pi is)+O\left(\frac{1}{n^{1+d}}\right).$$

(d) *For* $\lambda=0$

$$D_n(1;d)=\frac{1}{\Gamma(1-d)}\frac{1}{n^d}\left[1+O\left(\frac{1}{n}\right)\right].$$

In the above formulae, ${}_1F_1(a,b;z)$ *and* ${}_2F_1(a,b,c;z)$ *denote the confluent hypergeometric function and the hypergeometric function, respectively.*

From part (d), it follows that $D_n(1;d)$ differs from 0 by a term of $O\left(n^{-d}\right)$. From part (c), the same also applies to $D_n\left(e^{i\lambda_s};d\right)$ when s is fixed and $\lambda_s = \frac{2\pi is}{n} \to 0$. Of course, in the event that d is a positive integer, we have the following terminating polynomials

$$D_n(1;d) = \sum_{k=0}^{n} \frac{(-d)_k}{k!} = \sum_{k=0}^{d} \frac{(-d)_k}{k!} = \sum_{k=0}^{d} \binom{d}{k} (-1)^k = (1-1)^d = 0,$$

and

$$D_n\left(e^{i\lambda_s};d\right) = \sum_{k=0}^{n} \frac{(-d)_k\, e^{i\lambda_s k}}{k!} = \sum_{k=0}^{d} \binom{d}{k} \left(-e^{i\lambda_s}\right)^k = \left(1-e^{i\lambda_s}\right)^d$$

in this case.

Our next focus of interest is the correction term in (3.8) that involves $\tilde{X}_{\lambda n}(d)$. We are especially interested in deriving an asymptotic approximation to $\tilde{X}_{\lambda n}(d)$ at the fundamental frequencies λ_s. As in Lemma 3.1, the asymptotic behavior of $\tilde{X}_{\lambda_s,n}(d)$ is sensitive to the value of s in $\lambda_s = \frac{2\pi s}{n}$. In particular, when $d \in \left(\frac{1}{2},1\right)$, the asymptotic form of $\tilde{X}_{\lambda_s,n}(d)$ differs, depending on whether s is fixed or whether $s \to \infty$ as $n \to \infty$. In the latter case, $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s,n}(d) = o_p(1)$, while in the former $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s,n}(d) = O_p(1)$. On the other hand, when $d = 1$, $n^{-\frac{1}{2}}\tilde{X}_{\lambda_s,n}(d) = O_p(1)$ for all $s \neq 0$. The results are given in the following theorem.

4.3. Theorem

Suppose $d \in \left(\frac{1}{2},1\right)$. *Then*

(a) *For fixed* $\lambda \neq 0$ *as* $n \to \infty$,

$$\frac{\tilde{X}_{\lambda,n}(d)}{\sqrt{n}} = -\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}} \frac{X_n}{\sqrt{n}} + o_p\left(\frac{1}{n^{1-d}}\right) = O_p\left(\frac{1}{n^{1-d}}\right).$$

(b) *For* $\lambda = \lambda_s = \frac{2\pi s}{n} \to 0$ *and* $\frac{s}{n^{\alpha}} \to \infty$, *as* $n \to \infty$, *for some* $\alpha \in \left(\frac{1}{2},1\right)$

$$\begin{aligned}\frac{\tilde{X}_{\lambda_s,n}(d)}{\sqrt{n}} &= -\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}} \frac{X_n}{\sqrt{n}} + o_p\left(\frac{1}{s^{1-d}}\right) = -\frac{e^{i\lambda_s}}{\left(-2\pi is\right)^{1-d}} \frac{X_n}{n^{d-\frac{1}{2}}} + o_p\left(\frac{1}{s^{1-d}}\right)\\ &= O_p\left(\frac{1}{s^{1-d}}\right).\end{aligned}$$

(c) *For* $\lambda = \lambda_s = \frac{2\pi s}{n}$ *and s fixed, as* $n \to \infty$,

$$\frac{\tilde{X}_{\lambda_s n}(d)}{\sqrt{n}} = \frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr$$
$$-\frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right) = O_p(1),$$

where $X_{n,d}(r) = \dfrac{X_{\lfloor nr \rfloor}}{n^{d-\frac{1}{2}}}$.

(d) *When* $d=1$, *the equation*

$$\frac{\tilde{X}_{\lambda,n}(1)}{\sqrt{n}} = -e^{i\lambda}\frac{X_n}{\sqrt{n}} = O_p(1)$$

holds for λ *fixed, or* $\lambda = \lambda_s = \dfrac{2\pi s}{n} \to 0$ *with* $s \to \infty$, *or* $\lambda_s = \dfrac{2\pi s}{n} \to 0$ *with* s *fixed.*

In parts (a) and (b) of Theorem 4.3 the leading term in the asymptotic approximation of $n^{-\frac{1}{2}}\tilde{X}_{\lambda,n}(d)$ is the same and so, although the error order of magnitude differs, we may write

$$\frac{\tilde{X}_{\lambda,n}(d)}{\sqrt{n}} = -\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}} + o_p\left(\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}}\right),$$

for both these cases. Further, the leading term of $n^{-\frac{1}{2}}\tilde{X}_{\lambda,n}(d)$ is $O_p\left(\frac{1}{n^{1-d}}\right)$ for fixed $\lambda \neq 0$, is $O_p\left(\frac{1}{s^{1-d}}\right)$ for $\lambda_s = \frac{2\pi s}{n} \to 0$ and $\frac{s}{n^{\alpha}} \to \infty$, and is $O_p(1)$ for $\lambda_s = \frac{2\pi s}{n} \to 0$ with s fixed. Thus, the correction term $n^{-\frac{1}{2}}\tilde{X}_{\lambda,n}(d)$ is nonnegligible in a region around the origin when $d \in \left(\frac{1}{2},1\right)$. The asymptotic form of $n^{-\frac{1}{2}}\tilde{X}_{\lambda,n}(d)$ in that case (i.e., case (c), with $\lambda_s = \frac{2\pi s}{n}$, and s fixed) is more complicated than the other cases and it involves hypergeometric series. The representation given in case (c) actually includes $s=0$, for which we have the simpler form

$$\frac{\tilde{X}_{\lambda_0 n}(d)}{\sqrt{n}} = \frac{1}{\Gamma(1-d)}\int_0^1 X_{n,d}(r)dr - \frac{1}{\Gamma(1-d)}\int_0^1 r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right). \quad (4.3)$$

When $d=1$, the formula given in (d) is exact, as follows directly from (3.9).

Finally, we look at the correction term $\tilde{U}_{\lambda n}(f)$ that appears in (3.12). We concentrate on the interesting case where λ is in the vicinity of the origin and give the result corresponding to part (c) of Theorem 4.3.

4.4. Theorem

Suppose $d \in \left(\frac{1}{2},\frac{3}{2}\right)$ and $f = 1-d$. *Then, for* $\lambda = \lambda_s = \dfrac{2\pi s}{n}$ *and* s *fixed, as* $n \to \infty$

$$\frac{\tilde{U}_{\lambda_s n}(f)}{\sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(1-f)n^f}\left\{ {}_1F_1(1,1-f;-2\pi is)\int_0^1 e^{-2\pi isr}dX_n(1-r) \right. \\ \left. -\int_0^1 r^{-f}\ {}_1F_1(1,1-f;-2\pi isr)\,dX_n(1-r)\right\} + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{4.4}$$

where $X_n(r) = n^{-\frac{1}{2}}\sum_{t=0}^{\lfloor nr \rfloor} u_t$. *When* $f = 0, \tilde{U}_{\lambda_s n}(0) = 0$.

4.5. Approximations for $w_x(\lambda)$

Evaluating (3.8) at λ_s, we have

$$w_x(\lambda_s) = D_n\left(e^{i\lambda_s};d\right)^{-1}\left[w_u(\lambda_s) + \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda_s n}(d)\right].$$

We use Lemma 4.2 and Theorem 4.3 to obtain explicit expressions for $w_x(\lambda_s)$ in terms of $w_u(\lambda_s)$ and a correction term. When $d = 1$, the following exact form comes directly from (3.9)

$$w_x(\lambda_s) = \left(1-e^{i\lambda_s}\right)^{-1} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}}, \tag{4.5}$$

and holds for all $s = 1,2,\ldots$. When $d \in \left(\frac{1}{2},1\right)$, it is convenient to separate the following three cases:

(a) Case $\lambda_s \to \phi \neq 0$

Here, from Lemma 4.2 we have

$$D_n\left(e^{i\lambda_s};d\right) = \left(1-e^{i\lambda_s}\right)^d - \frac{1}{\Gamma(-d)n^{1+d}}\frac{e^{in\lambda_s}}{1-e^{i\lambda_s}}\left[1+O\left(\frac{1}{n}\right)\right] \\ = \left(1-e^{i\lambda_s}\right)^d + O\left(\frac{1}{n^{1+d}}\right),$$

uniformly for $\lambda_s \in \mathcal{B}_\phi = \left\{\phi - \frac{\pi}{M}, \phi + \frac{\pi}{M}\right\}$ where $M \to \infty$ as $n \to \infty$. Similarly, from Theorem 4.3,

$$\frac{\tilde{X}_{\lambda_s,n}(d)}{\sqrt{n}} = -\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}\frac{X_n}{\sqrt{n}} + o_p\left(\frac{1}{n^{1-d}}\right)$$

uniformly for $\lambda_s \in \mathcal{B}_\phi$. It follows that

$$w_x(\lambda_s) = \left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{1}{n^{1-d}}\right), \tag{4.6}$$

uniformly for $\lambda_s \in \mathcal{B}_\phi$.

(b) Case $\lambda_s = \dfrac{2\pi is}{n} \to 0$ and $s \to \infty$

From Lemma 4.2 (b) when $s \to \infty$ as $n \to \infty$

$$D_n\left(e^{i\lambda_s};d\right) = \left(1-e^{i\lambda_s}\right)^d + \frac{1}{\Gamma(-d)n^d}\frac{1}{2\pi is}\left[1+O\left(\frac{1}{s}\right)\right] + O\left(\frac{1}{n^{1+d}}\right).$$

And from Theorem 4.3 (b) with $\dfrac{s}{n^\alpha} \to \infty$ for some $\alpha \in \left(\dfrac{1}{2},1\right)$ as $n \to \infty$,

$$\frac{\tilde{X}_{\lambda_s,n}(d)}{\sqrt{n}} = -\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}\frac{X_n}{\sqrt{n}} + o_p\left(\frac{1}{s^{1-d}}\right).$$

It follows that if $\dfrac{s}{n} + \dfrac{n^\alpha}{s} \to 0$ as $n \to \infty$, for some $\alpha \in \left(\dfrac{1}{2},1\right)$, then

$$w_x(\lambda_s) = \left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{\left(1-e^{i\lambda_s}\right)^{-d}}{s^{1-d}}\right). \tag{4.7}$$

Observe that the first two terms of (4.6) and (4.7) are the same. Although the order of magnitude of the error differs in the two cases, we may write

$$w_x(\lambda_s) = \left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)}\frac{X_n}{\sqrt{n}}\right) \tag{4.8}$$

for both these cases, and (4.8) is valid for all $\lambda_s = \dfrac{2\pi s}{n}$ with $\dfrac{n^\alpha}{s} \to 0$.

(c) Case $\lambda_s = \dfrac{2\pi is}{n} \to 0$ and s fixed

From Lemma 4.2 (c) when s is fixed as $n \to \infty$, we have

$$D_n\left(e^{i\lambda_s};d\right) = \frac{1}{\Gamma(1-d)n^d}\ {}_1F_1(1,1-d;-2\pi is) + O\left(\frac{1}{n^{1+d}}\right), \tag{4.9}$$

and it follows that

$$\frac{1}{n^d}w_x(\lambda_s) = \frac{1}{n^d}\left[\frac{1}{\Gamma(1-d)n^d}\ {}_1F_1(1,1-d;-2\pi is) + O\left(\frac{1}{n^{1+d}}\right)\right]^{-1}$$
$$\times\left[w_u(\lambda_s) + \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda_s n}(d)\right],$$

giving

$$\frac{w_x(\lambda_s)}{n^d} = \frac{\Gamma(1-d)}{{}_1F_1(1,1-d;-2\pi is)}\left[w_u(\lambda_s) + \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda_s n}(d)\right] + O_p\left(\frac{1}{n}\right). \tag{4.10}$$

Further, from Theorem 4.3 (c),

$$\begin{aligned}\frac{\tilde{X}_{\lambda_s n}(d)}{\sqrt{n}} &= \frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr \\ &\quad - \frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right),\end{aligned}$$

so that

$$\begin{aligned}\frac{1}{n^d}w_x(\lambda_s) &= \frac{\Gamma(1-d)}{{}_1F_1(1,1-d;-2\pi is)}w_u(\lambda_s) + \frac{1}{\sqrt{2\pi}}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr \\ &\quad - \frac{(2\pi)^{-\frac{1}{2}}}{{}_1F_1(1,1-d;-2\pi is)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr \\ &\quad + O_p\left(\frac{1}{n^{1-d}}\right).\end{aligned} \tag{4.11}$$

Unlike (4.6) and (4.8), the term

$$\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} \tag{4.12}$$

does not figure directly in (4.11). In fact, as the alternate representation in the next section shows, the term (4.12) is absorbed into the series expression in (4.11), so it is still present and figures in the leading term of the dft $w_x(\lambda_s)$ when s is fixed.

(c) Case $\lambda_s = \dfrac{2\pi is}{n} \to 0$ and s fixed: alternate form.

Theorem 3.7 gives

$$w_x(\lambda_s)\left(1-e^{i\lambda_s}\right) = D_n\left(e^{i\lambda_s};f\right)w_u(\lambda_s) - \frac{1}{\sqrt{2\pi n}}\tilde{U}_{\lambda_s n}(f) - e^{i\lambda_s}\frac{X_n}{\sqrt{2\pi n}}, \tag{4.13}$$

with $f = 1-d$, Lemma 4.2 (c) gives

$$D_n\left(e^{i\lambda_s};f\right) = \frac{1}{\Gamma(1-f)n^f}\ {}_1F_1(1,1-f;-2\pi is) + O\left(\frac{1}{n^{1+f}}\right),$$

and Theorem 4.4 gives

$$\frac{\tilde{U}_{\lambda_s n}(f)}{\sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)n^f} \left\{ {}_1F_1(1,1-f;-2\pi is) \int_0^1 e^{-2\pi isr} dX_n(1-r) \right.$$
$$\left. - \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) dX_n(1-r) \right\} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Also,

$$w_u(\lambda_s) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n e^{2\pi si\frac{t}{n}} u_t = \frac{1}{\sqrt{2\pi}} \sum_{k=1}^n e^{2\pi si\frac{n-k}{n}} \frac{u_{n-k}}{\sqrt{n}}$$
$$= \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-2\pi isr} dX_n(1-r) + O_p\left(\frac{1}{n}\right).$$

Combining these last three representations in (4.13), we get

$$\begin{aligned}
& w_x(\lambda_s)\left(1-e^{i\lambda_s}\right) \\
&= \frac{1}{\Gamma(1-f)n^f} \; {}_1F_1(1,1-f;-2\pi is) \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-2\pi isr} dX_n(1-r) + O_p\left(\frac{1}{n}\right) \\
&\quad - \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)n^f} \; {}_1F_1(1,1-f;-2\pi is) \int_0^1 e^{-2\pi isr} dX_n(1-r) \\
&\quad + \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)n^f} \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) dX_n(1-r) - e^{i\lambda_s} \frac{X_n}{\sqrt{2\pi n}} \\
&\quad + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)n^f} \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) dX_n(1-r) - e^{i\lambda_s} \frac{X_n}{\sqrt{2\pi n}} + O_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

leading to

$$\frac{1}{n^d} w_x(\lambda_s) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)n\left(1-e^{i\lambda_s}\right)} \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) dX_n(1-r)$$
$$- \frac{1}{\sqrt{2\pi}} \frac{e^{i\lambda_s}}{n\left(1-e^{i\lambda_s}\right)} \frac{X_n}{n^{d-\frac{1}{2}}} + O_p\left(\frac{1}{n^{d-\frac{1}{2}}}\right), \tag{4.14}$$

which shows how (4.12) continues to play a role in the leading term of $w_x(\lambda_s)$.

4.6. Limit Theory

Under (2.4), partial sums of u_t satisfy the functional law

$$X_n(r) = \frac{1}{\sqrt{n}} \sum_{t=0}^{\lfloor nr \rfloor} u_t \to_d B(r), \tag{4.15}$$

where B is Brownian motion with variance $\omega^2 = \sigma^2 C(1)^2$, for example, Phillips and Solo (1992). There is a corresponding functional law for suitably standardized elements of the time series X_t. Akonom and Gouriéroux (1987) showed such a functional law for $n^{\frac{1}{2}-d} X_t$ when the components u_t follow a stationary ARMA process and the following simply extends their result to the linear process u_t.

4.7. Lemma

For u_t satisfying (2.4) and with ε_t iid$\left(0,\sigma^2\right)$ and $E\left|\varepsilon_t\right|^p < \infty$ for $p > \max\left(\frac{1}{d-\frac{1}{2}}, 2\right)$,

$$X_{n,d}(r) = \frac{X_{\lfloor nr \rfloor}}{n^{d-\frac{1}{2}}} \xrightarrow{d} B_{d-1}(r) = \frac{1}{\Gamma(d)} \int_0^r (r-s)^{d-1} dB(s), \tag{4.16}$$

a fractional Brownian motion where $B(s)$ is Brownian motion with variance ω^2.

Like X_t, the fractional Brownian motion $B_{d-1}(r)$ is initialized at the origin, and therefore has nonstationary increments, in contrast to the other fractional process

$$W_H(r) = \frac{1}{C(H)} \int_{-\infty}^{\infty} \left[\left\{ (r-s)_+ \right\}^{H-\frac{1}{2}} - \left\{ (-s)_+ \right\}^{H-\frac{1}{2}} \right] dB(s), \quad H = d - \frac{1}{2},$$

$$C(H) = \left\{ \frac{1}{2H} + \int_0^{\infty} \left[(1+s)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}} \right]^2 ds \right\}^{\frac{1}{2}}, \quad 0 < H < 1 \tag{4.17}$$

introduced by Mandelbrot and Van Ness (1968) and studied by Samorodnitsky and Taqqu (2017) in this form. Both processes reduce to Brownian motion for special cases of the parameters, viz., $d = 1$ for (4.16), and $H = \frac{1}{2}$ for (4.17).

These functional laws enable us to get limit representations of the correction term $n^{-\frac{1}{2}} \tilde{X}_{\lambda_s n}(d)$. The case where s is fixed as $n \to \infty$ is especially interesting, the other two cases following immediately from (4.16) and the respective expressions (4.6) and (4.7).

4.8. Lemma

For $\lambda_s = \frac{2\pi i s}{n} \to 0$ and s fixed

$$\frac{\tilde{X}_{\lambda_s n}(d)}{\sqrt{n}} \xrightarrow{d} \frac{1}{\Gamma(1-d)} \int_0^1 e^{2\pi i s r} B_{d-1}(r) dr\, {}_1F_1(1, 1-d; -2\pi i s) - \int_0^1 e^{2\pi i s r} dB(r). \tag{4.18}$$

The next result gives formulae for the stochastic Fourier integral $\int_0^r e^{2\pi siq} dB(q)$ that appears in (4.18) and (when $s = 0$) for the constituent Brownian motion B in terms of the fractional Brownian motion B_{d-1}.

4.9. Theorem

For fixed integer s

$$\int_0^r e^{-2\pi si(r-q)} dB(q) = \frac{1}{\Gamma(1-d)} \int_0^r {}_1F_1(1, 1-d; -2\pi is(r-q))(r-q)^{-d} B_{d-1}(q)dq, \tag{4.19}$$

and, in the special case where $s = 0$,

$$B(r) = \frac{1}{\Gamma(1-d)} \int_0^r (r-q)^{-d} B_{d-1}(q)dq. \tag{4.20}$$

The equality (4.20) is the inverse (integral) transform of the fractional Brownian motion $B_{d-1}(r)$. In effect, the right side of (4.20) is the $(1-d)$'th fractional integral of the $(d-1)$'th fractional derivative of Brownian motion. Formula (4.19) extends this representation to the case $s \neq 0$. When $r = 1$, (4.19) becomes

$$\int_0^1 e^{2\pi siq} dB(q) = \frac{1}{\Gamma(1-d)} \int_0^1 {}_1F_1(1, 1-d; -2\pi is(1-q))(1-q)^{-d} B_{d-1}(q)dq.$$

4.10. Theorem

Suppose $d \in \left(\frac{1}{2}, 1\right)$. *The following limit results apply.*

(a) *Let* $\phi > 0$ *and suppose* $\lambda_{s_j} \in \mathcal{B}_\phi = \left\{\phi - \frac{\pi}{2M}, \phi + \frac{\pi}{2M}\right\}$ *for a finite set of distinct integers* s_j $(j = 1, \ldots, J)$. *When* $M \to \infty$ *as* $n \to \infty$, *the family* $\{w_x(\lambda_{s_j})\}_{j=1}^J$ *are asymptotically independently distributed as complex normal* $N_c(0, f_x(\phi))$, *where* $f_x(\phi) = \left|1 - e^{i\phi}\right|^{-2d} f_u(\phi)$.

(b) *Let* $\{s_j\}_{j=1}^J$ *be distinct integers with* $0 < l < s_j < L$ *for each* j *and with* $\frac{L}{n} + \frac{n^\alpha}{l} \to 0$ *as* $n \to \infty$, *for some* $\alpha \in \left(\frac{1}{2}, 1\right)$. *The family* $\{(\lambda_{s_j})^d w_x(\lambda_{s_j})\}_{j=1}^J$ *are asymptotically independently distributed as* $N_c(0, f_u(0))$.

(c) *Let* $\{s_j\}_{j=1}^J$ *be a finite set of distinct positive integers which are fixed as* $n \to \infty$. *Then, for each* j

$$\frac{1}{n^d} w_x\left(\lambda_{s_j}\right) \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_0^1 e^{2\pi i s_j r} B_{d-1}(r) dr, \tag{4.21}$$

where B_{d-1} is the fractional Brownian motion given in (4.16). *Joint convergence also applies.*
When $d=1$, the following limits apply.

(d) *Let $\phi > 0$ and suppose $\lambda_{s_j} \in \mathcal{B}_\phi = \left\{\phi - \frac{\pi}{M}, \phi + \frac{\pi}{M}\right\}$ for a finite set of distinct integers s_j $(j = 1,...,J)$. When $M \to \infty$ as $n \to \infty$, the family $\{w_x(\lambda_{s_j})\}_{j=1}^J$ are asymptotically distributed as*

$$\left\{\frac{1}{1-e^{i\phi}}\xi_j - \frac{e^{i\phi}}{1-e^{i\phi}}\eta\right\}_{j=1}^J, \tag{4.22}$$

where the $\{\xi_j\}_{j=1}^J$ are iid $N_c\left(0, f_u(\phi)\right)$ and are independent of

$$\eta = \frac{B(1)}{\sqrt{2\pi}}, \tag{4.23}$$

where B is Brownian motion with variance ω^2.

(e) *Let $\{s_j\}_{j=1}^J$ be a finite set of distinct positive integers for which $\frac{s_j}{n} \to 0$ as $n \to \infty$. The family $\{\lambda_{s_j} w_x(\lambda_{s_j})\}_{j=1}^J$ are asymptotically distributed as*

$$i\left(\xi_j - \eta\right), \tag{4.24}$$

where ξ_j and η are as in (4.22) and (4.23).

(f) When s_j is fixed as $n \to \infty$, the ξ_j in (e) have the representation

$$\xi_j = \frac{1}{\sqrt{2\pi}}\int_0^1 e^{2\pi i s_j r} dB(r), \tag{4.25}$$

and

$$\frac{1}{n} w_x\left(\lambda_{s_j}\right) \xrightarrow{d} \frac{1}{\sqrt{2\pi}}\int_0^1 e^{2\pi i s_j r} B(r)\, dr, \tag{4.26}$$

which also holds for $s_j = 0$.

Parts (a) and (d) show that Hannan (1973) result for the limit theory of dfts of stationary processes extends to fractional processes at frequencies removed from the origin when $d \in \left(\frac{1}{2}, 1\right)$ but not when $d = 1$. In the latter case, the leakage from the zero frequency is so substantial that it affects the limit theory of the dft at all frequencies, although the limit distribution is still normal. Moreover, as is apparent from the form of (4.22), the limit variates are spatially correlated across frequency by virtue of the presence of the random component η, through which the leakage is transmitted.

Part (b) shows that, when $d \in \left(\frac{1}{2},1\right)$, a version of Hannan's result applies to the scaled transforms $\left(\frac{s_j}{n}\right)^d w_x\left(\lambda_{s_j}\right)$ in a (distant) vicinity of the origin where $\lambda_{s_j} = \frac{2\pi s_j i}{n} \to 0$ but $\frac{n^\alpha}{s_j} \to 0$ as $n \to \infty$, for some $\alpha \in \left(\frac{1}{2},1\right)$. However, when $d = 1$, the scaled transforms $\frac{s_j}{n} w_x\left(\lambda_{s_j}\right)$ are asymptotically dependent across frequency.

Part (c) shows that in the immediate vicinity of the origin (i.e., for $\lambda_{s_j} = \frac{2\pi s_j i}{n} \to 0$ with s_j fixed), the $n^{-d} w_x(\lambda_{s_j})$ are asymptotically dependent for $d \in \left(\frac{1}{2},1\right]$ and each converges weakly to an integral functional of fractional Brownian motion that involves the integer s_j. In earlier work, Akonom and Gouriéroux (1987) gave (4.21) in the case of ARMA u_t. An alternate expression for (4.21), which relates to (4.14) is

$$\frac{1}{n^d} w_x\left(\lambda_{s_j}\right) \xrightarrow{d} \frac{1}{\sqrt{2\pi}\Gamma(1+d)} \int_0^1 {}_1F_1\left(1,1+d;-2\pi i s_j r\right) r^d dB(1-r)$$

and can be obtained from the formula

$$\int_0^1 e^{2\pi i s r} B_{d-1}(r)\, dr = \frac{1}{\Gamma(1+d)} \int_0^1 {}_1F_1(1,1+d;-2\pi i s r) r^d dB(1-r),$$

which is proved in Lemma E in the Technical Appendix and Proofs.

The methods in the proof of Theorem 4.10 are used in (Phillips, 2007, Theorem 3.2) to extend existing theory showing the asymptotic independence of a finite collection of dfts of stationary time series (Hannan, 1973) to collections of a small (i.e., with less than sample size) infinity of dfts at Fourier frequencies.

5. STATISTICAL APPLICATIONS

5.1. Spectrum Estimation for Fractional Processes

The limit theory in Section 4.6 is useful in obtaining the asymptotic behavior of spectral estimates for fractional processes. We give some results for smoothed periodogram estimates for frequencies at the origin and away from the origin. The former are of interest in procedures that are used to estimate the memory parameter d. The latter reveal any leakage from low to high frequencies that occurs in spectrum estimation.

For frequencies away from the origin such as $\phi \neq 0$, the usual smoothed periodogram estimator of $f_x(\phi)$ is given by

$$\hat{f}_{xx}(\phi) = \frac{1}{m} \sum_{\lambda_s \in \mathcal{B}(\phi)} w_x(\lambda_s) w_x(\lambda_s)^*, \tag{5.1}$$

where $\mathcal{B}_m(\phi)=\left(\phi-\frac{\pi}{2M},\phi+\frac{\pi}{2M}\right]$ and M is the bandwidth parameter that determines the number of frequencies $m=\#\{\lambda_s\in\mathcal{B}_m(\phi)\}=[n/2M]$ used in the smoothing. At the zero frequency $\phi=0$, we consider a one-sided average of m periodogram ordinates at the origin

$$\hat{f}_{xx}(0)=\frac{1}{m}\sum_{s=0}^{m-1}w_x(\lambda_s)w_x(\lambda_s)^*. \tag{5.2}$$

The following theorem gives the asymptotic behavior of $\hat{f}_{xx}(\phi)$ for these two cases and for $d\in\left(\frac{1}{2},1\right)$ and $d=1$.

5.2. Theorem

(a) *For* $\phi\neq 0$ *and* $\frac{1}{2}<d<1$

$$\hat{f}_{xx}(\phi)\to_p f_x(\phi)=\frac{f_u(\phi)}{\left|1-e^{i\phi}\right|^{2d}}.$$

(b) *For* $\phi\neq 0$ *and* $d=1$

$$\hat{f}_{xx}(\phi)\to_d f_x(\phi)+\frac{1}{2\pi}\left|1-e^{i\phi}\right|^{-2}B(1)^2.$$

(c) *For* $\frac{1}{2}<d<1$ and m such that $\frac{m}{n^{\alpha}}\to\infty$ with $\alpha\geq\frac{1}{2d}$

$$\frac{m}{n^{2d}}\hat{f}_{xx}(0)\to_d \frac{1}{2\pi}\int_0^1 B_{d-1}(r)^2\,dr.$$

(d) *For* $d=1$ and m such that $\frac{m}{\sqrt{n}}\to\infty$

$$\frac{m}{n^2}\hat{f}_{xx}(0)\to_d \frac{1}{2\pi}\int_0^1 B(r)^2\,dr.$$

According to part (a), spectral estimates like $\hat{f}_{xx}(\phi)$ at frequencies removed from the origin are consistent for $f_x(\phi)=\left|1-e^{i\phi}\right|^{-2d}f_u(\phi)$ provided $d<1$. When $d=1$, the estimate is inconsistent and converges weakly to a random quantity. In this case, the leakage from low frequency behavior is strong enough to persist in the limit at all frequencies $\phi>0$. Part (d) was given in earlier work by Phillips (1991), where it was shown to be useful in analyzing regression in the frequency domain with integrated time series. A new and simpler derivation is given here based on the decomposition (3.9). Part (c) can be expected to be useful in similar regression contexts with fractional processes.

5.3. Semiparametric Estimation of d

We indicate some potential applications of the above theory for the estimation of the memory parameter d in (2.1). This is a large subject which goes beyond the scope of this chapter and for which theoretical development was undertaken after the original version of this chapter was completed in 1999. The main references will be reported in the following discussion. The presentation here focuses on the new ideas that led into these developments and not the technical details.

Concordant with the nonparametric approach, our concern is with the case where little is known about the short memory component u_t of (2.1) and its spectrum $f_u(\lambda)$ is treated nonparametrically. In both log periodogram estimation and local Whittle estimation, this is accomplished by working with the dft $w_x(\lambda_s)$ of the data X_t over a set of m Fourier frequencies $\left\{\lambda_s = \frac{2\pi s}{n} : s = 1,\ldots,m\right\}$ that shrink slowly to origin as the sample size $n \to \infty$ by virtue of a condition on m of the type $\frac{m}{n} \to 0$. It has been suggested that, in view of the asymptotic correlation of the ordinates in the vicinity of the origin (Künsch, 1986), it may be useful to trim this set of frequencies away from the origin and restrict attention to $\left\{\lambda_s = \frac{2\pi s}{n} : s = l,\ldots,m\right\}$ where l is a trimming number that satisfies $l \to \infty$ and $\frac{\sqrt{m}\log m}{l} \to 0$ (Robinson, 1995b), although it is now known that this trimming is not necessary (Hurvich et al., 1998).

From (4.7) we know that for $d \in \left(\frac{1}{2},1\right)$, the dft $w_x(\lambda_s)$

$$w_x(\lambda_s) = \left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{\left(1-e^{i\lambda_s}\right)^{-d}}{s^{1-d}}\right), \tag{5.3}$$

when $\frac{s}{n} + \frac{n^{\alpha}}{s} \to 0$ as $n \to \infty$, for some $\alpha \in \left(\frac{1}{2},1\right)$. The asymptotic behavior of $w_x(\lambda_s)$ is dominated by the first two terms of (5.3), and as $d \to 1$ the importance of the second term in (5.3), which is $O_p(n^d / s)$, rivals that of the first term, which is $O_p(n^d / s^d)$. Apparently, therefore, it would seem desirable to correct the dft $w_x(\lambda_s)$ for the effects of leakage in semiparametric estimation of d simply by adding the correction term supplied by the known form of the expansion (5.3). For log periodogram regression this amounts to using the quantity

$$v_x(\lambda_s) = w_x(\lambda_s) + \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} \tag{5.4}$$

in place of $w_x(\lambda_s)$ in the regression. Thus, in place of the usual least squares regression (over $s = 1,\ldots,m$)

$$\ln\left(I_x(\lambda_s)\right) = \hat{c} - \hat{d}\ln\left|1-e^{i\lambda_s}\right|^2 + \text{error}$$

that is inspired by the form of the moment relation (2.6) in the frequency domain, the argument above suggests the linear least squares regression

$$\ln\left(I_v(\lambda_s)\right) = \tilde{c} - \tilde{d}\ln\left|1-e^{i\lambda_s}\right|^2 + \text{error}, \tag{5.5}$$

in which the periodogram ordinates, $I_x(\lambda_s)$, are replaced by $I_v(\lambda_s) = v_x(\lambda_s)v_x(\lambda_s)^*$. We call this procedure *modified log periodogram regression*. This replacement is inspired by (5.3), which approximates the data generating process of the dft $w_x(\lambda_s)$ over the relevant set of frequencies as $m \to \infty$ in the regression. In place of the "regression model":

$$\ln\left(I_x(\lambda_s)\right) = c - d\ln\left|1-e^{i\lambda_s}\right|^2 + u(\lambda_s),$$

with $c = \ln\left(f_u(0)\right)$ and

$$u(\lambda_s) = \ln\left[I_x(\lambda_s)/f_x(\lambda_s)\right] + \ln\left(f_u(\lambda_s)/f_u(0)\right),$$

as in (2.7), we now have from (5.3)

$$\begin{aligned} I_v(\lambda_s) &= \left|\left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) + o_p\left(\frac{n^d}{s}\right)\right|^2 \\ &= \left|1-e^{i\lambda_s}\right|^{-2d} I_u(\lambda_s)\left[1+\left(1-e^{i\lambda_s}\right)^d w_u(\lambda_s)^{-1} o_p\left(\frac{n^d}{s}\right)\right] \\ &= \left|1-e^{i\lambda_s}\right|^{-2d} I_u(\lambda_s)\left|\left[1+o_p\left(\frac{1}{s^{1-d}}\right)\right]\right|^2, \end{aligned}$$

which leads to the new regression model

$$\ln\left(I_v(\lambda_s)\right) = c - d\ln\left|1-e^{i\lambda_s}\right|^2 + a(\lambda_s), \tag{5.6}$$

with

$$a(\lambda_s) = \ln\left[I_u(\lambda_s)/f_u(\lambda_s)\right] + \ln\left(f_u(\lambda_s)/f_u(0)\right) + O_p\left(\frac{1}{s^{1-d}}\right). \tag{5.7}$$

This relationship holds for frequencies λ_s satisfying $\frac{s}{n} + \frac{n^\alpha}{s} \to 0$ as $n \to \infty$, in view of (5.3).

The new regression (5.5) seems likely to be most useful in cases where nonstationarity is suspected. Note, however, that when $d < \frac{1}{2}$, the correction term in (5.4) is $o_p(1)$ when $\frac{\sqrt{n}}{s} \to 0$, so that use of (5.5) can also be expected to be satisfactory in the stationary case. When $d = 1$, the correction is exact for all frequencies, as is clear from (3.9). In that case, therefore, (5.6) is an exact regression relation whose error is given by

$$a(\lambda_s) = \ln\left[I_u(\lambda_s)/f_u(\lambda_s)\right] + \ln\left(f_u(\lambda_s)/f_u(0)\right). \tag{5.8}$$

It is then a relatively straightforward matter to show that the modified log periodogram estimator has the following limit theory

$$\sqrt{m}\left(\tilde{d}-d\right)\xrightarrow{d} N\left(0,\frac{\pi^2}{24}\right), \tag{5.9}$$

that is the same limit distribution as the log periodogram estimator in the stationary case (Hurvich et al., 1998; Robinson, 1995b). By contrast, the usual log periodogram estimator $\hat{d}$ has a mixed normal limit theory when $d = 1$, as shown in Phillips (1999b, 2007). The mixed normal limit arises in this case because of the presence of the term $(2\pi)^{-\frac{1}{2}} e^{i\lambda_s} n^{-\frac{1}{2}} X_n$ in (3.9) which is $O_p(1)$ and does not vanish as $n \to \infty$. Moreover, the usual log periodogram estimator $\hat{d}$ is inconsistent and converges in probability to unity when $d \in (1,2)$ as shown in Kim and Phillips (2006), which makes use of some of the present methods.

The modified regression (5.5) appears to be even more useful in the nonstationary case when $d > 1$. In that case, the usual estimator $\hat{d}$ is inconsistent, and $\hat{d} \to_p 1$, a fact that can be established using the expansions obtained in Sections 2 and 3, whereas $\tilde{d}$ is consistent and has the same limit distribution as that shown in (5.9). Further analysis of this modified log periodogram estimator, together with an empirical application to the Nelson-Plosser data (Nelson & Plosser, 1982), was given in Kim and Phillips (2003).

The intuition leading to the modified regression (5.5) can also be employed in the case of the local Whittle estimator Künsch (1987) and Robinson (1995a). We will not go into details here. Suffice to remark that we would simply replace $I_v(\lambda_s;d)$ in the extremum estimation problem (5.16)–(5.18) given below by $I_v(\lambda_s)$, which can be computed from $v_x(\lambda_s)$ as in (5.4). The resulting estimator is a modified local Whittle estimator, and, like the modified log periodogram regression estimator in (5.5), its asymptotic properties can be expected to be the same for stationary and nonstationary values of the memory parameter, including those for which $d > 1$.

Our theory also suggests some other possibilities. In particular, we may build on the idea noted above that (5.6) gives an exact relationship when $d = 1$ with error (5.8). Indeed, the decomposition (3.8) implies the following exact relation between the transforms $w_x(\lambda_s)$ and $w_u(\lambda_s)$

$$w_x(\lambda_s) = D_n\left(e^{i\lambda_s};d\right)^{-1}\left[w_u(\lambda_s) + \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda n}(d)\right].$$

Define the new transform

$$v_x(\lambda_s;d) = w_x(\lambda_s) - D_n\left(e^{i\lambda_s};d\right)^{-1}\frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda n}(d), \tag{5.10}$$

which is dependent on the memory parameter d and for which the equation

$$v_x(\lambda_s;d) = D_n\left(e^{i\lambda_s};d\right)^{-1} w_u(\lambda_s) \tag{5.11}$$

holds exactly. Extending the ideas that led to (5.6) above, we have the exact periodogram relation

$$\ln\left(I_v(\lambda_s;d)\right) = c + \ln\left|D_n\left(e^{i\lambda_s};d\right)\right|^{-2} + a(\lambda_s), \tag{5.12}$$

with $I_v(\lambda_s;d) = v_x(\lambda_s;d)\,v_x(\lambda_s;d)^*$, and

$$a(\lambda_s) = \ln\left[I_u(\lambda_s)/f_u(\lambda_s)\right] + \ln\left(f_u(\lambda_s)/f_u(0)\right),$$

just as in (5.8). In place of linear least squares regression, it is now possible to apply nonlinear regression directly to the regression model (5.12). Let $Y_s(d) = \ln\left(I_v(\lambda_s;d)\right)$, and $A_s = \ln\left|D_n\left(e^{i\lambda_s};d\right)\right|^{-2}$. Then, nonlinear regression leads to the following extremum estimator

$$d^{\#} = \arg\min_d Q_m(d),$$

where

$$Q_m(d) = \frac{1}{m}\sum_{s=1}^{m}\left[\left\{Y_s(d) - \overline{Y_s(d)}\right\} - d\left\{A_s - \overline{A_s}\right\}\right]\left[\left\{Y_s(d) - \overline{Y_s(d)}\right\} - d\left\{A_s - \overline{A_s}\right\}\right]^*,$$

and $\overline{A_s} = m^{-1}\sum_{s=1}^{m} A_s$, $\overline{Y_s(d)} = m^{-1}\sum_{s=1}^{m} Y_s(d)$. The advantage of $d^{\#}$ is that it is the natural estimator of d that emerges from the exact formulation of the regression model in the frequency domain, that is, (5.12). Its disadvantage is that it is more complicated to compute than the conventional log periodogram regression estimator $\hat{d}$ and the modified estimator $\tilde{d}$, neither of which require numerical methods. Some simplifications in computation can be obtained by using some of the approximations developed in Sections 3 and 4.

Finally, we remark that the exact relationship (5.11) can be used to obtain an exact form of local Whittle estimator under Gaussian assumptions about u_t. The local Whittle likelihood suggested by Künsch (1987) and studied by Robinson (1995a) has the form

$$K_m(G,d) = \frac{1}{m}\sum_{s=1}^{m}\left[\log\left(G\lambda_s^{-2d}\right) + \frac{\lambda_s^{2d}}{G} I_x(\lambda_s)\right], \tag{5.13}$$

and is minimized jointly with respect to the parameters (G,d), where $G_0 = f_u(0)$ is the true value of G. The (negative) Whittle likelihood (e.g., Hannan & Deistler, 2012, pp. 224–225) based on frequencies up to λ_m and up to scale multiplication is

$$\sum_{s=1}^{m}\log f_u(\lambda_s)+\sum_{s=1}^{m}\frac{I_u(\lambda_s)}{f_u(\lambda_s)}. \tag{5.14}$$

The objective function (5.13) is derived from (5.14) by using the approximate relationship

$$w_x(\lambda_s)\sim\left(1-e^{i\lambda_s}\right)^{-d}w_u(\lambda_s)\sim(-\lambda_s)^{-d}w_u(\lambda_s),$$

or

$$I_u(\lambda_s)\sim\lambda_s^{2d}I_x(\lambda_s),$$

to transform (5.14) to be data dependent, in conjunction with the local approximation $f_u(\lambda_s)\sim G_0$. We may now proceed to transform (5.14) using the exact relationship between $w_u(\lambda_s)$ and $w_x(\lambda_s)$ that is given by (5.11) and (5.10). We get

$$\frac{1}{2}\sum_{s=1}^{m}\log\left\{\left|D_n\left(e^{i\lambda_s};d\right)\right|^{-2}f_u(\lambda_s)\right\}+\frac{1}{2}\sum_{s=1}^{m}\frac{\left|D_n\left(e^{i\lambda_s};d\right)\right|^{2}I_v(\lambda_s;d)}{f_u(\lambda_s)},$$

and this leads directly to the following "exact" version of the local Whittle likelihood

$$L_m(G,d)=\frac{1}{m}\sum_{s=1}^{m}\left[\log\left(\left|D_n\left(e^{i\lambda_s};d\right)\right|^{-2}G\right)+\frac{\left|D_n\left(e^{i\lambda_s};d\right)\right|^{2}}{G}I_v(\lambda_s;d)\right]. \tag{5.15}$$

The new estimates are obtained from the joint minimization

$$\left(G^{**},d^{**}\right)=\arg\min_{d,G}L_m(G,d).$$

Concentrating out G, we find that d^{**} satisfies

$$d^{**}=\arg\min_{d}R_m(d), \tag{5.16}$$

with

$$R_m(d)=\log G^{**}(d)-2\frac{1}{m}\sum_{j=1}^{m}\log\left|D_n\left(e^{i\lambda_s};d\right)\right|, \tag{5.17}$$

where

$$G^{**}(d)=\frac{1}{m}\sum_{j=1}^{m}\left|D_n\left(e^{i\lambda_s};d\right)\right|^{2}I_v(\lambda_s;d). \tag{5.18}$$

The estimator d^{**} would seem to offer an attractive semiparametric procedure because it is based on likelihood principles and involves the exact data generating mechanism for the dft. This procedure is more computationally intensive than the usual Whittle estimator but no impediment to practical use. A full analytic investigation of the exact local Whittle estimator was conducted and reported in Shimotsu and Phillips (2005) showing that the same asymptotic properties of the local Whittle estimator apply to the exact local Whittle estimator over a full range of stationary and nonstationary values of the memory parameter d. This approach enables consistent estimation of d and the construction of valid confidence intervals for d for both stationary and nonstationary long memory time series. The procedure has proved popular in empirical research. Further work on nonstationarity-extended Whittle estimation has been done by Abadir et al. (2007) and Shao (2010).

5.4. Final Remarks

Fractional processes conveniently embody in a single memory parameter d an index that measures the extent of long range dependence in an observed time series. When a nonparametric formulation is employed for the innovations that drive the observed process, a great deal of model generality is achieved. Integer values of d include integrated processes and the special value $d = \frac{1}{2}$ provides a simple boundary between stationary and nonstationary cases. This flexibility has enabled a fundamental extension of the simple ARIMA models popularized in the 1970s wherein variate differencing became a common method of dealing with nonstationarity. The flexibility of long memory also enriched the concept of cointegration by allowing for fractional possibilities in long run equilibrium errors, thereby narrowing the differential (between variables and errors) that distinguishes a cointegrating relationship among observable integrated time series. In view of this generality, semiparametric methods and frequency domain methods such as those used in the present work have been found to be very useful in estimation, inference, and asymptotic analysis of long memory systems.

In spite of the generality that long range dependence brings to empirical analysis, it is worth remembering that some important cases are not included in its orbit. Explosive and mildly explosive time series are prime examples that have particular relevance in economics and finance where exuberance and speculation are not uncommon in real estate and financial asset markets. A simple autoregressive time series with an explosive root is not rendered stationary by differencing or fractional differencing, just as differentiating an exponential function produces a derivative that simply reproduces the exponential. Parameterizations of nonstationarity using simple autoregressive coefficients and the tests that are so enabled by such formulations therefore offer possibilities that are not encompassed in the notion of long range dependence. While autoregressions and long memory systems provide a dual parametric source of unit root dynamics, these parameterizations deliver alternative departures from unit roots that help enrich our capacity to model different types of nonstationary time series behavior and evolution.

6. TECHNICAL APPENDIX AND PROOFS

6.1. Preliminary Results

We provide some technical lemmas that are useful throughout this chapter. Lemmas A and B provide results on binomial coefficients and hypergeometric functions that are either standard (e.g., Erdélyi, 1953) or follow from standard results. We give them here to facilitate our own derivations and to make this chapter more accessible. Lemmas C and D provide some more specific results on sinusoidal polynomials and hypergeometric functions of sinusoids that are immediately relevant to formulae in this chapter. Lemma E gives a useful inverse transform of fractional Brownian motion, an inverse transform for a hypergeometric series of fractional Brownian motion and some useful relationships between certain integral functionals of fractional Brownian motion and Brownian motion. Lemma F provides a new asymptotic expansion for hypergeometric series that allows for increasing coefficients as well as an argument that tends to unity. The expansion should be useful in other work with hypergeometric series.

6.1.1. Lemma A

(a) $\binom{d}{k} = (-1)^k \dfrac{(-d)_k}{k!}$.

(b) $(p+a)_j = \dfrac{(j+a)_p (a)_j}{(a)_p}$, $(a)_{j+k} = (a)_j (a+j)_k$.

(c) $\sum_{k=0}^{n} \dfrac{(-d)_k}{k!} = \dfrac{(1-d)_n}{n!}\mathbf{1}(d \neq 0,1,..) + \sum_{k=0}^{d} \dfrac{(-d)_k}{k!}\mathbf{1}(d = 0,1,..)$.

(d) $\dfrac{\Gamma(n+\alpha)}{\Gamma(n+\beta)} = n^{\alpha-\beta}\left[1+O\left(\dfrac{1}{n}\right)\right]$.

Proof

Part (a) is immediate from the definition

$$\binom{d}{k} = \frac{d!}{(d-k)!k!} = \frac{d(d-1)\ldots(d-k+1)}{k!} = (-1)^k \frac{(-d)\ldots(-d+k-1)}{k!}$$
$$= (-1)^k \frac{(-d)_k}{k!}.$$

The second formula in Part (b) is immediate from the definition of the forward factorial. The first formula in Part (b) follows from

$$(p+a)_j = \frac{\Gamma(p+a+j)}{\Gamma(p+a)} = \frac{\Gamma(p+a+j)}{\Gamma(j+a)} \frac{\Gamma(j+a)/\Gamma(a)}{\Gamma(p+a)/\Gamma(a)}$$
$$= (j+a)_p \frac{(a)_j}{(a)_p}.$$

For part (c), we write the sum as a terminating hypergeometric function, and use Lemma B (a) and (c) to obtain

$$\begin{aligned}\sum_{k=0}^{n}\frac{(-d)_k}{k!} &= \frac{(-d)_n}{n!}\ {}_2F_1(-n,1;d-n+1;1) \\ &= \frac{(-d)_n}{n!}\frac{\Gamma(d)\Gamma(d-n+1)}{\Gamma(d+1)\Gamma(d-n)} = \frac{\Gamma(-d+n)}{\Gamma(-d)n!}\frac{d-n}{d}, \\ &= \frac{\Gamma(-d+n+1)}{\Gamma(-d+1)n!} = \frac{(1-d)_n}{n!}\end{aligned}$$

for $d \neq 0,1,2,..$, while for $d = 0,1,..$ the sum $\sum_{k=0}^{n}\frac{(-d)_k}{k!}$ simply terminates at $k = d$.

Part (d) is a standard result that follows from the Stirling approximation, for example, Erdélyi (1953, p. 47).

6.1.2. Lemma B

In the following formulae, ${}_2F_1(a,b,c;z) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{k!(c)_k}z^k$ *is the hypergeometric function.*

(a) $\sum_{k=0}^{n}\frac{(-d)_k}{k!}z^k = \frac{(-d)_n}{n!}z^n\ {}_2F_1\left(-n,1;d-n+1;z^{-1}\right)\mathbf{1}(d \neq 0,1,..)$
$+\sum_{k=0}^{d}\frac{(-d)_k z^k}{k!}\mathbf{1}(d = 0,1,..).$

(b) $\sum_{t=m+1}^{\infty}\frac{(-d)_t}{t!}z^t = z^{m+1}\frac{(-d)_{m+1}}{(m+1)!}\,{}_2F_1(m+1-d,1;m+2;z).$

(c) ${}_2F_1(a,b,c;1) = \Gamma(c)\Gamma(c-a-b)/[\Gamma(c-a)\Gamma(c-b)]$ *for* $\mathrm{Re}(c-a-b) > 0$ and $c \neq 0,-1,-2,.....$

(d) *If* $|z| < 1$ and $|z/(z-1)| < 1$

$${}_2F_1(a,b;c;z) = (1-z)^{-a}\ {}_2F_1(a,c-b;c;z/(z-1)), \tag{6.1}$$

the right hand side giving an analytic continuation of the hypergeometric function to the half-plane $\mathrm{Re}(z) < \frac{1}{2}$.

(e) $\sum_{k=0}^{n}\frac{(-d)_k\left(e^{-i\lambda}\right)^k}{k!} = \frac{(1-d)_n e^{-i\lambda n}}{n!}\ {}_2F_1(-n,1;1-d;1-e^{i\lambda})\mathbf{1}(d \neq 0,1,..)$
$+\sum_{k=0}^{d}\frac{(-d)_k e^{-i\lambda k}}{k!}\mathbf{1}(d = 0,1,..).$

(f) *If* $\mathrm{Re}(c) > \mathrm{Re}(b) > 0$

$$ {}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\, dt, \tag{6.2}$$

which gives an analytic continuation of ${}_2F_1(a,b;c;z)$ *to the entire z plane cut along* $[1,\infty]$ *that is, to all z for which* $\arg(1-z) < \pi$.

Proof
Part (a) is given in Erdélyi (1953, pp. 87 and 101) in terms of binomial coefficients. Using the form given there and Lemma A (a), we have for $d \neq 0,1,...$

$$\begin{aligned}
\sum_{k=0}^{n} \frac{(-d)_k}{k!} z^k &= \sum_{k=0}^{n} \binom{d}{k} (-z)^k \\
&= \binom{d}{n} (-z)^n \;\, {}_2F_1\left(-n,1;d-n+1;z^{-1}\right) \\
&= \frac{(-d)_n}{n!} z^n \;\, {}_2F_1\left(-n,1;d-n+1;z^{-1}\right).
\end{aligned}$$

When $d = 0,1,..$ the sum simply terminates at $k = d$ and the stated result follows.

For part (b) we have

$$\begin{aligned}
\sum_{k=m+1}^{\infty} \frac{(-d)_k}{k!} x^k &= x^{m+1} \sum_{k=0}^{\infty} \frac{(-d)_{m+1+k}}{(m+1+k)!} x^k \\
&= x^{m+1} \sum_{k=0}^{\infty} \frac{\Gamma(m+1+k-d)}{\Gamma(-d)\Gamma(m+2+k)} x^k \\
&= x^{m+1} \sum_{k=0}^{\infty} \frac{(m+1-d)_k}{(m+2)_k} \frac{\Gamma(m+1-d)}{\Gamma(-d)\Gamma(m+2)} x^k \\
&= x^{m+1} \frac{\Gamma(m+1-d)}{\Gamma(-d)\Gamma(m+2)} \sum_{k=0}^{\infty} \frac{(m+1-d)_k}{(m+2)_k\, k!} x^k \qquad (6.3) \\
&= x^{m+1} \frac{\Gamma(m+1-d)}{\Gamma(-d)\Gamma(m+2)} \;\, {}_2F_1(m+1-d,1;m+2;x) \\
&= x^{m+1} \frac{(-d)_{m+1}}{(m+1)!} \;\, {}_2F_1(m+1-d,1;m+2;x).
\end{aligned}$$

The hypergeometric function ${}_2F_1(a,b,c;z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{k!(c)_k} z^k$ is absolutely convergent for all $|z| \leq 1$ when $\mathrm{Re}(a+b-c) < 0$ (Erdélyi, 1953, p. 57). Hence, the series in (6.3) converges absolutely for all $|z| \leq 1$ when $d > 0$.

Part (c) is a well-known summation formula (Erdélyi, 1953, p. 61). Part (d) is Euler's formula (Erdélyi, 1953, pp. 64 & 105). The series for ${}_2F_1(a,b,c;z)$

converges absolutely for all $|z|<1$ and converges absolutely for $|z|=1$ when $\mathrm{Re}(c-a-b)>0$ Erdélyi, 1953, p. 57). The series for ${}_2F_1(a,c-b;c;z/(z-1))$ converges for $|z/(z-1)|<1$. Since the latter inequality holds for all z for which $\mathrm{Re}(z)<\frac{1}{2}$, it follows that the right side of (6.1) gives the analytic continuation of ${}_2F_1(a,b;c;z)$ to the half-plane $\mathrm{Re}(z)<\frac{1}{2}$ (Erdélyi, 1953, p. 64).

Part (e) is obtained by direct calculation. Using (a), we proceed as follows for the case $d \neq 0,1,\ldots$:

$$\sum_{k=0}^{n}\frac{(-d)_k\left(e^{-i\lambda}\right)^k}{k!}=\frac{(-d)_n}{n!}\left(e^{-i\lambda}\right)^n\ {}_2F_1\left(-n,1;d-n+1;e^{i\lambda}\right) \tag{6.4}$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{j=0}^{n}\frac{(-n)_j(1)_j\left[1+\left(e^{i\lambda}-1\right)\right]^j}{j!(d-n+1)_j}$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{j=0}^{n}\frac{(-n)_j}{(d-n+1)_j}\sum_{q=0}^{j}\binom{j}{q}\left(e^{i\lambda}-1\right)^q$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{j=0}^{n}\frac{(-n)_j}{(d-n+1)_j}\sum_{q=0}^{j}\frac{j!}{(j-q)!q!}\left(e^{i\lambda}-1\right)^q$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{1}{q!}\left(e^{i\lambda}-1\right)^q\sum_{j=q}^{n}\frac{(-n)_j\, j!}{(d-n+1)_j(j-q)!}$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{1}{q!}\left(e^{i\lambda}-1\right)^q\sum_{s=0}^{n-q}\frac{(-n)_{s+q}(s+q)!}{(d-n+1)_{s+q}\, s!}. \tag{6.5}$$

Since $(-n)_{q+s}=(-n)_q(-n+q)_s$, and $(d-n+1)_{s+q}=(d-n+1)_q(d-n+1+q)_s$ from Lemma A (b), (6.5) becomes

$$\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q}{(d-n+1)_q}\left(e^{i\lambda}-1\right)^q\sum_{s=0}^{n-q}\frac{(q-n)_s(q+1)_s}{(d-n+1+q)_s\, s!}$$

$$=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q}{(d-n+1)_q}\left(e^{i\lambda}-1\right)^q\ {}_2F_1\left(q-n,q+1;q-n+d+1;1\right). \tag{6.6}$$

In this expression, the ${}_2F_1$ series terminates, so Lemma B (c) holds and (6.6) sums to

$$\begin{aligned}
&\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q}{(d-n+1)_q}\left(e^{i\lambda}-1\right)^q\frac{\Gamma(q-n+d+1)\Gamma(d-q)}{\Gamma(d+1)\Gamma(d-n)}\\
&=\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q(1)_q}{q!}\left(e^{i\lambda}-1\right)^q\frac{\Gamma(d-n+1)\Gamma(d-q)}{\Gamma(d+1)\Gamma(-n+d)}\\
&=\frac{d-n}{d}\frac{(-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q(1)_q}{q!}\left(e^{i\lambda}-1\right)^q\frac{\Gamma(d-q)}{\Gamma(d)}\\
&=\frac{(1-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q(1)_q}{q!}\left(e^{i\lambda}-1\right)^q\frac{1}{(d-1)(d-2)...(d-q)}\\
&=\frac{(1-d)_n e^{-i\lambda n}}{n!}\sum_{q=0}^{n}\frac{(-n)_q(1)_q}{q!}\left(e^{i\lambda}-1\right)^q\frac{(-1)^q}{(1-d)_q}\\
&=\frac{(1-d)_n e^{-i\lambda n}}{n!}\ {}_2F_1(-n,1,1-d;1-e^{i\lambda}),
\end{aligned}$$

giving the stated result for the case $d \neq 0,1,\ldots$. The result for $d = 0,1,..$ follows immediately because the series terminates at $k = d$. An alternative and more direct proof of the result makes use in (6.4) of the fact that

$$ {}_2F_1\left(-n,1;d-n+1;e^{i\lambda}\right)=\frac{(d-n)_n}{(d-n+1)_n}\ {}_2F_1\left(-n,1;1-d;1-e^{i\lambda}\right) \tag{6.7}$$

employing the linear transformation formula ${}_2F_1(-m,b;c;z)=\frac{(c-b)_m}{(c)_m}$ ${}_2F_1(-m,b;b-c-m;1-z)$ for terminating hypergeometric series – see Olver et al. (2010, Formula 15.8.7, p. 390).

Part (f) is a standard result (Erdélyi, 1953, p. 59).

6.1.3. Lemma C

Assume $d \neq 0,1,\ldots$ Then:

(a) *For fixed $\lambda \neq 0$ as $n \to \infty$*

$$\sum_{k=n+1}^{\infty}\frac{(-d)_k}{k!}e^{i\lambda k}=O\left(\frac{1}{n^{1+d}}\right).$$

(b) *For $\lambda_s = \frac{2\pi s}{n} \to 0$ and $s \to \infty$ as $n \to \infty$*

$$\sum_{k=n+1}^{\infty}\frac{(-d)_k}{k!}e^{i\lambda_s k}=-\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^d s}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right).$$

(c) *For* $\lambda_s = \dfrac{2\pi s}{n} \to 0$ *and s fixed as* $n \to \infty$

$$\sum_{k=n+1}^{\infty} \frac{(-d)_k}{k!} e^{i\lambda_s k} = O\left(\frac{1}{n^d}\right).$$

Proof

Using Lemma B (b), Lemma A (d), and Lemma F (b), given below, we get

$$\begin{aligned}
&\sum_{k=n+1}^{\infty} \frac{(-d)_k}{k!} e^{i\lambda k} \\
&= \left(e^{i\lambda}\right)^{n+1} \frac{\Gamma(n+1-d)}{\Gamma(-d)\Gamma(n+2)} \; {}_2F_1\left(n+1-d,1;n+2,e^{i\lambda}\right) \qquad (6.8)\\
&= e^{i\lambda(n+1)} \frac{1}{\Gamma(-d)n^{1+d}}\left[1+O\left(\frac{1}{n}\right)\right]\frac{1}{1-e^{i\lambda}}\left[1+O\left(\frac{1}{n}\right)\right] = O\left(\frac{1}{n^{1+d}}\right),
\end{aligned}$$

giving part (a). For $\lambda = \lambda_s = \dfrac{2\pi s}{n} \to 0$ and $s \to \infty$ as $n \to \infty$ we have, using Lemma F (a),

$$\begin{aligned}
&\left(e^{i\lambda_s}\right)^{n+1} \frac{\Gamma(n+1-d)}{\Gamma(-d)\Gamma(n+2)} \; {}_2F_1\left(n+1-d,1;n+2;e^{i\lambda_s}\right) \\
&= \frac{e^{i\lambda_s}}{\Gamma(-d)n^{1+d}}\left[1+O\left(\frac{1}{n}\right)\right] \; {}_2F_1\left(n+1-d,1;n+2;e^{i\lambda_s}\right) \\
&= \frac{e^{i\lambda_s}}{\Gamma(-d)n^{1+d}}\left\{\frac{1}{1-e^{i\lambda_s}}\sum_{j=0}^{k-1}\frac{(1+d)_j\,(1)_j}{j!}\left(\frac{1}{2\pi i s}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j\right. \\
&\qquad \left. +O\left(\left(\frac{1}{2\pi i s}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right)\right\}\left[1+O\left(\frac{1}{n}\right)\right] \\
&= \frac{1}{\Gamma(-d)n^{1+d}}\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\left[1+O\left(\frac{1}{s}\right)+O\left(\frac{1}{n}\right)\right] \\
&= \frac{1}{\Gamma(-d)n^{1+d}}\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right) \\
&= -\frac{1}{\Gamma(-d)n^{d}}\frac{\left[1+O\left(\frac{s}{n}\right)\right]}{2\pi s i}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right) \qquad (6.9)\\
&= -\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^{d}s}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right)
\end{aligned}$$

giving part (b). Finally, for s fixed as $n \to \infty$, we have

$$\sum_{k=n+1}^{\infty} \frac{(-d)_k}{k!} \left(e^{i\lambda_s}\right)^k = O\left(\sum_{k=n+1}^{\infty} \frac{1}{k^{1+d}}\right) = O\left(\frac{1}{n^d}\right),$$

giving part (c).

6.1.4. Lemma D

Assume $d \neq 1, 2, \ldots$, let $r \in (0,1)$ and let $\lambda_s = \frac{2\pi s}{n} \to 0$ *with s fixed as* $n \to \infty$. *Then*:

$$_2F_1(-\lfloor nr \rfloor, 1, 1-d; 1-e^{i\lambda_s}) = {}_1F_1(1, 1-d; 2\pi isr) + O\left(n^{-1}\right), \tag{6.10}$$

$$_2F_1(-\lfloor nr \rfloor, 1, 1-d; e^{-i\lambda_s} - 1) = {}_1F_1(1, 1-d; 2\pi isr) + O\left(n^{-1}\right), \tag{6.11}$$

and for nonnegative integer $p \leq n$

$$_2F_1(-p, 1, 1-d; 1-e^{-i\lambda_s}) = {}_1F_1\left(1, 1-d; -2\pi is\frac{p}{n}\right) + O\left(p^{-1}\right). \tag{6.12}$$

Proof

The same argument gives both results (6.10) and (6.11). We prove (6.10).

$$\begin{aligned}
&{}_2F_1(-\lfloor nr \rfloor, 1, 1-d; 1-e^{i\lambda_s}) \\
&= \sum_{j=0}^{\lfloor nr \rfloor} \frac{(-\lfloor nr \rfloor)_j}{(1-d)_j} \left(-\frac{2\pi is}{n} + O\left(n^{-2}\right)\right)^j \\
&= \sum_{j=0}^{\lfloor nr \rfloor} \frac{(1)_j \dfrac{(-\lfloor nr \rfloor)_j}{(-\lfloor nr \rfloor)^j}}{(1-d)_j\, j!} \left(2\pi isr + O\left(n^{-1}\right)\right)^j \\
&= \sum_{j=0}^{\infty} \frac{(1)_j}{(1-d)_j\, j!} (2\pi isr)^j + O\left(n^{-1}\right) - \sum_{j=\lfloor nr \rfloor+1}^{\infty} \frac{(1)_j}{(1-d)_j\, j!} (2\pi isr)^j \qquad (6.13) \\
&= {}_1F_1(1, 1-d; 2\pi isr) + O\left(n^{-1}\right)
\end{aligned}$$

because

$$\begin{aligned}
\sum_{j=N+1}^{\infty} \frac{(1)_j\, x^j}{(1-d)_j\, j!} &= x^{N+1} \sum_{k=0}^{\infty} \frac{x^k}{(1-d)_{k+N+1}} \\
&= \frac{x^{N+1}}{\Gamma(1-d)^{-1}} \sum_{k=0}^{\infty} \frac{x^k}{\Gamma(k+2+N-d)} \\
&= \frac{x^{N+1}\Gamma(1-d)}{\Gamma(N+2-d)} \sum_{k=0}^{\infty} \frac{x^k\, (1)_k}{(2+N-d)_k\, k!} \\
&= \frac{x^{N+1}\Gamma(1-d) e^{N+1-d}}{\sqrt{2\pi}\, (N+2-d)^{N+1-d}} \left[\sum_{k=0}^{\infty} \frac{x^k\, (1)_k}{(2+N-d)_k\, k!}\right] \left[1 + O\left(\frac{1}{N}\right)\right] \\
&= O\left(\frac{1}{N^{N-\delta}}\right)
\end{aligned}$$

for all $\delta > 0$ and all finite x. Line (6.13) above follows because, for $1 \le j \le \lfloor nr \rfloor$,

$$\left|1-\frac{(-\lfloor nr\rfloor)_j}{(-\lfloor nr\rfloor)^j}\right| = \left|1-(1)\left(1-\frac{1}{\lfloor nr\rfloor}\right)\ldots\left(1-\frac{j-1}{\lfloor nr\rfloor}\right)\right| \le \left|1-\left(1-\frac{j-1}{\lfloor nr\rfloor}\right)^j\right| = O\left(\frac{j^2}{\lfloor nr\rfloor}\right),$$

and

$$\begin{aligned}
&\frac{1}{n}\sum_{j=0}^{\lfloor nr\rfloor}\frac{(1)_j\, j^2}{(1-d)_j\, j!}(2\pi isr)^j \\
&= O\left(\frac{1}{n}\sum_{j=0}^{\lfloor nr\rfloor}\frac{(3)_j}{(1-d)_j\, j!}(2\pi isr)^j\right) \\
&= O\left(\frac{1}{n}\ {}_1F_1(3,1-d,2\pi isr)\right) = O\left(\frac{1}{n}\right),
\end{aligned} \tag{6.14}$$

since the ${}_1F_1$ function is everywhere convergent.

Next, for (6.12) we have

$$\begin{aligned}
&{}_2F_1\left(-p,1,1-d;1-e^{-i\lambda_s}\right) \\
&= \sum_{j=0}^{p}\frac{(-p)_j}{(1-d)_j}\left(\frac{2\pi is}{n}+O\left(n^{-2}\right)\right)^j \\
&= 1+\frac{(-p)}{1-d}\left(\frac{2\pi is}{n}+O\left(n^{-2}\right)\right)+\frac{(-p)(-p+1)}{(1-d)_2}\left(\frac{2\pi is}{n}+O\left(n^{-2}\right)\right)^2 \\
&\quad +\ldots+\frac{(-p)_p}{(1-d)_p}\left(\frac{2\pi is}{n}+O\left(n^{-2}\right)\right)^p \\
&= 1+\frac{(-1)}{1-d}\left(2\pi isr+O\left(n^{-1}\right)\right)+\frac{(-1)\left(-1+O\left(p^{-1}\right)\right)}{(1-d)_2}\left(2\pi is\frac{p}{n}+O\left(n^{-1}\right)\right)^2 \\
&\quad +\ldots+\frac{\left(-1+O\left(p^{-1}\right)\right)^p}{(1-d)_p}\left(2\pi is\frac{p}{n}+O\left(n^{-1}\right)\right)^p
\end{aligned} \tag{6.15}$$

$$\begin{aligned}
&= \sum_{j=0}^{p}\frac{(1)_j\left(1+O\left(\frac{j}{p}\right)\right)^j}{(1-d)_j\, j!}\left(-2\pi is\frac{p}{n}+O\left(n^{-1}\right)\right)^j \\
&= \sum_{j=0}^{\infty}\frac{(1)_j}{(1-d)_j\, j!}\left(-2\pi is\frac{p}{n}\right)^j+O\left(p^{-1}\right)++O\left(n^{-1}\right) \\
&\quad -\sum_{j=p+1}^{\infty}\frac{(1)_j}{(1-d)_j\, j!}\left(-2\pi is\frac{p}{n}\right)^j \\
&= {}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right)+O\left(n^{-1}\right)+O\left(p^{-1}\right),
\end{aligned}$$

giving (6.12). Again, line (6.15) above follows because

$$\frac{1}{p}\sum_{j=0}^{p}\frac{(1)_j\, j^2}{(1-d)_j\, j!}\left(-2\pi is\frac{p}{n}\right)^j = O\left(\frac{1}{p}\sum_{j=0}^{p}\frac{(3)_j}{(1-d)_j\, j!}\left(-2\pi is\frac{p}{n}\right)^j\right) = O\left(\frac{1}{p}\right).$$

6.1.5. Lemma E

(a) *For* $j = 1,2,\ldots$

$$\Gamma(j+1-d)^{-1}\int_0^r (r-s)^{j-d} B_{d-1}(s)\,ds = \Gamma(j)^{-1}\int_{q=0}^r (r-q)^{j-1} B(q)\,dq$$

and for $j = 0,1,2,\ldots$

$$\Gamma(j+1-d)^{-1}\int_0^r (r-s)^{j-d} B_{d-1}(s)\,ds = \Gamma(j+1)^{-1}\int_{q=0}^r (r-q)^j\, dB(q).$$

(b) $\Gamma(1-d)^{-1}\int_0^r {}_1F_1(1,1-d;-2\pi is(r-q))(r-q)^{-d}\, B_{d-1}(q)\,dq$

$= \int_{q=0}^r e^{-2\pi is(r-q)}\,dB(q).$

(c) $\dfrac{1}{\Gamma(1+d)}\displaystyle\int_0^1 {}_1F_1(1,1+d;-2\pi isr)\, r^d\, dB(1-r)$

$= \dfrac{1}{\Gamma(1-f)(-2\pi si)}\displaystyle\int_0^1 r^{-f}\ {}_1F_1(1,1-f;-2\pi isr)\,dB(1-r) + \dfrac{1}{(-2\pi si)}B_{d-1}(1).$

(d) $\dfrac{1}{\Gamma(1+d)}\displaystyle\int_0^1 {}_1F_1(1,1+d;-2\pi isr)\, r^d\, dB(1-r) = \int_0^1 e^{2\pi isr} B_{d-1}(r)\,dr.$

In the above formulae, B is Brownian motion with variance ω^2 and $B_{d-1}(r) = \dfrac{1}{\Gamma(d)}\displaystyle\int_0^r (r-s)^{d-1}dB(s)$ is a fractional Brownian motion initialized at the origin, as in Lemma 3.4.

Proof

To prove part (a) we use an operator approach with $D = \dfrac{d}{dx}$ and allow for fractional powers of D with a Weyl integral interpretation (see Lovoie et al., 1976; Phillips et al., 1986a) for the approach used here). The operator e^{qD} is treated at the translation operator, so that $e^{qD} f(x) = f(x+q)$. Setting $B_{d-1}(s) = 0$ for all $s \le 0$ we have

$$\frac{1}{\Gamma(j+1-d)}\int_0^r (r-s)^{j-d} B_{d-1}(s)\,ds = \frac{1}{\Gamma(j+1-d)}\int_{q=0}^\infty q^{j-d} B_{d-1}(r-q)\,dq$$

$$= \frac{1}{\Gamma(j+1-d)}\int_{q=0}^\infty q^{j-d} e^{-qD} B_{d-1}(r)\,dq = D^{d-j-1}B_{d-1}(x)|_{x=r}$$

$$= D^{d-j-1}D^{1-d}B(x)|_{x=r} = D^{-j}B(x)|_{x=r}$$

$$= \Gamma(j)^{-1}\int_{q=0}^r q^{j-1}B(r-q)\,dq = \Gamma(j)^{-1}\int_{q=0}^r (r-q)^{j-1}B(q)\,dq, \tag{6.16}$$

giving the first of the stated results and, consequently,

$$\int_0^r (r-s)^{j-d} B_{d-1}(s)\,ds = \frac{\Gamma(1-d)(1-d)_j}{\Gamma(j)} \int_{q=0}^r (r-q)^{j-1} B(q)\,dq.$$

To obtain the second form of the result we use integration by parts to give

$$\Gamma(j)^{-1} \int_{q=0}^r (r-q)^{j-1} B(q)\,dq = j^{-1}\Gamma(j)^{-1} \int_{q=0}^r (r-q)^j dB(q)$$
$$= \Gamma(j+1)^{-1} \int_{q=0}^r (r-q)^j dB(q). \tag{6.17}$$

Combining (6.16) and (6.17), we have

$$\frac{1}{\Gamma(j+1-d)} \int_0^r (r-s)^{j-d} B_{d-1}(s)\,ds = \Gamma(j+1)^{-1} \int_{q=0}^r (r-q)^j dB(q)$$

which holds also when $j = 0$, giving the inverse relation

$$\frac{1}{\Gamma(1-d)} \int_0^r (r-s)^{-d} B_{d-1}(s)\,ds = B(r), \tag{6.18}$$

(see Theorem 4.9). An alternate weak convergence proof of (6.18) is given in the proof of Theorem 4.9 below and, from this result, (6.17) can alternatively be obtained by subsequent integration.

To prove part (b) we proceed as follows:

$$\frac{1}{\Gamma(1-d)} \int_0^r {}_1F_1(1, 1-d; -2\pi is(r-q))(r-q)^{-d}\, B_{d-1}(q)\,dq$$
$$= \frac{1}{\Gamma(1-d)} \sum_{j=0}^{\infty} \frac{(1)_j(-1)^j}{(1-d)_j\, j!} \int_0^r (2\pi is(r-q))^j (r-q)^{-d}\, B_{d-1}(q)\,dq$$
$$= \frac{1}{\Gamma(1-d)} \sum_{j=0}^{\infty} \frac{(-2\pi is)^j}{(1-d)_j} \int_0^r (r-q)^{j-d} B_{d-1}(q)\,dq$$
$$= \sum_{j=0}^{\infty} \frac{(-2\pi is)^j (1-d)_j}{(1-d)_j\, \Gamma(j)} \int_{q=0}^r (r-q)^{j-1} B(q)\,dq$$
$$= \sum_{j=0}^{\infty} \frac{(-2\pi is)^j}{\Gamma(j)} \int_{q=0}^r (r-q)^{j-1} B(q)\,dq$$
$$= \sum_{j=0}^{\infty} \frac{(-2\pi is)^j}{j!} \int_{q=0}^r (r-q)^j dB(q) = \int_{q=0}^r e^{-2\pi is(r-q)} dB(q),$$

using (6.17) in the penultimate line. This proves part (b).

To prove part (c), we expand the ${}_1F_1$ function on the right side of the formula and use

$$B_{d-1}(1) = \frac{1}{\Gamma(d)} \int_0^1 (1-s)^{d-1} dB(s) = -\frac{1}{\Gamma(d)} \int_0^1 r^{d-1} dB(1-r),$$

to get

$$\begin{aligned}
&\frac{1}{\Gamma(1-f)(-2\pi si)}\int_0^1 r^{-f}\ {}_1F_1(1,1-f;-2\pi isr)\,dB(1-r)+\frac{1}{(-2\pi si)}B_{d-1}(1)\\
&=\frac{1}{\Gamma(1-f)(-2\pi si)}\sum_{j=0}^{\infty}\frac{(1)_j(-2\pi si)^j}{j!(1-f)_j}\int_0^1 r^{j-f}\,dB(1-r)\\
&\quad-\frac{1}{(-2\pi si)}\frac{1}{\Gamma(1-f)}\int_0^1 r^{-f}\,dB(1-r)\\
&=\frac{1}{\Gamma(1-f)(-2\pi si)}\sum_{j=1}^{\infty}\frac{(1)_j(-2\pi si)^j}{j!(1-f)_j}\int_0^1 r^{j-f}\,dB(1-r)\\
&=\sum_{j=1}^{\infty}\frac{(-2\pi si)^{j-1}}{\Gamma(j+1-f)}\int_0^1 r^{j-f}\,dB(1-r)\\
&=\sum_{k=0}^{\infty}\frac{(-2\pi si)^{k}}{\Gamma(k+1+d)}\int_0^1 r^{k+d}\,dB(1-r)\\
&=\frac{1}{\Gamma(1+d)}\sum_{k=0}^{\infty}\frac{(1)_k(-2\pi si)^k}{k!(1+d)_k}\int_0^1 r^{k+d}\,dB(1-r)\\
&=\frac{1}{\Gamma(1+d)}\int_0^1 {}_1F_1(1,1+d;-2\pi isr)\,r^d\,dB(1-r),
\end{aligned}$$

giving the stated result.

To prove part (d) we use the exponential expansion for $e^{2\pi isr}$ in the integral on the right side, giving

$$\begin{aligned}
\int_0^1 e^{2\pi isr}B_{d-1}(r)\,dr&=\int_0^1 e^{2\pi is(1-r)}B_{d-1}(1-r)\,dr=\int_0^1 e^{-2\pi isr}B_{d-1}(1-r)\,dr\\
&=\sum_{j=0}^{\infty}\frac{(-2\pi si)^j}{j!}\int_0^1 r^jB_{d-1}(1-r)\,dr\\
&=\sum_{j=0}^{\infty}\frac{(-2\pi si)^j}{j!}\int_0^1 (1-r)^jB_{d-1}(r)\,dr. \qquad (6.19)
\end{aligned}$$

From part (a) we have

$$\Gamma(j+1-d)^{-1}\int_0^r (r-s)^{j-d}B_{d-1}(s)\,ds=\Gamma(j+1)^{-1}\int_{q=0}^r (r-q)^j\,dB(q),$$

and setting $k=j-d$ and $r=1$ gives the formula

$$\Gamma(k+1)^{-1}\int_0^1 (1-s)^kB_{d-1}(s)\,ds=\Gamma(k+d+1)^{-1}\int_{q=0}^1 (1-q)^{k+d}\,dB(q),$$

or

$$\Gamma(k+1)^{-1}\int_0^1 s^k B_{d-1}(1-s)\,ds=\Gamma(k+d+1)^{-1}\int_{q=0}^1 q^{k+d}\,dB(1-q). \quad (6.20)$$

Using (6.20) in (6.19) we get

$$\begin{aligned}\int_0^1 e^{2\pi isr}B_{d-1}(r)\,dr &= \sum_{j=0}^{\infty}\frac{(-2\pi si)^j}{j!}\int_0^1 r^j B_{d-1}(1-r)\,dr\\ &= \sum_{j=0}^{\infty}\frac{(-2\pi si)^j}{j!}\frac{\Gamma(j+1)}{\Gamma(j+d+1)}\int_0^1 q^{j+d}\,dB(1-q)\\ &= \sum_{j=0}^{\infty}\frac{(-2\pi si)^j}{j!}\frac{(1)_j}{\Gamma(j+d+1)}\int_0^1 q^{j+d}\,dB(1-q)\\ &= \frac{1}{\Gamma(d+1)}\int_0^1\sum_{j=0}^{\infty}\frac{(-2\pi siq)^j}{j!}\frac{(1)_j}{(1+d)_j}q^d\,dB(1-q)\\ &= \frac{1}{\Gamma(d+1)}\int_0^1 {}_1F_1(1,1+d;-2\pi isq)\,q^d\,dB(1-q),\end{aligned}$$

giving the stated result.

6.1.6. Lemma F

Let α and β be constants for which $\mathrm{Re}(\beta),\mathrm{Re}(\beta-\alpha)>0$. The following asymptotic expansions to some given order k hold

(a) *Let* $\lambda_s=\dfrac{2\pi s}{n}$. *If* $\dfrac{s}{n}\to 0$ as $n\to\infty$ *and* $s\to\infty$, *then*

$$\begin{aligned}&{}_2F_1\left(\alpha,n-\beta;n;e^{i\lambda_s}\right)\\ &=(1-e^{i\lambda_s})^{-\alpha}\left[\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+O\left(\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right)\right]\\ &=(1-e^{i\lambda_s})^{-\alpha}\left[\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+O\left(\frac{1}{s^k}\right)\right].\end{aligned}$$

(b) *Let* $\lambda\neq 0$ *be fixed as* $n\to\infty$. *Then*

$$\begin{aligned}&{}_2F_1\left(\alpha,n-\beta;n;e^{i\lambda}\right)\\ &=(1-e^{i\lambda})^{-\alpha}\left[\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{n}\frac{e^{i\lambda}}{e^{i\lambda}-1}\left[1+O\left(\frac{1}{n}\right)\right]\right)^j+O\left(\frac{1}{n^k}\right)\right].\end{aligned}$$

(c) *Let* $\lambda_s = \dfrac{2\pi s}{n}$. *If* $\dfrac{s}{n} + \dfrac{n}{sp} \to 0$ as $n, s, p \to \infty$, *then*

$$\begin{aligned} &{}_2F_1\left(\alpha, p-\beta; p; e^{i\lambda_s}\right) \\ &= (1-e^{i\lambda_s})^{-\alpha}\left[\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j + O\left(\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right)\right]. \end{aligned}$$

Proof

Since $\mathrm{Re}(\beta-\alpha) > 0$, the series for ${}_2F_1\left(\alpha, n-\beta; n; e^{i\lambda_s}\right)$ converges absolutely for all λ_s. Using (6.1) from Lemma B (d), we write

$${}_2F_1\left(\alpha, n-\beta; n; e^{i\lambda_s}\right) = (1-e^{i\lambda_s})^{-\alpha}\ {}_2F_1\left(\alpha, \beta; n; \frac{e^{i\lambda_s}}{e^{i\lambda_s}-1}\right), \tag{6.21}$$

where the right side has a convergent series representation for suitable λ_s, viz., when $|e^{i\lambda_s}/(e^{i\lambda_s}-1)| < 1$, or $\cos(\lambda_s) < \dfrac{1}{2}$. Although the domain of convergence of the series on the right side series is restricted, the right hand side has a valid asymptotic expansion for large n that applies to all λ_s as we shall now show.

First observe that as $n, s \to \infty$ with $\dfrac{s}{n} \to 0$, the complex quantity

$$Z_{ns} = \frac{e^{i\lambda_s}}{e^{i\lambda_s}-1} = \frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right] = \frac{n}{2\pi is}[1+o(1)] \tag{6.22}$$

lies inside the plane cut along $[1,\infty]$, that is, $|\arg(1-Z_{ns})| < \pi$. Hence, we may use the analytic continuation of the right hand side of (6.21) based on the following integral representation (Erdélyi, 1953, p. 59; Lemma B(f)):

$${}_2F_1(\beta, \alpha; n; Z_{ns}) = \frac{\Gamma(n)}{\Gamma(\alpha)\Gamma(n-\alpha)}\int_0^1 t^{\alpha-1}(1-t)^{n-\alpha-1}(1-tZ_{ns})^{-\beta}\,dt. \tag{6.23}$$

An asymptotic series that is valid even for $|Z_{ns}| > 1$ for large n may now be obtained using a method due to MacRobert (see Erdélyi, 1953, p. 76) as follows. Expand the last binomial factor in (6.23) in MacLaurin's expansion up to k terms with remainder as

$$(1-tZ_{ns})^{-\beta} = \sum_{j=0}^{k-1}\frac{(\beta)_j}{j!}(tZ_{ns})^j + \frac{(\beta)_k}{k!}(tZ_{ns})^k\int_0^1 k(1-q)^{k-1}(1-qtZ_{ns})^{-\beta-k}\,dq.$$

Now scale this expansion by $\frac{\Gamma(n)}{\Gamma(\alpha)\Gamma(n-\alpha)}t^{\alpha-1}(1-t)^{n-\alpha-1}$ and integrate term by term, using the formula

$$\frac{\Gamma(n)}{\Gamma(\alpha)\Gamma(n-\alpha)}\int_0^1 t^{\alpha+j-1}(1-t)^{n-\alpha-1}\,dt=\frac{\Gamma(n)}{\Gamma(\alpha)\Gamma(n-\alpha)}\frac{\Gamma(\alpha+j)\Gamma(n-\alpha)}{\Gamma(n+j)}=\frac{(\alpha)_j}{(n)_j}.$$

This leads to

$$\begin{aligned}
{}_2F_1(\beta,\alpha;n;Z_{ns}) &= \sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{(n)_j\,j!}Z_{ns}^j+R_{kn}\\
&=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{(n)_j\,j!}\left(\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+R_{kn}\\
&=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\left[1+O\left(n^{-1}\right)\right]\right)^j+R_{kn}\\
&=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+R_{kn},
\end{aligned}\tag{6.24}$$

where

$$\begin{aligned}
R_{kn}&=\frac{(\beta)_k}{k!\,B(\alpha,n-\alpha)}\int_0^1 t^{\alpha-1}(1-t)^{n-\alpha-1}(tZ_{ns})^k\int_0^1 k(1-q)^{k-1}(1-qtZ_{ns})^{-\beta-k}\,dq\,dt\\
&=\frac{k(\alpha)_k(\beta)_k\left(\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k}{(n)_k\,k!\,B(\alpha+k,n-\alpha)}\\
&\quad\times\int_0^1 t^{\alpha+k-1}(1-t)^{n-\alpha-1}\int_0^1(1-q)^{k-1}\left(1-qt\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^{-\beta-k}dq\,dt,
\end{aligned}$$

since the beta function factors as follows

$$\begin{aligned}
\frac{1}{B(\alpha,n-\alpha)}&=\frac{\Gamma(n)}{\Gamma(\alpha)\Gamma(n-\alpha)}\\
&=\frac{\Gamma(\alpha+k)\Gamma(n)}{\Gamma(\alpha)\Gamma(n+k)}\frac{\Gamma(n+k)}{\Gamma(\alpha+k)\Gamma(n-\alpha)}=\frac{(\alpha)_k}{(n)_k\,B(\alpha+k,n-\alpha)}.
\end{aligned}$$

In view of (6.22) there exists a constant $c>0$ for which $\left|\mathrm{Im}(Z_{ns})\right|\ge c$. Then, for any given β and k, there exists an M, independent of n and s, such that

$$\sup_{t,q\in[0,1]}\left|(1-qtZ_{ns})^{-\beta-k}\right|<M.$$

Then,

$$\left|R_{kn}\right| \leq M \frac{k(\alpha)_k(\beta)_k\left(\dfrac{n}{2\pi is}\left[1+O\left(\dfrac{s}{n}\right)\right]\right)^k}{(n)_k k!B(\alpha+k,n-\alpha)} \int_0^1 t^{\alpha+k-1}(1-t)^{n-\alpha-1}\int_0^1(1-q)^{k-1}dq$$

$$= M \frac{k(\alpha)_k(\beta)_k\left(\dfrac{n}{2\pi is}\left[1+O\left(\dfrac{s}{n}\right)\right]\right)^k}{(n)_k k!B(\alpha+k,n-\alpha)} B(\alpha+k,n-\alpha)B(k,1)$$

$$= M \frac{k(\alpha)_k(\beta)_k\left(\dfrac{n}{2\pi is}\left[1+O\left(\dfrac{s}{n}\right)\right]\right)^k}{(n)_k k!} \frac{\Gamma(k)}{\Gamma(k+1)}$$

$$= M \frac{(\alpha)_k(\beta)_k\left(\dfrac{1}{2\pi is}\left[1+O\left(\dfrac{s}{n}\right)\right]\left[1+O\left(\dfrac{1}{n}\right)\right]\right)^k}{k!}$$

$$= M \frac{(\alpha)_k(\beta)_k}{k!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k,$$

so that R_{kn} has the same order of magnitude as the first neglected term in the expansion (6.24). Thus, (6.24) is a valid asymptotic expansion of the form

$$_2F_1\left(\beta,\alpha;n;\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)$$

$$=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+O\left(\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right),$$

giving the required result for part (a). Part (b) follows in an identical manner using

$$Z=\frac{e^{i\lambda}}{e^{i\lambda}-1}$$

in place of Z_{ns}.

To prove part (c) we proceed as in the proof of part (a), setting $Z_{ns}=\dfrac{e^{i\lambda_s}}{e^{i\lambda_s}-1}$ as in (6.22). Then

$$_2F_1(\beta,\alpha;p;Z_{ns})=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{(p)_j j!}Z_{ns}^j+R_{knp}$$

$$=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{(p)_j j!}\left(\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+R_{knp}$$

$$=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\left[1+O\left(p^{-1}\right)\right]\right)^j+R_{knp}$$

$$=\sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+R_{knp},$$

since $\frac{ps}{n} \to \infty$. The remainder is

$$R_{knp} = \frac{(\beta)_k}{k!B(\alpha, p-\alpha)} \int_0^1 t^{\alpha-1}(1-t)^{p-\alpha-1}(tZ_{ns})^k \int_0^1 k(1-q)^{k-1}(1-qtZ_{ns})^{-\beta-k}\,dqdt$$

$$= \frac{k(\alpha)_k(\beta)_k\left(\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k}{(p)_k k!B(\alpha+k, p-\alpha)}$$

$$\times \int_0^1 t^{\alpha+k-1}(1-t)^{p-\alpha-1}\int_0^1(1-q)^{k-1}\left(1-qt\frac{n}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^{-\beta-k} dqdt.$$

As in the case of R_{kn}, we have

$$\left|R_{knp}\right| \le M\frac{(\alpha)_k(\beta)_k}{k!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\left[1+O\left(p^{-1}\right)\right]\right)^k$$

$$= M\frac{(\alpha)_k(\beta)_k}{k!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k,$$

again since $\frac{ps}{n} \to \infty$. Thus, R_{knp} has the same order as the first neglected term in the series and we get the asymptotic expansion

$${}_2F_1\left(\beta,\alpha;p;\frac{e^{i\lambda_s}}{e^{i\lambda_s}-1}\right)$$

$$= \sum_{j=0}^{k-1}\frac{(\alpha)_j(\beta)_j}{j!}\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j + O\left(\left(\frac{n}{2\pi isp}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right),$$

which leads to the stated result.

6.2. Proofs of Main Lemmas and Theorems

6.2.1. Proof of Lemma 3.1

See Phillips and Solo (1992, formula (32)).

6.2.2. Proof of Theorem 3.2

From (3.2) we have the following alternate form for the model (2.1) for all $t \le n$

$$u_t = (1-L)^d X_t = D_n(L;d)X_t = D_n\left(e^{i\lambda};d\right)X_t + \tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)\left(e^{-i\lambda}L-1\right)X_t. \quad (6.25)$$

Observe that

$$\tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)\left(e^{-i\lambda}L-1\right)X_t = \left(e^{-i\lambda}L-1\right)\tilde{X}_{\lambda t} = e^{-i\lambda}\tilde{X}_{\lambda t-1}(d) - \tilde{X}_{\lambda t}(d), \quad (6.26)$$

where $\tilde{X}_{\lambda t}(d) = \tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)X_t = \sum_{p=0}^{n-1}\tilde{d}_{\lambda p}e^{-ip\lambda}X_{t-p}$. Since the right side of (6.26) is a telescoping Fourier sum, taking dfts of (6.26) leaves us with $\frac{1}{\sqrt{2\pi n}}\left(\tilde{X}_{\lambda 0}(d) - e^{in\lambda}\tilde{X}_{\lambda n}(d)\right)$. It follows that when we take dfts of expression (6.25) we have

$$\left[D_n\left(e^{i\lambda};d\right)\right]w_x(\lambda_s) + \frac{1}{\sqrt{2\pi n}}\left(\tilde{X}_{\lambda 0}(d) - e^{in\lambda}\tilde{X}_{\lambda n}(d)\right) = w_u(\lambda), \tag{6.27}$$

giving the required formula (3.3).

6.2.3. Proof of Theorem 3.7

Equation (3.11) follows immediately from the definition $(1-L)X_t = z_t$ and (3.9). Equation (3.12) follows by applying (3.8) to $z_t = (1-L)^{1-d}u_t$.

6.2.4. Proof of Lemma 4.2

Using the hypergeometric series representation from Lemma B (b), and the asymptotic expansion in Lemma A (d), we have for $d > 0$

$$\begin{aligned}
&D_n\left(e^{i\lambda};d\right) \\
&= \sum_{k=0}^{n}\frac{(-d)_k}{k!}e^{ik\lambda} = \left(\sum_{k=0}^{\infty} - \sum_{k=n+1}^{\infty}\right)\frac{(-d)_k}{k!}e^{ik\lambda} \\
&= \left(1-e^{i\lambda}\right)^d - e^{i(n+1)\lambda}\frac{\Gamma(n+1-d)}{\Gamma(-d)(n+1)!}\;{}_2F_1\left(n+1-d,1;n+2;e^{i\lambda}\right) \\
&= \left(1-e^{i\lambda}\right)^d - \frac{e^{i(n+1)\lambda}}{\Gamma(-d)n^{1+d}}\left[1+O\left(\frac{1}{n}\right)\right]\;{}_2F_1\left(n+1-d,1;n+2;e^{i\lambda}\right),
\end{aligned} \tag{6.28}$$

giving (4.1). Formula (4.2) follows immediately from Lemma B (d), noting that $|e^{i\lambda}/(e^{i\lambda}-1)| < 1s$ when $2\cos(\lambda) < 1$.

Next, using Lemma F (b), we have for fixed $\lambda \neq 0$,

$${}_2F_1\left(n+1-d,1;n+2;e^{i\lambda}\right) = (1-e^{i\lambda})^{-1}\left[1+O\left(\frac{1}{n}\right)\right]. \tag{6.29}$$

It follows from (6.28) and (6.29) that as $n \to \infty$ and for fixed $\lambda \neq 0$

$$D_n\left(e^{i\lambda};d\right) = \left(1-e^{i\lambda}\right)^d - \frac{1}{\Gamma(-d)n^{1+d}}\frac{e^{i(n+1)\lambda}}{1-e^{i\lambda}}\left[1+O\left(\frac{1}{n}\right)\right],$$

giving part (a).

When $\lambda_s = \frac{2\pi i s}{n} \to 0$ as $n \to \infty$ and $s \to \infty$, we proceed as follows. Using Lemma F (a) in the hypergeometric factor in the second term of (6.28), we have

$$
\begin{aligned}
&{}_2F_1\left(n+1-d,1;n+2;e^{i\lambda_s}\right)\\
&=\frac{1}{1-e^{i\lambda_s}}\sum_{j=0}^{k-1}\frac{(1+d)_j\,(1)_j}{j!}\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^j+O\left(\left(\frac{1}{2\pi is}\left[1+O\left(\frac{s}{n}\right)\right]\right)^k\right). \qquad (6.30)
\end{aligned}
$$

Then, as in the argument leading to (6.9), the second term of (6.28) admits the following valid asymptotic expansion for $\lambda=\lambda_s\to 0$ as $n\to\infty$ and $s\to\infty$:

$$
\begin{aligned}
&\frac{e^{i\lambda_s}}{\Gamma(-d)n^{1+d}}\left[1+O\left(\frac{1}{n}\right)\right]\ {}_2F_1\left(n+1-d,1;n+2;e^{i\lambda_s}\right)\\
&=-\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^d s}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right), \qquad (6.31)
\end{aligned}
$$

and so from (6.28) and (6.31) we get

$$
D_n\left(e^{i\lambda_s};d\right)=\left(1-e^{i\lambda_s}\right)^d+\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^d s}\left[1+O\left(\frac{1}{s}\right)\right]+O\left(\frac{1}{n^{1+d}}\right),
$$

giving part (b). The result can also be shown directly by noting from Lemma C (b) that

$$
\begin{aligned}
D_n\left(e^{i\lambda_s};d\right)&=\left(1-e^{i\lambda_s}\right)^d-\sum_{k=n+1}^{\infty}\frac{(-d)_k}{k!}e^{ik\lambda_s}\\
&=\left(1-e^{i\lambda_s}\right)^d+\frac{1}{2\pi i}\frac{1}{\Gamma(-d)n^d s}\left(1+O\left(\frac{1}{s}\right)\right)+O\left(\frac{1}{n^{1+d}}\right).
\end{aligned}
$$

For part (c), we start by using the following summation formula from Lemma B (e)

$$
\sum_{k=0}^{n}\frac{(-d)_k\left(e^{i\lambda_s}\right)^k}{k!}=\frac{(1-d)_n\,e^{i\lambda_s n}}{n!}\ {}_2F_1(-n,1,1-d;1-e^{-i\lambda_s}).
$$

Since s is fixed, we have from Lemma D (6.12) with $p=n$

$$
{}_2F_1(-n,1,1-d;1-e^{-i\lambda_s})={}_1F_1\left(1,1-d;-2\pi is\right)+O\left(n^{-1}\right).
$$

It follows that

$$
\begin{aligned}
\sum_{k=0}^{n}\frac{(-d)_k\left(e^{i\lambda_s}\right)^k}{k!}&=\frac{(1-d)_n\,e^{i\lambda_s n}}{n!}\left[{}_1F_1\left(1,1-d;-2\pi is\right)+O\left(n^{-1}\right)\right]\\
&=\frac{(1-d)_n}{n!}\ {}_1F_1\left(1,1-d;-2\pi is\right)+O\left(\frac{1}{n^{1+d}}\right). \qquad (6.32)
\end{aligned}
$$

and, then, for fixed s as $n\to\infty$, we have

$$D_n\left(e^{i\lambda_s};d\right)=\sum_{k=0}^{n}\frac{(-d)_k\, e^{i\lambda_s k}}{k!}=\frac{1}{\Gamma(1-d)n^d}\ {}_1F_1(1,1-d;-2\pi is)+O\left(\frac{1}{n^{1+d}}\right), \tag{6.33}$$

as required for part (c).

Part (d) follows as a special case of formula (6.33) with $s=0$. We also get the result directly from Lemma A (c), viz.,

$$D_n(1;d)=\sum_{k=0}^{n-1}\frac{(-d)_k}{k!}=\frac{(1-d)_{n-1}}{(n-1)!}=\frac{1}{\Gamma(1-d)}\frac{1}{n^d}\left[1+O\left(\frac{1}{n}\right)\right].$$

It follows that $D_n(1;d)$ differs from zero by a term of $O\left(n^{-d}\right)$.

6.2.5. Proof of Theorem 4.3

Parts (a) and (b). We write $\tilde{X}_{\lambda n}(d)$ as the sum of two components, the first involving $L+1$ components, with $1<L<n$ and where the choice of L will be discussed below. We have:

$$\tilde{X}_{\lambda n}(d)=\tilde{D}_{n\lambda}\left(e^{-i\lambda}L;d\right)X_n=\sum_{p=0}^{n-1}\tilde{d}_{\lambda p}e^{-ip\lambda}X_{n-p}=\sum_{p=0}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}X_{n-p}$$

$$=\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}X_{n-p}+\sum_{p=L+1}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}X_{n-p}.$$

Then

$$\frac{\tilde{X}_{\lambda n}(d)}{\sqrt{n}}=\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}$$

$$+\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}. \tag{6.34}$$

Next, look at the sinusoidal sum $\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}$ that appears in (6.34). We use the truncated binomial series formula from Lemma B (b) in this sum, giving

$$\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{i\lambda k}=\sum_{k=p+1}^{\infty}\frac{(-d)_k}{k!}e^{i\lambda k}-\sum_{k=n+1}^{\infty}\frac{(-d)_k}{k!}e^{i\lambda k}$$

$$=\left(e^{i\lambda}\right)^{p+1}\frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d,1;p+2,e^{i\lambda}\right)$$

$$-\left(e^{i\lambda}\right)^{n+1}\frac{(-d)_{n+1}}{(n+1)!}\ {}_2F_1\left(n+1-d,1;n+2,e^{i\lambda}\right). \tag{6.35}$$

For large n and fixed $\lambda\neq 0$, we have, using Lemma C (a),

$$\sum_{k=n+1}^{\infty}\frac{(-d)_k}{k!}e^{i\lambda k}=O\left(\frac{1}{n^{1+d}}\right), \tag{6.36}$$

while for $\lambda = \lambda_s = \dfrac{2\pi i s}{n} \to 0$ and $s \to \infty$ as $n \to \infty$ we have from Lemma C (b)

$$\sum_{k=n+1}^{\infty} \frac{(-d)_k}{k!} e^{i\lambda_s k} = -\frac{1}{\Gamma(-d)n^d} \frac{1}{2\pi i s}\left[1 + O\left(\frac{1}{s}\right)\right] + O\left(\frac{1}{n^{1+d}}\right). \tag{6.37}$$

So, neglecting the second term of (6.35) in view of (6.37), we get

$$\sum_{t=p+1}^{n} \frac{(-d)_t}{t!}\left(e^{i\lambda_s}\right)^t = \left(e^{i\lambda_s}\right)^{p+1} \frac{\Gamma(p+1-d)}{\Gamma(-d)(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) + O\left(\frac{1}{n^d s}\right) \tag{6.38}$$

for all $s \to \infty$, as $n \to \infty$. Finally, for s fixed as $n \to \infty$, we have from Lemma C (c)

$$\sum_{k=n+1}^{\infty} \frac{(-d)_k}{k!}\left(e^{i\lambda_s}\right)^k = O\left(\frac{1}{n^d}\right),$$

so that (6.38) also holds with s fixed.

Using (6.38), we deduce that

$$\begin{aligned}
&\frac{1}{n^{1-d}} \sum_{p=0}^{L}\left(\sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda_s}\right) e^{-ip\lambda_s} \\
&= e^{i\lambda_s}\ \frac{1}{n^{1-d}} \sum_{p=0}^{L} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) + O\left(\frac{L}{ns}\right) \\
&= e^{i\lambda_s}\ \frac{1}{n^{1-d}} \sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) \\
&\quad - e^{i\lambda_s}\ \frac{1}{n^{1-d}} \sum_{p=L+1}^{\infty} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) + O\left(\frac{L}{ns}\right).
\end{aligned} \tag{6.39}$$

Now

$$\begin{aligned}
&\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda}\right) \\
&= \sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \sum_{k=0}^{\infty} \frac{(1+p-d)_k (1)_k}{k!(p+2)_k} e^{i\lambda k} \\
&= \sum_{k=0}^{\infty} \sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1+p-d)_k}{(p+2)_k} e^{i\lambda k} \\
&= \sum_{k=0}^{\infty} \sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1-d+k)_p (2)_p}{(1-d)_p (k+2)_p} \frac{(1-d)_k}{(2)_k} e^{i\lambda k} \\
&= \sum_{k=0}^{\infty}\left[\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1-d+k)_p (2)_p}{(1-d)_p (k+2)_p}\right] \frac{(1-d)_k}{(2)_k} e^{i\lambda k}.
\end{aligned} \tag{6.40}$$

Next, since $(2)_p = (p+1)!$ and

$$(-d)_{p+1} = \frac{\Gamma(1-d+p)}{\Gamma(-d)} = \frac{(-d)\Gamma(1-d+p)}{\Gamma(1-d)} = (-d)(1-d)_p$$

we have

$$\begin{aligned}
&\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1-d+k)_p (2)_p}{(1-d)_p (k+2)_p} = (-d)\sum_{p=0}^{\infty} \frac{(1-d+k)_p}{(k+2)_p} \\
&= (-d)\sum_{p=0}^{\infty} \frac{(1-d+k)_p (1)_p}{(k+2)_p\, p!} = (-d)\ {}_2F_1(k+1-d, 1; k+2; 1) \\
&= (-d)\frac{\Gamma(k+2)\Gamma(d)}{\Gamma(k+1)\Gamma(1+d)} = -(k+1), \qquad (6.41)
\end{aligned}$$

where the explicit representation in the last line follows by the summation formula of Lemma B (c). Using (6.41) in (6.40) we get

$$\begin{aligned}
&\sum_{k=0}^{\infty} \left[\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1-d+k)_p (2)_p}{(1-d)_p (k+2)_p} \right] \frac{(1-d)_k}{(2)_k} e^{i\lambda k} = -\sum_{k=0}^{\infty} \frac{(k+1)(1-d)_k}{(2)_k} e^{i\lambda k} \\
&= -\sum_{k=0}^{\infty} \frac{(1-d)_k}{k!} e^{i\lambda k} = -\frac{1}{\left(1-e^{i\lambda}\right)^{1-d}}. \qquad (6.42)
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) \\
&= \sum_{k=0}^{\infty} \left[\sum_{p=0}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{(1-d+k)_p (2)_p}{(1-d)_p (k+2)_p} \right] \frac{(1-d)_k}{(2)_k} e^{i\lambda_s k} = -\frac{1}{\left(1-e^{i\lambda_s}\right)^{1-d}}. \qquad (6.43)
\end{aligned}$$

Next, using Lemma F (c) we find that for $\frac{s}{n} + \frac{n}{Ls} \to 0$ (which holds under the conditions on s and L that are given below),

$$\begin{aligned}
&\sum_{p=L+1}^{\infty} \frac{(-d)_{p+1}}{(p+1)!}\ {}_2F_1\left(1+p-d, 1; p+2, e^{i\lambda_s}\right) \\
&= O\left(\sum_{p=L+1}^{\infty} \frac{(-d)_{p+1}}{(p+1)!} \frac{1}{1-e^{i\lambda_s}} \left[1 + O\left(\frac{n}{sp}\right) \right] \right) \\
&= O\left(\frac{1}{1-e^{i\lambda_s}} \sum_{p=L+1}^{\infty} \frac{1}{p^{1+d}} \left[1 + O\left(p^{-1}\right)\right] \left[1 + O\left(\frac{n}{sp}\right) \right] \right) \\
&= O\left(\frac{1}{L^d} \frac{1}{1-e^{i\lambda_s}} \right). \qquad (6.44)
\end{aligned}$$

It follows from (6.39), (6.43), and (6.44) that

$$\begin{aligned}
\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda_s} &= -\frac{1}{n^{1-d}}\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}+O\left(\frac{L}{ns}\right)\\
&\quad+\frac{1}{n^{1-d}}\sum_{p=L+1}^{\infty}\frac{(-d)_{p+1}}{(p+1)!}\,{}_2F_1\left(1+p-d,1;p+2,e^{i\lambda_s}\right)\\
&= -\frac{1}{n^{1-d}}\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}+O\left(\frac{L}{ns}\right)+O\left(\frac{n^d}{L^d}\frac{1}{s}\right). \qquad (6.45)
\end{aligned}$$

The first term in (6.45) is $O\left(\frac{1}{s^{1-d}}\right)$ and dominates the second term. The first term also dominates the third term when $\frac{n}{Ls}\to 0$, which will be the case when $\frac{s}{n^\alpha}\to\infty$, as $n\to\infty$, for some $\alpha\in(0,1)$ and $L=\lfloor n^{1-\alpha}\rfloor$ and when $d<1$. (Note that for s fixed the last term of (6.45) does matter, and this distinguishes the s fixed case, which will be considered below in the proof of part (c).) Hence, when $n\to\infty$, $\lambda_s\to 0$, and $\frac{s}{n^\alpha}\to\infty$ (with L chosen as $L=\lfloor n^{1-\alpha}\rfloor$), we have

$$\begin{aligned}
&\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\\
&=\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\left[\frac{X_n}{n^{d-\frac{1}{2}}}+o_p(1)\right] \qquad (6.46)\\
&=-\frac{1}{n^{1-d}}\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}\frac{X_n}{n^{d-\frac{1}{2}}}+o_p\left(\frac{1}{s^{1-d}}\right)\\
&=-\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}\frac{X_n}{\sqrt{n}}+o_p\left(\frac{1}{s^{1-d}}\right). \qquad (6.47)
\end{aligned}$$

Line (6.46) above is justified by a separate argument, which we now develop. We use the fact, from Lemma 4.7, that $n^{\frac{1}{2}-d}X_{n-p}=O_p(1)$ and $p\le L=\lfloor n^{1-\alpha}\rfloor$. We proceed as follows. Select $K=\lfloor n^{1-\eta}\rfloor\to\infty$ with $0<\eta<\alpha$ (we will place a further condition on η below). Then, $\frac{L}{K}+\frac{K}{n}\to 0$ and we may write (for large n)

$$\frac{X_{n-p}}{n^{d-\frac{1}{2}}} = \frac{1}{n^{d-\frac{1}{2}}}\sum_{j=0}^{n-p}\frac{(d)_j}{j!}u_{n-p-j} = \frac{1}{n^{d-\frac{1}{2}}}\sum_{j=K+1}^{n-p}\frac{(d)_j}{j!}u_{n-p-j} + \frac{1}{n^{d-\frac{1}{2}}}\sum_{j=0}^{K}\frac{(d)_j}{j!}u_{n-p-j}$$

$$= \sum_{j=K+1}^{n-p}\frac{1}{\left(\frac{j}{n}\right)^{1-d}}\frac{u_{n-p-j}}{\sqrt{n}}\left[1+O_p\left(\frac{1}{K}\right)\right] + \left(\frac{K}{n}\right)^{d-\frac{1}{2}}\frac{1}{K^{d-\frac{1}{2}}}\sum_{j=0}^{K}\frac{(d)_j}{j!}u_{n-p-j}$$

$$= \sum_{j=K+1}^{n-p}\frac{1}{\left(\frac{j}{n}\right)^{1-d}}\frac{u_{n-p-j}}{\sqrt{n}} + O_p\left(\frac{1}{K}\right) + O_p\left(\left(\frac{K}{n}\right)^{d-\frac{1}{2}}\right)$$

$$= \sum_{j=K+1}^{n-p}\frac{1}{\left(\frac{j+p}{n}\right)^{1-d}}\frac{u_{n-p-j}}{\sqrt{n}}\left(\frac{j+p}{j}\right)^{1-d} + O_p\left(\frac{1}{K}\right) + O_p\left(\left(\frac{K}{n}\right)^{d-\frac{1}{2}}\right)$$

$$= \sum_{k=K+p+1}^{n}\frac{1}{\left(\frac{k}{n}\right)^{1-d}}\frac{u_{n-k}}{\sqrt{n}}\left(\frac{k}{k-p}\right)^{1-d} + O_p\left(\frac{1}{K}\right) + O_p\left(\left(\frac{K}{n}\right)^{d-\frac{1}{2}}\right)$$

$$= \sum_{k=K+p+1}^{n}\frac{1}{\left(\frac{k}{n}\right)^{1-d}}\frac{u_{n-k}}{\sqrt{n}} - \left(1+O\left(\frac{p}{k}\right)\right)^{d-1} + \sum_{k=1}^{K+p}\frac{1}{\left(\frac{k}{K+p}\right)^{1-d}}\frac{u_{n-k}}{\sqrt{K+p}}$$

$$= +O_p\left(\frac{p}{n^{d-\frac{1}{2}}}\sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}u_{n-k}\right) + O_p\left(\frac{1}{K}\right) + O_p\left(\left(\frac{K}{n}\right)^{d-\frac{1}{2}}\right). \tag{6.48}$$

Observe that for any $\delta > 0$, $\sum_{k=1}^{\infty}\frac{1}{k^{1+\delta}}u_{n-k}$ converges almost surely since $\sum_{k=1}^{\infty}\frac{1}{k^{1+\delta}}E\,|\,u_{n-k}\,| < \infty$. Then,

$$E\left|\sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}u_{n-k}\right| \le \sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}E\left|u_{n-k}\right| \le \sum_{k=K+p+1}^{\infty}\frac{1}{k^{2-d}}E\left|u_{n-k}\right|$$
$$\le \frac{1}{K^{1-d-\delta}}\sum_{k=K+p+1}^{\infty}\frac{1}{k^{1+\delta}}E\left|u_{n-k}\right| = o\left(\frac{1}{K^{1-d+\delta}}\right),$$

and so

$$\sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}u_{n-k} = o_p\left(\frac{1}{K^{1-d-\delta}}\right).$$

It follows that

$$\frac{p}{n^{d-\frac{1}{2}}}\sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}u_{n-k}=o_p\left(\frac{p}{n^{d-\frac{1}{2}}}\frac{1}{n^{(1-\eta)(1-d-\delta)}}\right)=o_p\left(\frac{L}{n^{\frac{1}{2}-\eta(1-d-\delta)-\delta}}\right)$$
$$=o_p\left(\frac{\sqrt{n}}{n^{\alpha-\eta(1-d-\delta)-\delta}}\right)$$

uniformly for $p\le L$. For $K=\lfloor n^{1-\eta}\rfloor$ and with η satisfying

$$0<\eta<\min\left(\alpha,\frac{\alpha-\frac{1}{2}-\delta}{1-d-\delta}\right),$$

and choosing δ such that $0<\delta<\alpha-\frac{1}{2}$, we have

$$\frac{p}{n^{d-\frac{1}{2}}}\sum_{k=K+p+1}^{n}\frac{1}{k^{2-d}}u_{n-k}=o_p(1), \tag{6.49}$$

uniformly for $p\le L$.

Using (6.49), we find that (6.48) can be written as

$$\frac{X_{n-p}}{n^{d-\frac{1}{2}}}=\left[\frac{1}{n^{d-\frac{1}{2}}}\sum_{k=0}^{n}\frac{(d)_k}{k!}u_{n-k}+o_p(1)\right]+O_p\left(\frac{K+p}{n}\right)^{d-\frac{1}{2}}+o_p(1)+O_p\left(\frac{1}{K}\right)$$
$$+O_p\left(\left(\frac{K}{n}\right)^{d-\frac{1}{2}}\right)$$
$$=\frac{X_n}{n^{d-\frac{1}{2}}}+O_p\left(\frac{K}{n}\right)^{d-\frac{1}{2}}+O_p\left(\frac{1}{K}\right)+o_p(1)=\frac{X_n}{n^{d-\frac{1}{2}}}+o_p(1),$$

uniformly for $p\le L=n^{1-\alpha}$ with $\alpha>\frac{1}{2}$, thereby establishing (6.46).

When $n\to\infty$ with fixed $\lambda\neq 0$, we have, in view of the use of (6.36) rather than (6.37) in the above arguments, the same expression but with an $o_p\left(n^{-(1-d)}\right)$ error. Specifically,

$$\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}$$
$$=-\frac{1}{n^{1-d}}\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{n^{d-\frac{1}{2}}}+o_p\left(\frac{1}{n^{1-d}}\right)+O\left(\frac{1}{n}\frac{n^d}{L^d}\frac{1}{1-e^{i\lambda}}\right)+O\left(\frac{1}{n^{1-d}}\frac{1}{n^d}\right)$$
$$=-\frac{1}{n^{1-d}}\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{n^{d-\frac{1}{2}}}+o_p\left(\frac{1}{n^{1-d}}\right). \tag{6.50}$$

In both cases the dominant approximation is given by the first term and we can write

$$\frac{1}{n^{1-d}}\sum_{p=0}^{L}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}=-\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}}+o_p\left(\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}}\right).$$

It remains to show that we may neglect the second term of (6.34). Using Lemma C (b), Lemma 4.7, (6.38), and Lemma F (c), we have, when $n\to\infty$, $\lambda_s\to 0$, $\frac{s}{n^{\alpha}}\to\infty$, and $L=n^{1-\alpha}$

$$\begin{aligned}
&\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\\
&=\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\left[\frac{e^{i\lambda_s(p+1)}(-d)_{p+1}}{(p+1)!}\,{}_2F_1\left(1+p-d,1;p+2,e^{i\lambda_s}\right)+O\left(\frac{1}{n^d s}\right)\right]e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\\
&=\frac{e^{i\lambda_s}}{n^{1-d}}\sum_{p=L+1}^{n-1}\left(\frac{(-d)_{p+1}}{(p+1)!}\,{}_2F_1\left(1+p-d,1;p+2,e^{i\lambda_s}\right)\right)\frac{X_{n-p}}{n^{d-\frac{1}{2}}}+O_p\left(\frac{1}{s}\right)\\
&=O\left(\frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\left(\frac{(-d)_{p+1}}{(p+1)!}\left[1+O\left(\frac{n}{sp}\right)\right]\right)\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\right)+O_p\left(\frac{1}{s}\right)\\
&=O_p\left(\frac{n^d}{L^d s}\right)+O_p\left(\frac{1}{s}\right),
\end{aligned}\tag{6.51}$$

which is $o_p\left(\frac{1}{s^{1-d}}\right)$ since $\frac{n}{Ls}\to 0$.

For the case of fixed $\lambda\neq 0$ and with $L=n^{1-\alpha}$ we get

$$\begin{aligned}
\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}&=O_p\left(\frac{1}{n^{1-d}}\sum_{p=L+1}^{n-1}\frac{1}{p^{1+d}}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\right)\\
&=O_p\left(\frac{1}{n^{1-d}L^d}\right)=O_p\left(\frac{1}{n^{1-\alpha d}}\right)=o_p\left(\frac{1}{n^{1-d}}\right).
\end{aligned}\tag{6.52}$$

In both cases (6.51) and (6.52), the order is smaller than the leading term of (6.47) and (6.50), respectively. Hence, for both fixed $\lambda\neq 0$ and $\lambda_s\to 0$ and $\frac{s}{n^{\alpha}}\to\infty$ as $n\to\infty$, we have

$$\begin{aligned}
\frac{\tilde{X}_{\lambda n}(d)}{\sqrt{n}}&=\frac{1}{n^{1-d}}\sum_{p=0}^{n}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}\\
&=-\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}}+o_p\left(\frac{e^{i\lambda}}{\left(1-e^{i\lambda}\right)^{1-d}}\frac{X_n}{\sqrt{n}}\right),
\end{aligned}$$

giving the required results.

Part (c). Our interest is in

$$\frac{\tilde{X}_{\lambda_s n}(d)}{\sqrt{n}} = \frac{1}{n^{1-d}} \sum_{p=0}^{n-1} \left(\sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda_s} \right) e^{-ip\lambda_s} \frac{X_{n-p}}{n^{d-\frac{1}{2}}}.$$

From Lemma B (e) we have

$$\sum_{k=0}^{m} \frac{(-d)_k \left(e^{i\lambda_s}\right)^k}{k!} = \frac{(1-d)_m e^{i\lambda_s m}}{m!} \; {}_2F_1(-m,1,1-d;1-e^{-i\lambda_s}). \tag{6.53}$$

Since s is fixed, $1-e^{-i\lambda_s} = \dfrac{2\pi is}{n} + O\left(n^{-2}\right)$ and using Lemma D and (6.53) we get

$$\sum_{k=0}^{n} \frac{(-d)_k \left(e^{i\lambda_s}\right)^k}{k!} = \frac{(1-d)_n}{n!} \; {}_1F_1(1,1-d;-2\pi is) + O\left(\frac{1}{n^{1+d}}\right). \tag{6.54}$$

Using (6.53) with $m = p$ and Lemma D again we obtain

$$\begin{aligned} \sum_{k=0}^{p} \frac{(-d)_k \left(e^{i\lambda_s}\right)^k}{k!} &= \frac{(1-d)_p e^{i\lambda_s p}}{p!} \; {}_2F_1(-p,1,1-d;1-e^{-i\lambda_s}) \\ &= \frac{(1-d)_p e^{i\lambda_s p}}{p!} \; {}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right) + O\left(\frac{1}{p^{1+d}}\right). \end{aligned} \tag{6.55}$$

Now $n^{\frac{1}{2}-d} X_{n-p} = O_p(1)$, uniformly in $p \le n$, so that

$$\frac{1}{n^{1-d}} \sum_{p=0}^{n} \left(\sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda_s} \right) e^{-ip\lambda_s} \frac{X_{n-p}}{n^{d-\frac{1}{2}}} = \left[\frac{1}{n^{1-d}} \sum_{p=0}^{n} \left(\sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda_s} \right) e^{-ip\lambda_s} \right] O_p(1).$$

Using (6.54) and (6.55) and noting that $\sum_{p=0}^{n} p^{-1-d} = O(1)$, we have

$$\begin{aligned} &\frac{1}{n^{1-d}} \sum_{p=0}^{n} \left(\sum_{k=p+1}^{n} \frac{(-d)_k}{k!} e^{ik\lambda_s} \right) e^{-ip\lambda_s} \frac{X_{n-p}}{n^{d-\frac{1}{2}}} \\ &= \frac{1}{n^{1-d}} \sum_{p=0}^{n} \left\{ \frac{(1-d)_{n1} F_1(1,1-d;-2\pi is)}{n!} - \frac{(1-d)_p e^{i\lambda_s p} \; {}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right)}{p!} \right\} e^{-ip\lambda_s} \frac{X_{n-p}}{n^{d-\frac{1}{2}}} \\ &\quad + O_p\left(\frac{1}{n^{1-d}}\right). \end{aligned}$$

Next observe that, since s is fixed as $n \to \infty$,

$$\frac{1}{n^{1-d}}\frac{(1-d)_n}{n!}\sum_{p=0}^{n} e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}} = \frac{1}{\Gamma(1-d)}\frac{1}{n}\sum_{p=0}^{n} e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}} + O_p\left(\frac{1}{n}\right)$$

$$= \frac{1}{\Gamma(1-d)}\frac{1}{n}\sum_{p=0}^{n} e^{-2\pi is\frac{p}{n}}\frac{X_{n-p}}{n^{d-\frac{1}{2}}} + O_p\left(\frac{1}{n}\right) = \frac{1}{\Gamma(1-d)}\int_0^1 e^{-2\pi isr}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n}\right)$$

$$= \frac{1}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr + O_p\left(\frac{1}{n}\right).$$

Further,

$$\frac{1}{n^{1-d}}\sum_{p=0}^{n}\frac{(1-d)_p\ {}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right)}{p!}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}$$

$$= \frac{1}{\Gamma(1-d)}\frac{1}{n^{1-d}}\sum_{p=1}^{n}\frac{{}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right)}{p^d}\frac{X_{n-p}}{n^{d-\frac{1}{2}}} + O_p\left(\frac{1}{n^{1-d}}\right)$$

$$= \frac{1}{\Gamma(1-d)}\frac{1}{n}\sum_{p=1}^{n}\frac{{}_1F_1\left(1,1-d;-2\pi is\frac{p}{n}\right)}{\left(\frac{p}{n}\right)^d}\frac{X_{n-p}}{n^{d-\frac{1}{2}}} + O_p\left(\frac{1}{n^{1-d}}\right)$$

$$= \frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right).$$

We deduce that

$$\frac{1}{n^{1-d}}\sum_{p=0}^{n}\left(\sum_{k=p+1}^{n}\frac{(-d)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}$$

$$= \frac{1}{n^{1-d}}\sum_{p=0}^{n}\left\{\frac{(1-d)_{n1}F_1(1,1-d;-2\pi is)}{n!} - \frac{(1-d)_p e^{i\lambda_s p}\ {}_1F_1(1,1-d;-2\pi isr)}{p!}\right\}e^{-ip\lambda_s}\frac{X_{n-p}}{n^{d-\frac{1}{2}}}$$

$$+ O_p\left(\frac{1}{n^{1-d}}\right)$$

$$= \frac{1}{\Gamma(1-d)}\left[\int_0^1 e^{-2\pi isr}\ {}_1F_1(1,1-d;-2\pi is)X_{n,d}(1-r)dr\right.$$

$$\left.- \int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr\right] + O_p\left(\frac{1}{n^{1-d}}\right)$$

$$= \frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr$$

$$- \frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right),$$

giving the stated result.

Part (d). When $d = 1$ the series expression for $n^{-\frac{1}{2}}\tilde{X}_{\lambda n}(d)$ terminates because $(-d)_k = 0$ for all $k > 1$, so that only the term involving $p = 0$ is retained. We then have

$$\frac{\tilde{X}_{\lambda n}(1)}{\sqrt{n}} = -e^{i\lambda}\frac{X_n}{\sqrt{n}},$$

which holds for all λ.

6.2.6. Proof of Theorem 4.4

By definition, $z_t = (1-L)^{1-d} u_t = (1-L)^f u_t$, and from Theorem 3.7 we have

$$\begin{aligned} w_x(\lambda)\left(1-e^{i\lambda}\right) &= w_z(\lambda) - e^{i\lambda}\frac{X_n}{\sqrt{2\pi n}} \\ &= D_n\left(e^{i\lambda};f\right)w_u(\lambda) - \frac{e^{i\lambda n}}{\sqrt{2\pi n}}\tilde{U}_{\lambda n}(f) - e^{i\lambda}\frac{X_n}{\sqrt{2\pi n}}, \end{aligned}$$

where

$$\tilde{U}_{\lambda n}(f) = \tilde{D}_{n\lambda}\left(e^{-i\lambda}L;f\right)u_n = \sum_{p=0}^{n-1}\tilde{f}_{\lambda p}e^{-ip\lambda}u_{n-p}, \quad \text{and} \quad \tilde{f}_{\lambda p} = \sum_{k=p+1}^{n}\frac{(-f)_k}{k!}e^{ik\lambda}.$$

Now, as in Lemma B (e), we have

$$\begin{aligned} \frac{\tilde{U}_{\lambda_s n}(f)}{\sqrt{2\pi n}} &= \frac{1}{\sqrt{2\pi n}}\sum_{p=0}^{n-1}\tilde{f}_{\lambda_s p}e^{-ip\lambda_s}u_{n-p} \\ &= \frac{1}{\sqrt{2\pi}}\sum_{p=0}^{n-1}\left(\sum_{k=p+1}^{n}\frac{(-f)_k}{k!}e^{ik\lambda_s}\right)e^{-ip\lambda_s}\frac{u_{n-p}}{\sqrt{n}} \\ &= \frac{1}{\sqrt{2\pi}}\sum_{p=0}^{n}\left\{\frac{(1-f)_n e^{i\lambda_s n}}{n!}\,{}_2F_1\left(-n,1,1-f;1-e^{-i\lambda_s}\right)\right. \\ &\quad \left. -\frac{(1-f)_p e^{i\lambda_s p}}{p!}\,{}_2F_1(-p,1,1-f;1-e^{-i\lambda_s})\right\}e^{-ip\lambda_s}\frac{u_{n-p}}{\sqrt{n}}. \end{aligned}$$

As in the proof of Theorem 4.3 and using the fact that $\sum_{p=1}^{n} p^{-1-f}u_{n-p} = O_p(1)$ as $n \to \infty$, we proceed as follows

$$
\begin{aligned}
&\frac{\tilde{U}_{\lambda_s n}(f)}{\sqrt{2\pi n}} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{p=0}^{n} \left\{ \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \right. \\
&\quad \left. - \frac{(1-f)_p \, e^{i\lambda_s p} \; {}_1F_1\left(1,1-f;-2\pi is \frac{p}{n}\right)}{p!} + O\left(\frac{1}{p^{1+f}}\right) \right\} e^{-ip\lambda_s} \frac{u_{n-p}}{\sqrt{n}} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{p=0}^{n} \left\{ \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \right. \\
&\quad \left. - \frac{(1-f)_p \, e^{i\lambda_s p} \; {}_1F_1\left(1,1-f;-2\pi is \frac{p}{n}\right)}{p!} \right\} e^{-ip\lambda_s} \frac{u_{n-p}}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \sum_{p=0}^{n} e^{-ip\lambda_s} \frac{u_{n-p}}{\sqrt{n}} \\
&\quad - \frac{1}{\sqrt{2\pi}} \sum_{p=0}^{n} \frac{(1-f)_p \; {}_1F_1\left(1,1-f;-2\pi is \frac{p}{n}\right)}{p!} \frac{u_{n-p}}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \int_0^1 e^{-2\pi isr} dX_n(1-r) \\
&\quad - \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f)} \sum_{p=0}^{n} \frac{1}{p^f} \left[1 + O\left(\frac{1}{p}\right)\right] {}_1F_1\left(1,1-f;-2\pi is \frac{p}{n}\right) \frac{u_{n-p}}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \int_0^1 e^{-2\pi isr} dX_n(1-r) \\
&\quad - \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f) n^f} \sum_{p=0}^{n} \frac{1}{\left(\frac{p}{n}\right)^f} \, {}_1F_1\left(1,1-f;-2\pi is \frac{p}{n}\right) \frac{u_{n-p}}{\sqrt{n}} + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{(1-f)_n \; {}_1F_1(1,1-f;-2\pi is)}{n!} \int_0^1 e^{-2\pi isr} dX_n(1-r) \\
&\quad - \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f) n^f} \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) \, dX_n(1-r) + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(1-f) n^f} \left\{ {}_1F_1(1,1-f;-2\pi is) \int_0^1 e^{-2\pi isr} dX_n(1-r) \right. \\
&\quad \left. - \int_0^1 r^{-f} \; {}_1F_1(1,1-f;-2\pi isr) \, dX_n(1-r) \right\} + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

So we have

$$\frac{\tilde{U}_{\lambda_s n}(f)}{\sqrt{2\pi n}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(1-f)n^f}\left\{ {}_1F_1(1,1-f;-2\pi is)\int_0^1 e^{-2\pi isr} dX_n(1-r) \right.$$
$$\left. -\int_0^1 r^{-f}\ {}_1F_1(1,1-f;-2\pi isr)\, dX_n(1-r)\right\} + O_p\left(\frac{1}{\sqrt{n}}\right),$$

as required. Note that when $f = 0$, we get

$${}_1F_1(1,1;-2\pi is) = e^{-2\pi si} = 1, \quad {}_1F_1(1,1;-2\pi isr) = e^{-2\pi isr},$$

and $\tilde{U}_{\lambda_s n}(0) = 0$.

6.2.7. Proof of Lemma 4.7

Akonom and Gouriéroux (1987) prove the result when u_t follows a stationary and invertible ARMA process. Using the device in Phillips and Solo (1992), we write

$$u_t = C(L)\varepsilon_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t,$$

where $\tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t = \sum_{j=0}^{\infty} \tilde{c}_j \varepsilon_{t-j}$ and $\tilde{c}_j = \sum_{k=j+1}^{\infty} c_k$. Under (2.4), $\tilde{\varepsilon}_t$ is stationary with mean zero and finite variance $\sigma^2 \sum_{j=0}^{\infty} \tilde{c}_j^2$. Then

$$X_t = (1-L)^{-d} u_t = C(1)(1-L)^{-d}\varepsilon_t - (1-L)^{1-d}\tilde{\varepsilon}_t.$$

Now for $\frac{1}{2} < d \le 1$, $\xi_t = (1-L)^{1-d}\tilde{\varepsilon}_t$ is stationary with mean zero and finite variance, so that $n^{\frac{1}{2}-d}\xi_{\lfloor nr \rfloor} \to_p 0$. On the other hand, $X_t^{\varepsilon} = (1-L)^{-d}\varepsilon_t$ is a fractional process constructed from $iid(0,\sigma^2)$ innovations with $E|\varepsilon_t|^p < \infty$, and so from Akonom and Gouriéroux (1987)

$$X_{n,d}^{\varepsilon}(r) = \frac{1}{n^{d-\frac{1}{2}}} X_{\lfloor nr \rfloor}^{\varepsilon} \xrightarrow{d} \frac{\sigma}{\Gamma(d)}\int_0^r (r-s)^{d-1} dW(s).$$

It follows that

$$X_{n,d}(r) = \frac{1}{n^{d-\frac{1}{2}}} X_{\lfloor nr \rfloor} \xrightarrow{d} B_{d-1}(r) = \frac{\sigma C(1)}{\Gamma(d)}\int_0^r (r-s)^{d-1} dW(s)$$
$$= \frac{1}{\Gamma(d)}\int_0^r (r-s)^{d-1} dB(s),$$

as stated.

6.2.8. Proof of Lemma 4.8

By Theorem 4.3 (c), Lemma 4.7 and the continuous mapping theorem we have

$$
\begin{aligned}
\frac{\tilde{X}_{\lambda n}(d)}{\sqrt{n}} &= \frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}X_{n,d}(r)dr \\
&\quad -\frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}X_{n,d}(1-r)dr+O_p\left(\frac{1}{n^{1-d}}\right) \\
&\xrightarrow{d} \frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{-2\pi isr}B_{d-1}(1-r)dr \\
&\quad -\frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}B_{d-1}(1-r)dr.
\end{aligned}
\tag{6.56}
$$

In the above, we can replace $X_{n,d}(r)$ by a continuous polygonal version up to an $o_p(1)$ error uniformly over $r\in[0,1]$. The continuous mapping theorem then applies since the mapping $f\mapsto\int_0^1 r^{-d}f(1-r)dr$ is continuous when $d<1$ for all continuous functions f, and since the confluent hypergeometric function ${}_1F_1(a,c;x)$ is an entire function of x.

Now observe from Lemma E that

$$
\Gamma(1-d)^{-1}\int_0^1 {}_1F_1(1,1-d;-2\pi is(1-q))(1-q)^{-d}\,B_{d-1}(q)dq=\int_{q=0}^1 e^{-2\pi is(1-q)}dB(q).
$$

It follows that (6.56) is

$$
\begin{aligned}
&\frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{-2\pi isr}B_{d-1}(1-r)dr \\
&\quad -\frac{1}{\Gamma(1-d)}\int_0^1 {}_1F_1(1,1-d;-2\pi isr)r^{-d}B_{d-1}(1-r)dr \\
&=\frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi is(1-r)}B_{d-1}(1-r)dr-\int_{q=0}^1 e^{-2\pi is(1-q)}dB(q) \\
&=\frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}B_{d-1}(r)dr-\int_{q=0}^1 e^{2\pi isq}dB(q).
\end{aligned}
\tag{6.57}
$$

Then,

$$
\frac{\tilde{X}_{\lambda n}(d)}{\sqrt{n}}\xrightarrow{d}\frac{{}_1F_1(1,1-d;-2\pi is)}{\Gamma(1-d)}\int_0^1 e^{2\pi isr}B_{d-1}(r)dr-\int_{q=0}^1 e^{2\pi isq}dB(q),
\tag{6.58}
$$

giving the first stated result.

6.2.9. Proof of Theorem 4.9

We offer two proofs of (4.20). The first is by operational techniques and is given in the Proof of Lemma E (a) – see (6.18). The second is by way of weak convergence of the two sides of (3.8) as $n \to \infty$. At $\lambda_s = 0$, (3.8) is

$$\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} u_t = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} X_t D_n(1,d) - \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda_0,n}(d). \tag{6.59}$$

From Lemma A (c) for $d \in \left(\frac{1}{2},1\right]$

$$\begin{aligned} D_n(1,d) &= \sum_{k=0}^{n}\frac{(-d)_k}{k!} = \frac{(1-d)_n}{n!} \\ &= \frac{1}{\Gamma(1-d)n^d}\left[1+O\left(n^{-1}\right)\right], \end{aligned}$$

so that

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t D_n(1,d) &= \frac{1}{\Gamma(1-d)}\frac{1}{n}\sum_{t=1}^{n}\frac{X_t}{n^{d-\frac{1}{2}}}\left[1+O\left(n^{-1}\right)\right] \\ &\xrightarrow{d} \frac{1}{\Gamma(1-d)}\int_0^1 B_{d-1}(r)dr. \end{aligned} \tag{6.60}$$

From Theorem 4.3 (c), (4.3), Lemma 4.7, and the continuous mapping theorem we have

$$\begin{aligned} \frac{\tilde{X}_{\lambda_0 n}(d)}{\sqrt{n}} &= \frac{1}{\Gamma(1-d)}\int_0^1 X_{n,d}(r)dr - \frac{1}{\Gamma(1-d)}\int_0^1 r^{-d}X_{n,d}(1-r)dr + O_p\left(\frac{1}{n^{1-d}}\right) \\ &\xrightarrow{d} \frac{1}{\Gamma(1-d)}\left[\int_0^1 B_{d-1}(r)dr - \int_0^1 r^{-d}B_{d-1}(1-r)dr\right]. \end{aligned} \tag{6.61}$$

It follows from (6.59), (6.60), and (6.61) that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \xrightarrow{d} B(1) = \frac{1}{\Gamma(1-d)}\int_0^1 (1-r)^{-d}B_{d-1}(r)dr, \tag{6.62}$$

Applying the same argument to the relation

$$\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{\lfloor nr\rfloor} u_t = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{\lfloor nr\rfloor} X_t D_n(1,d) - \frac{1}{\sqrt{2\pi n}}\tilde{X}_{\lambda_0,\lfloor nr\rfloor}(d),$$

instead of (6.59), we obtain the more general formula

$$B(r) = \frac{1}{\Gamma(1-d)}\int_0^r (r-q)^{-d}B_{d-1}(q)dq.$$

To prove (4.19), we can proceed in the same way using (3.8) and Theorem 4.3 (c). Or we can employ operational techniques, as in the Proof of Lemma E (b), which gives the stated result directly.

6.2.10. Proof of Theorem 4.10

Part (a) follows from the representation (4.6) and standard results on the asymptotic behavior of the dft of a stationary process whose spectrum is continuous. Indeed, from (4.6) and using Lemma 4.7 we have

$$\begin{aligned} w_x\left(\lambda_{s_j}\right) &= \left(1-e^{i\lambda_{s_j}}\right)^{-d} w_u\left(\lambda_{s_j}\right) - \frac{e^{i\lambda_{s_j}}}{1-e^{i\lambda_{s_j}}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{1}{n^{1-d}}\right) \\ &= \left(1-e^{i\phi}\right)^{-d} w_u\left(\lambda_{s_j}\right)\left[1+O\left(\frac{1}{M}\right)\right] + O_p\left(\frac{1}{n^{1-d}}\right), \end{aligned}$$

where the error magnitudes hold uniformly for $\lambda_{s_j} \in \mathcal{B}_\phi = \left\{\phi - \frac{\pi}{M}, \phi + \frac{\pi}{M}\right\}$. Theorem 3 of Hannan (1973) implies that the quantities $\{w_u(\lambda_{s_j})\}_{j=1}^J$ are asymptotically independent and distributed with the same complex normal distribution $N_c(0, f_u(\phi))$ as $n \to \infty$. The stated result for the quantities $\{w_x(\lambda_{s_j})\}_{j=1}^J$ follows directly.

Part (b) proceeds as follows. From (4.7) we have

$$w_x\left(\lambda_{s_j}\right) = \left(1-e^{i\lambda_{s_j}}\right)^{-d} w_u\left(\lambda_{s_j}\right) - \frac{e^{i\lambda_{s_j}}}{1-e^{i\lambda_{s_j}}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{\left(1-e^{i\lambda_{s_j}}\right)^{-d}}{s_j^{1-d}}\right).$$

Then,

$$\begin{aligned} \left(\lambda_{s_j}\right)^d w_x\left(\lambda_{s_j}\right) &= \left(\lambda_{s_j}\right)^d \left(1-e^{i\lambda_{s_j}}\right)^{-d} w_u\left(\lambda_{s_j}\right) - \left(\lambda_{s_j}\right)^d \frac{e^{i\lambda_{s_j}}}{1-e^{i\lambda_{s_j}}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{s_j^d\left(1-e^{i\lambda_{s_j}}\right)^{-d}}{n^d s_j^{1-d}}\right) \\ &= \left(-\frac{1}{i}\right)^d w_u\left(\lambda_{s_j}\right)\left[1+O\left(\frac{L}{n}\right)\right] \\ &\quad + \left(\frac{2\pi s_j}{n}\right)^d \frac{n}{2\pi i s_j}\left[1+O\left(\frac{L}{n}\right)\right]\frac{1}{\sqrt{2\pi}n^{1-d}}\frac{X_n}{n^{d-\frac{1}{2}}} + o_p\left(\frac{1}{n^{\alpha(1-d)}}\right) \\ &= e^{\frac{\pi d i}{2}} w_u\left(\lambda_{s_j}\right) + O_p\left(\frac{L}{n}\right) + o_p\left(\frac{1}{n^{\alpha(1-d)}}\right) \end{aligned}$$

uniformly over s_j. It follows that the family $\left\{\left(\lambda_{s_j}\right)^d w_x\left(\lambda_{s_j}\right)\right\}_{j=1}^J$ are asymptotically distributed as $\left\{e^{\frac{\pi d i}{2}} w_u\left(\lambda_{s_j}\right)\right\}_{j=1}^J$, that is the members of the family are asymptotically independent and have the same complex normal distribution, $e^{\frac{\pi di}{2}} N_c(0, f_u(0))$ or simply $N_c(0, f_u(0))$, as $n \to \infty$.

For part (c) note that for each j

$$\frac{1}{n^d} w_x\left(\lambda_{s_j}\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{n} \sum_{t=1}^{n} \frac{X_t}{n^{d-\frac{1}{2}}} e^{2\pi s_j i \frac{t}{n}} = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{2\pi i s_j r} X_{n,d}(r) dr + o_p(1),$$

and so, by the continuous mapping theorem,

$$\frac{1}{n^d} w_x\left(\lambda_{s_j}\right) \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_0^1 e^{2\pi i s_j r} B_{d-1}(r) dr,$$

giving the stated result for each s_j. It is clear from the Cramér-Wold device that joint convergence for $\{n^{-d} w_x(\lambda_{s_j}) : j = 1, \ldots, J\}$ also applies. Another approach to this result is to note from (4.10) that (dropping the subscript on s_j)

$$\frac{w_x\left(\lambda_s\right)}{n^d} = \frac{\Gamma(1-d)}{{}_1F_1\left(1, 1-d; -2\pi i s\right)} \left[w_u\left(\lambda_s\right) + \frac{1}{\sqrt{2\pi n}} \tilde{X}_{\lambda_s n}(d) \right] + O_p\left(\frac{1}{n}\right). \tag{6.63}$$

Now

$$w_u\left(\lambda_s\right) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} u_t e^{2\pi s i \frac{t}{n}} = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{2\pi s i r} dX_n(r) + o_p(1), \tag{6.64}$$

where $X_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} u_t$, and from (6.58) it follows that we may write

$$\frac{\tilde{X}_{\lambda_s n}(d)}{\sqrt{n}} = \frac{{}_1F_1\left(1, 1-d; -2\pi i s\right)}{\Gamma(1-d)} \int_0^1 e^{2\pi i s r} X_{n,d}(r) dr - \int_{q=0}^{1} e^{2\pi i s q} dX_n(q) + o_p(1). \tag{6.65}$$

Combining (6.64) and (6.65) in (6.63) we get

$$\frac{w_x\left(\lambda_s\right)}{n^d} = \int_0^1 e^{2\pi i s r} X_{n,d}(r) dr + o_p(1) \xrightarrow{d} \frac{1}{\sqrt{2\pi}} \int_0^1 e^{2\pi i s r} B_{d-1}(r) dr,$$

as above.

Part (d) follows from (3.9) and (4.15). Explicitly,

$$\begin{aligned} w_x\left(\lambda_{s_j}\right) &= \left(1 - e^{i\lambda_{s_j}}\right)^{-1} w_u\left(\lambda_{s_j}\right) - \frac{e^{i\lambda_{s_j}}}{1 - e^{i\lambda_{s_j}}} \frac{X_n}{\sqrt{2\pi n}} \\ &= \left[\left(1 - e^{i\phi}\right)^{-1} w_u\left(\lambda_{s_j}\right) - \frac{e^{i\phi}}{1 - e^{i\phi}} \frac{X_n}{\sqrt{2\pi n}} \right] \left[1 + O\left(\frac{1}{M}\right) \right] \\ &\xrightarrow{d} \left(1 - e^{i\phi}\right)^{-1} \xi_j - \frac{e^{i\phi}}{1 - e^{i\phi}} \eta \end{aligned} \tag{6.66}$$

where the family $\{\xi_j\}_{j=1}^J$ are *iid* $N_c\left(0, f_u(\phi)\right)$ as in part (a), and the ξ_j are independent of

$$\eta = \frac{B(1)}{\sqrt{2\pi}}, \tag{6.67}$$

where B is the Brownian motion in (4.15), since the ordinates $w_u(\lambda_{s_j})$ are asymptotically independent of $w_u(\lambda_0)$ for all $s_j \neq 0$.

For part (e), (3.9) yields

$$\begin{aligned}
\left(\lambda_{s_j}\right) w_x\left(\lambda_{s_j}\right) &= \left(\lambda_{s_j}\right)\frac{1}{1-e^{i\lambda_{s_j}}} w_u\left(\lambda_{s_j}\right) - \left(\lambda_{s_j}\right)\frac{e^{i\lambda_{s_j}}}{1-e^{i\lambda_{s_j}}}\frac{X_n}{\sqrt{2\pi n}} \\
&= -\frac{1}{i} w_u\left(\lambda_{s_j}\right)\left[1+O\left(\frac{1}{n}\right)\right] + \frac{1}{i}\left[1+O\left(\frac{1}{n}\right)\right]\frac{1}{\sqrt{2\pi}}\frac{X_n}{\sqrt{n}} \\
&\xrightarrow{d} i\left(\xi_j - \eta\right),
\end{aligned}$$

where the family $\{\xi_j\}_{j=1}^J$ are *iid* $N_c\left(0, f_u(0)\right)$, and the ξ_j are independent of η, which has the same form as in (6.67) above. Finally, when s_j is fixed, (4.15) and the continuous mapping theorem imply that

$$\frac{1}{n} w_x\left(\lambda_{s_j}\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{n}\sum_{t=1}^{n}\frac{X_t}{\sqrt{n}} e^{2\pi i s_j \frac{t}{n}} \xrightarrow{d} \frac{1}{\sqrt{2\pi}}\int_0^1 e^{2\pi i s_j r} B(r)\,dr, \tag{6.68}$$

which gives (4.26). Since $e^{2\pi i s_j r}$ is continuously differentiable we may apply by integration by parts to (6.68), giving

$$\frac{1}{\sqrt{2\pi}}\left[\left.\frac{e^{2\pi i s_j r} B(r)}{2\pi i s_j}\right|_0^1 - \frac{1}{2\pi i s_j}\int_0^1 e^{2\pi i s_j r}\,dB(r)\right] = \frac{1}{\sqrt{2\pi}}\frac{1}{2\pi i s_j}\left[B(1) - \int_0^1 e^{2\pi i s_j r}\,dB(r)\right],$$

which leads to the representation

$$\xi_j = \frac{1}{\sqrt{2\pi}}\int_0^1 e^{2\pi i s_j r}\,dB(r),$$

giving (4.25). Obviously, (6.68) also holds for $s_j = 0$, and part (f) is proved.

6.2.11. Proof of Theorem 5.2

From (4.6) and Lemma 4.7 we have

$$\begin{aligned}
w_x(\lambda_s) &= \left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{1}{n^{1-d}}\right) \\
&= \left(1-e^{i\phi}\right)^{-d} w_u(\lambda_s)\left[1+O\left(\frac{1}{M}\right)\right] + O_p\left(\frac{1}{n^{1-d}}\right),
\end{aligned}$$

where the error magnitudes hold uniformly for $\lambda_s \in \mathcal{B}_\phi = \left\{\phi - \frac{\pi}{M}, \phi + \frac{\pi}{M}\right\}$. Then, as $n \to \infty$ with $\frac{M}{n} \to 0$, we have

$$\begin{aligned}\hat{f}_{xx}(\phi) &= \frac{1}{m}\sum_{\lambda_s\in\mathcal{B}(\phi)} w_x(\lambda_s)w_x(\lambda_s)^* \\ &= \frac{1}{\left|1-e^{i\phi}\right|^{2d}}\frac{1}{m}\sum_{\lambda_s\in\mathcal{B}(\phi)} w_u(\lambda_s)w_u(\lambda_s)^* + O_p\left(\frac{1}{M}\right) + O_p\left(\frac{1}{n^{1-d}}\right) \\ &\xrightarrow{p} \frac{1}{\left|1-e^{i\phi}\right|^{2d}} f_u(\phi),\end{aligned} \tag{6.69}$$

by virtue of the consistency of the smoothed periodogram estimate in the stationary (linear process) case (e.g., Hannan, 1970, Ch. IV), giving part (a).

For part (b), when $d = 1$ we have from (6.66)

$$w_x(\lambda_s) = \left[\left(1-e^{i\phi}\right)^{-1} w_u(\lambda_s) - \frac{e^{i\phi}}{1-e^{i\phi}}\frac{X_n}{\sqrt{2\pi n}}\right]\left[1+O\left(\frac{1}{M}\right)\right],$$

and, as $n \to \infty$ with $\frac{M}{n} \to 0$, we have

$$\begin{aligned}\hat{f}_{xx}(\phi) &= \frac{1}{m}\sum_{\lambda_s\in\mathcal{B}(\phi)} w_x(\lambda_s)w_x(\lambda_s)^* \\ &= \frac{1}{\left|1-e^{i\phi}\right|^{2}}\frac{1}{m}\sum_{\lambda_s\in\mathcal{B}(\phi)} w_u(\lambda_s)w_u(\lambda_s)^* - \frac{2}{\left|1-e^{i\phi}\right|^{2}}\mathrm{Re}\left(\frac{1}{m}\sum_{\lambda_s\in\mathcal{B}(\phi)} w_u(\lambda_s)\frac{e^{-i\phi}X_n}{\sqrt{2\pi n}}\right) \\ &\quad + \frac{1}{\left|1-e^{i\phi}\right|^{2}}\left(\frac{X_n}{\sqrt{2\pi n}}\right)^2 + O_p\left(\frac{1}{M}\right), \\ &\xrightarrow{d} \frac{1}{\left|1-e^{i\phi}\right|^{2d}} f_u(\phi) + \frac{1}{\left|1-e^{i\phi}\right|^{2}}\left(\frac{B(1)}{\sqrt{2\pi}}\right)^2\end{aligned}$$

in view of (6.69) and (4.15).

To prove part (c), we write the sum (5.2) as the sum over the full set of frequencies $\{\lambda_s\}_{s=0}^{n-1}$ and a residual, that is,

$$\begin{aligned}\frac{m}{n^{2d}}\hat{f}_{xx}(0) &= \sum_{s=0}^{m-1}\frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d} \\ &= \sum_{s=0}^{n-1}\frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d} - \sum_{s=m}^{n-1}\frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d} \\ &= \frac{1}{2\pi}\sum_{t=1}^{n}\left(\frac{X_t}{n^d}\right)^2 - \sum_{s=m}^{n-1}\frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d} \\ &= \frac{1}{2\pi}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{n^{d-\frac{1}{2}}}\right)^2 - \sum_{s=m}^{n-1}\frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d}.\end{aligned} \tag{6.70}$$

Since $\frac{m}{n^{\alpha}} \to \infty$ we have by (4.8)

$$\frac{1}{n^d} w_x(\lambda_s) = \frac{1}{n^d}\left[\left(1-e^{i\lambda_s}\right)^{-d} w_u(\lambda_s) - \frac{e^{i\lambda_s}}{1-e^{i\lambda_s}}\frac{X_n}{\sqrt{2\pi n}} + o_p\left(\frac{e^{i\lambda_s}}{\left(1-e^{i\lambda_s}\right)^{1-d}}\frac{X_n}{\sqrt{n}}\right)\right]$$
$$= O_p\left(\frac{1}{m^d}\right),$$

uniformly for $s \geq m$. When m is such that $\frac{m}{n^{\alpha}} \to \infty$, it follows that

$$\frac{1}{n^d} w_x(\lambda_s) = o_p\left(\frac{1}{n^{\alpha d}}\right),$$

and then

$$\sum_{s=m}^{n-1} \frac{w_x(\lambda_s)}{n^d}\frac{w_x(\lambda_s)^*}{n^d} = o_p\left(\frac{n}{n^{2\alpha d}}\right) = o_p(1), \tag{6.71}$$

for α chosen such $\alpha \geq \frac{1}{2d}$. We deduce from (6.70), (6.71), (4.15) and the continuous mapping theorem that

$$\frac{m}{n^{2d}}\hat{f}_{xx}(0) = \frac{1}{2\pi}\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_t}{n^{d-\frac{1}{2}}}\right)^2 + o_p(1) \to_d \frac{1}{2\pi}\int_0^1 B_{d-\frac{1}{2}}(r)^2\, dr,$$

giving the stated result in part (c). Part (d) follows in an analogous fashion with $d = 1$ and $\alpha \geq \frac{1}{2}$.

7. NOTATION

$\to_{\text{a.s.}}$	Almost sure convergence
$=_d$	Distributional equivalence
$:=$	Definitional equality
$o_{\text{a.s.}}(1)$	Tends to zero almost surely
$o_p(1)$	Tends to zero in probability
$\to_p$	Convergence in probability
$\overset{d}{\to}, \to_d$	Weak convergence
$\lfloor \cdot \rfloor$	Integer part of
$(a)_k$	$(a)(a+1)\ldots(a+k-1)$ forward factorial
${}_1F_1(a,c;z)$	$\sum_{k=0}^{\infty}\frac{(a)_k}{(c)_k\, k!}z^k$, confluent hypergeometric function
${}_2F_1(a,b,c;z)$	$\sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k\, k!}z^k$ hypergeometric function
$\mathbf{1}(A)$	indicator of A

$X_n(r)$	$n^{-\frac{1}{2}}\sum_{t=0}^{\lfloor nr \rfloor} u_t$
$X_{n,d}(r)$	$n^{\frac{1}{2}-d} X_{\lfloor nr \rfloor}$
$\Gamma(z)$	$\int_0^{\infty} e^{-t} t^{z-1} dt$ gamma function $(\mathrm{Re}(z) > 0)$
$B(z,w)$	$\frac{\Gamma(z)\Gamma(w)}{\Gamma(z+w)}$ beta function
$w_a(\lambda)$	$\frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} a_t e^{it\lambda}$ discrete Fourier transform

NOTES

1. See Aldrich (1995) for an overview of early research on correlation, including nonsense correlations where, as Aldrich aptly puts it, "there are more ways of going wrong than going right."

2. Zeros everywhere in $f_u(\lambda)$ are ruled out if the last condition of (2.4) is strengthened to $C(e^{i\lambda}) \neq 0$ for all $\lambda \in [0, \pi]$.

3. Here, and elsewhere in this chapter, where fractional powers of a complex variable are given they are taken to be evaluated at their principal values.

REFERENCES

Abadir, K. M., Distaso, W., & Giraitis, L. (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics*, *141*, 1353–1384.

Akonom, J., & Gouriéroux, C. (1987). A functional limit theorem for fractional processes. Technical Report No. 8801. CEPREMAP.

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical Science*, *10*(4), 364–376.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, *17*, 21–86.

Corbae, D., Ouliaris, S., & Phillips, P. C. B. (2002). Band spectral regression with trending data. *Econometrica*, *70*, 1067–1109.

Davidson, J., & Hashimzade, N. (2009). Type I and type II fractional Brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, *53*, 2089–2106.

Durlauf, S. N., & Phillips, P. C. B. (1988). Trends versus random walks in time series analysis. *Econometrica*, *56*(6), 1333–1354.

Erdélyi, A. (1953). *Higher transcendental functions* (Vol. 1). McGraw Hill.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.

Hannan, E. J. (1970). *Multiple time series*. John Wiley & Sons.

Hannan, E. J. (1973). Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *26*, 157–170.

Hannan, E. J., & Deistler, M. (2012). *The statistical theory of linear systems*. SIAM.

Hooker, R. H. (1901). Correlation of the marriage-rate with trade. *Journal of the Royal Statistical Society*, *64*, 485–492.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, *116*, 770–799.

Hurst, H. E. (1956). Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers*, *5*, 519–543.

Hurvich, C. M., Deo, R., & Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, *19*, 19–46.

Kim, C. S., & Phillips, P. C. B. (2003). *Modified log periodogram regression*. Working Paper. Yale University (pp. 1–50).

Kim, C. S., & Phillips, P. C. B. (2006). *Log periodogram regression: The nonstationary case*. Cowles Foundation Discussion Paper No. 1547, Yale University.

Künsch, H. (1986). Discrimination between long range dependence and monotonic trends. *Journal of Applied Probability*, *23*, 1025–1030.

Künsch, H. (1987). Statistical aspects of self-similar processes. *Bernoulli*, *1*, 67–74.

Lovoie, J., Osler, T. J., & Tremblay, R. (1976). Fractional derivatives and special functions. *SIAM Review*, *18*, 240–268.

Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, *10*, 422–437.

Marinucci, D., & Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, *80*, 111–122.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics*, *10*, 139–162.

Olver, F. W., Lozier, D. W., Boisvert, R. F., & Clark, C. W. (2010). *NIST handbook of mathematical functions – Hardback and CD-ROM*. Cambridge University Press.

Park, J. Y., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*, 468–497.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*, 95–131.

Pearson, K., & Elderton, E. M. (1923). On the variate difference method. *Biometrika*, *14*, 281–310.

Phillips, P. C. B. (1986a). Fractional matrix calculus and the distribution of multivariate tests. In I. B. MacNeill & G. J. Umphrey (Eds.), *Time series and econometric modelling* (pp. 219–234). D. Reidel.

Phillips, P. C. B. (1986b). Understanding spurious regressions in econometrics. *Journal of Econometrics*, *33*, 311–340.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, *55*, 277–301.

Phillips, P. C. B. (1988). Regression theory for near-integrated time series. *Econometrica*, *56*, 1021–1043.

Phillips, P. C. B. (1991). Spectral regression for cointegrated time series. In J. P. Barnett & G. Tauchen (Eds.), *Nonparametric and semiparametric methods in economics and statistics* (pp. 413–435). Cambridge University Press.

Phillips, P. C. B. (1999a). *Discrete Fourier transforms of fractional processes*. Cowles Foundation Discussion Paper No. 1243. Yale University (pp. 1–58).

Phillips, P. C. B. (1999b). *Unit root log periodogram regression*. Cowles Foundation Discussion Paper No. 1244. Yale University (pp. 1–23).

Phillips, P. C. B. (2007). Unit root log periodogram regression. *Journal of Econometrics*, *138*, 104–124.

Phillips, P. C. B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics*, *178*, 210–224.

Phillips, P. C. B., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, *53*, 473–495.

Phillips, P. C. B., & Shimotsu, K. (2004). Local Whittle estimation in nonstationary and unit root cases. *The Annals of Statistics*, *32*, 656–692.

Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, *20*, 971–1001.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, *23*(5), 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, *23*(3), 1048–1072.

Samorodnitsky, G., & Taqqu, M. S. (2017). *Stable non-Gaussian random processes: Stochastic models with infinite variance: Stochastic modeling*. Routledge.

Shao, X. (2010). Nonstationarity-extended Whittle estimation. *Econometric Theory*, *26*, 1060–1087.

Shimotsu, K., & Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, *33*, 1890–1933.
Shimotsu, K., & Phillips, P. C. B. (2006). Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, *130*, 209–233.
Solo, V. (1992). Intrinsic random functions and the paradox of 1/f noise. *SIAM Journal on Applied Mathematics*, *52*, 270–291.
Sun, Y., & Phillips, P. C. B. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, *115*, 355–389.
Sun, Y., & Phillips, P. C. B. (2004). Understanding the Fisher equation. *Journal of Applied Econometrics*, *19*, 869–886.
Surgailis, D., Koul, H. L., & Giraitis, L. (2012). *Large sample inference for long memory processes*. World Scientific Publishing Company.
Velasco, C. (1999). Non-stationary log-periodogram regression. *Journal of Econometrics*, *91*, 325–371.
Velasco, C., & Robinson, P. M. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, *95*, 1229–1243.
Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, *89*, 1–63.

This page intentionally left blank

CHAPTER 2

ASYMPTOTIC PROPERTIES OF THE LEAST SQUARES ESTIMATOR IN LOCAL TO UNITY PROCESSES WITH FRACTIONAL GAUSSIAN NOISE

Xiaohu Wang[a], Weilin Xiao[b] and Jun Yu[c]

[a]*School of Economics, Fudan University, Shanghai, China, and Shanghai Institute of International Finance and Economics, Shanghai, China*
[b]*School of Management, Zhejiang University, Hangzhou, China*
[c]*School of Economics and Lee Kong Chian School of Business, Singapore Management University, Singapore, Singapore*

ABSTRACT

This chapter derives asymptotic properties of the least squares (LS) estimator of the autoregressive (AR) parameter in local to unity processes with errors being fractional Gaussian noise (FGN) with the Hurst parameter $\mathrm{H} \in (0,1)$. *It is shown that the estimator is consistent for all values of* $\mathrm{H} \in (0,1)$. *Moreover, the rate of convergence is* n^{-1} *when* $\mathrm{H} \in [0.5,1)$. *The rate of convergence is* n^{-2H} *when* $\mathrm{H} \in (0,0.5)$. *Furthermore, the limiting distribution of the centered LS estimator depends on* H. *When* $\mathrm{H} = 0.5$, *the limiting distribution is the same as that obtained in Phillips (1987a) for the local to unity model with errors for which the standard functional central limit theorem is applicable. When* $\mathrm{H} > 0.5$ *or when* $\mathrm{H} < 0.5$, *the limiting distributions are new to the*

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 73–95

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A002

literature. The asymptotic properties of the LS estimator with fitted intercept are also derived. Simulation studies are performed to check the reliability of the asymptotic approximation for different values of sample size.

Keywords: Least squares; local to unity; fractional Brownian motion; fractional Ornstein–Uhlenbeck process; fitted intercept; functional central limit theorem
JEL classification: C22

1. INTRODUCTION

In this chapter, we consider the following model:

$$X_t = \rho_n X_{t-1} + \varepsilon_t, \rho_n = \exp(-c/n),\ t = 1,...,n, \tag{1}$$

where c is a constant, $\varepsilon_t = \sigma u_t$, u_t is a FGN that has mean zero, variance one, and covariance function as

$$\gamma_u(k) := \mathbb{E}(u_t u_s) = \frac{1}{2}\left[(k+1)^{2H} + (k-1)^{2H} - 2k^{2H}\right] \text{with } k = |t-s|, \tag{2}$$

and $H \in (0,1)$. The parameter H is known as the Hurst parameter in the literature. When $H = 0.5$, it has $\gamma_u(k) = 0$ for any $k \neq 0$, in which case $\{u_t\}$ form a sequence of independent and identically distributed (i.i.d.) variables with the standard normal distribution $N(0, 1)$. However, when $H \neq 0.5$, it has $\gamma_u(k) \neq 0$ for any k, meaning that $\{u_t\}$ have serial dependence. Moreover, it has

$$\gamma_u(k) \sim H(2H-1)k^{2H-2}, \text{ for large } k. \tag{3}$$

That is $\gamma_u(k)$ decays at a hyperbolic rate as k goes to infinity. As a result, for the case of $H > 0.5$, it has $\gamma_u(k) > 0$ and $\sum_{k=-\infty}^{\infty} \gamma_u(k) = \infty$, giving rise to the terminology of "long-range-dependent" errors. In contrast, for the case of $H < 0.5$, it has $\gamma_u(k) < 0$ for $k \neq 0$ and $\sum_{k=-\infty}^{\infty} \gamma_u(k) = 0$, giving rise to the terminology of "anti-persistent" errors.

The FGN u_t has the same distribution as the increment of the fractional Brownian motion (fBm) $B^H(t)$ that is a zero-mean Gaussian process with the covariance function

$$\mathrm{Cov}\left(B^H(t), B^H(s)\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)\ \forall t, s \geq 0. \tag{4}$$

That is $u_t \sim B^H(t) - B^H(t-1)$ where $\sim$ stands for equivalence in distribution.

Model (1) is related to the local to unity model of Phillips (1987a) and Chan and Wei (1987) by replacing the noise where the classical central limit theorem is applicable with FGNs. Model (1) is also related to the fractional unit root model of Sowell (1990) by replacing the AR coefficient of unity with the AR coefficient of local to unity. Although we replace the $I(d)$ noise of Sowell (1990) with the FGN, the results in this chapter also apply to $I(d)$ errors as it will become clear later. Model (1) is also related to the model of Park (2003) where $\rho_n = 1 - m/n$ if we assume m is fixed in his model.

We consider two regressions to estimate the AR root ρ_n in Model (1). The first is an AR regression without intercept fitted, which leads to the LS estimator of ρ_n as

$$\hat{\rho}_n = \sum_{t=1}^{n} X_{t-1} X_t / \sum_{t=1}^{n} X_{t-1}^2 = \rho_n + \sum_{t=1}^{n} X_{t-1}\varepsilon_t / \sum_{t=1}^{n} X_{t-1}^2. \tag{5}$$

The second is an AR regression with intercept fitted, giving the following LS estimator of ρ_n

$$\begin{aligned} \tilde{\rho}_n &= \sum_{t=1}^{n} \left(X_{t-1} - \bar{X}_{-1}\right) X_{t-1} / \sum_{t=1}^{n} \left(X_{t-1} - \bar{X}_{-1}\right)^2 \\ &= \rho_n + \sum_{t=1}^{n} \left(X_{t-1} - \bar{X}_{-1}\right)\varepsilon_t / \sum_{t=1}^{n} \left(X_{t-1} - \bar{X}_{-1}\right)^2 \end{aligned}, \tag{6}$$

where $\bar{X}_{-1} = \frac{1}{n}\sum_{t=1}^{n} X_{t-1}$.

The goal of this chapter is to derive the asymptotic properties of the two estimators $\hat{\rho}_n$ and $\tilde{\rho}_n$ under the assumption of $n \to \infty$. As it is well expected for local to unity model, the initial value of X_t significantly affects the finite sample distribution of $\hat{\rho}_n$ and $\tilde{\rho}_n$. To capture the impact of the initial value on asymptotics, we set the initial value of X_t to be $X_0 = O_p\left(n^H\right)$ and

$$n^{-H}\frac{X_0}{\sigma} \xrightarrow{p} \pi_0,$$

where π_0 is a constant (such as zero) or $O_p(1)$.

The rest of this chapter is organized as follows. Section 2 reviews the results in the literature. The asymptotic properties of the normalized $\hat{\rho}_n - \rho_n$ are developed in Section 3. Section 4 extends the results to the case when the intercept is fitted. Section 5 examines the finite sample properties of the normalized $\hat{\rho}_n - \rho_n$ and $\tilde{\rho}_n - \rho_n$. Section 6 concludes. The Appendix collects proofs of the main results.

Throughout the chapter, we use $\xrightarrow{p}, \xrightarrow{d}, \Rightarrow$, and $\sim$ to denote convergence in probability, convergence in distribution, convergence in functional space, and equivalence in distribution, respectively. The notation $[nr]$ represents the integer part of nr.

2. A LITERATURE REVIEW

Phillips (1987a) considers the following local to unity model

$$X_t = \rho_n X_{t-1} + v_t,\ \rho_n = \exp(-c/n),\ X_0 = O_p(1), \tag{7}$$

where $\{v_t\}$ is a strong mixing sequence with mixing coefficients α_m that satisfies $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$ and $\sup_t |v_t|^{\beta+\delta} < \infty$ for some $\beta > 2$ and $\delta > 0$. There are two important features in Model (7). First, since $\rho_n = 1 - c/n + O(n^{-2})$, the AR coefficient depends on n and converges to unity as $n \to \infty$. Second, the functional central limit theorem is applicable to $\{v_t\}$. An interesting special case of Model (7) is when $\{v_t\}$ are i.i.d. with $E|v_t|^{\beta} < \infty$ for some $\beta > 2$. In this case, as $n \to \infty$, it has

$$n(\hat{\rho}_n - \rho_n) \xrightarrow{d} \frac{\int_0^1 J_c(r) dW(r)}{\int_0^1 J_c(r)^2 dr} = \frac{\left\{J_c(1)^2 + 2c\int_0^1 J_c(r)^2 dr - 1\right\}/2}{\int_0^1 J_c(r)^2 dr}. \tag{8}$$

where $J_c(r)$ denotes an Ornstein–Uhlenbeck (OU) process defined by the stochastic differential equation

$$dJ_c(r) = -cJ_c(r)dr + dW(r),\ J_c(0) = 0, \tag{9}$$

with $W(r)$ being a standard Brownian motion.

Sowell (1990) considers the following unit root model with $\rho = 1$:

$$X_t = \rho X_{t-1} + \sigma v_t,\ v_t = (1-L)^{-d}\epsilon_t,\ \epsilon_t \overset{i.i.d.}{\sim} (0,1),\ \ X_0 = O_p(1), \tag{10}$$

where L is the lag operator with $(1-L)^{-d}$ defined as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j \text{ for } d \in (-0.5, 0.5).$$

In this model, the error term v_t is assumed to follow a fractionally integrated process of order d, which is referred to as an $I(d)$ process in the literature. With $\hat{\rho}$ being the LS estimator of ρ, Sowell (1990) and Marinucci and Robinson (1999) show that, as $n \to \infty$,

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 W(r) dW(r)}{\int_0^1 W(r)^2 dr}, \text{ if } d = 0, \tag{11}$$

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{\frac{1}{2} B^H(1)^2}{\int_0^1 B^H(r)^2 dr}, \text{ if } d > 0, \tag{12}$$

$$n^{2H}\left(\hat{\rho}-1\right)\xrightarrow{d}-\frac{H\dfrac{\Gamma(0.5+H)}{\Gamma(1.5-H)}}{\int_0^1 B^H(r)^2\,dr}, \text{ if } d<0, \tag{13}$$

where $H = d + 0.5$.[1]

Setting $c = 0$ in (8) or setting $d = 0$ in (11) can lead to the well-known result for the unit root model obtained in Phillips (1987b) as

$$n\left(\hat{\rho}-1\right)\xrightarrow{d}\frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2\,dr}=\frac{\frac{1}{2}\left(W(1)^2-1\right)}{\int_0^1 W(r)^2\,dr}.$$

3. ASYMPTOTIC PROPERTIES

To develop the asymptotic properties of the centered LS estimator $\hat{\rho}_n - \rho_n$ defined in (5), we first introduce the limit behavior of the partial sum process $\sum_{t=1}^{[nr]} u_t$ for any $r \in [0,1]$. As $u_t = B^H(t) - B^H(t-1)$, we have

$$\begin{aligned} n^{-H}\sum_{t=1}^{[nr]} u_t &\sim n^{-H}\sum_{t=1}^{[nr]}\left\{B^H(t)-B^H(t-1)\right\} \\ &= n^{-H}B^H([nr]) \\ &\sim B^H\left(\frac{[nr]}{n}\right) \\ &\Rightarrow B^H(r), \quad \text{as } n\to\infty \end{aligned}, \tag{14}$$

where equivalence in distribution comes from the self-similarity property of the fBm $B^H(t)$. Note that the sample path of $n^{-H}\sum_{t=1}^{[nr]} u_t$ is a function of $r \in [0,1]$ that is right-continuous with left limits. Hence, the convergence result of the partial sum sequence is built up in the space of $D[0,1]$, which is the space of all real valued functions on $[0,1]$ that are right-continuous with finite left limits, equipped with the Skorokhod topology. The convergence results obtained in the rest of this chapter are all considered in the space of $D[0,1]$ with the same topology.

The convergence result in (14) is the source of the asymptotic theory developed in this chapter. Sowell (1990) gives a similar weak convergence result for the partial sum process $\sum_{t=1}^{[nr]} u_t$ when $u_t \sim I(d)$; see also Marinucci and Robinson (1999). Therefore, all the results in this chapter applies to the case where $u_t \sim I(d)$. It is important to note that Sowell (1990) uses the result of Davydov (1970) to establish the weak convergence while we do not need to resort to Davydov (1970) as our errors are normally distributed.

The result in (14) compares with Donsker's functional central limit theorem, which states that,

$$n^{-0.5}\sum_{t=1}^{[nr]}\epsilon_t \Rightarrow W(r) = B^{0.5}(r), \text{ as } n \to \infty, \tag{15}$$

where ϵ_t is a sequence of i.i.d. random variables with mean zero and variance one.

Define a fractional OU (fOU) process through the following stochastic differential equation

$$dJ_c^H(t) = -cJ_c^H(t)dt + dB^H(t), \text{ with } J_c^H(0) = 0. \tag{16}$$

Cheridito et al. (2003) proved that, for $t > 0$, the differential equation (16) has a unique solution, taking the form of

$$J_c^H(t) = \int_0^t e^{-c(t-s)} dB^H(s),$$

where the integral is a path-wise Riemann–Stieltjes integral. It is worthwhile to mention that, when $H = 0.5$, $J_c^H(t)$ becomes the traditional OU process considered in Phillips (1987a). If in addition, $c = 0$, the process $J_c^H(t)$ becomes a standard Brownian motion.

Lemma 1. *Let $\{X_t\}$ be the time series generated by Model (1). Then, as $n \to \infty$,*

1. $n^{-H}X_{[nr]} \Rightarrow \sigma J_c^H(r) + e^{-cr}\sigma\pi_0$;

2. $n^{-1-H}\sum_{t=1}^{n} X_t \Rightarrow \sigma\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right]dr$;

3. $n^{-1-2H}\sum_{t=1}^{n} X_t^2 \Rightarrow \sigma^2\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr$;

4. $n^{-2H}\sum_{t=1}^{n} X_{t-1}\varepsilon_t$

$$\Rightarrow \begin{cases} \sigma^2\left(\left[J_c(1) + e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1 \left[J_c(r) + e^{-cr}\pi_0\right]^2 dr - 1\right)/2, & \text{if } H = 0.5 \\ \sigma^2\left(\left[J_c^H(1) + e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr\right)/2, & \text{if } H > 0.5 \end{cases};$$

5. $n^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \xrightarrow{p} -\sigma^2/2$, *if* $H < 0.5$.

Remark 1. *This lemma is related to Lemma 1 in Phillips (1987a) with several differences. First, the initial condition π_0, which is the limit of $n^{-H}X_0/\sigma$, plays explicit roles in all the limits except the last one. Second, compared with Lemma 1.a–1.c of*

Phillips (1987a), $J_c(r)$ is replaced with $J_c^H(r)$ in our Lemma, and the orders of $X_{[nr]}, \sum_{t=1}^{n} X_t$, and $\sum_{t=1}^{n} X_t^2$ becomes n^{H}, n^{-1+H}, and n^{1+2H}, respectively. Third, both the order and the limit of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ depend on H. When $H \geq 0.5$, the order of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ is n^{2H}, whereas, when $H < 0.5$, the order becomes n. In addition, the limit of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ has one more term (i.e., $-\sigma^2/2$) when $H = 0.5$ than when $H > 0.5$, and three more terms than $H < 0.5$. These differences root in the distinct properties of the FGN, u_{t}, when the Hurst parameter H takes different values. For example, when $H = 0.5$, the limit of $n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2$ is σ^2, whereas, when $H > 0.5$, the limit of $n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2$ is zero.

Remark 2. *When $H = 0.5$, it has $J_c^H(r) = J_c(r)$. If we further let $\pi_0 = 0$, the results in Parts 1–3 of Lemma 1 above becomes exactly the same as those in Lemma 1.a–1.c in Phillips (1987a). Moreover, the result in Part 4 of Lemma 1 above can be written as*

$$n^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \Rightarrow \sigma^2\left(J_c(1)^2 + 2c\int_0^1 J_c(r)^2\,dr - 1\right)/2 = \sigma^2\int_0^1 J_c(r)dW(r),$$

which is the same as that in Lemma 1.d of Phillips (1987a).

Remark 3. *The convergence result in Part 1 of Lemma 1 is the key to the development of the results in the rest of the Lemma. With slight adjustments, the result in Part 1 can be extended to the case where u_{t} becomes an I(d) process. When $u_t \sim I(d)$, $n^{-H}\left(\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\right)^{-1/2}\sum_{t=1}^{[nr]} u_t \Rightarrow B^H(r)$ as $n \to \infty$. This is a special case of a more general result obtained in Taqqu (1975). Consequently, with the use of the continuous mapping theorem, it can be proved easily that*

$$\begin{aligned} &n^{-H}\left(\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\right)^{-1/2} X_{[nr]} \\ \Rightarrow\quad &\sigma J_c^H(r) + \left(\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\right)^{-1/2} e^{-cr}\sigma\pi_0 \end{aligned}.$$

Theorem 2. *Let $\{X_t\}$ be the time series generated by (1) and (2). Then, as $n \to \infty$, if $H = 0.5$,*

$$n(\hat{\rho}_n - \rho_n) \Rightarrow \frac{\left(\left[J_c(1) + e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1\left[J_c(r) + e^{-cr}\pi_0\right]^2 dr - 1\right)/2}{\int_0^1\left[J_c(r) + e^{-cr}\pi_0\right]^2 dr}; \quad (17)$$

if $H > 0.5$

$$n\left(\hat{\rho}_n-\rho_n\right)\Rightarrow\frac{\left(\left[J_c^H(1)+e^{-c}\pi_0\right]^2-\pi_0^2+2c\int_0^1\left[J_c^H(r)+e^{-cr}\pi_0\right]^2dr\right)/2}{\int_0^1\left[J_c^H(r)+e^{-cr}\pi_0\right]^2dr}; \tag{18}$$

if $H<0.5$,

$$n^{2H}\left(\hat{\rho}_n-\rho_n\right)\Rightarrow\frac{-1/2}{\int_0^1\left[J_c^H(r)+e^{-cr}\pi_0\right]^2dr}. \tag{19}$$

Remark 4. *When we compare Theorem 2 in this chapter to Theorem 1 in Phillips (1987a), we have a few observations. First, the initial condition* π_0 *plays significant roles in all the limits in Theorem 2. Second, when* $H=0.5$ *and* $\pi_0=0$, $\hat{\rho}_n-\rho_n$ *has the same convergence rate and the same limiting distribution as those in Phillips (1987a). Third, there is a discontinuity in our limit theory when H passes 0.5. When* $H>0.5$, *the convergence rate of* $\hat{\rho}_n-\rho_n$ *is n, which is the same as that when* $H=0.5$. *However, the numerator of the limit has one term less comparing to the case of* $H=0.5$. *Furthermore, When* $H<0.5$, *the rate of convergence of* $\hat{\rho}_n-\rho_n$ *becomes* n^{2H}, *which is slower than that when* $H\geq 0.5$. *The numerator in the limit has three terms less than that when* $H=0.5$.

Remark 5. *If* $c=0$, *then* $\rho_n=\exp(-c/n)=1$. *In this case, Model (1) gives a unit root process with FGNs. With the further assumption of* $X_0=0$ *that leads to* $\pi_0=0$, *the results in Theorem 2 become*

$$n\left(\hat{\rho}_n-\rho_n\right)\Rightarrow\frac{\frac{1}{2}B^H(1)^2}{\int_0^1 B^H(r)^2dr},\ \text{when } H>0.5, \tag{20}$$

and

$$n^{2H}\left(\hat{\rho}_n-\rho_n\right)\Rightarrow\frac{-1/2}{\int_0^1 B^H(r)^2dr},\ \text{when } H<0.5. \tag{21}$$

The result in (20) is the same as that developed in Sowell (1990) and Marinucci and Robinson (1999) for the unit root process with I(d) errors when $d=H-1/2>0$. *However, when* $H<0.5$ *our limiting result in (21) is slightly different with that obtained in Sowell (1990) and Marinucci and Robinson (1999) when* $d=H-1/2<0$; *see (13) in this chapter. The difference arises because the I(d) process used in Sowell (1990) has different variance and long-run variance from those of the FGN. The variance and the long-run variance of an I(d) process is* $\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}$ *and* $O(n^{2H})\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}$, *respectively. The ratio of* $\frac{\Gamma(1-2d)}{\Gamma(1-d)^2}$ *and* $\frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}$, *divided by 2, gives*

$$\frac{(1+2d)\Gamma(1+d)}{2\Gamma(1-d)} = \frac{H\Gamma(0.5+H)}{\Gamma(1.5-H)},$$

which is the numerator of the limit in (12) that has been derived by Marinucci and Robinson (1999).

4. ASYMPTOTIC PROPERTIES WITH FITTED INTERCEPT

In this section, we assume that, while the data are generated from Model (1), it is not known apriori that the intercept is zero. Hence, an AR regression with an intercept is estimated as

$$X_t = \alpha + \rho_n X_{t-1} + \varepsilon_t, \tag{22}$$

which leads to the LS estimator of ρ_n as in (6).

Theorem 3 presents the large sample theory of $\tilde{\rho}_n$ for various values of H.

Theorem 3. *Let $\{X_t\}$ be the time series generated by Model (1) and $\tilde{\rho}_n$ be the estimator of the AR root from Model (22) with fitted intercept. Then, as $n \to \infty$, if $H = 0.5$,*

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{\left(\left[J_c(1)+e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1 \left[J_c(r)+e^{-cr}\pi_0\right]^2 dr - 1\right)/2 - B^H(1)\int_0^1 \left[J_c(r)+e^{-cr}\pi_0\right]dr}{\int_0^1 \left[J_c(r)+e^{-cr}\pi_0\right]^2 dr - \left(\int_0^1 \left[J_c(r)+e^{-cr}\pi_0\right]dr\right)^2}; \tag{23}$$

if $H > 0.5$,

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{\left(\left[J_c^H(1)+e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]^2 dr\right)/2 - B^H(1)\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]dr}{\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]^2 dr - \left(\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]dr\right)^2}; \tag{24}$$

if $H < 0.5$,

$$n^{2H}(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{-1/2}{\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]^2 dr - \left(\int_0^1 \left[J_c^H(r)+e^{-cr}\pi_0\right]dr\right)^2}. \tag{25}$$

Remark 6. *When $H = 0.5$ and $\pi_0 = 0$, it has $B^H(r) = W(r)$ and*

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{\left(J_c(1)^2 + 2c\int_0^1 J_c(r)^2\, dr - 1\right)/2 - B^H(1)\int_0^1 J_c(r)dr}{\int_0^1 [J_c(r)]^2 dr - \left(\int_0^1 J_c(r)dr\right)^2}$$

Appling a result in Phillips (1987a) that

$$\left(J_c(1)^2 + 2c\int_0^1 J_c(r)^2\, dr - 1\right)/2 = \int_0^1 J_c(r)dW(r),$$

we obtain the limiting distribution of $n(\tilde{\rho}_n - \rho_n)$ as

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{\int_0^1 J_c(r)dW(r) - B^H(1)\int_0^1 J_c(r)dr}{\int_0^1 [J_c(r)]^2 dr - \left(\int_0^1 J_c(r)dr\right)^2} = \frac{\int_0^1 \bar{J}_c(r)dW(r)}{\int_0^1 \bar{J}_c(r)^2\, dr},$$

where $\bar{J}_c(r) = J_c(r) - \int_0^1 J_c(s)ds$ is the de-meaned OU process. This limiting distribution is the same as that given in Remark 3 of Mikusheva (2015) for the local to unity model with weakly dependent errors.

Remark 7. *Theorem 3 shows that the large sample theory of the centered LS estimator $\tilde{\rho}_n - \rho_n$ not only depends on the values of* H, *but also on the initial condition* π_0.

In Corollary 4, it is shown that, when $c = 0$ that makes the Model (1) a unit root process, the limiting distributions of $\tilde{\rho}_n - \rho_n$ becomes independent of the initial condition.

Corollary 4. *Let $\{X_t\}$ be the time series generated from Model (1) with $c = 0$. In this case, it has $J_c^H(r) = B^H(r)$. Then, as $n \to \infty$,*

if $H = 0.5$,

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{\left(W(1)^2 - 1\right)/2 - W(1)\int_0^1 W(r)dr}{\int_0^1 W(r)^2\, dr - \left(\int_0^1 W(r)dr\right)^2}; \tag{26}$$

if $H > 0.5$

$$n(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{B^H(1)^2/2 - B^H(1)\int_0^1 B^H(r)\, dr}{\int_0^1 B^H(r)^2\, dr - \left(\int_0^1 B^H(r)\, dr\right)^2}; \tag{27}$$

if $H < 0.5$,

$$n^{2H}(\tilde{\rho}_n - \rho_n) \Rightarrow \frac{-1/2}{\int_0^1 B^H(r)^2\, dr - \left(\int_0^1 B^H(r)\, dr\right)^2}. \tag{28}$$

Remark 8. *The large sample result in (26) of Corollary 4 is well-known in the literature; see, for example, Equation (17.4.28) in Hamilton (1994). Loosely speaking, the large sample results in (27) and (28) of Corollary 4 extend those of Sowell (1990) from the estimated model without fitted intercept to the estimated model with fitted intercept.*

5. MONTE CARLO STUDIES

To check how well the limit distribution perform in finite sample, we carry out several Monte Carlo studies. In all studies, we simulate data from Model (1). Four different sample sizes are considered, namely, $n = 32$, 512, 2,048, and 8,192. Three values are considered for H, namely $H = 0.5$, 0.9, and 0.1.[2] Two values are considered for c, namely, $c = 10$ and 5.

5.1. Without Fitted Intercept

For each time series simulated, we estimate ρ_n by $\hat{\rho}_n$ and calculate $n(\hat{\rho}_n - \rho_n)$ when $H \geq 0.5$ and $n^{2H}(\hat{\rho}_n - \rho_n)$ when $H < 0.5$. The 200,000 replications are used to obtain density of $n(\hat{\rho}_n - \rho_n)$ or $n^{2H}(\hat{\rho}_n - \rho_n)$.

Figs. 1 and 2 display the density of $n(\hat{\rho}_n - \rho_n)$ when $H = 0.5$ and $c = 10$ and 5. In each of the two values of c, the densities are almost identical for all n, suggesting the limit distribution provides accurate approximations to the finite sample distribution when the sample size is as small as 32. In all cases, the density is left-skewed.

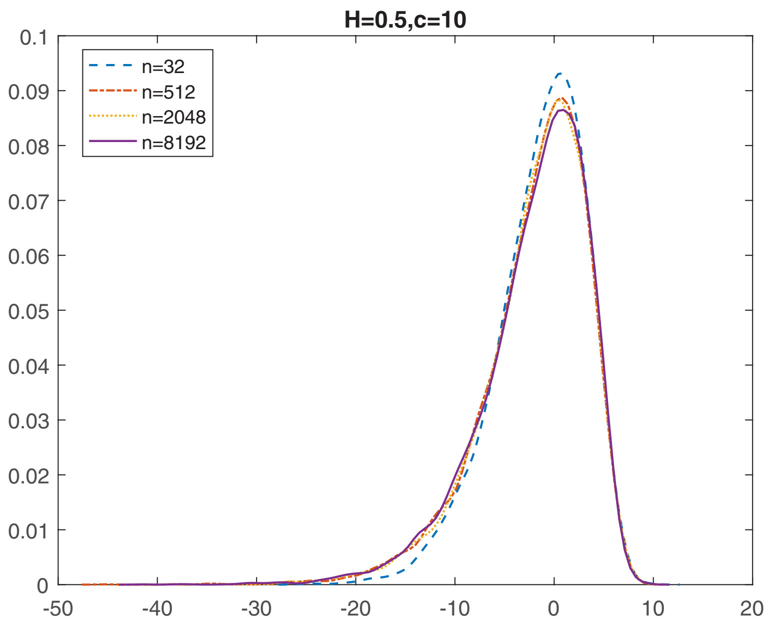


Fig. 1. The Density of $n(\hat{\rho}_n - \rho_n)$.

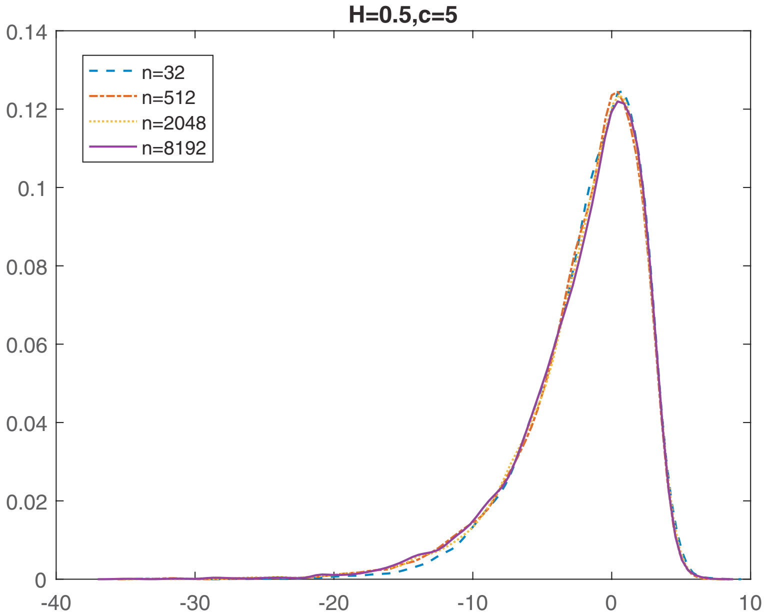


Fig. 2. The Density of $n(\hat{\rho}_n - \rho_n)$.

Figs. 3 and 4 display the density of $n(\hat{\rho}_n - \rho_n)$ when $H = 0.9$ and $c = 10$ and 5. For both values of c, the density when $n = 32$ is very different from that when $n = 8{,}192$. The density for $n = 2{,}048$ is very close to that for $n = 8{,}192$. For small values of n, the density is left-skewed. Interestingly, the density becomes right-skewed when n is larger. Although the same rate applies to $H = 0.5$ and $H > 0.5$, the convergence of the density is much slower when $H > 0.5$ than that when $H = 0.5$. This study indicates that the asymptotic distribution approximates the finite sample distribution less accurately when $H > 0.5$ than when $H = 0.5$ if n is small.

Figs. 5 and 6 display the density of $n^{2H}(\hat{\rho}_n - \rho_n)$ when $H = 0.1$ and $c = 10$ and 5. For both values of c, the density when $n = 32$ is hugely different from those for other values of n, suggesting one would make a terrible mistake by using the limit distribution to approximate the finite sample distribution when $n = 32$. However, the densities for $n = 2{,}048$, 8,192 are close to each other. For all values of n, the density of $n^{2H}(\hat{\rho}_n - \rho_n)$ is symmetric.

5.2. With Fitted Intercept

For each time series simulated, we now estimate ρ_n by $\tilde{\rho}_n$ and calculate $n(\tilde{\rho}_n - \rho_n)$ when $H \geq 0.5$ and $n^{2H}(\tilde{\rho}_n - \rho_n)$ when $H < 0.5$. The 200,000 replications are used to obtain density of $n(\tilde{\rho}_n - \rho_n)$ or $n^{2H}(\tilde{\rho}_n - \rho_n)$.

Figs. 7 and 8 display the densities of $n(\tilde{\rho}_n - \rho_n)$ when $H = 0.5$ and $c = 10$ and 5. For every value of c, the densities are almost identical for all n, suggesting that

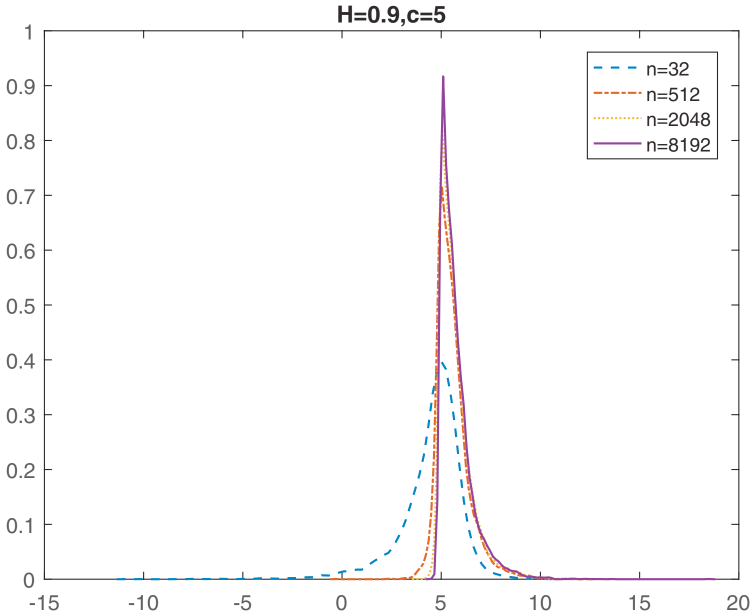


Fig. 3. The Density of $n(\hat{\rho}_n - \rho_n)$.

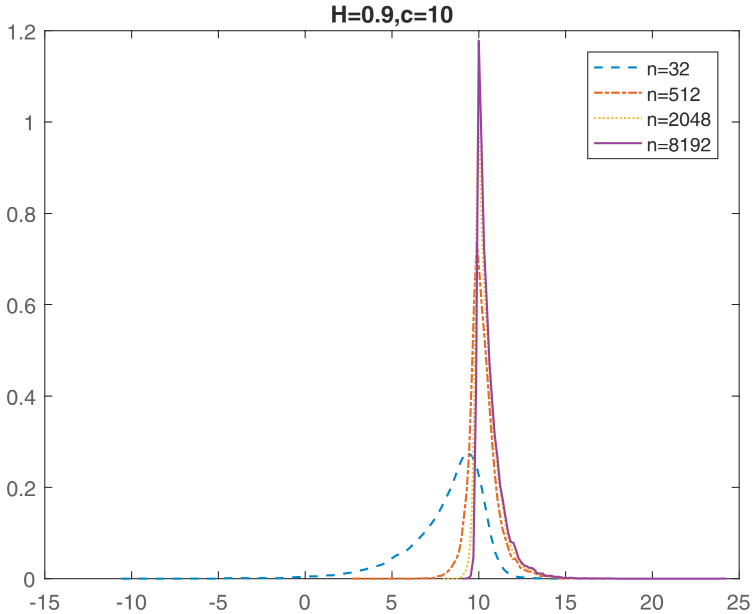


Fig. 4. The Density of $n(\hat{\rho}_n - \rho_n)$.

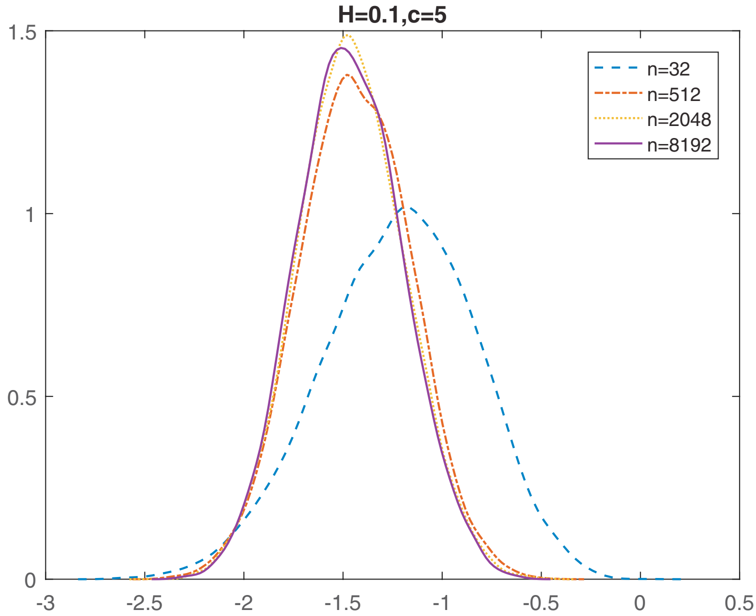


Fig. 5. The Density of $n^{0.2}\left(\hat{\rho}_n - \rho_n\right)$.

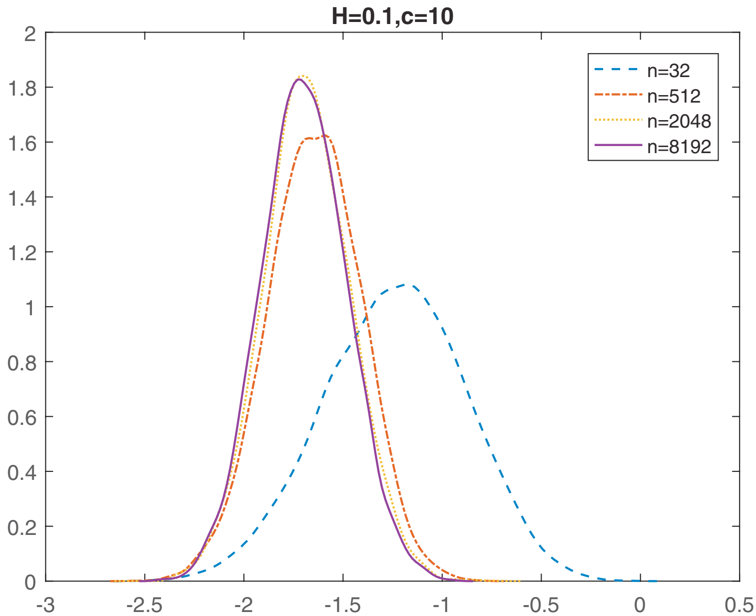


Fig. 6. The Density of $n^{0.2}\left(\hat{\rho}_n - \rho_n\right)$.

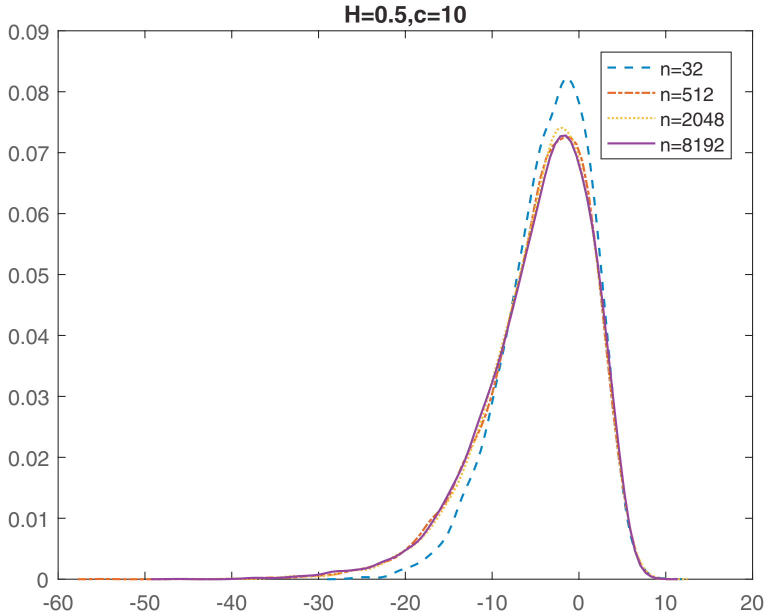


Fig. 7. The Density of $n(\tilde{\rho}_n - \rho_n)$.

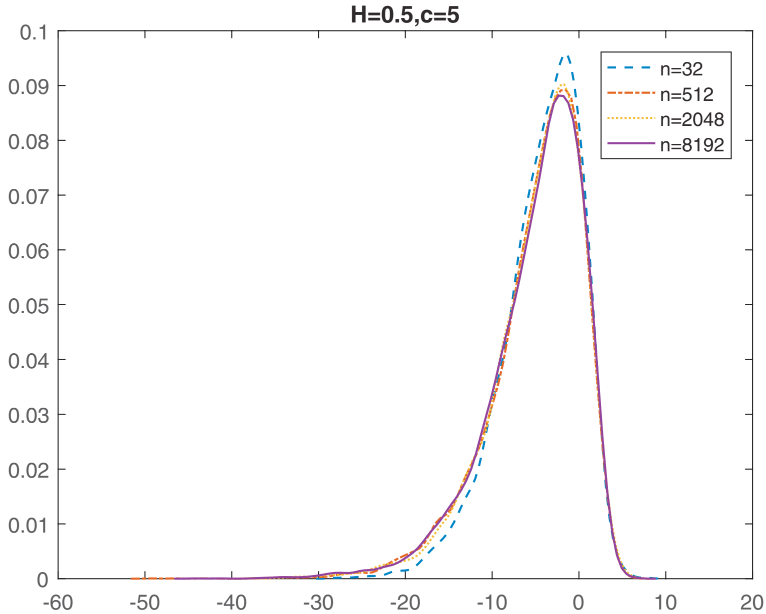


Fig. 8. The Density of $n(\tilde{\rho}_n - \rho_n)$.

the limiting distribution provides accurate approximations to the finite sample distribution when the sample size is as small as 32. In all cases, the density is left-skewed. Compared with Figs. 1 and 2, the densities in Figs. 7 and 8 are more spread. This is expected as the intercept is also fitted.

Figs. 9 and 10 display the densities of $n(\tilde{\rho}_n - \rho_n)$ when $H = 0.9$ and $c = 10$, 5. For both values of c, the density when $n = 32$ is very different from that when $n = 8{,}192$. The density for $n = 2{,}048$ is very close to that for $n = 8{,}192$. The densities are right-skewed for all n. Although the same rate applies to $H = 0.5$ and $H > 0.5$, the convergence in density is much slower when $H > 0.5$ than when $H = 0.5$. This study indicates that the asymptotic distribution approximates the finite sample distribution less accurately when $H > 0.5$ than when $H = 0.5$ if n is small. Compared with Figs. 3 and 4, the densities in Figs. 9 and 10 are more spread, as expected.

Figs. 11 and 12 display the density of $n^{2H}(\tilde{\rho}_n - \rho_n)$ when $H = 0.1$ and $c = 10$ and 5. For both values of c, the density when $n = 32$ is hugely different from those for other values of n, suggesting that one would make a terrible mistake by using the limiting distribution to approximate the finite sample distribution when $n = 32$. However, the densities for $n = 2{,}048$ and 8,192 are nearly identical. The density is symmetric for all n. Compared with Figs. 5 and 6, the densities in Figs. 11 and 12 are more spread, as expected.

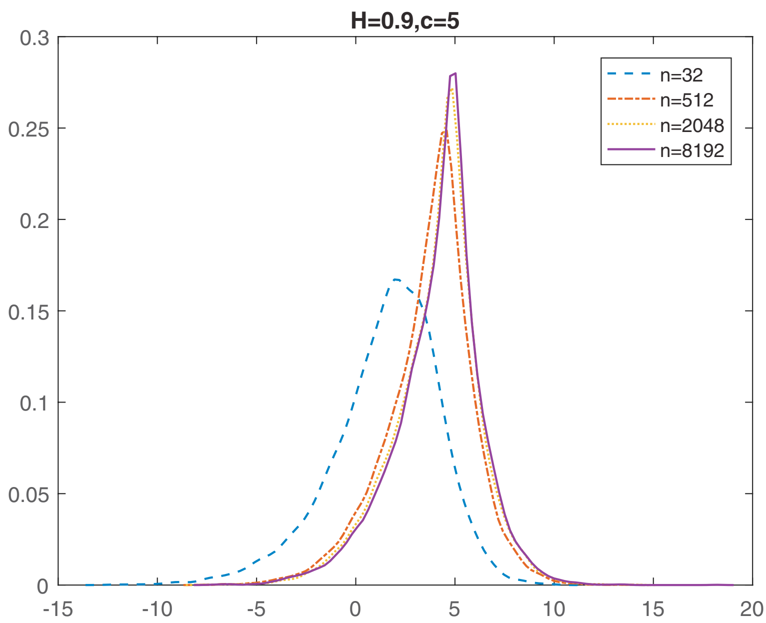


Fig. 9. The Density of $n(\tilde{\rho}_n - \rho_n)$.

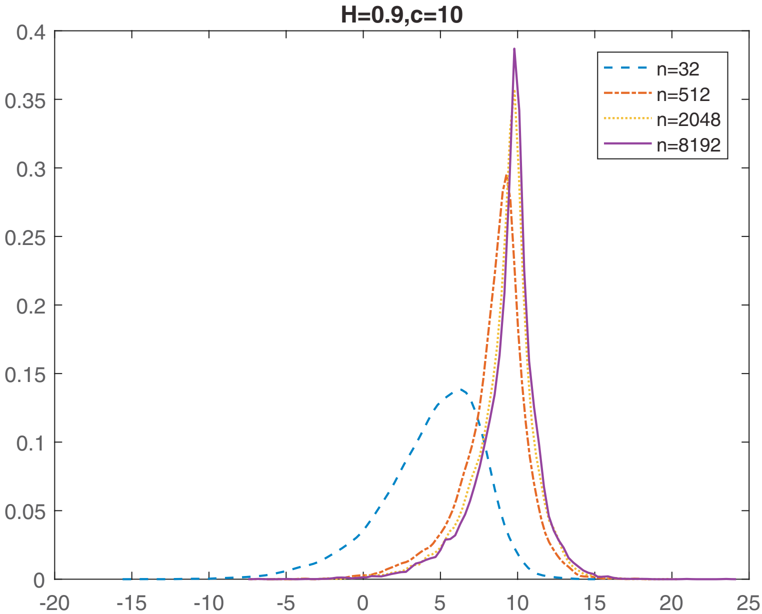


Fig. 10. The Density of $n\left(\tilde{\rho}_n - \rho_n\right)$.

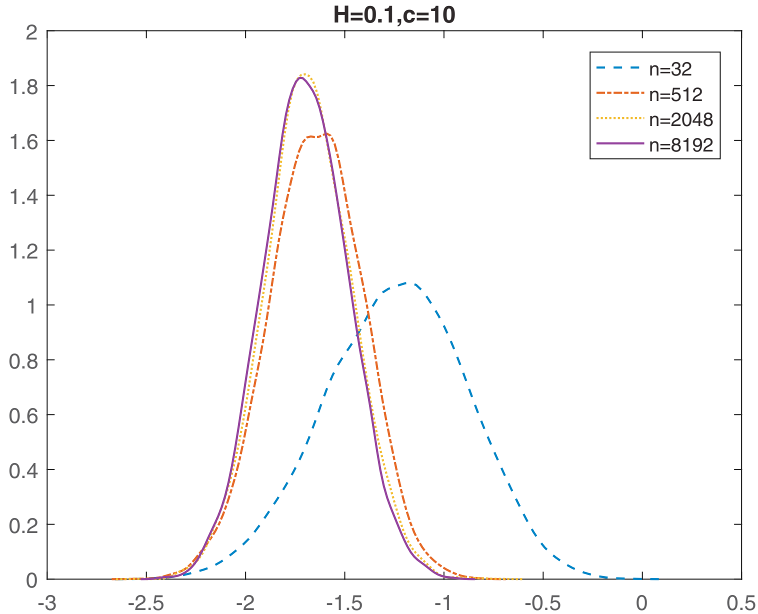


Fig. 11. The density of $n^{0.2}\left(\tilde{\rho}_n - \rho_n\right)$.

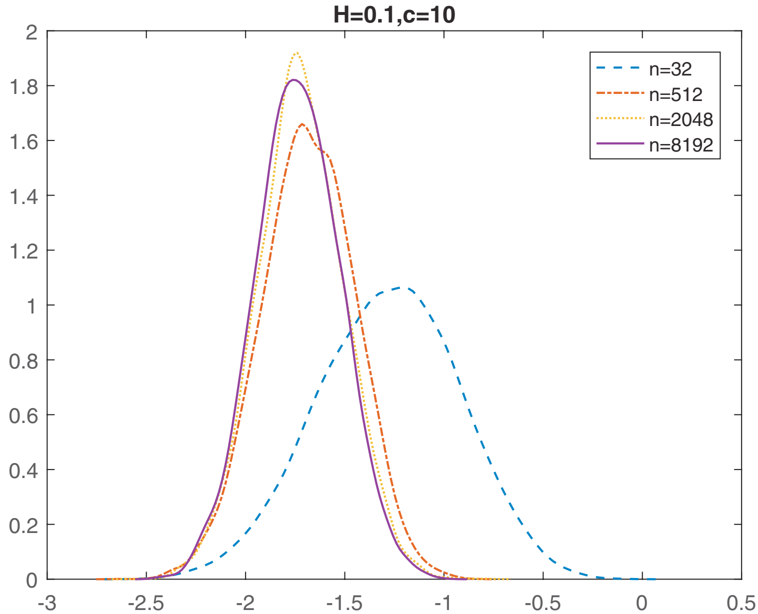


Fig. 12. The Density of $n^{0.2}\left(\tilde{\rho}_n - \rho_n\right)$.

6. CONCLUSIONS

In this chapter, we study the properties of the LS estimator (with and without the intercept fitted) of the AR parameter in local to unity processes when errors are assumed to be FGNs with the Hurst parameter H. It is shown that the estimator is consistent when $H \in (0,1)$. Moreover, the rate of convergence is n when $H \in [0.5,1)$, whereas the rate of convergence is n^{2H} when $H \in (0,0.5)$. This result suggests that the estimator has a slower rate of consistency when $H \in (0,0.5)$ than when $H \in [0.5,1)$.

Furthermore, the limiting distribution of the centered LS estimator depends on H. When $H = 0.5$, the limiting distribution is the same as that obtained in Phillips (1987a) for the local to unity model with errors for which the standard functional central theorem is applicable. When $H > 0.5$ or when $H < 0.5$, the limiting distributions are new to the literature. The limiting distribution for $H > 0.5$ has one term less than that for $H = 0.5$. The limiting distribution for $H < 0.5$ has three terms less than that for $H = 0.5$. Simulation studies are performed to check the reliability of the asymptotic approximation. When $H > 0.5$, a large sample size is needed for the limiting distribution to provide an accurate approximation to the finite sample distribution. When $H = 0.5$, a small sample size is enough for the limiting distribution to provide an accurate approximation to the finite sample distribution. When $H < 0.5$, a moderate sample size is needed for the limiting distribution to approximate the finite sample distribution accurately.

APPENDIX

Proof of Lemma 1. To prove Lemma 1.1, we first note that

$$\begin{aligned} X_t &= \rho_n X_{t-1} + \varepsilon_t = \rho_n^t X_0 + \sum_{j=0}^{t-1} \rho_n^j \varepsilon_{t-j} = \rho_n^t X_0 + \sum_{s=1}^{t} \rho_n^{t-s} \varepsilon_s \\ &\sim \rho_n^t X_0 + \sigma \sum_{s=1}^{t} \rho_n^{t-s} \left[B^H(s) - B^H(s-1) \right] \\ &\sim \sim \rho_n^t X_0 + n^H \sigma \sum_{s=1}^{t} \rho_n^{t-s} \left[B^H\left(\frac{s}{n}\right) - B^H\left(\frac{s-1}{n}\right) \right] \\ &= \rho_n^t X_0 + n^H \sigma \sum_{s=1}^{t} \rho_n^{t-s} \int_{(s-1)/n}^{s/n} dB^H(r) \end{aligned},$$

where the fifth equation is from the similarity property of the fBm. We then have

$$\begin{aligned} n^{-H} X_t &\sim e^{-ct/n} \frac{X_0}{n^H} + \sigma \sum_{s=1}^{t} \int_{(s-1)/n}^{s/n} e^{-c(t-s)/n} dB^H(r) \\ &= e^{-ct/n} \frac{X_0}{n^H} + \sigma \sum_{s=1}^{t} \int_{(s-1)/n}^{s/n} e^{-c(t/n-r)} e^{-c(r-s/n)} dB^H(r) \\ &= e^{-ct/n} \frac{X_0}{n^H} + \sigma \sum_{s=1}^{t} \int_{(s-1)/n}^{s/n} e^{-c(t/n-r)} \left[1 + O(1/n) \right] dB^H(r) \\ &= e^{-ct/n} \frac{X_0}{n^H} + \sigma \int_0^{t/n} e^{-c(t/n-r)} dB^H(r) + O_p(1/n) \\ &= \sigma e^{-ct/n} \left[\pi_0 + o_p(1) \right] + \sigma \int_0^{t/n} e^{-c(t/n-r)} dB^H(r) + O_p(1/n) \\ &= e^{-ct/n} \sigma \pi_0 + \sigma J_c^H(t/n) + o_p(1) \end{aligned}$$

where the third equation is from the Taylor expansion of $e^{-c(r-s/n)}$ and the last equation comes from the definition of the fOU process $J_c^H(t/n)$ given in (16). Hence, for any $r \in [0,1]$.

$$n^{-H} X_{[nr]} \sim \exp\left\{ -c \frac{[nr]}{n} \right\} \sigma \pi_0 + \sigma J_c^H\left(\frac{[nr]}{n} \right) + o_p(1) \Rightarrow e^{-cr} \sigma \pi_0 + \sigma J_c^H(r), \text{ as } n \to \infty.$$

Since $n^{-H} X_{[nr]}$ is a Gaussian process with a finite first-order absolute moment, it is easy to show that the above result holds uniformly in $r \in [0,1]$ under the Skorokhod topology. This proves Lemma 1.1.

Then, the convergence results in Lemmas 1.2 and 1.3 can be obtained straightforwardly by using the continuous mapping theorem (Billingsley, 1968, p. 30).

To prove the results in Lemmas 1.4 and 1.5, we first have

$$\begin{aligned} X_t^2 &= \left(\rho_n X_{t-1} + \varepsilon_t\right)^2 = \rho_n^2 X_{t-1}^2 + 2\rho_n X_{t-1}\varepsilon_t + \varepsilon_t^2 \\ &= X_{t-1}^2 + \left(\rho_n^2 - 1\right) X_{t-1}^2 + 2\rho_n X_{t-1}\varepsilon_t + \varepsilon_t^2 \end{aligned},$$

and

$$\sum_{t=1}^{n} X_{t-1}\varepsilon_t = \frac{1}{2\rho_n}\left\{X_n^2 - X_0^2 - \left(\rho_n^2 - 1\right)\sum_{t=1}^{n} X_{t-1}^2 - \sum_{t=1}^{n}\varepsilon_t^2\right\}.$$

From the results in Lemmas 1.1–1.3, we have

$$\frac{X_n^2 - X_0^2}{n^{2H}} \Rightarrow \sigma^2\left\{\left[J_c^H(1) + e^{-c}\pi_0\right]^2 - \pi_0^2\right\}$$

and

$$n^{-2H}\left(\rho_n^2 - 1\right)\sum_{t=1}^{n} X_{t-1}^2 \Rightarrow -2c\sigma^2\int_0^1\left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr.$$

It is crucially important to note that $\sum_{t=1}^{n}\varepsilon_t^2 = O_p(n)$ for all values of $H \in (0,1)$ and

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 &\sim n^{-1}\sigma^2\sum_{t=1}^{n}\left[B^H(t) - B^H(t-1)\right]^2 \\ &\sim n^{-1+2H}\sigma^2\sum_{t=1}^{n}\left[B^H\left(\frac{t}{n}\right) - B^H\left(\frac{t-1}{n}\right)\right]^2 \xrightarrow{p} \sigma^2 \end{aligned},$$

where the convergence result is from Proposition 4.2 in Viitasaari (2019). As a result,

$$n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2 \xrightarrow{p} \begin{cases} 0 \text{ when } H > 0.5 \\ \sigma^2 \text{ when } H = 0.5 \\ +\infty \text{ when } H = 0.5 \end{cases}.$$

This is the reason why $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ having distinct asymptotic behaviors when H takes various values.

Now, it can be seen clearly that, when $H = 0.5$, the four items in the decomposition of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ have the same order and, as $n \to \infty$,

$$\begin{aligned} & n^{-2H}\sum_{t=1}^{n} X_{t-1}\varepsilon_t \\ = & \frac{1}{2\rho_n}\left\{\frac{X_n^2 - X_0^2}{n^{2H}} - n\left(\rho_n^2 - 1\right)\frac{1}{n^{1+2H}}\sum_{t=1}^{n} X_{t-1}^2 - n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2\right\} \\ \Rightarrow & \frac{\sigma^2}{2}\left\{\left[J_c^H(1) + e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1\left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr - 1\right\} \end{aligned}.$$

Whereas, when $H > 0.5$, it has $2H > 1$. Thus, $\sum_{t=1}^{n} \varepsilon_t^2 = O_p(n)$ is asymptotically dominated by the other terms in the decomposition of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$. Hence, it disappears in the limit of $n^{-2H}\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ that takes the form of

$$\begin{aligned} n^{-2H}\sum_{t=1}^{n} X_{t-1}\varepsilon_t &= \frac{1}{2\rho_n}\left\{\frac{X_n^2 - X_0^2}{n^{2H}} - n\left(\rho_n^2 - 1\right)\frac{1}{n^{1+2H}}\sum_{t=1}^{n} X_{t-1}^2 - n^{-2H}\sum_{t=1}^{n}\varepsilon_t^2\right\} \\ &\Rightarrow \frac{\sigma^2}{2}\left\{\left[J_c^H(1) + e^{-c}\pi_0\right]^2 - \pi_0^2 + 2c\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr\right\} \end{aligned}.$$

In contrast, when $H < 0.5$, $2H < 1$, thereby, $\sum_{t=1}^{n} \varepsilon_t^2 = O_p(n)$ asymptotically dominates the other terms in the decomposition of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$. Hence,

$$n^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t = \frac{1}{2\rho_n}\left\{o_p(1) - n^{-1}\sum_{t=1}^{n}\varepsilon_t^2\right\} \xrightarrow{p} -\frac{\sigma^2}{2}.$$

The Proof of Lemma 1 is completed.

Proof of Theorem 2: The theorem is the direct consequence of Lemmas 1.3–1.5. In particular, (17) and (18) follow from Lemmas 1.3 and 1.4 and (19) follow from Lemmas 1.3 and 1.5.

Proof of Theorem 3: The centered LS estimator given in (6) has the following representation,

$$\tilde{\rho}_n - \rho_n = \frac{\sum_{t=1}^{n} X_{t-1}\varepsilon_t - n^{-1}\left(\sum_{t=1}^{n} X_{t-1}\right)\left(\sum_{t=1}^{n}\varepsilon_t\right)}{\sum_{t=1}^{n} X_{t-1}^2 - n^{-1}\left(\sum_{t=1}^{n} X_{t-1}\right)^2}.$$

From the results in Lemmas 1.2 and 1.3, when $n \to \infty$, the large sample theory of the denominator of $\tilde{\rho}_n - \rho_n$ is obtained as

$$\begin{aligned} & n^{-1-2H}\sum_{t=1}^{n} X_{t-1}^2 - n^{-2-2H}\left(\sum_{t=1}^{n} X_{t-1}\right)^2 \\ \Rightarrow\ & \sigma^2\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right]^2 dr - \sigma^2\left(\int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right] dr\right)^2 \end{aligned}.$$

The numerator of $\tilde{\rho}_n - \rho_n$ has two components, $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ and $-n^{-1}\left(\sum_{t=1}^{n} X_{t-1}\right)\left(\sum_{t=1}^{n}\varepsilon_t\right)$, respectively. The second term, based on the results in Equation (14) and in Lemma 1.2, has the order of n^{2H} and the following large sample property as $n \to \infty$:

$$-n^{-1-2H}\left(\sum_{t=1}^{n} X_{t-1}\right)\left(\sum_{t=1}^{n} \varepsilon_t\right) \Rightarrow -\sigma^2 \int_0^1 \left[J_c^H(r) + e^{-cr}\pi_0\right] dr \cdot B^H(1).$$

Whereas the order of the first term, that is $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$, is n^{2H} when $H \geq 0.5$, and n when $H < 0.5$, as proved in Lemmas 1.4 and 1.5. Therefore, when $H \geq 0.5$ the two terms in the numerator of $\tilde{\rho}_n - \rho_n$ have the same magnitude and are equally important as $n \to \infty$. In this case, the numerator has the order of n^{2H} and the limit can be obtained straightforwardly from Lemma 1.4. Note that the limits of $\sum_{t=1}^{n} X_{t-1}\varepsilon_t$ are different with each other when $H = 0.5$ and $H > 0.5$.

In contrast, when $H < 0.5$, the first term in the numerator of $\tilde{\rho}_n - \rho_n$ dominates the second term. In this case, the numerator has the order of n and the following limit as $n \to \infty$:

$$\begin{aligned} & n^{-1}\left[\sum_{t=1}^{n} X_{t-1}\varepsilon_t - n^{-1}\left(\sum_{t=1}^{n} X_{t-1}\right)\left(\sum_{t=1}^{n} \varepsilon_t\right)\right] \\ = \; & n^{-1}\sum_{t=1}^{n} X_{t-1}\varepsilon_t + o_p(1) \xrightarrow{p} -\sigma^2/2 \end{aligned},$$

where the last limit comes from the result in Lemma 1.5.

From the limits obtained above of the numerator and the denominator of the estimator $\tilde{\rho}_n - \rho_n$, the large sample theory presented in Theorem 3 can be obtained straightforwardly. The proof of Theorem 3 is completed.

NOTES

1. Equations (11)–(13) are different from those reported in Theorem 3 in Sowell (1990). This is because, as remarked in Section 3 of Marinucci and Robinson (1999), the partial sum of an $I(d)$ process, adjusted an appropriate normalizing term, should converge to the Type I fBm denoted by $B^H(t)$ in this chapter, not to the Type II fBm adopted in Sowell (1990).
2. The choice of $H = 0.1$ is empirically relevant for modeling logarithmic realized volatility, as found in Gatheral et al. (2018) and Wang et al. (2021).

ACKNOWLEDGMENTS

Wang acknowledges the support by Shanghai Pujiang Program under No.22PJC022. Yu would like to acknowledge that this research/project is supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002).

REFERENCES

Billingsley, P. (1968). *Convergence of probability measures*. Wiley.

Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, *15*(3), 1050–1063

Cheridito, P., Kawaguchi, H., & Maejima, M. (2003). Fractional Ornstein–Uhlenbeck processes. *Electronical Journal Probability*, *8*(3), 1–14.

Davydov, Y. A. (1970). The invariance principle for stationary processes. *Theory of Probability and its Applications*, *15*, 487–489.

Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, *18*(6), 933–949.

Hamilton, J. (1994). *Time series analysis*. Princeton University Press.

Marinucci, D., & Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, *80*, 111–122.

Mikusheva, A. (2015). Second order expansion of t-statistic in autoregressive models. *Econometric Theory*, *31*(3), 426–448.

Park, J. Y. (2003). *Weak unit roots*. Working Paper. Department of Economics, Rice University.

Phillips, P. C. B. (1987a). Toward a unified asymptotic theory for autoregression. *Biometrika*, *74*, 533–547.

Phillips, P. C. B. (1987b). Time series regression with a unit root. *Econometrica*, *55*, 277–301.

Sowell, F. (1990). The fractional unit root distribution. *Econometrica*, *58*, 495–505.

Taqqu, M. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *31*, 287–302.

Viitasaari, L. (2019). Necessary and sufficient conditions for limit theorems for quadratic variations of Gaussian sequences. *Probability Survey*, *16*, 62–98.

Wang, X., Xiao, W., & Yu, J. (2021). Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process. *Journal of Econometrics*, forthcoming. doi:10.1016/j.jeconom.2021.08.001

CHAPTER 3

POWERFUL SELF-NORMALIZING TESTS FOR STATIONARITY AGAINST THE ALTERNATIVE OF A UNIT ROOT

Uwe Hassler[a] and Mehdi Hosseinkouchack[b]

[a]Goethe-Universität Frankfurt, Frankfurt am Main, Germany
[b]EBS Universität, Oestrich-Winkel, Germany

ABSTRACT

The authors propose a family of tests for stationarity against a local unit root. It builds on the Karhunen–Loève (KL) expansions of the limiting CUSUM process under the null hypothesis and a local alternative. The variance ratio type statistic $\mathcal{VR}_q$ is a ratio of quadratic forms of q weighted Gaussian sums such that the nuisance long-run variance cancels asymptotically without having to be estimated. Asymptotic critical values and local power functions can be calculated by standard numerical means, and power grows with q. However, Monte Carlo experiments show that q may not be too large in finite samples to obtain tests with correct size under the null. Balancing size and power results in a superior performance compared to the classic KPSS test.

Keywords: I(0); KPSS; variance ratio tests; Karhunen-Loève; scale invariance; local alternative
JEL classifications: C12 (hypothesis testing); C22 (time series)

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 97–114

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A003

1. INTRODUCTION

Stationarity of economic and financial time series is an underlying assumption of many models, and more specifically integration of order 0, $I(0)$, is a prerequisite of many econometric techniques. For this reason, the so-called KPSS test by Kwiatkowski et al. (1992) enjoys great popularity in empirical economics. It relies on consistent estimation of the so-called long-run variance under the null hypothesis, which is a notoriously difficult issue; for a recent treatment in a continuous time framework under high frequency, see Lu and Park (2019) and Jiang et al. (2020).

To circumvent long-run variance estimation one may call on the principle of self-normalization or scale invariance, see the low-frequency stationary test (LFST) by Müller and Watson (2008). It relies on the likelihood ratio statistic of weighted averages of the data characterized by different covariance matrices under the null hypothesis and the local alternative. The alternative assumes a so-called local level model, testing for $I(0)$ against a local random walk. The long-run variance cancels from the likelihood ratio statistic that hence enjoys the property of self-normalization, at least asymptotically. Critical values have been determined by simulation.

In this chapter, we adopt the local level model and the idea of computing ratios of weighted averages. However, we follow the proposal by Hassler and Hosseinkouchack (2019) and use weighting schemes obtained from the respective KL expansion. To this end, we derive the eigenstructure of the autocovariance kernels of the respective limiting processes under different assumptions on the deterministic component in Propositions 1–3. For testing we consider only one of the two factors of the likelihood ratio statistic corresponding to a kind of variance ratio of quadratic forms ($\mathcal{VR}_q$ say) in (asymptotically) normal variates, which allows to compute critical values and local power function analytically from Hassler and Hosseinkouchack (2019, Thm. 1). Here, q is the number of weighted averages included in the statistic, which provides a family of tests, each having asymptotically correct size for any finite q, while asymptotic power grows with q. Monte Carlo experiments, however, show that q may not be too large to control the size of the test under the null hypothesis in small samples. A choice of $q = 25$ results in a procedure that outperforms KPSS and LFST in terms of size and power for many realistic combinations of sample size and strength of persistence under the alternative.

The rest of this chapter is organized as follows. Section 2 becomes precise on the models and assumptions and briefly reviews KPSS. Section 3 introduces our tests as self-normalizing alternatives to KPSS. Asymptotic and finite sample power is addressed in Section 4, while some conclusions are drawn in the Section 5. Mathematical proofs are relegated to the Appendix.

A word on notation before we begin. Weak convergence in the Skorohod space of cadlag functions is denoted by $\Rightarrow$ as the sample size T goes off to infinity. Integrals are taken from 0 to 1, unless indicated otherwise. $\lfloor x \rfloor$ returns the largest integer smaller than or equal to some real x, $x > 0$. And sinh and cosh denote the hyperbolic sine and cosine, respectively.

2. KPSS

We adopt the local level model maintained by Müller and Watson (2008),

$$y_t = d_t + \frac{\sigma_\eta}{T}\sum_{j=1}^{t}\eta_j + e_t, \quad t = 1,\ldots,T, \tag{1}$$

with the null hypothesis $\sigma_\eta = 0$ and the (local) alternative $\sigma_\eta > 0$. We assume three models $M0$ through $M2$, where $M0$ stands for the case without deterministics ($d_t = 0$), while $M1$ and $M2$ allow for a constant term and a linear time trend, respectively, $d_t = \mu$ and $\mu + \delta t$. The least squares residuals under the null hypothesis are denoted accordingly, $\hat{u}_{t,0} = y_t$ and $\hat{u}_{t,1} = y_t - \bar{y}$ for $M0$ and $M1$, respectively, and $\hat{u}_{t,2} = y_t - \hat{\mu} - \hat{\delta} t$ for $M2$. The KPSS statistics build on the partial sums under the null hypothesis,

$$S_{t,k} = \sum_{j=1}^{t}\hat{u}_{j,k}, \quad k = 0,1,2, \tag{2}$$

and become

$$\mathcal{KPSS}_k = \frac{\sum_{t=1}^{T} S_{t,k}^2}{T^2\hat{\omega}_{e,k}^2}, \quad k = 0,\ 1,\ 2. \tag{3}$$

Here, $\hat{\omega}_{e,k}^2$ is a consistent estimator of the long-run variance from Assumption 1. The limiting distributions and critical values are discussed in Kwiatkowski et al. (1992). For the rest of this chapter, we will work under assumptions parallel to Kwiatkowski et al. (1992) and Müller and Watson (2008).

Assumption 1. *Let $\{e_t, \eta_t\}'$ be a zero mean vector process with*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\begin{pmatrix} e_t \\ \eta_t \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_e W_e(r) \\ W_\eta(r) \end{pmatrix} \quad \text{for } r \in [0,1] \text{ as } T \to \infty,$$

where W_e and W_η are standard Wiener processes independent of each other, and $0 < \omega_e^2 < \infty$.

Kwiatkowski et al. (1992) did not consider model $M0$. It is, however, straightforward under $\sigma_\eta = 0$ with Assumption 1 that

$$\text{KPSS}_0 = \frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\hat{u}_{j,0}\right)^2}{T^2\hat{\omega}_{e,0}^2} \Rightarrow \int_0^1 W_e^2(r)\,\mathrm{d}r;$$

critical values are available from MacNeill (1978, Table 2) and Nyblom (1989, Table 1).

Table 1. Critical Values $\ell_q(\alpha)$ for Model *Mk*, $k = 0, 1, 2$.

		$q=10$	$q=15$	$q=25$	$q=50$	$q=100$
$k=0$	$\ell_q(0.01)$	2.3628	1.7128	1.3608	1.1606	1.0760
	$\ell_q(0.05)$	1.8156	1.4447	1.2317	1.1052	1.0502
	$\ell_q(0.10)$	1.6268	1.3460	1.1819	1.0831	1.0398
	θ_1	5.2583	4.7661	4.4503	4.2477	4.1555
$k=1$	$\ell_q(0.01)$	3.0622	1.9622	1.4589	1.1975	1.0922
	$\ell_q(0.05)$	2.3833	1.6627	1.3226	1.1409	1.0662
	$\ell_q(0.10)$	2.1282	1.5456	1.2674	1.1173	1.0553
	θ_1	10.7880	9.1443	8.2355	7.6971	7.4618
$k=2$	$\ell_q(0.01)$	5.3490	2.4653	1.6187	1.2511	1.1145
	$\ell_q(0.05)$	4.1877	2.0936	1.4683	1.1920	1.0880
	$\ell_q(0.10)$	3.7225	1.9402	1.4045	1.1664	1.0763
	θ_1	25.0919	17.7328	14.7969	13.2726	12.6451

Note: Values for θ_1 are all calculated such that the tests have a rejection rate of 50% at a 5% nominal size when the alternative hypothesis is $\theta = \theta_1$.

KPSS builds on the Lagrange Multiplier principle, see Tanaka (1983), and becomes locally best invariant under Gaussianity and absence of serial correlation, while no uniformly most powerful tests exist, see Nyblom and Mäkeläinen (1983) and similarly Nabeya and Tanaka (1988, Sect. 4.1) for the case without deterministics. Under serial correlation, KPSS is plagued by the long-run variance ω_e^2 that has to be removed. Of course, this is a standard problem often labeled HAC (heteroskedasticity and autocorrelation consistent) variance estimation in econometrics. The vast majority of HAC estimators falls into the class of kernel estimators, see the pioneering papers by Newey and West (1987) and Andrews (1991), cf. also Liu and Wu (2010). All these estimators are subject to the critique by Müller (2007, p. 1338)

> […] that any consistent long-run variance estimator is necessarily a discontinuous function […], i.e. sample paths […] that are close in the sup norm do not in general lead to similar long-run variance estimates.[1]

To circumvent this non-robust estimation we call on the principle of self-normalization or scale invariance and discuss ratio tests.

3. VARIANCE RATIO TESTS

3.1. Limiting Processes

Let us stick to the normalized partial sum process from the residuals computed under H_0:

$$x_{\lfloor rT \rfloor,k} := \frac{S_{\lfloor rT \rfloor,k}}{\sqrt{T}} = \frac{\sum_{j=1}^{\lfloor rT \rfloor} \hat{u}_{j,k}}{\sqrt{T}}, \quad k = 0,1,2.$$

Table 2. Experimental Size at Level 5% for $M1$.

ϕ	KPSS_1	LFST_1	$\text{VR}_{1,15}$	$\text{VR}_{1,25}$	$\text{VR}_{1,50}$	$\text{VR}_{1,100}$
$T = 100$						
−0.75	0.58	5.55	4.45	3.91	1.21	0.00
−0.50	1.30	5.56	4.34	3.52	1.21	0.00
−0.25	2.69	5.78	4.54	4.05	1.95	0.01
0.00	4.74	5.89	4.60	4.59	3.69	1.08
0.25	6.75	6.99	5.82	6.19	9.36	12.29
0.50	9.73	8.55	7.61	11.95	29.62	53.86
0.75	15.94	17.08	17.94	37.54	75.64	94.87
$T = 200$						
−0.75	1.59	5.78	4.68	4.76	3.84	1.30
−0.50	2.89	5.95	4.95	4.75	3.90	1.32
−0.25	3.35	6.35	5.50	4.66	4.35	1.96
0.00	4.72	5.89	4.84	4.66	4.42	3.46
0.25	6.41	6.44	5.16	5.38	6.54	10.21
0.50	7.39	6.26	5.28	6.52	12.16	30.31
0.75	12.67	9.82	9.05	16.59	42.68	80.34
$T = 500$						
−0.75	2.61	6.44	5.05	5.16	4.93	3.99
−0.50	3.35	6.28	4.90	5.08	4.83	4.35
−0.25	3.71	6.16	5.00	5.39	4.87	4.37
0.00	5.11	6.03	4.73	4.84	4.65	4.58
0.25	6.00	5.83	5.17	5.06	5.51	6.11
0.50	6.87	6.34	5.43	5.41	6.78	10.56
0.75	10.69	6.67	5.55	6.94	12.88	33.80
$T = 1{,}000$						
−0.75	2.50	5.88	4.79	4.92	4.97	4.81
−0.50	3.04	5.95	5.08	5.01	5.30	5.12
−0.25	3.76	5.53	4.35	4.53	4.46	4.30
0.00	5.02	5.78	4.72	4.71	4.86	4.91
0.25	5.54	6.11	5.18	5.18	5.00	5.46
0.50	6.18	6.03	5.14	5.28	5.45	6.65
0.75	8.48	6.60	5.48	5.55	7.21	13.61

Note: The data are generated under H_0 with $e_t = \phi e_{t-1} + \varepsilon_t$.

Under Assumption 1, it then holds for $M0$ that

$$x_{\lfloor rT \rfloor,0} \Rightarrow \omega_e X_{0,\theta}(r), \quad X_{0,\theta}(r) := W_e(r) + \theta \int_0^r W_\eta(s)\,\mathrm{d}s \ \text{ with } \ \theta := \frac{\sigma_\eta}{\omega_e}. \tag{4}$$

For $M1$, $x_{\lfloor rT \rfloor,1}$ converges to the corresponding tied down process: $x_{\lfloor rT \rfloor,1} \Rightarrow \omega_e \left(X_{0,\theta}(r) - r\, X_{0,\theta}(1) \right)$ with θ defined in (4). This can be rewritten as

$$X_{1,\theta}(r) := X_{0,\theta}(r) - rX_{0,\theta}(1) = W_e(r) - rW_e(1) + \theta\left(\int_0^r W_\eta(s)\,\mathrm{d}s - r\int_0^1 W_\eta(s)\,\mathrm{d}s \right).$$

Finally, under $M2$ one obtains $x_{\lfloor rT \rfloor,2} \Rightarrow \omega_e X_{2,\theta}(r)$ where

$$X_{2,\theta}(r) := X_{0,\theta}(r) + (2r - 3r^2)X_{0,\theta}(1) + 6(r^2 - r)\int_0^1 X_{0,\theta}(a)\,\mathrm{d}a,$$

with θ as before. Of course, $X_{2,\theta}$ parallels the so-called second-level Brownian bridge from Kwiatkowski et al. (1992, eq. 16).

Due to the independence maintained under Assumption 1, the autocovariance kernel of $X_{0,\theta}$ is

$$K_0(s,t;\theta) = \min(s,t) + \theta^2 K(s,t),$$

where we define $K(s, t)$ as the kernel of the integrated component in (4):

$$K(s,t) := \frac{\min^2(s,t)(3\max(s,t) - \min(s,t))}{6}.$$

The autocovariance kernel of $X_{1,\theta}$ becomes

$$K_1(s,t;\theta) = K_0(s,t;0) - st + \theta^2\left[K(s,t) - sK(1,t) - tK(s,1) + stK(1,1)\right].$$

For $X_{2,\theta}$ one obtains

$$\begin{aligned} K_2(s,t;\theta) = {} & K_1(s,t;0) - 3st(1-s)(1-t) \\ & + \theta^2\Big[\frac{1}{3}g_1(s)g_1(t) + \frac{25}{4}g_2(s)g_2(t) + \frac{5}{4}g_1(s)g_2(t) + \frac{5}{4}g_1(t)g_2(s) \\ & - \frac{1}{4}g_2(t)g_4(s) - \frac{1}{4}g_2(s)g_4(t) - \frac{1}{6}g_1(s)g_3(t) - \frac{1}{6}g_1(t)g_3(s) + k(s,t)\Big] \end{aligned}$$

where $g_1(s) = 4s - 3s^2$, $g_2(s) = s^2 - s$, $g_3(s) = 3s^2 - s^3$, $g_4(s) = 6s^2 - s^4$.

Testing for the null hypothesis $\sigma_\eta = 0$ in model $M0$ through $M2$ amounts to testing for $\theta = \theta_0 = 0$ in $X_{k,\theta}(r)$, k = 0, 1, 2. The following tests are directed against specific alternatives, $\theta = \theta_1 > 0$. In order to denote these two values, we will write θ_h with $h \in \{0,1\}$.

3.2. Test Statistics

Obviously, $X_{0,\theta}(r)$ and $X_{1,\theta}(r)$ share the additive structure maintained in Hassler and Hosseinkouchack (2019, Ass. 2), and one can show that this holds true for $X_{2,\theta}(r)$, too. Hence, we can follow their route and build statistics from the eigenstructures of the kernels. They consist of the eigenfunctions $f_{j,\theta_h,k}(s)$ and eigenvalues $\lambda_{j,\theta_h,k}$ under the null and the alternative, $h \in \{0,1\}$. They are the non-trivial solutions to the Fredholm integral equation ($j = 1,2,\ldots$)

$$f_{j,\theta_h,k}(t) = \lambda_{j,\theta_h,k}\int_0^1 K_k(s,t;\theta_h) f_{j,\theta_h,k}(s)\,\mathrm{d}s.$$

The eigenvalues are the zeros of the Fredholm determinant $D_{\theta_h,k}(\lambda)$ defined as

$$D_{\theta_h,k}(\lambda) = \lim_{T\to\infty} \det\left(I_T - \frac{\lambda}{T}\left[K_k\left(\frac{j}{T},\frac{\ell}{T};\theta_h\right)\right]_{j,\ell=1,\ldots,T}\right),$$

where I_T denotes the identity matrix.

The eigenstructure provides the ingredients for the KL expansion,

$$X_{k,\theta}(r) = \sum_{j=1}^{\infty} \frac{f_{j,\theta,k}(r)}{\lambda_{j,\theta,k}^{1/2}} Z_j,$$

where $\{Z_j\}$ is a sequence of independent standard normal variates. In general, such infinite expansions are not unique, see, for example, the discussion in Phillips (1998). But a truncated KL expansion,

$$X_{k,\theta}(r) = \sum_{j=1}^{q} \frac{f_{j,\theta,k}(r)}{\lambda_{j,\theta,k}^{1/2}} Z_j + \text{error}_{k,\theta}(r),$$

minimizes the total mean squared error for any q. This suggests to compute q weighted sums of the normalized partial sum process under the null and the alternative:

$$X_{j,\theta_h,k} := \sum_{t=1}^{T} \left[\int_{(t-1)/T}^{t/T} f_{j,\theta_h,k}(s)\, \mathrm{d}s \right] \frac{S_{t,k}}{\sqrt{T}}, \quad j = 1, \ldots, q, \quad k = 0, 1, 2. \tag{5}$$

Due to the continuous mapping theorem it holds that

$$X_{j,\theta_h,k} \Rightarrow \omega_e \int_0^1 f_{j,\theta_h,k}(s) X_{k,\theta_h}(s)\, \mathrm{d}s.$$

The long-run variance cancels from the following variance ratio type statistics along the lines of Hassler and Hosseinkouchack (2019):

$$\mathcal{VR}_{k,q} = \frac{\sum_{j=1}^{q} \lambda_{j,\theta_0,k} X_{j,\theta_0,k}^2}{\sum_{j=1}^{q} \lambda_{j,\theta_1,k} X_{j,\theta_1,k}^2}, \quad k = 0, 1, 2. \tag{6}$$

In the limit, one obtains quadratic forms of normal variates independent of a nuisance scaling parameter, say $\mathcal{L}_k(\theta) = \mathcal{L}_{k,q,\theta_0}(\theta) / \mathcal{L}_{k,q,\theta_1}(\theta)$. Of course, $\mathcal{L}_k(\theta)$ depend on q, too, but this is suppressed for convenience. The cumulative distribution function can be computed by numerical means, see Hassler and Hosseinkouchack (2019, Thm. 1), which allows to compute critical values as well as asymptotic power functions.

Remark 1. *Computation of the variance ratio type statistic requires a choice of q. Notice that q does not affect the asymptotic size of the test. The number q does not parallel the bandwidth when computing the KPSS statistic, which has to grow with the sample size to ensure consistent long-run variance estimation. We rather propose a family of asymptotically valid tests for any finite q. Asymptotically, the choice of q affects only power, see Section 4.1. In finite samples, however, it is correct that q has to be small relative to the sample size, see Section 4.2.*

Let $\ell_q(\alpha)$ denote the $(1-\alpha)$ -quantiles of the null distribution $\mathcal{L}_k(\theta_0)$; they are the critical values when testing at level α. The 5% critical values will be determined jointly with θ_1 such that the asymptotic power equals 50% at θ_1. Note that critical values and θ_1 depend on *k, k* = 0, 1, 2, as well as on *q*, which we suppress for notational convenience. These same values for θ_1 are employed to determine critical values at level 1% and 10%, too. A selection is provided in Table 1. The eigenstructures that are required to compute the test statistics are given next.

3.3. Eigenstructure

As in Table 1 we will suppress for simplicity that the eigenstructures depend on the model tested, $k = 0, 1, 2$. We begin with model *M*0 without intercept.

Proposition 1. *For the eigenstructure of* $K_0(s,t;\theta)$ *from M0 it holds that the eigenvalues are the positive roots of* $D_\theta(\lambda) = 0$ *with*

$$D_\theta(\lambda) = \frac{2\theta^2 + \left(2\theta^2 + \lambda\right)\cos\mu_1 \cosh\mu_2 - \mu_1\mu_2 \sin\mu_1 \sinh\mu_2}{4\theta^2 + \lambda}$$

where $\mu_1 = \mu_1(\lambda;\theta)$ *and* $\mu_2 = \mu_2(\lambda;\theta)$ *are short for*

$$\mu_1 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} + \lambda}{2}} \quad \textit{and} \quad \mu_2 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} - \lambda}{2}}.$$

The eigenfunctions are given as

$$f(t) = \alpha_2\left(\alpha_1 \cos(\mu_1 t) + \sin(\mu_1 t) + \alpha_3 \exp(\mu_2 t) + \alpha_4 \exp(-\mu_2 t)\right),$$

with α_1 *through* α_4 *defined in the Appendix.*

Proof: See Appendix.

Note that $\lim_{\theta\to 0}\mu_1 = \sqrt{\lambda}$ and $\lim_{\theta\to 0}\mu_2 = 0$ such that $\lim_{\theta\to 0} D_\theta(\lambda) = \cos\sqrt{\lambda}$ with $\lambda_j = (j - 1/2)^2\pi^2$, which reproduces of course the well-known eigenvalues of the standard Wiener process. Further, careful expansions of α_1 through α_4 around $\theta = 0$ show that $\lim_{\theta\to 0} f(\lambda) = \sqrt{2}\sin(\sqrt{\lambda}t)$.

Next, we turn to the model with intercept.

Proposition 2. *For the eigenstructure of* $K_1(s,t;\theta)$ *from M1 it holds that the eigenvalues are the positive roots of* $D_\theta(\lambda) = 0$ *with*

$$D_\theta(\lambda) = \frac{\sin\mu_1 \sinh\mu_2}{\mu_1\mu_2},$$

where $\mu_1 = \mu_1(\lambda;\theta)$ *and* $\mu_2 = \mu_2(\lambda;\theta)$ *are from Proposition 1. The eigenfunctions are given as*

$$f(t) = \sqrt{2}\sin(\mu_1 t).$$

Proof: See Appendix.

Note that $\lim_{\theta\to 0} D(\lambda) = \frac{\sin\sqrt{\lambda}}{\sqrt{\lambda}}$ with $\lambda_j = j^2\pi^2$, and $\lim_{\theta\to 0} f(t) = \sqrt{2}\sin(j\pi t)$, which of course reproduces the case of a Brownian bridge (see, e.g., Phillips, 1998, p. 1303).

Remark 2. *Somehow surprisingly, the case with intercept is simpler than the case without. Note that for k = 1,* μ_1 *has to be a multiple of π, such that* $f(t) = \sqrt{2}\sin(j\pi t)$ *irrespective of the value of θ. Hence, the weights in (5) are identical,* $f_{j,\theta_0,1}(s) = f_{j,\theta_1,1}(s)$, *which is not the case for k = 0. Further, the eigenvalues required for (6) become for k = 1 simply* $\lambda_{j,\theta_h,1} = \frac{j^4\pi^4}{j^2\pi^2 + \theta_h^2}$.

Finally, we turn to the case with intercept and a linear time trend.

Proposition 3. *For the eigenstructure of* $K_2(s,t;\theta)$ *from M2 it holds that the eigenvalues are the positive roots of* $D_\theta(\lambda) = 0$ *with*

$$D_\theta(\lambda) = \frac{144(2 - 2\cos\mu_1 - \mu_1\sin\mu_1)(2 - 2\cosh\mu_2 + \mu_2\sinh\mu_2)}{\mu_1^4\mu_2^4}$$

where $\mu_1 = \mu_1(\lambda;\theta)$ *and* $\mu_2 = \mu_2(\lambda;\theta)$ *are from Proposition 1. The eigenfunctions are given as*

$$f(t) = \begin{cases} \sqrt{2}\left(\cot\frac{\mu_1}{2}\cos(\mu_1 t) + \sin(\mu_1 t) - \cot\frac{\mu_1}{2}\right) & \text{if } \mu_1\cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2} = 0, \\ \sqrt{2}\sin(\mu_1 t) & \text{if } \sin\frac{\mu_1}{2} = 0. \end{cases}$$

Proof: See Appendix.

By expansions of $x\sinh x$ and 2–2 $\cosh x$ it is not hard to verify that $\lim_{\theta\to 0} D_\theta(\lambda) = \frac{12}{\lambda^2}\left(2 - 2\cos\sqrt{\lambda} - \sqrt{\lambda}\sin\sqrt{\lambda}\right)$, which is the result for the second-level Brownian bridge by Nabeya and Tanaka (1988, Thm. 6).

Remark 3. *The computation of the test statistic from (6) is facilitated by the following considerations. Note that for k = 2,* μ_1 *solves* $\sin\frac{\mu_1}{2}\left(\mu_1\cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2}\right) = 0$. *Let us call these solutions* $\mu_{1,j}$, *where* $\mu_{1,j}$ *is either an even multiple of π or it* *solves* $\mu_1 \cos\frac{\mu_1}{2} - 2\sin\frac{\mu_1}{2} = 0$ *. Since the eigenfunctions are just a function of* μ_1*, we have again that* $f_{j,\theta_0,2}(s) = f_{j,\theta_1,2}(s)$ *implying that the weights in (5) are identical. Further, it holds that* $\lambda_{j,\theta_h,2} = \dfrac{\mu_{1,j}^4}{\mu_{1,j}^2 + \theta_h^2}$.

4. POWER

4.1. Asymptotic

Local power of $\mathcal{VR}_{k,q}$ can be computed along the lines of Hassler and Hosseinkouchack (2019, Thm. 1). Moreover, using Tanaka (1996, Thm. 5.11) we obtain the limiting local power function of the KPSS tests. Since it holds that

$$\mathrm{KPSS}_k \Rightarrow \int_0^1 \left(X_{k,\theta}(r)\right)^2 \mathrm{d}r,$$

the characteristic function $\phi_{\mathrm{KPSS},k}$ of this limit is given by

$$\phi_{\mathrm{KPSS},k}(t) = \frac{1}{\sqrt{D_{k,\theta}(2\mathrm{i}t)}},$$

where $D_{k,\theta}(\cdot)$ are the Fredholm determinants from Propositions 2 and 3 for $k = 1$ and $k = 2$, respectively. Using Lévy's inversion theorem it is only a numerical problem to compute distribution functions (see Figs. 1 and 2). While the KPSS

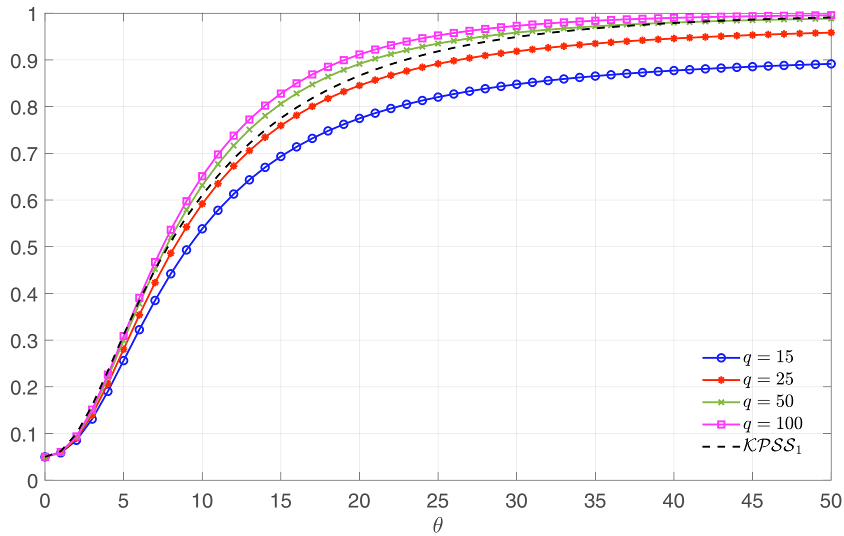


Fig. 1. Limiting Power for Model with Constant ($k = 1$) at Size 5%.

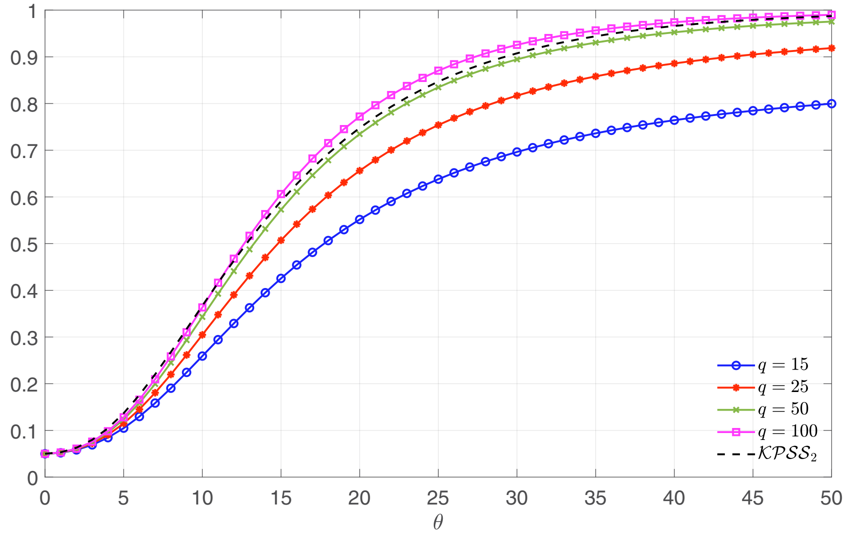


Fig. 2. Limiting Power for Model with Trend ($k = 2$) at Size 5%.

tests are locally best (having the steepest slope of the power function at the null), they are beaten for large enough q as we move away from the null hypothesis.

4.2. Experimental Evidence

All computer experiments were performed with MATLAB and rejection frequencies are computed from 10^4 replications when testing at nominal level of 5%. The errors from models $M0$ through $M2$ are autoregressive of order 1, $e_t = \phi e_{t-1} + \varepsilon_t$, and independent of the random walk. Both white noise innovations are standard normal, that is,

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \text{i.i. } \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

When a long-run variance has to be estimated, we use the Bartlett (or triangular) window popularized in econometrics by Newey and West (1987) and Kwiatkowski et al. (1992) with data-driven bandwidth selection according to Andrews (1991, eq. 6.2), which builds on the optimal rate. Sample sizes are $T \in \{100, 200, 500, 1{,}000\}$. We only report results for the empirically most relevant case $k = 1$ with intercept. Evidence for $k = 0$ and $k = 2$ is available upon request. In particular, the case of detrending produces very similar findings in term of ranking of the tests and choice of q, only that the power functions are generally flatter, as one would expect from Figs. 1 and 2.

For completeness, we also include the LFST test by Müller and Watson (2008). For the model with intercept ($k = 1$) they suggested weighted sums of the residuals, where the weights are derived from the autocovariance kernel of a demeaned Wiener process. Hence,

$$Y_{j,k} := \sum_{t=1}^{T} \left[\int_{(t-1)/T}^{t/T} \sqrt{2} \cos(\pi js)\, \mathrm{d}s \right] \hat{u}_{t,1}, \quad j = 1, \ldots, 13, \tag{7}$$

where the number of weights is restricted to 13, since Müller and Watson (2008) were only interested in lower than business cycle variability. Again, the nuisance parameter ω_e^2 cancels from the statistic asymptotically,

$$\mathrm{LFST}_1 := \frac{\sum_{j=1}^{13} Y_{j,k}^2}{\sum_{j=1}^{13} \frac{Y_{j,k}^2}{1+(g_1 / j\pi)^2}}, \tag{8}$$

where $g_1 = 10$.

From Table 2 we learn that for $\phi = 0$, all tests are correctly sized or even undersized in small samples. For $\phi > 0$, the case of particular interest, we find: KPSS_1 is mildly oversized even in larger samples and outperformed by LFST_1. The size-distortion of $\mathrm{VR}_{1,}q$ is growing with q. $\mathrm{VR}_{1,15}$ and $\mathrm{VR}_{1,25}$ do not display notable distortions, being less distorted than KPSS_1, while $\mathrm{VR}_{1,50}$ and $\mathrm{VR}_{1,100}$ are distorted when the sample size is too small or the autocorrelation too strong. Generally, the closer ϕ is to a unit root the smaller has q to be (for fixed T) in order to control size.

For a power analysis we focus on $T = 500$ and $\phi \in \{0, 0.5\}$. For a fair comparison, size-adjusted power is presented in Fig. 3. For $\phi = 0$, $\mathrm{VR}_{1,50}$ is more powerful

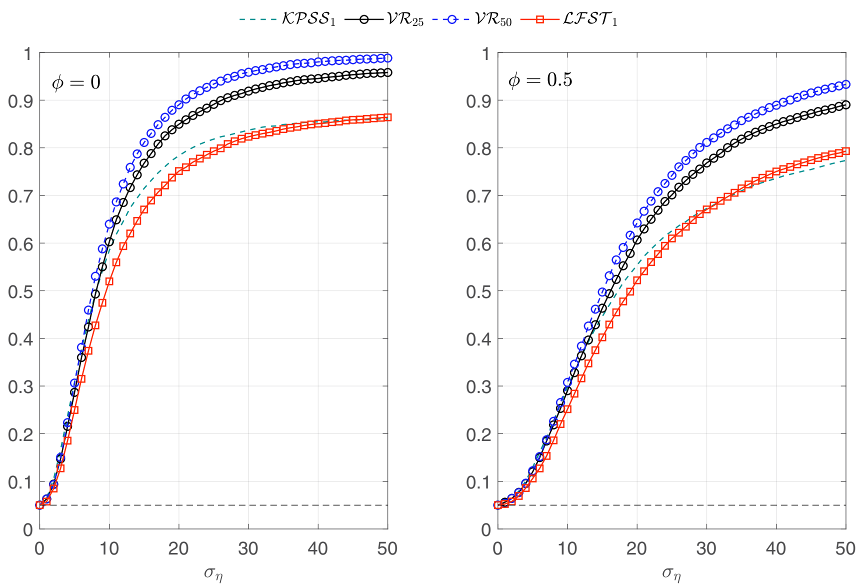


Fig. 3. Size-adjusted Power for $T = 500$ With Demeaning; The Process is From (1) With $e_t = \phi e_{t-1} + \varepsilon_t$.

than $\mathrm{VR}_{1,25}$, and both dominate KPSS_1, which is more powerful than LFST_1. The identical ranking is found for $\phi = 0.5$, only that the power curves are closer.

We conclude that the improved size performance of LFST_1 relative to KPSS_1 comes at a price in terms of power, while $\mathrm{VR}_{1,25}$ has similar size properties to LFST_1 and is more powerful than KPSS_1 at the same time.

5. SUMMARY

We introduced a family of tests for stationarity against local alternatives of a random walk. It builds on variance ratios VRq, where q is the number of weighted sums of CUSUM processes under three considered models: without deterministics, with a constant term, and with a linear time trend. The long-run variance as nuisance parameter cancels asymptotically from the ratios, which makes our procedure self-normalizing. The limit distribution involves quadratic forms of normal variates, and critical values have been computed and are provided in Table 1, and local power functions are available, too.

Asymptotically, q does not affect the size of the test, while the limiting power grows with q. Monte Carlo experiments, however, show that q may not be too large relative to a finite sample size and relative to the persistence in the stationary process under the null in order to obtain a test with correct size. While $q = 100$ or $q = 50$ may result in oversized tests, $q = 25$ balances size distortion and power over a wide range of autocorrelation and sample size. VR_{25} is superior to the classic competitor by Kwiatkowski et al. (1992), not only in terms of power but also in terms of size distortion.

NOTE

1. The estimation of the long-run variance (being proportional to the spectrum of a stationary process at the origin) has been classified as an "ill-posed" problem before. Faust (1996, Prop. 2) argued that any bounded confidence interval for ω_e^2 has necessarily a coverage probability of zero if the spectrum is not sufficiently smooth over the parameter space, since convergence of a pointwise consistent estimator is not uniform; see Pötscher (2002, Thm. 4.3) for a reinforcement.

ACKNOWLEDGMENTS

Earlier versions of this chapter were presented at the 13th International Conference on Computational and Financial Econometrics (London, 2019), the 2nd Italian Workshop of Econometrics and Empirical Economics (Venice, 2020). We are grateful to Ye Lu, Morten Nielsen, and Martin Wagner for helpful comments. We further thank an anonymous referee and Isaac Miller (Editor) for thoughtful and very helpful suggestions.

APPENDIX

First through fourth derivatives of f are denoted as f', f'', f''' and $f^{(4)}$.

Proof of Proposition 1

The integral equation associated with $K_0(s,t;\theta)$ is

$$f(t) = \lambda \int_0^1 f(s) K_0(s,t;\theta)\, \mathrm{d}s,$$

which is equivalent to

$$f^{(4)}(t) = -\lambda f''(t) + \lambda\theta^2 f(t),$$

with $f(0) = 0, f''(1) = -\lambda f(1), f''(0) = -\left(\lambda + \theta^2\right) f(0)$, and $f'''(1) = -\lambda f'(1)$.

The solution of the latter differential equation reads

$$f(t) = c_1 \cos\left(\mu_1 t\right) + c_2 \sin\left(\mu_1 t\right) + c_3 \exp\left(\mu_2 t\right) + c_4 \exp\left(-\mu_2 t\right),$$

where

$$\mu_1 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} + \lambda}{2}} \text{ and } \mu_2 = \sqrt{\frac{\sqrt{\lambda}\sqrt{4\theta^2 + \lambda} - \lambda}{2}}.$$

Using the boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda) = \frac{2\theta^2 + \left(2\theta^2 + \lambda\right)\cos\mu_1 \cosh\mu_2 - \mu_1\mu_2 \sin\mu_1 \sinh\mu_2}{4\theta^2 + \lambda}.$$

We can verify the following conditions:

$$\begin{aligned}
\lim_{\lambda\to 0} D_\theta(\lambda) &= 1,\\
\int K_0(t,t;\theta)\,\mathrm{d}t &= \frac{1}{2} + \frac{1}{12}\theta^2 = -\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda} D_\theta(\lambda),\\
\int\int K_0^2(s,t;\theta)\,\mathrm{d}s\,\mathrm{d}t &= \frac{1}{6} + \frac{11}{180}\theta^2 + \frac{11}{1{,}680}\theta^4\\
&= -\lim_{\lambda\to 0}\frac{\partial^2}{\partial\lambda^2} D_\theta(\lambda) + \left(\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda} D_\theta(\lambda)\right)^2,\\
\lim_{R\to\infty}\frac{\log\log\left|D_\theta\left(Re^{i\varphi}\right)\right|}{\log R} &< 1 \text{ (due to hyperbolic behavior).}
\end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_0(s,t;\theta)$.

Now, using the first three boundary conditions together with $\int_0^1 f(t)^2\,dt = 1$ we obtain

$$f(t) = \alpha_2\left(\alpha_1\cos(\mu_1 t) + \sin(\mu_1 t) + \alpha_3\exp(\mu_2 t) + \alpha_4\exp(-\mu_2 t)\right),$$

where

$$\begin{aligned}
\alpha_1 &= \frac{\left(\mu_2^2 - 2\theta^2\right)\left(1+\cos\mu_1\cosh\mu_2\right)}{\mu_1^2\cosh\mu_2\sin\mu_1 + \mu_1\mu_2\cos\mu_1\sinh\mu_2},\\
\alpha_2 &= \left(\frac{e^{-2\mu_2}}{4\mu_1\mu_2\left(\mu_1^2+\mu_2^2\right)}\right)^{-1/2} \times\\
&\quad \Big[2\alpha_4^2\left(-1+e^{2\mu_2}\right)\mu_1\left(\mu_1^2+\mu_2^2\right) + 8\alpha_4 e^{2\mu_2}\mu_1\mu_2\left(\mu_1+\alpha_3\mu_1^2+\mu_2\left(\alpha_1+\alpha_3\mu_2\right)\right)\\
&\quad +2e^{2\mu_2}\Big(4\alpha_3\mu_1\mu_2\left(\mu_1-\alpha_1\mu_2\right)+\alpha_3^2\left(-1+e^{2\mu_2}\right)\mu_1\left(\mu_1^2+\mu_2^2\right)\\
&\quad +\left(\alpha_1+\mu_1+\alpha_1^2\mu_1\right)\mu_2\left(\mu_1^2+\mu_2^2\right)\Big)\\
&\quad -8\alpha_4 e^{\mu_2}\mu_2\mu_1\left(\left(\mu_1+\alpha_1\mu_2\right)\cos\mu_1+\left(-\alpha_1\mu_1+\mu_2\right)\sin\mu_1\right)\\
&\quad +8\alpha_3 e^{3\mu_2}\mu_2\mu_1\left(\left(-\mu_1+\alpha_1\mu_2\right)\cos\mu_1+\left(\alpha_1\mu_1+\mu_2\right)\sin\mu_1\right)\\
&\quad +e^{2\mu_2}\mu_2\left(\mu_1^2+\mu_2^2\right)\left(-2\alpha_1\cos\left(2\mu_1\right)+\left(-1+\alpha_1^2\right)\sin\left(2\mu_1\right)\right)\Big]^{-1/2},\\
\alpha_3 &= \frac{e^{\mu_2}\mu_1\left(-\lambda+\mu_1^2\right)\mu_2\left(\theta^2+\lambda+\mu_2^2\right)+\mu_1\left(\theta^2-\lambda+\mu_1^2\right)\left(\lambda+\mu_2^2\right)\left(\mu_2\cos\mu_1-\mu_1\sin\mu_1\right)}{\mu_2\left(\lambda+\mu_2^2\right)\left(\theta^2+\lambda+\mu_2^2\right)\left(\left(-1+e^{2\mu_2}\right)\mu_2\cos\mu_2+\left(1+e^{2\mu_2}\right)\mu_1\sin\mu_1\right)},\\
\alpha_4 &= \frac{e^{\mu_2}\mu_1\left(-\left(\lambda-\mu_1^2\right)\mu_2\left(\theta^2+\lambda+\mu_2^2\right)+e^{\mu_2}\left(\theta^2+\lambda-\mu_1^2\right)\left(\lambda+\mu_2^2\right)\left(\mu_2\cos\mu_1+\mu_1\sin\mu_1\right)\right)}{\mu_2\left(\lambda+\mu_2^2\right)\left(\theta^2+\lambda+\mu_2^2\right)\left(\left(-1+e^{2\mu_2}\right)\mu_2\cos\mu_1+\left(1+e^{2\mu_2}\right)\mu_1\sin\mu_1\right)}.
\end{aligned}$$

Hence, the proof is complete.

Proof of Proposition 2

The Fredholm integral equation for $K_1(s,t;\theta)$ is equivalent to

$$f^{(4)}(t) = -\lambda f''(t) + \lambda\theta^2 f(t),$$

with boundary conditions $f(0)=0$, $f(1)=0$, $f''(0)=-\lambda f(0)$, $f''(1)=-\lambda f(1)$. The solution to the latter differential equation reads

$$f(t)=c_1\cos(\mu_1 t)+c_2\sin(\mu_1 t)+c_3\exp(\mu_2 t)+c_4\exp(-\mu_2 t),$$

where μ_1 and μ_2 are as defined under the Proof of Proposition 1. Using the boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda)=\frac{\sin\mu_1\sinh\mu_2}{\mu_1\mu_2}.$$

We can verify the following conditions:

$$\begin{aligned}
\lim_{\lambda\to 0} D_\theta(\lambda) &= 1,\\
\int K_1(t,t;\theta)\,\mathrm{d}t &= \frac{1}{6}+\frac{1}{90}\theta^2=-\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}D_\theta(\lambda),\\
\int\int K_1^2(s,t;\theta)\,\mathrm{d}s\,\mathrm{d}t &= \frac{1}{90}+\frac{2}{945}\theta^2+\frac{1}{9{,}450}\theta^4\\
&= -\lim_{\lambda\to 0}\frac{\partial^2}{\partial\lambda^2}D_\theta(\lambda)+\left(\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}D_\theta(\lambda)\right)^2,\\
\lim_{R\to\infty}\frac{\log\log\left|D_\theta\left(Re^{i\varphi}\right)\right|}{\log R} &< 1 \text{ (due to hyperbolic behavior).}
\end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_1(s,t;\theta)$.

Further, using the first three boundary conditions together with $\int_0^1 f^2(t)\,\mathrm{d}t=1$ we obtain

$$f(t)=\sqrt{2}\sin(\mu_1 t),$$

as required to complete the proof.

Proof of Proposition 3

The Fredholm integral equation for $K_2(s,t;\theta)$ is equivalent to

$$\begin{aligned}
f^{(4)}(t) &= -\lambda f''(t)+\lambda\theta^2 f(t)-6\theta^2\lambda a,\\
a &= \int_0^1 s(1-s)f(s)\,\mathrm{d}s,
\end{aligned}$$

with boundary conditions $f(0)=0$, $f(1)=0$ and

$$\begin{aligned}
f''(0) &= -\lambda f(0)+\lambda\int_0^1 \frac{1}{10}(-1+s)(-5s^2\theta^2+5s^3\theta^2+s(-60+\theta^2))f(s)\,\mathrm{d}s,\\
f'''(0) &= -\lambda f'(0)+\lambda\theta^2\int_0^1(1-3s)(s-1)f(s)\,\mathrm{d}s,\\
f'''(1) &= -\lambda f'(1)+\lambda\theta^2\int_0^1(3s-2)sf(s)\,\mathrm{d}s.
\end{aligned}$$

The solution to the latter differential equation reads

$$f(t)=c_1\cos(\mu_1 t)+c_2\sin(\mu_1 t)+c_3\exp(\mu_2 t)+c_4\exp(-\mu_2 t)+6a.$$

Using the first boundary condition, $f(0)=0$, we eliminate a from the rest of boundary conditions. Using the new boundary conditions we obtain as candidate for the Fredholm determinant

$$D_\theta(\lambda)=-\frac{144(-2+2\cos\mu_1+\mu_1\sin\mu_1)(2-2\cosh\mu_2+\mu_2\sinh\mu_2)}{\mu_1^4\mu_2^4}.$$

We can verify the following conditions:

$$\begin{aligned}
\lim_{\lambda\to 0}D_\theta(\lambda) &= 1,\\
\int K_2(t,t;\theta)\,\mathrm{d}t &= \frac{1}{15}+\frac{11}{12{,}600}\theta^2=-\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}D_\theta(\lambda),\\
\int\int K_2^2(s,t;\theta)\,\mathrm{d}s\,\mathrm{d}t &= \frac{11}{12{,}600}+\frac{1}{27{,}000}\theta^2+\frac{509}{1{,}164{,}240{,}000}\theta^4\\
&= -\lim_{\lambda\to 0}\frac{\partial^2}{\partial\lambda^2}D_\theta(\lambda)+\left(\lim_{\lambda\to 0}\frac{\partial}{\partial\lambda}D_\theta(\lambda)\right)^2,\\
\lim_{R\to\infty}\frac{\log\log\left|D_\theta\left(Re^{i\varphi}\right)\right|}{\log R} &< 1 \text{ (due to hyperbolic behavior).}
\end{aligned}$$

Therefore the conditions of Nabeya (2001, Theorem 5) are satisfied, which implies that the candidate $D_\theta(\lambda)$ is indeed the Fredholm determinant of $K_2(s,t;\theta)$.

Finally using the boundary conditions together with $\int_0^1 f^2(t)\,\mathrm{d}t=1$ we obtain

$$f(t)=\begin{cases}\sqrt{2}\left(\cot\frac{\mu_1}{2}\cos(\mu_1 t)+\sin(\mu_1 t)-\cot\frac{\mu_1}{2}\right) & \text{if } \mu_1\cos\frac{\mu_1}{2}-2\sin\frac{\mu_1}{2}=0,\\ \sqrt{2}\sin(\mu_1 t) & \text{if } \sin\frac{\mu_1}{2}=0.\end{cases}$$

as required to complete the proof.

REFERENCES

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*, 817–858.

Faust, J. (1996). Near observational equivalence and theoretical size problems with unit root tests. *Econometric Theory*, *12*, 724–731.

Hassler, U., & Hosseinkouchack, M. (2019). Ratio tests under limiting normality. *Econometric Reviews*, *38*, 793–813.

Jiang, B., Lu, Y., & Park, J. Y. (2020). Testing for stationarity at high frequency. *Journal of Econometrics*, *215*, 341–374.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*, 159–178.

Liu, W., & Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory*, *26*, 1218–1245.

Lu, Y., & Park, J. Y. (2019). Estimation of longrun variance of continuous time stochastic process using discrete sample. *Journal of Econometrics*, *210*, 236–267.

MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics*, *6*, 422–433.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, *141*, 1331–1352.

Müller, U. K., & Watson, M. W. (2008). Testing models of low-frequency variability. *Econometrica*, *76*, 979–1016.

Nabeya, S. (2001). Approximation to the limiting distribution of t- and F-statistics in testing for seasonal unit roots. *Econometric Theory*, *17*, 711–737.

Nabeya, S., & Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Annals of Statistics*, *16*, 218–235.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, *84*, 223–230.

Nyblom, J., & Mäkeläinen, T. (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model. *Journal of the American Statistical Association*, *78*, 856–864.

Phillips, P. C. B. (1998). New tools for understanding spurious regressions. *Econometrica*, *66*, 1299–1325.

Pötscher, 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*, 1035–1065.

Tanaka, K. (1983). Non-normality of the Lagrange multiplier statistic for testing the constancy of regression coefficients. *Econometrica*, *51*, 1577–1582.

Tanaka, K. (1996). *Time series analysis: Nonstationary and noninvertible distribution theory*. Wiley.

CHAPTER 4

A SEQUENTIAL TEST FOR A UNIT ROOT IN MONITORING A *p*-TH ORDER AUTOREGRESSIVE PROCESS

Kohtaro Hitomi[a], Keiji Nagai[b], Yoshihiko Nishiyama[c] and Junfan Tao[c]

[a]*Kyoto Institute of Technology, Kyoto, Japan*
[b]*Yokohama National University, Yokohama, Japan*
[c]*Kyoto University, Kyoto, Japan*

ABSTRACT

In this study, the authors investigate methods of sequential analysis to test prospectively for the existence of a unit root against stationary or explosive states in a p-th order autoregressive (AR) process monitored over time. Our sequential sampling schemes use stopping times based on the observed Fisher information of a local-to-unity parameter. In contrast to the Dickey–Fuller (DF) test statistic, the sequential test statistic has asymptotic normality. The authors derive the joint limit of the test statistic and the stopping time, which can be characterized using a 3/2-dimensional Bessel process driven by a time-changed Brownian motion. The authors obtain their limiting joint Laplace transform and density function under the null and local alternatives. In addition, simulations are conducted to show that the theoretical results are valid.

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 115–153

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A004

Keywords: Sequential sampling; unit root test; observed Fisher information; DDS Brownian motion; Bessel process; functional central limit theorem in $D[0,\infty)$; local asymptotic normality

MSC 2000 subject classifications: Primary 62M10; 62L10; Secondary 60J70; 60F17

1. INTRODUCTION

Unit roots have been an important topic of econometric research since the late 1970s. Time series processes behave quite differently depending on whether a unit root is present or not. White (1958) considered the asymptotic properties of the ordinary least square estimator (LSE) of an AR (1) unit root process. Dickey (1976) proposed using it to test the null hypothesis of the unit root against stationarity for AR(1) processes; therefore, it is called the DF test. Dickey and Fuller (1979) and Said and Dickey (1984) extended the DF test to AR(p) and AR moving average (ARMA) models. Instead of estimating the ARMA model, Said and Dickey (1984) used the long autoregression model, which does not require the non-linear optimization used for the ARMA model estimation. This feature of Said and Dickey's test makes it one of the most frequently used unit root tests, and it is often called the augmented Dickey–Fuller (ADF) test. Chang and Park (2002) relaxed the conditions on errors and lag length of AR(p) approximation. The Phillips–Perron test (Phillips, 1987a; Phillips & Perron, 1988) is another widely used unit root test that assumes AR(1) processes with general stationary error terms. Chang and Park (2004) and Chang (2012) proposed asymptotically normal unit root tests using a subset of observations for time series and panel data, respectively. Many authors have considered local alternatives to examine the statistical properties and performance of the test under a near-unit-root process, for example, Bobkoski (1983), Cavanagh (1985), Chan (1988), Chan and Wei (1987), and Phillips (1987b). There are a vast number of studies related to unit root tests (see, e.g., Choi, 2015; Stock, 1994).

This study considers unit root tests for AR processes of order p, AR(p), under sequential sampling schemes. We are often interested in the existence of a unit root or bubble after a particular event and seek to figure it out as quickly as possible. Asset bubbles are often described as explosive processes. The following is a motivating example. Suppose that the government announces a policy of providing a lump-sum payment to all households. Financial markets have been normal up to now, but there is a possibility of a bubble starting today because of this policy, then fund managers would like to find out whether or not it occurs as soon as possible using data from today onward. In such a monitoring situation, the method of sequential unit root tests proposed in this chapter is useful. Policy-makers also may want to detect asset bubbles as soon as possible to avoid macroeconomic chaos (see Blanchard, 2016).

Sequential analysis, initiated by Wald (1947), is a prospective study in which statistical inference is made by monitoring incoming observations. The stopping rule is set to collect information from the observed data until precise statistical inferences can be made, and sample size determination and statistical decision-making are carried out simultaneously. Therefore, the sample size is a random variable determined by a stopping rule. Especially, sequential sampling schemes are desirable for the quick detection of anomalies.

Lai and Siegmund (1983) considered the sequential estimation of an AR(1) process. Their approach is "purely sequential." Purely sequential analysis is a prospective study that makes no assumptions about existing observations, monitors arrivals of future observations, and determine a sample size only by a stopping rule using the information obtained from the incoming observations. This study also deals with a purely sequential sampling method.

Lai and Siegmund (1983) proposed to continue sampling the AR(1) process $x_n = \beta x_{n-1} + \epsilon_n$ with $\epsilon_n \sim i.i.d.(0,\sigma^2)$ until the observed Fisher information reaches level c, that is, for any $c > 0$, Lai and Siegmund's stopping time N_c is defined as

$$N_c = \inf\left\{m > 1 : I_m \geq c\right\} \tag{1}$$

where $I_m = \sum_{n=2}^{m} x_{n-1}^2 / \sigma^2$ is the observed Fisher information of β under normal disturbances. Then, at the stopping time N_c, we have a sample $(x_1,\ldots,x_{N_c})$, and I_{N_c}, computed from the sample, is close to c. They showed that if $|\beta| < 1$, $N_c / c \to 1 - \beta^2$ and if $|\beta| = 1$,

$$N_c / \sqrt{c} \Rightarrow \inf\left\{t : \int_0^t W_s^2 ds = 1\right\} \tag{2}$$

where W_s is a standard Brownian motion and $\Rightarrow$ indicates weak convergence. In ordinary inference with a non-random sample size, asymptotic theory is constructed by increasing the sample size. In sequential analysis, since the sample size is the stopping time, the asymptotic theory is derived by increasing c, and, accordingly, increasing the random stopping time N_c. The sequential estimator $\hat{\beta}_{N_c} = \sum_{n=2}^{N_c} x_n x_{n-1} / \sum_{n=2}^{N_c} x_{n-1}^2$ possesses the uniform asymptotic normality under some assumptions:

$$\sqrt{c}(\hat{\beta}_{N_c} - \beta) \xrightarrow{d} N(0,1) \quad \text{uniformly for } |\beta| \leq 1 \text{ as } c \to \infty. \tag{3}$$

The standard error of the sequential estimator $\hat{\beta}_{N_c}$ is $1/\sqrt{c}$, which implies that c represents the accuracy of the sequential estimation. Therefore, the sequential estimation using a stopping rule based on the observed Fisher information is called a "fixed accuracy estimation."

Since σ^2 is unknown in practice, Lai and Siegmund (1983) defined a feasible stopping time in the following way. Replacing σ^2 with the estimator $s_m^2 = \sum_{n=2}^{m}\left(x_n - \hat{\beta}_m x_{n-1}\right)^2 / (m-2)$ in (1), one gets

$$\hat{N}_c = \inf\left\{m > 2 : \hat{I}_m \geq c\right\}. \tag{4}$$

where $\hat{I}_m = \sum_{n=2}^{m} x_{n-1}^2 / s_m^2$ is the feasible observed Fisher information. Here, every time a new observation x_m arrives at time m, the estimators $\hat{\beta}_m$ and s_m^2 are recomputed and the feasible observed Fisher information $\hat{I}_m$ is updated. Each time, $\hat{I}_m$ is compared with c, and $\hat{N}_c$ is the first time that $\hat{I}_m$ exceeds c.

Konev and Pergamenshchicov (1986) and Mukhopadhyay and Sriram (1992) examined the stochastic properties of sequential estimators of stationary vector AR(1) processes. Sriram (2001) proposed a risk-efficient estimator for a stationary threshold AR(1) model (see also Sriram & Iaci, 2014). Galtchouk and Konev (2003, 2004, 2005) consider the sequential estimation of stationary AR parameters. Galtchouk and Konev (2006, 2011) proved the uniform joint normality of the AR(2) and AR(p) parameter estimators, respectively, treating both stationary and non-stationary cases. Nagai et al. (2018) considered sequential unit root tests for AR(1) processes and some of their results are used in this study.

The basic idea of our test is to use the ADF test for an AR(p) process under a sequential sampling scheme. For this purpose, we propose a sequential unit root test using a stopping time based on the observed Fisher information. We stop sampling if we obtain enough observed Fisher information, then perform a unit root test. We derive the joint limit of the test statistics and the stopping time, which can be characterized using a 3/2-dimensional Bessel process driven by a time-changed Brownian motion. To the best of our knowledge, previous research in this field only obtained the marginal limits. We obtain their limiting joint Laplace transform and density function under the null and local alternatives. To obtain these results, we use a diffusion approximation on $D[0,\infty)$ and a time-changed Brownian motion.

The remainder of this chapter is organized as follows. Section 2 describes the models, stopping times, and test statistic. The joint asymptotic distribution of the estimator and the stopping time are included in Section 3. In Section 4 we present the testing procedure and compare the theoretical values of size, power, and average sample size with the values obtained from Monte Carlo simulations. We also discuss simulations using estimated models with different lag lengths than the true model. All the proofs are given in Appendix.

2. SEQUENTIAL TEST FOR NEAR-UNIT-ROOT AR(p) PROCESS

Consider a near-unit-root AR(p) process $\{x_n\}$ under a complete probability space $(\Omega, \mathcal{F}, P)$ with root α_c near unity and initial values $x_1, \ldots, x_p \in L^2$;

$$\alpha_c = 1 + \delta / \sqrt{c}, \qquad \delta \in (-\infty, \infty), \quad c > 0 \tag{5}$$

$$(1 - \alpha_c L)\Psi(L) x_n = \epsilon_n, \qquad n = p+1, p+2, \ldots, \tag{6}$$

where δ is a local parameter, $\epsilon_n \sim i.i.d.(0,\sigma^2)$ are independent of $x_1,\ldots,x_p$, and $\Psi(L)$ is a polynomial in the lag operator L;

$$\Psi(L) = 1 - \psi_1 L - \cdots - \psi_{p-1} L^{p-1} = (1-\alpha_1 L)\cdots(1-\alpha_{p-1}L). \tag{7}$$

We assume that the roots of $\Psi(z)$ lie outside the closed unit disc and

$$|\alpha_i| < |\alpha_c| \le 1 \quad \text{or} \quad |\alpha_i| < 1 \le |\alpha_c| \quad (i = 1,\ldots,p-1). \tag{8}$$

In this study, we investigate the asymptotic properties of the sequential procedures for testing

$$H_0 : \delta = 0 \quad \text{versus} \quad H_1 : \delta > 0 \text{ or } H_1 : \delta < 0. \tag{9}$$

Once ϵ_n are assumed to be $\epsilon_n \sim i.i.d.N(0,\sigma^2)$, the observed Fisher information of α_c is identified as $I_m^{(p)} = \sum_{n=p+1}^{m} \left(\Psi(L)x_{n-1}\right)^2 / \sigma^2$. Using this information, the stopping time is defined as

$$\tau_c = \inf\left\{m > p : I_m^{(p)} \ge c\right\} \tag{10}$$

which is reduced to the stopping time (2) defined by Lai and Siegmund (1983) when $p = 1$. Thus, the stopping time τ_c is the first time at which the observed Fisher information reaches level c. Unlike Chang and Park (2004) and Chang (2012), our approach is a purely sequential prospective study in the sense that we do not set a maximum sample size and do not restrict the possible values of the stopping time. In ordinary inference with a non-random sample size N, asymptotic theory is constructed by increasing N. In sequential analysis, the counterpart to the sample size is τ_c, and thus the asymptotic theory is constructed by increasing c and, accordingly, increasing the random stopping time τ_c. Replacing the parameters with their estimators, a stopping time using the feasible observed Fisher information corresponding to (4) is defined as

$$\hat{\tau}_c = \inf\left\{m > p+1 : \hat{I}_m^{(p)} \ge c\right\} \tag{11}$$

where $\hat{I}_m^{(p)} = \sum_{n=p+1}^{m} \left(\hat{\Psi}_m(L)x_{n-1}\right)^2 / s_m^2$ with the estimators of $\Psi(L)$ and σ^2;

$$\hat{\Psi}_m(L) = 1 - \hat{\psi}_{m,1} L - \cdots - \hat{\psi}_{m,p-1} L^{p-1} \tag{12}$$

using the LSE $\hat{\psi}_{m,1},\ldots,\hat{\psi}_{m,p-1}$ defined in (25) and the estimator s_m^2 defined in (24). Here, when a new observation x_m arrives at time m, the estimators $\hat{\Psi}_m(L)$ and s_m^2 are recomputed and the feasible observed Fisher information $\hat{I}_m^{(p)}$ is updated. Each time, $\hat{I}_m^{(p)}$ is compared with c, and $\hat{\tau}_c$ is defined as the first time that $\hat{I}_m^{(p)}$ exceeds c. Note that τ_c or $\hat{\tau}_c$ is thought to be an "augmented" version of the stopping time introduced by Lai and Siegmund (1983) for an AR(1) model. We will see that τ_c and $\hat{\tau}_c$ determine the accuracy of inference for the detection of a unit root.

Fig. 1 visualizes the procedures of the sequential sampling and stopping rule by flowcharts. (A) presents the case when we know the parameter values $\psi_1, \cdots, \psi_{p-1}, \sigma^2$, whereas (B) includes estimation of the nuisance parameters.

Note that the expression of the local parameter is arbitrary in the sense that one can set $\alpha_c = 1 + \delta / c^{\eta}$ for any $\eta > 0$. It determines the formulation of the functional central limit theorem (FCLT). For example, in the AR(1) model $x_n = \alpha_c x_{n-1} + \epsilon_n$, the functional central limit is written as follows:

$$\frac{1}{c^{\eta/2}\sigma}\sum_{n=2}^{\lfloor c^{\eta} t \rfloor} \epsilon_n \Rightarrow W_t \qquad \text{as} \quad c \to \infty.$$

Then, one should define the stopping time as

$$\tau_c = \inf\left\{m > 1 : I_m \geq c^{2\eta}\right\}$$

where $I_m = \sum_{n=2}^{m} x_{n-1}^2 / \sigma^2$ is the observed Fisher information. As shown in the same way as Lai and Siegmund (1983) or our Theorem 2, the asymptotic property of τ_c under the null hypothesis is reduced to be

$$\tau_c / c^{\eta} \Rightarrow \inf\left\{t : \int_0^t W_s^2 ds = 1\right\}.$$

To align our stopping time formulation with (1) of Lai and Siegmund (1983), we have chosen $\eta = 1/2$.

2.1. A Reparameterization of Regression for Near-Unit-Root AR(p) Process

In this subsection we will obtain a suitable representation for the regression analysis of the near-unit-root AR(p). Defining the difference operator as $\Delta = 1 - L$, we can write (6) as

$$\Psi(L)\left(\Delta x_n - \frac{\delta}{\sqrt{c}} x_{n-1}\right) = \epsilon_n$$

and obtain the difference equation of the AR(p) process x_n in (6);

$$\Delta x_n - \sum_{i=1}^{p-1} \psi_i \Delta x_{n-i} - \frac{\delta}{\sqrt{c}}\left(x_{n-1} - \sum_{i=1}^{p-1} \psi_i x_{n-i-1}\right) = \epsilon_n. \tag{13}$$

Using a telescoping relation $x_{n-j-1} = x_{n-1} - \sum_{i=1}^{j} \Delta x_{n-i}$, we have

$$\sum_{j=1}^{p-1} \psi_j x_{n-j-1} = \sum_{j=1}^{p-1} \psi_j x_{n-1} - \sum_{i=1}^{p-1}\sum_{j=i}^{p-1} \psi_j \Delta x_{n-i}.$$

Then (13) can be represented as follows:

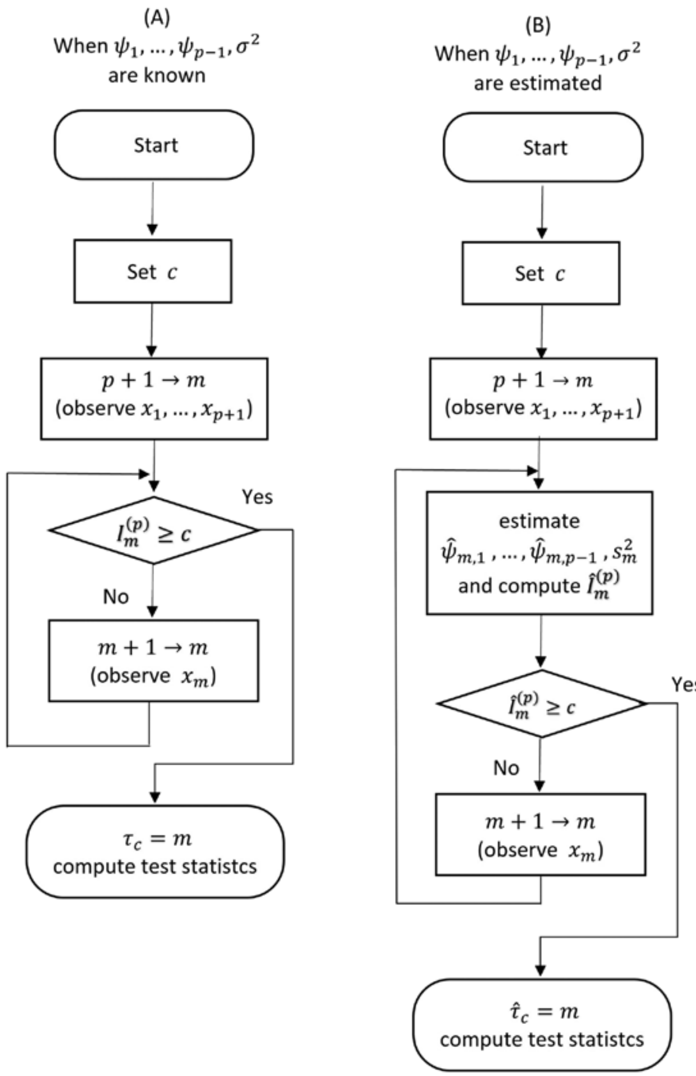


Fig. 1. The Stopping Rule of the Sequential Unit Root Tests.

$$\Delta x_n - \frac{\delta}{\sqrt{c}}\Psi(1)x_{n-1} - \sum_{i=1}^{p-1}\left(\psi_i + \frac{\delta}{\sqrt{c}}\sum_{j=i}^{p-1}\psi_j\right)\Delta x_{n-i} = \epsilon_n, \tag{14}$$

where $\Psi(1) = 1 - \psi_1 - \cdots - \psi_{p-1}$.

Now we can reparameterize a regression model for a near-unit-root AR(p) process. We use $'$ for the matrix transpose. Put $\boldsymbol{\phi}_2^c = \left(\phi_2^c, \phi_3^c, \cdots, \phi_p^c\right)'$, $\boldsymbol{\psi} = \left(\psi_1, \psi_2, \cdots, \psi_{p-1}\right)'$, and $(p-1)\times(p-1)$ matrix

$$A_c = \begin{pmatrix} \alpha_c & \alpha_c - 1 & \cdots & \alpha_c - 1 & \alpha_c - 1 \\ 0 & \alpha_c & \cdots & \alpha_c - 1 & \alpha_c - 1 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha_c & \alpha_c - 1 \\ 0 & 0 & \cdots & 0 & \alpha_c \end{pmatrix}. \tag{15}$$

Let the lag polynomial of (6) be $\Phi^c(\lambda) = \left(1 - \alpha_c \lambda\right)\Psi(\lambda)$ and write

$$\Phi^c(\lambda) = (1-\lambda) - \phi_1^c \lambda - \phi_2^c (1-\lambda)\lambda - \cdots - \phi_p^c (1-\lambda)\lambda^{p-1}, \tag{16}$$

where

$$\phi_1^c = \frac{\delta}{\sqrt{c}}\Psi(1), \quad \boldsymbol{\phi}_2^c = A_c \boldsymbol{\psi}. \tag{17}$$

The model (6) becomes

$$\Phi^c(L)x_n = \Delta x_n - \phi_1^c x_{n-1} - \phi_2^c \Delta x_{n-1} - \cdots - \phi_p^c \Delta x_{n-p+1} = \epsilon_n. \tag{18}$$

2.2. Normal Equation and Sequential Estimator

For the sample $x_1, \ldots, x_m$, the vector representation of the AR(p) in (18) is

$$\Delta \boldsymbol{x_m} - \boldsymbol{X_m} \boldsymbol{\phi}^c = \boldsymbol{\epsilon_m}, \tag{19}$$

where $\boldsymbol{\Delta x_m} = \left(\Delta x_{p+1}, \Delta x_{p+2}, \cdots, \Delta x_m\right)'$, $\boldsymbol{\phi}^c = \left(\phi_1^c, \phi_2^c, \cdots, \phi_p^c\right)'$, $\boldsymbol{\epsilon_m} = \left(\epsilon_{p+1}, \epsilon_{p+2}, \cdots, \epsilon_m\right)'$, and the $(m-p)\times p$ matrix

$$\boldsymbol{X_m} = \left(x_{n-1}, \Delta x_{n-1}, \cdots, \Delta x_{n-p+1}\right)_{n=p+1,\ldots,m}.$$

The normal equation for the LSE $\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^c = \left(\hat{\phi}_{1,m}^c, \hat{\phi}_{2,m}^c, \cdots, \hat{\phi}_{p,m}^c\right)'$ is

$$\boldsymbol{X}_{\boldsymbol{m}}' \boldsymbol{X_m} \hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^c = \boldsymbol{X}_{\boldsymbol{m}}' \boldsymbol{\Delta x_m}, \tag{20}$$

which has the form

$$\begin{pmatrix} \sum_{n=p+1}^{m} x_{n-1}^2 & \left(\sum_{n=p+1}^{m} x_{n-1}\Delta x_{n-j}\right)_j \\ \left(\sum_{n=p+1}^{m} x_{n-1}\Delta x_{n-i}\right)_i' & \left(\sum_{n=p+1}^{m} \Delta x_{n-i}\Delta x_{n-j}\right)_{i,j} \end{pmatrix} \hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^c = \begin{pmatrix} \sum_{n=p+1}^{m} x_{n-1}\Delta x_n \\ \left(\sum_{n=p+1}^{m} \Delta x_{n-i}\Delta x_n\right)_i' \end{pmatrix} \tag{21}$$

where *i, j* run over $1,2,\ldots,p-1$. For the estimator of the lag polynomial $\Phi^c(\lambda)$ in (16), let

$$\hat{\Phi}^c_m(\lambda)=(1-\lambda)-\hat{\phi}^c_{1,m}\lambda-\hat{\phi}^c_{2,m}(1-\lambda)\lambda-\cdots-\hat{\phi}^c_{p,m}(1-\lambda)\lambda^{p-1} \tag{22}$$

and $\hat{\alpha}_{c,m}$ be the LSE of AR(1) model $x_n=\alpha_c x_{n-1}+v_n$;

$$\hat{\alpha}_{c,m}=\sum_{n=p+1}^{m}x_{n-1}x_n \Big/ \sum_{n=p+1}^{m}x_{n-1}^2 \tag{23}$$

where v_n is defined in (27).

As to the stopping time $\hat{\tau}_c$ in (11), we obtain the sequential estimator $\hat{\boldsymbol{\phi}}^c_{\hat{\tau}_c}$ and the sequential unit root test statistics $\hat{\delta}_{\hat{\tau}_c}$ in the following manner. Put

$$s_m^2=\left(\boldsymbol{\Delta x_m}-\boldsymbol{X_m}\hat{\boldsymbol{\phi}}^c_{\boldsymbol{m}}\right)'\left(\boldsymbol{\Delta x_m}-\boldsymbol{X_m}\hat{\boldsymbol{\phi}}^c_{\boldsymbol{m}}\right)/(m-p) \tag{24}$$

for the consistent estimator of σ^2. In view of (17), let

$$\hat{\boldsymbol{\psi}}_{\boldsymbol{m}}=\hat{A}^{-1}_{c,m}\hat{\boldsymbol{\phi}}^c_{\boldsymbol{2,m}}, \tag{25}$$

where $\hat{\boldsymbol{\phi}}^c_{\boldsymbol{2,m}}=\left(\hat{\phi}^c_{2,m},\hat{\phi}^c_{3,m},\cdots,\hat{\phi}^c_{p,m}\right)'$ and $\hat{A}_{c,m}$ is designated by replacing α_c with $\hat{\alpha}_{c,m}$ in A_c defined in (15).

Letting $\hat{\Psi}_m(1)=1-\hat{\psi}_{1,m}-\cdots-\hat{\psi}_{p-1,m}$, we can set $\hat{\tau}_c$ in (11) and obtain the sequential estimators, $\hat{\Psi}_{\hat{\tau}_c}(1)$ and $\hat{\boldsymbol{\phi}}^c_{\hat{\tau}_c}$, and the sequential unit root test statistic,

$$\hat{\delta}_{\hat{\tau}_c}=\frac{\sqrt{c}\,\hat{\phi}^c_{1,\hat{\tau}_c}}{\hat{\Psi}_{\hat{\tau}_c}(1)}. \tag{26}$$

3. THE ASYMPTOTIC PROPERTIES OF $\hat{\tau}_c$ AND THE SEQUENTIAL PROCEDURES FOR A NEAR-UNIT-ROOT AR(*p*)

In this section, we consider the asymptotic properties of the stopping times $\hat{\tau}_c$ in (11), the sequential unit root test statistics $\hat{\delta}_{\hat{\tau}_c}$, and the sequential estimator $\hat{\boldsymbol{\phi}}^c_{\hat{\tau}_c}$ defined in (20) when *c* goes to ∞. To deal with asymptotic results relating to the stopping times, we use the weak convergence theorem in $D[0,\infty)$ and Skorohod's representation theorem (Billingsley, 1999; Theorem 6.7). For brief descriptions, see Nagai et al. (2018).

Let

$$z_n=\Psi(L)x_n,\quad \text{and}\quad v_n=(1-\alpha_c L)x_n. \tag{27}$$

Then z_n is a near-unit-root AR(1) process with initial value $z_p = x_p - \psi_1 x_{p-1} - \cdots - \psi_{p-1} x_1$ and v_n is an asymptotically stationary AR(p − 1) process with initial values $v_i = x_i - \alpha_c x_{i-1}$ $(i = 2, \ldots p)$. Using them, write

$$\Delta z_n = \frac{\delta}{\sqrt{c}} z_{n-1} + \epsilon_n, \quad \text{and} \quad \Psi(L) v_n = \epsilon_n, \qquad n = p+1, p+2, \ldots. \tag{28}$$

3.1. Convergence to an Ornstein–Uhlenbeck (OU) Process on D[0,∞)

For z_n defined in (27), let

$$Z_c(t) = z_{\lfloor \sqrt{c} t \rfloor} / c^{1/4} \sigma. \tag{29}$$

Let $S_n = \epsilon_{p+1} + \epsilon_{p+2} + \cdots + \epsilon_n$ with $S_1 = \cdots = S_{p-1} = 0$, W be a Brownian motion, and

$$W_c(t) = S_{\lfloor \sqrt{c} t \rfloor} / c^{1/4} \sigma. \tag{30}$$

Then, by the FCLT (Billingsley, 1999; Theorem 18.2), we have as $c \uparrow \infty$

$$W_c \Rightarrow W \quad \text{in} \quad D[0, \infty) \tag{31}$$

where ⇒ stands for weak convergence and $D[0,\infty)$ is the space of right continuous functions with left limits on $[0,\infty)$. We then obtain

$$Z_c \Rightarrow Z \quad \text{in} \quad D[0, \infty). \tag{32}$$

Here Z is the OU process with the Brownian motion W in (31):

$$dZ_t = \delta Z_t dt + dW_t. \tag{33}$$

The test hypothesis $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$ under the OU process Z becomes

$$H_0 : dZ_t = dW_t \quad \text{versus} \quad H_1 : dZ_t = \delta Z_t dt + dW_t.$$

3.2. Asymptotic Properties of the Components of the Normal Equation

Next, we consider a functional limit theorem in $D[0,\infty)$ with respect to the normal equations (20) and (21) with proper normalization. Let $\boldsymbol{D}_c$ be the $p \times p$ matrix $\operatorname{diag}\left(\sqrt{c}, c^{1/4}, \ldots, c^{1/4}\right)$. Multiplying the normal equation by $\boldsymbol{D}_c^{-1}$, we have

$$\boldsymbol{D}_c^{-1} X'_{\lfloor\sqrt{c}t\rfloor} X_{\lfloor\sqrt{c}t\rfloor} \boldsymbol{D}_c^{-1} \boldsymbol{D}_c \left(\hat{\boldsymbol{\phi}}^c_{\lfloor\sqrt{c}\boldsymbol{t}\rfloor} - \boldsymbol{\phi}^c\right) = \boldsymbol{D}_c^{-1} X'_{\lfloor\sqrt{c}t\rfloor} \boldsymbol{\epsilon}_{\lfloor\sqrt{c}\boldsymbol{t}\rfloor} \tag{34}$$

which has the form

$$\begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 / c & \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-j} / c^{3/4}\right)_j \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i} / c^{3/4}\right)'_i & \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\Delta x_{n-j} / \sqrt{c}\right)_{i,j} \end{pmatrix} \begin{pmatrix} \sqrt{c}(\hat{\phi}^c_{1,\lfloor\sqrt{c}t\rfloor} - \phi^c_1) \\ c^{1/4}\left(\hat{\boldsymbol{\phi}}^c_{2,\lfloor\sqrt{c}\boldsymbol{t}\rfloor} - \hat{\boldsymbol{\phi}}^c_2\right) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n / \sqrt{c} \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n / c^{1/4}\right)'_i \end{pmatrix} \quad (i, j = 1, \ldots, p-1). \tag{35}$$

The next proposition shows the main convergence results in (35).

Proposition 1. *The asymptotic autocovariance function* $\gamma_v(h)$ *of* v_n *exists for* $h = 0, 1, \ldots;$

$$\lim_{m\to\infty} \sum_{n=p+1}^{m} v_n v_{n+h} / (m-p) = \gamma_v(h) \ a.s., \tag{36}$$

and the FCLT with respect to martingale differences $v_{n-i}\epsilon_n$ *for* $i = 1, \ldots, p-1$ *holds; in the sense of* $D[0,\infty)$,

$$\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-i}\epsilon_n / \left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i \Rightarrow \boldsymbol{W}_t^{(1)} \quad \text{as} \quad c \to \infty, \tag{37}$$

where $\boldsymbol{W}^{(1)}$ *is a* $(p-1)$*-dimensional Brownian motion with the correlation coefficient matrix* $\boldsymbol{\rho} = \left(\gamma_v(|i-j|) / \gamma_v(0)\right)_{i,j=1,\ldots,p-1}$ *independent of W in (31).*

In the sense of $D[0,\infty)$, *as* $c \to \infty$,

$$\begin{pmatrix} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 / \left(c\sigma^2\right) \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\Delta x_{n-i} / \left(c^{3/4}\sigma^2\right)\right)_i' \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\Delta x_{n-j} / \left(\sqrt{c}\sigma^2\right)\right)_{i,j} \\ \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n / \left(\sqrt{c}\sigma^2\right) \\ \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n / \left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i' \end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^t Z_u^2 du / \Psi(1)^2 \\ \mathbf{0} \\ \left(\gamma_v\left(|i-j|\right)t\right)_{i,j} \\ \int_0^t Z_u dW_u / \Psi(1) \\ \boldsymbol{W}_t^{(1)} \end{pmatrix}, \tag{38}$$

where Z is the OU process in (33).

3.3. Main Theorems

Since the stopping time τ_c in (10) has the deformation $\tau_c = \inf\left\{\sqrt{c}t > p : I_{\lfloor\sqrt{c}t\rfloor}^{(p)} \geq c\right\}$ following asymptotic property; as $c \to \infty$,

$$\tau_c / \sqrt{c} = \inf\left\{t > p/\sqrt{c} : \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}^2 / c\sigma^2 \geq 1\right\} \Rightarrow \inf\left\{t \geq 0 : \int_0^t Z_u^2 du = 1\right\},$$

which is a similar result as Lai and Siegmund (1983) obtained for the stopping time N_c in (1). Therefore, we modify the expression $\hat{\tau}_c$ as follows for large c:

$$\hat{\tau}_c / \sqrt{c} = \inf\left\{t > 0 : \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\hat{\Psi}_{\lfloor\sqrt{c}t\rfloor}(L)x_{n-1}\right)^2 / cs_{\lfloor\sqrt{c}t\rfloor}^2 \geq 1\right\} \tag{39}$$

where $\hat{\Psi}_m(L) = 1 - \hat{\psi}_{1,m}L - \cdots - \hat{\psi}_{p-1,m}L^{p-1}$ and $\hat{\psi}_{i,m}$ are defined in (25). Letting $t = \hat{\tau}_c / \sqrt{c}$ in Proposition 1, the main theorem is obtained. The proofs of Proposition 1 and Theorem 2 are given in the Appendices 6.1 and 6.2.

Theorem 2. *Consider the stopping time $\hat{\tau}_c$ in (11), the sequential estimator $\hat{\phi}_{\hat{\tau}_c}^c$ in (20) and $\hat{\psi}_{\hat{\tau}_c}$ in (25) by letting $m = \hat{\tau}_c$. Let Z be the OU process as defined in (33) with $Z_0 = 0$. Define the martingale M_t and its quadratic variation $\langle M\rangle_t$ as*

$$M_t = \int_0^t Z_u dW_u \quad \text{and} \quad \langle M\rangle_t = \int_0^t Z_u^2 du. \tag{40}$$

Then, we define $U_s = \langle M \rangle_s^{-1}$,

$$U_s = \inf\left\{t \ge 0 : \int_0^t Z_u^2 du = s\right\}, \tag{41}$$

and the time-changed (Dambis–Dubins–Schwarz) Brownian motion

$$B_s = M_{U_s}. \tag{42}$$

Let $\rho_s = Z_{U_s}^2 / 2$, *then* ρ_t *becomes a 3/2-dimensional Bessel process with drift* δ *and initial value* $\rho_0 = 0$

$$\rho_t = \rho_0 + B_s + \int_0^t \left(\frac{1}{4\rho_s} + \delta\right) ds \quad \text{and} \quad U_t = \int_0^t \frac{1}{2\rho_s} ds. \tag{43}$$

The asymptotic behavior of the stopping time $\hat{\tau}_c$ *and the sequential unit root test statistics* $\hat{\delta}_{\hat{\tau}_c}$ *are as follows. As* $c \to \infty$, $s_{\hat{\tau}_c}^2 \to_p \sigma^2$,

$$\left(\hat{\delta}_{\hat{\tau}_c}, \hat{\tau}_c / \sqrt{c}\right) \Rightarrow \left(\delta + \int_0^{U_1} Z_u dW_u, U_1\right) = \left(\delta + B_1, \int_0^1 \frac{1}{2\rho_s} ds\right), \tag{44}$$

$$\left(\sqrt{\hat{\tau}_c}\left(\hat{\phi}_{2,\hat{\tau}_c}^c - \phi_2^c\right) \right) = \left(\sqrt{\hat{\tau}_c}\left(\hat{\psi}_{\hat{\tau}_c} - \psi\right) \right) + o_p(1) \Rightarrow \left(\frac{1}{\sigma\sqrt{\gamma_v(0)}} \boldsymbol{\rho}^{-1} \boldsymbol{W}_1^{(1)} \right) \tag{45}$$

where $\boldsymbol{W}^{(1)}$ *is a* $(p-1)$*-dimensional Brownian motion independent of W, with the correlation coefficient matrix* $\boldsymbol{\rho} = \left(\gamma_v(|i-j|) / \gamma_v(0)\right)_{i,j=1,\dots,p-1}$.

Note that, since $\boldsymbol{W}^{(1)}$ is independent of W, the sequential unit root test statistic $\hat{\delta}_{\hat{\tau}_c}$ and the coefficient estimator $\hat{\psi}_{\hat{\tau}_c}$ of the stationary AR process v_n written in (28) are independent.

The following two theorems give the joint density of the sequential estimators and the stopping time. The inverse Laplace transform $\mathcal{L}_\gamma^{-1}\{\cdot\}(y)$ is defined as $\mathcal{L}_\lambda^{-1}\left\{\int_0^\infty \exp(-\gamma y) F(y) dy\right\}(y) = F(y)$. See the Appendix 6.3 for the proof.

Theorem 3. *Let* B_v *be a standard Brownian motion,* ρ_v *be a Bessel process defined as (43) with* $\rho_0 = \delta = 0$, *and* $U_v = \int_0^v \frac{1}{2\rho_s} ds$. *Then, the joint density of* (B_v, U_v) *under the null hypothesis has the following form for* $u > 0$, $z > -u/2$:

$$P_{H_0}\left(B_v \in dz, U_v \in du\right) = 2\sqrt{\frac{2z+u}{2\pi}}\, \mathrm{es}_v\left(\frac{1}{2}, \frac{1}{2}, u, 0, \frac{2z+u}{2}\right) dz du, \tag{46}$$

where es_y *is defined by*

$$\begin{aligned}\mathrm{es}_y(\mu,\nu,t,x,z) &= \mathscr{L}_\gamma^{-1}\left(\left(\frac{(2\gamma)^{\mu/2}}{\sinh^\nu\left(t\sqrt{2\gamma}\right)}\right)\exp\left(-x\sqrt{2\gamma}-z\sqrt{2\gamma}\coth\left(t\sqrt{2\gamma}\right)\right)\right)(y)\\ &= \sum_{k=0}^{\infty}\frac{(-z)^k}{k!}\,\mathrm{s}_y(\mu+k,\nu+k,t,x+z+kt)\end{aligned} \tag{47}$$

and

$$\mathrm{s}_y(\mu,\nu,t,z)=2^\nu\sum_{l=0}^{\infty}\frac{\Gamma(\nu+l)\exp\left(-(\nu t+z+2lt)^2/4y\right)}{\sqrt{2\pi}\,y^{1+\mu/2}\,\Gamma(\nu)\,l!}D_{\mu+1}\left(\frac{\nu t+z+2lt}{\sqrt{y}}\right). \tag{48}$$

Here, $D_{\mu+1}$ *are the parabolic cylinder functions.*

See Borodin and Salminen (2002) on the definitions of es_y, s_y, and $D_{\mu+1}$ for details. Applying Girsanov's theorem to $\delta v+B_v$ in Theorem 3, we derive the explicit form of the joint density as follows. The relation in (49) indicates the local asymptotic normality property. See Le Cam (1986).

Theorem 4. *Let* B_v *be a standard Brownian motion,* ρ_v *be a Bessel process with drift* δ *and* $\rho_0=0$ *defined as (43), and* $U_v=\int_0^v\frac{1}{2\rho_s}ds$ *. The joint density of* $(\delta v+B_v,U_v)$ *under the alternative hypothesis in Theorem 2 becomes*

$$P_{H_1}(\delta v+B_v\in dz,U_v\in du)=\exp\left(\delta z-\delta^2 v/2\right)P_{H_0}(B_v\in dz,U_v\in du) \tag{49}$$

where $P_{H_0}(B_v\in dz,U_v\in du)$ *is (46) in Theorem 3.*

From the Bessel bridges in Pitman and Yor (1982), we can compute the joint moments of the stopping time and test statistic through the following theorem under the null hypothesis.

Theorem 5. *Let* W_t *be a standard Brownian motion and* ρ_t *be a Bessel process defined as (43) with* $\rho_0=\delta=0$. *Using time-change via* $t=U_v=\int_0^v\frac{1}{2\rho_s}ds$, *the joint Laplace transform of* $\left(W_t^2,\int_0^t W_s^2ds\right)$ *can be written as a modified Laplace transform of* (ρ_v,U_v)*;*

$$\begin{aligned}&\int_0^\infty e^{-\beta t}E_{H_0}\left[\exp\left(-\alpha W_t^2-\gamma\int_0^t W_s^2ds\right)\right]dt\\ &=\int_0^\infty e^{-\gamma v}E_{H_0}\left[\exp\left(-\alpha 2\rho_v-\beta U_v\right)/2\rho_v\right]dv.\end{aligned} \tag{50}$$

This relation implies

$$E_{H_0}\left[\exp(-\alpha 2\rho_v - \beta U_v)/2\rho_v\right]$$
$$= \mathscr{L}_\gamma^{-1}\left[\frac{2^{7/4}\gamma^{1/4}\,{}_2F_1\left(\frac{1}{2},\frac{1}{4}\left(\frac{\beta\sqrt{2}}{\sqrt{\gamma}}+1\right);\frac{1}{4}\left(\frac{\beta\sqrt{2}}{\sqrt{\gamma}}+5\right);\frac{4\alpha}{2\alpha+\sqrt{2}\sqrt{\gamma}}-1\right)}{\sqrt{2\alpha+\sqrt{2}\sqrt{\gamma}}\left(2\beta+\sqrt{2}\sqrt{\gamma}\right)}\right](v) \tag{51}$$

$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\alpha^k\beta^l}{k!\,l!}\int_0^1 J(t,v,k,l)dt, \tag{52}$$

where ${}_2F_1(a,b;c;z)=\sum_{k=0}^{\infty}(a)_k(b)_k/(c)_k\, z^k/k!$ *and*

$$J(s,v,k,l)=\frac{2^{\frac{1}{2}(k-3l-2)}\left(-\frac{1}{2}\right)^{(k)}v^{\frac{1}{2}(k+l-1)}}{\Gamma\left(\frac{1}{2}(k+l+1)\right)}s^{-\frac{3}{4}}(1-s)^k(1+s)^{-k-\frac{1}{2}}\log^l(s) \tag{53}$$

with $x^{(m)}$ *being the factorial power* $x(x-1)\ldots(x-(m-1))$.

The joint Laplace transform under the alternative hypothesis is given as follows.

Theorem 6. *Let ρ_t be a Bessel process defined as (43) with $\rho_0 = 0$ and $\delta \neq 0$. The joint Laplace transform of (ρ_1, U_1) under the alternative H_1 becomes*

$$\begin{aligned} & E_{H_1}\left[\exp(-\alpha 2\rho_1 - \beta U_1)/2\rho_1\right] \\ = \; & E_{H_0}\left[\exp\left(-\left(\alpha-\frac{\delta}{2}\right)2\rho_1-\left(\beta+\frac{\delta}{2}\right)U_1-\frac{\delta^2}{2}\right)/2\rho_1\right]dv \\ = \; & \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{\alpha^k\beta^l}{k!\,l!}m(\delta,v,k,l) \end{aligned} \tag{54}$$

where

$$m(\delta,v,k,l)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(-1)^i\frac{(\delta/2)^{i+j}}{i!\,j!}\int_0^1 J(t,v,i+k,j+l)dt \tag{55}$$

with $J(t,v,k,l)$ *defined in (53).*

See Appendix 6.4 for the proofs of the above two theorems.

4. TESTING PROCEDURE, THEORETICAL VALUES, AND SIMULATION RESULTS

4.1. Testing Procedure

The procedures for only the one-sided test against the alternatives of $\delta < 0(\alpha_c < 1)$ or $\delta > 0(\alpha_c > 1)$ are conducted. Let α be a significance level, and select c. The sequential unit root test employs the asymptotic normality of $\hat{\delta}_{\hat{\tau}_c}$ as shown in Theorem 2 as $c \to \infty$. Since the null hypothesis is $\delta = 0$ and its asymptotic variance equals 1, the sequential unit root test simply looks at $\hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c} / \hat{\Psi}_{\hat{\tau}_c}(1)$. Let z_α be the α quantile of the standard normal distribution. With the stationary alternative $\delta < 0$, we reject the null hypothesis when

$$\hat{\delta}_{\hat{\tau}_c} < z_\alpha$$

for the left tailed test. With the explosive alternative $\delta > 0$, we reject the null hypothesis when

$$\hat{\delta}_{\hat{\tau}_c} > z_{1-\alpha}$$

for the right tailed test.

In practice, $m > p+1$ in the definition of $\hat{\tau}_c$ in (11) does not work well because the sampling may stop too early for m close to $p + 1$, which will lead to unstable parameter estimation. Replacing it by $m > m_0$ for some suitably large $m_0(> p+1)$, we set the stopping time as

$$\hat{\tau}_c = \inf\left\{m > m_0 : \hat{I}^{(p)}_m \geq c\right\}. \tag{56}$$

We should choose m_0 large enough to stabilize the forecast $\hat{I}^{(p)}_m$ of the observed Fisher information. Thereby, we see that the simulation results are consistent with the theory of this study, even when c is relatively small.

4.2. Theoretical Values for Moments of $\hat{\tau}_c$

Based on the asymptotic property of $\hat{\tau}_c$ provided in Theorem 2, Tables 1 and 2 show the theoretical values of the expectation and standard deviation of $\hat{\tau}_c$ under the null and alternative hypotheses.

The expected value and standard deviation of U_1 under the null hypothesis are calculated using the Laplace transform given in Theorem 5 as follows:

$$E(U_1) = 2.09210 \qquad sd(U_1) = 1.08758.$$

Novikov (1972) also obtained $E(U_1)$ and $E(U_1^2)$. In Table 1, for each value of $\sqrt{c}$, the theoretical values of $E(\hat{\tau}_c)$ and $sd(\hat{\tau}_c)$ are computed from the asymptotic relations $E(\hat{\tau}_c) \approx \sqrt{c} \times E(U_1)$ and $sd(\hat{\tau}_c) \approx \sqrt{c} sd(U_1)$. We immediately see that the average sample size increases as c increases.

Table 2 provides the theoretical values of the expectation and the standard deviation of U_1 under alternative hypotheses. For each value of the local

Table 1. Theoretical Expected Values and Standard Deviations of $\hat{\tau}_c$ Under Null.

$\sqrt{c}$	$E(\hat{\tau}_c) \approx \sqrt{c}E(U_1)$	$sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1)$
50	104.6	54.4
100	209.2	108.8
150	313.8	163.1
200	418.4	217.5

parameter δ in the first column, the second and third columns provide the expected values and standard deviations of U_1 calculated from the Laplace transform given in Theorem 6. In the simulations of the two succeeding sections, we choose the numbers 0.99 and 1.01 for α_c, corresponding to the stationary and explosive alternatives, respectively. In those cases, c is determined from the relation $\alpha = 1 + \delta / \sqrt{c}$. For example, the eighth row gives the case, when $\delta = 1.5$ and $\sqrt{c} = 150$, and thus $\alpha_c = 1 + \delta / \sqrt{c} = 1.01$. Then, the expected value and standard error are $E(\hat{\tau}_c) \approx \sqrt{c}E(U_1) = 150 \times 1.19234 = 178.9$, and $sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1) = 150 \times 0.58689 = 88.0$ respectively.

4.3. The Simulation Settings and Results When Lag Length is Known

We conducted simulations to examine the performance of the sequential unit root test with the correct lag length. We investigate the sizes and powers of the test and the expected sample sizes and standard deviations of the stopping times. As explained at (3) in the introduction, $1/c$ represents the accuracy of the sequential estimation. As c increases, the stopping time becomes larger, and better estimates of the parameters will be obtained.

The data generating process (DGP) is based on the following AR(3) process

$$(1 - \alpha_c L)\Psi(L)x_n = \epsilon_n \quad n = 4, 5, \ldots \tag{57}$$

where $\Psi(L) = 1 - \psi_1 L - \psi_2 L^2 = (1 - \alpha_1 L)(1 - \alpha_2 L)$ and $\epsilon_n \sim i.i.d.N(0,1)$. The initial values $x_1, x_2, x_3 \sim i.i.d.N(0,1)$ are independent of $\epsilon_4, \epsilon_5, \ldots$.

There are 10,000 iterations in the simulation. We set $\alpha_1 = 0.5$, $\alpha_2 = 0.3$, then $\psi_1 = \alpha_1 + \alpha_2 = 0.8, \psi_2 = -\alpha_1\alpha_2 = -0.15$. For $\delta < 0(\alpha_c < 1)$, we consider the

Table 2. Theoretical Moments of $\hat{\tau}_c$ Under Alternatives.

Δ	$E(U_1)$	$sd(U_1)$	$\sqrt{c}$	α_c	$E(\hat{\tau}_c) \approx \sqrt{c}E(U_1)$	$sd(\hat{\tau}_c) \approx \sqrt{c}sd(U_1)$
-0.5	2.59506	1.28939	50	0.99	129.8	64.5
-1	3.21232	1.47454	100		321.2	147.5
-1.5	3.93244	1.62767	150		589.9	244.2
-2	4.73487	1.74208	200		947.0	348.4
0.5	1.70132	0.89481	50	1.01	85.1	44.7
1	1.40833	0.72450	100		140.8	72.4
1.5	1.19234	0.58689	150		178.9	88.0
2	1.03252	0.48170	200		206.5	96.3

stationary alternative of $\alpha_c = 0.99$. We use the settings $\delta = -0.5, -1, -1.5, -2$ corresponding to $\sqrt{c} = 50, 100, 150, 200$, respectively, which are the same as the upper half of Table 2. For the explosive alternative $\delta > 0(\alpha_c > 1)$, we set $\alpha_c = 1.01$ and have $\delta = 0.5, 1, 1.5, 2$ for $\sqrt{c} = 50, 100, 150, 200$ from the relationship $\delta = \sqrt{c}(1.01 - 1)$, respectively. We set $m_0 = 30$ for the definition of $\hat{\tau}_c$ in (56). We also perform simulations with $\sqrt{c} = 800$ to see at which level of $\sqrt{c}$ the size distortion disappears.

Table 3 shows the simulation results along with the theoretical values in parentheses for: (i) the value of $\sqrt{c}$; (ii) the sizes of the left and right tailed tests; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$. Tables 4 and 5 correspond to the alternative hypotheses $\alpha = 0.99$ and 1.01, respectively, and show the results for: (i) the values of $\sqrt{c}$ and δ; (ii) the powers of the sequential unit root test; and (iii) and (iv), which are equivalent to those in Table 3. The theoretical values of the mean and standard deviation of $\hat{\tau}_c$ are consulted from the Tables 1 and 2.

Let us examine the performance of the sequential unit root test by looking at the columns for Size in Table 3 and the columns for Power in Tables 4 and 5. For the left tailed test, both the simulated sizes in Table 3 and powers in Table 4 perform well for all c in the sense that the theoretical values are close to the simulation results. In the case of the right tailed test, when c is relatively small, the simulated sizes in Table 3 and powers in Table 5 are quite different from the theoretical values. When the null hypothesis is true, even with $\sqrt{c} = 200$, its size (7.61% of the right size in Table 3) does not match the nominal size of 5%. However, for relatively large enough c, for example, $\sqrt{c} = 800$, the simulated results of sizes approximate the theoretical values well. When the alternative

Table 3. Sizes of Left and Right Tailed Tests and Moments Under the Null.

$\alpha_c = 1$	Size (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}$	Left	Right	Mean	sd	Mean	sd	Mean	sd	Mean	sd
	(5%)	(5%)	$\left(\sqrt{c} \times 2.0921\right)$	$\left(\sqrt{c} \times 1.0876\right)$	(1)	$\left(1/\sqrt{c}\right)$	(0.8)	(0.989)	(−0.15)	(0.989)
50	5.20	12.95	101.9	57.4	1.019	0.0184	0.763	1.040	−0.170	0.985
			(104.6)	(54.4)						
100	4.62	9.83	205.1	110.8	1.009	0.0086	0.782	1.016	−0.161	0.981
			(209.2)	(108.8)						
150	5.19	8.22	310.0	167.4	1.006	0.0054	0.788	1.019	−0.156	0.988
			(313.8)	(163.1)						
200	5.04	7.19	413.9	218.8	1.004	0.0039	0.791	0.995	−0.154	0.981
			(418.4)	(217.5)						
800	4.87	5.23	1,668.8	871.8	1.001	0.0009	0.797	0.986	−0.151	0.994
			(1,673.7)	(870.1)						

Table 3 shows the simulation results along with the theoretical values in parentheses for: (i) the value of $\sqrt{c}$; (ii) the sizes of the left and right tailed tests; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

Table 4. Powers and Moments Under Stationary Alternatives (Significance Level: 5%).

$\alpha_c = 0.99$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}, \delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	12.50	126.0	68.4	1.012	0.0169	0.765	1.040	−0.168	1.005
$\delta = -0.5$	(12.61)	(129.8)	(64.5)	(0.99)	(0.02)	(0.8)	(0.9887)	(−0.15)	(0.989)
100	25.00	312.0	154.1	1.003	0.0069	0.782	1.011	−0.155	0.996
$\delta = -1$	(25.95)	(321.2)	(147.5)	(0.99)	(0.01)	(0.8)	(0.9887)	(−0.15)	(0.989)
150	42.64	575.8	257.6	1.000	0.0039	0.788	1.012	−0.153	0.998
$\delta = -1.5$	(44.24)	(589.9)	(244.2)	(0.99)	(0.0067)	(0.8)	(0.9887)	(−0.15)	(0.989)
200	61.88	933.5	364.2	0.999	0.0023	0.791	0.991	−0.151	0.996
$\delta = -2$	(63.88)	(947.0)	(348.4)	(0.99)	(0.005)	(0.8)	(0.9887)	(−0.15)	(0.989)

Table 4 corresponds to the alternative hypotheses $\alpha = 0.99$ and shows the simulation results along with the theoretical values in parentheses for: (i) the values of $\sqrt{c}$ and δ; (ii) the powers of the left tailed test; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

hypothesis is true, the power approaches the theoretical value as c increases. The simulated powers get closer to the theoretical powers from $\sqrt{c} = 100$ under the explosive alternative. In Tables 1–3, the results listed in the third to sixth columns clearly indicate the good performance of the stopping times and sequential estimators even when c is relatively small.

4.4. The Simulation Results When Lag Length is Unknown

We use simulations to investigate the performance of the sequential unit root test when the lag length of the AR process is unknown. We find the followings. When

Table 5. Powers and Moments Under Explosive Alternatives (Significance Level: 5%).

$\alpha_c = 1.01$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}, \delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	19.71	83.4	45.3	1.026	0.0190	0.764	1.045	−0.175	0.981
$\delta = 0.5$	(12.61)	(85.1)	(44.7)	(1.01)	(0.02)	(0.8)	(0.989)	(−0.15)	(0.989)
100	28.99	140.3	72.4	1.016	0.0098	0.779	1.029	−0.163	0.981
$\delta = 1$	(25.95)	(140.8)	(72.4)	(1.01)	(0.01)	(0.8)	(0.989)	(−0.15)	(0.989)
150	46.66	179.1	88.7	1.014	0.0066	0.787	1.002	−0.162	0.980
$\delta = 1.5$	(44.24)	(178.9)	(88.0)	(1.01)	(0.0067)	(0.8)	(0.989)	(−0.15)	(0.989)
200	64.88	208.8	96.0	1.012	0.0049	0.789	1.004	−0.158	0.986
$\delta = 2$	(63.88)	(206.5)	(96.3)	(1.01)	(0.005)	(0.8)	(0.989)	(−0.15)	(0.989)

Table 5 corresponds to the alternative hypotheses $\alpha = 1.01$ and shows the simulation results along with the theoretical values in parentheses for: (i) the values of $\sqrt{c}$ and δ; (ii) the powers of the right tailed test; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

Table 6. Rejection Rates (RRs) and Moments Under the Null (DGP: AR(4), Estimated Model: AR(3), Significance Level: 5%).

	RR (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$	
$\sqrt{c} = 200$	Left	Right	Mean	sd	Mean	sd
$\alpha_c = 1$	1.13	23.95	297.1	162.1	1.008	0.0045
	(5)	(5)	(418.4)	(217.5)	(1)	$(1/\sqrt{c})$
$\alpha_c = 0.99$	16.27		521.5	242.6	1.003	0.0066
	(63.88)		(881.4)	(384.4)	(0.99)	(0.005)
$\alpha_c = 1.01$		74.23	174.1	86.5	1.015	0.0078
		(63.88)	(206.5)	(95.3)	(1.01)	(0.005)

Table 6 shows the simulation results along with the theoretical values in parentheses for: (i) the sizes and powers of the left and right tailed test; (ii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iii) the mean and standard deviation of the sequential estimator $\hat{\alpha}_{c,\hat{\tau}_c}$.

the lag length is insufficient, there exists severe size distortion in both left tailed and right tailed tests. When sufficiently long lag lengths are used, the rejection rates are comparable to those when the correct lag lengths are used.

Table 6 gives simulation results when the lag length of the model is insufficient. The DGP is the AR(4) process with $\alpha_1 = 0.7, \alpha_2 = 0.7, \alpha_3 = 0.6$, but AR(3) is employed as the estimation model for the unit root test. The values in parentheses are the theoretical values when the AR process with true lag length, that is, AR(3), is used as the estimated model. We set $m_0 = 30$ for the definition of $\hat{\tau}_c$ in (56). The simulated sizes of left and right tailed test in Table 6 are 1.13% and 23.95%, respectively, though the nominal size is 5%. Because of the misspecification, it may be inevitable that the estimated parameters are biased and the sizes are distorted.

Tables 7 and 8 report the results when we apply models with sufficient lag length. From Theorem 2, we expect redundant lag lengths have no effect on the distribution of $\hat{\delta}_{\hat{\tau}_c}$ asymptotically. We confirm by simulation that the test yields similar results to those obtained when using true lag lengths. We now consider a unit root test with a lag length 3 in the estimated model under the DGP of the AR(2) process with $\alpha_1 = 0.5$. Tables 7 and 8, respectively, list the sizes and powers along with the results of the sequential estimation. For the left and right tailed tests, we see that the size properties are almost the same as when the lag length is correctly specified. In the left tailed test, the size characteristics in Table 7 are very good, as in Table 3. In the right tailed test, when $c \to \infty$, the size distortions get better as $11.72, 8.54, \ldots, 5.29$ from the top in Table 7, which is almost the same as $12.95, 9.83, \ldots, 5.23$ from the top in Table 3. The powers are also as good as those in Tables 4 and 5. The sequential estimation of the parameters of the remaining stationary process v_n with sufficiently long lag length is also well done.

We also conducted simulation studies using a lag length of 3 under the DGP of the AR(1) process. The results are similar to the case using the DGP of the AR(2).

Table 7. Sizes of Left and Right Tailed Tests and Moments Under the Null (DGP: AR(2), Estimated Model: AR(3), Significance Level: 5%).

$\alpha_c = 1$	Size (%)		$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c}$	Left	Right	Mean	sd	Mean	sd	Mean	sd	Mean	sd
	(5%)	(5%)	$\sqrt{c}\times 2.0921$	$\sqrt{c}\times 1.0876$	(1)	$(1/\sqrt{c})$	(0.5)	(1)	(0)	(1)
50	4.67	11.72	102.3 (104.6)	56.0 (54.4)	1.016	0.0181	0.468	1.053	−0.030	1.006
100	4.80	8.54	207.6 (209.2)	111.7 (108.8)	1.008	0.0085	0.483	1.026	−0.013	1.012
150	5.36	7.32	314.4 (313.8)	166.2 (163.1)	1.005	0.0055	0.490	1.026	−0.010	1.013
200	4.57	6.62	415.6 (418.4)	217.3 (217.5)	1.004	0.0039	0.493	1.017	−0.008	1.005
800	4.84	5.29	1664.4 (1673.7)	867.1 (870.1)	1.001	0.0009	0.499	1.003	−0.002	1.008

Table 7 shows the simulation results along with the theoretical values in parentheses for: (i) the value of $\sqrt{c}$; (ii) the sizes of the left and right tailed tests; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

Table 8. Powers and Moments Under Alternatives (DGP: AR(2), Estimated Model: AR(3), Significance Level: 5%).

$\alpha_c = 0.99$	Power (%)	$\hat{\tau}_c$		$\hat{\alpha}_{c,\hat{\tau}_c}$		$\hat{\psi}_{1,\hat{\tau}_c}$		$\hat{\psi}_{2,\hat{\tau}_c}$	
$\sqrt{c},\delta$		Mean	sd	Mean	sd	Mean	sd	Mean	sd
50	11.90	126.5	68.3	1.009	0.0168	0.469	1.055	−0.027	1.009
$\delta = -0.5$	(12.61)	(129.8)	(70.7)	(0.99)	(0.02)	(0.5)	(1)	(0)	(1)
100	25.47	316.2	154.6	1.001	0.0069	0.486	1.025	−0.010	1.015
$\delta = -1$	(25.95)	(321.2)	(147.5)	(0.99)	(0.01)	(0.5)	(1)	(0)	(1)
150	43.01	582.4	254.8	0.999	0.0039	0.492	1.010	−0.005	1.018
$\delta = -1.5$	(44.24)	(589.8)	(241.1)	(0.99)	(0.0067)	(0.5)	(1)	(0)	(1)
200	61.45	928.4	363.8	0.998	0.0026	0.493	1.005	−0.003	1.007
$\delta = -2$	(63.88)	(947.0)	(345.0)	(0.99)	(0.005)	(0.5)	(1)	(0)	(1)
$\alpha_c = 1.01$									
50	19.34	85.4	46.5	1.024	0.0189	0.466	1.068	−0.032	1.008
$\delta = 0.5$	(12.61)	(85.1)	(44.7)	(1.01)	(0.02)	(0.5)	(1)	(0)	(1)
100	29.99	142.7	74.4	1.016	0.0098	0.483	1.037	−0.019	0.999
$\delta = 1$	(25.95)	(140.8)	(72.5)	(1.01)	(0.01)	(0.5)	(1)	(0)	(1)
150	45.99	180.9	89.2	1.013	0.0064	0.487	1.015	−0.013	1.011
$\delta = 1.5$	(44.24)	(178.9)	(88.0)	(1.01)	(0.0067)	(0.5)	(1)	(0)	(1)
200	64.13	209.1	98.1	1.012	0.0049	0.490	1.029	−0.012	1.014
$\delta = 2$	(63.88)	(206.5)	(95.3)	(1.01)	(0.005)	(0.5)	(1)	(0)	(1)

Table 8 shows the simulation results along with the theoretical values in parentheses for: (i) the value of $\sqrt{c}$ and δ; (ii) the powers of the left and right tailed tests for $\alpha = 0.99$ and 1.01; (iii) the expected value and standard deviation of the stopping time $\hat{\tau}_c$; and (iv) the mean and standard deviation of the sequential estimators $\left(\hat{\alpha}_{c,\hat{\tau}_c}, \hat{\psi}_{1,\hat{\tau}_c}, \hat{\psi}_{2,\hat{\tau}_c}\right)$.

5. CONCLUDING REMARKS

This study considers testing for the existence of a unit root for AR(p) models under sequential sampling. Using a time change via the observed Fisher information, we obtain the joint asymptotic densities and Laplace transforms of the sequential test statistic and the stopping time both under the null and local alternatives from the theory of Bessel bridges in Pitman and Yor (1982). The null distribution of the stopping time is characterized by a 3/2-dimensional Bessel process, whereas the distribution under the local alternatives is represented in terms of a 3/2-dimensional Bessel process with a drift.

In this study, the simulations have shown that the sequential unit root test performs well when the estimated model has a sufficiently long lag length. However, we also find that it does not work when the lag length of the estimated model is insufficient.

There are some extensions for future research. In this study, we do not consider the upper limit of sample size, but in reality, there may be an upper limit to the acceptable sample size due to cost and budget constraints on sampling. We have developed a theory that also incorporates the upper limit of the sample size in a concurrent study. We assume that the disturbances constitute a sequence of $i.i.d.(0,\sigma^2)$ random variables, but this assumption could be relaxed to disturbances of martingale differences satisfying some additional conditions. The extension to sequential analysis of non-parametric AR processes and sequential tests for structural break or change point problems can be considered. We will consider a sequential analysis to determine the lag length consistent with the unit root test.

ACKNOWLEDGMENTS

We are grateful for the helpful comments and suggestions from the editor and an anonymous referee, which have significantly improved the structure and readability of this chapter. This study was supported by JSPS KAKENHI Grants Numbers JP17K03656, JP18K01543, JP19F19312, JP19H01473, JP19K21691, and JP20K01589.

REFERENCES

Billingsley, P. (1999). *Convergence of probability measures* (2nd ed.). John Wiley and Sons.

Blanchard, O. (2016). *Macroeconomics* (7th ed.). Pearson.

Bobkoski, M. J. (1983). *Hypothesis testing in nonstationary time series*. Ph.D. thesis, University of Wisconsin-Madison.

Borodin, A. N., & Salminen, P. (2002). *Handbook of Brownian motion – Facts and formulae* (2nd ed.). Birkhäuser.

Cavanagh, C. (1985). *Roots local to unity.* Department of Economics, Harvard University.

Chan, N. H. (1988). On the parameter inference for nearly nonstationary time series. *Journal of the American Mathematical Society*, *83*, 857–862.

Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *The Annals of Statistics*, *15*, 1050–1063.

Chang, Y. (2012). Taking a new contour: A novel approach to panel unit root test. *Journal of Econometrics*, *169*, 15–28.

Chang, Y., & Park, J. Y. (2002). On the asymptotics of ADF tests for unit roots. *Econometric Reviews*, *21*, 431–447.

Chang, Y., & Park, J. Y. (2004). *Taking a new contour: A novel view on unit root tests*. Rice University.

Choi, I. (2015). *Almost all about unit roots*. Cambridge University Press.

Dickey, D. A. (1976). *Estimation and hypothesis testing in nonstationary time series*. Ph.D. thesis, Iowa State University.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.

Galtchouk, L., & Konev, V. (2003). On uniform asymptotic normality of sequential least squares estimators for the parameters in a stable AR(1). *Sequential Analysis*, *22*, 31–54.

Galtchouk, L., & Konev, V. (2004). On uniform asymptotic normality of sequential least squares estimators for the parameters in a stable AR(p). *Journal of Multivariate Analysis*, *91*, 119–142.

Galtchouk, L., & Konev, V. (2005). On sequential least squares estimates of autoregressive parameters. *Sequential Analysis*, *24*, 335–364.

Galtchouk, L., & Konev, V. (2006). Sequential estimation of the parameters in unstable AR(2). *Sequential Analysis*, *25*, 25–43.

Galtchouk, L., & Konev, V. (2011). On asymptotic normality of sequential LS-estimates of unstable autoregressive processes. *Sequential Analysis*, *30*, 117–144.

Konev, V. V., & Pergamenshchicov, S. M. (1986). On the duration of sequential estimation of parameters of stochastic processes in discrete time. *Stochastics*, *18*, 133–154.

Lai, T. L., & Siegmund, D. (1983). Fixed accuracy estimation of an autoregressive parameter. *The Annals of Statistics*, *11*, 478–485.

Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer.

Mukhopadhyay, N., & Sriram, T. N. (1992). On sequential comparisons of means of first-order autoregressive models. *Metrika*, *39*, 155–164.

Nagai, K., Nishiyama, Y., & Hitomi, K. (2018). Sequential Tests of Unit Roots in the AR(1) Model. *KIER Discussion Paper 1003*, Kyoto University.

Novikov, A. A. (1972). Sequential estimation of the parameters of diffusion processes. *Mathematical Notes of the Academy of Sciences of the USSR*, *12*(5), 812–818.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica*, *55*(2), 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, *71*, 535–547.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*, 335–346.

Pitman, J., & Yor, M. (1982). A decomposition of bessel bridges. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *59*, 425–457.

Said, E. S., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*, 599–607.

Sriram, T. N. (2001). Fixed size confidence regions for parameters of threshold AR(1) models. *Journal of Statistical Planning and Inference*, *97*, 293–304.

Sriram, T. N., & Iaci, R. (2014). Sequential estimation for time series models. *Sequential Analysis*, *33*, 136–157.

Stock, J. H. (1994). *Unit roots, structural breaks and trends. Handbook of Econometrics* (Vol. 4). Elsevier.

Wald, A. (1947). *Sequential analysis*. John Wiley and Sons.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, *29*, 1188–1197.

6. APPENDIX

6.1. Proof of Proposition 1

Lemma 7 proves (36) and (37) in Proposition 1. Lemmas 10–14 provide the proofs of the convergences of the five components of (38) in Proposition 1. Lemma 8 transforms the strong law of large numbers for discrete-time random variables into almost sure convergence of the corresponding continuous-time stochastic process in $D[0,\infty)$. Lemma 9 provides key reformulations of the processes x_n and Δx_n for the proofs. Using Lemmas 7–9, we prove Lemmas 10–14.

Lemma 7. *For* α_i $(i=1,\dots,p-1)$ *in (7) with Assumption (8), the asymptotic autocovariance function* $\gamma_v(h)$ *of* v_n *exists for* $h=0,1,\dots$ *and satisfies (36) and (37).*

Proof. For v_n in (27), letting $\boldsymbol{v}_n=(v_n,v_{n-1},\cdots,v_{n-p+2})'$, $\boldsymbol{\epsilon_n}=(\epsilon_n,0,\cdots,0)'$, and

$$G=\begin{pmatrix} \psi_1 & \psi_2 & \cdots & \psi_{p-2} & \psi_{p-1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix},$$

we have

$$\boldsymbol{v}_n=G\boldsymbol{v}_{\boldsymbol{n-1}}+\boldsymbol{\epsilon}_n=\sum_{k=0}^{n-p-1}G^k\boldsymbol{\epsilon}_{\boldsymbol{n-k}}+G^{n-p}\boldsymbol{v}_{\mathrm{p}}. \tag{58}$$

We now consider another probability space $(\Omega',\mathcal{F}',P')$ on which independent variables $\epsilon_p,\epsilon_{p-1},\dots,\epsilon_0,\epsilon_{-1},\dots$ have the same distribution as ϵ_{p+1} on $(\Omega,\mathcal{F},P)$. Let $(\Omega^*,\mathcal{F}^*,P^*)$ be the completed probability space of the product space $(\Omega\times\Omega',\mathcal{F}\times\mathcal{F}',P\times P')$. Let v_n^* be the ergodic, stationary AR($p-1$) process in $(\Omega^*,\mathcal{F}^*,P^*)$;

$$=\sum_{k=0}^{\infty}G^k\boldsymbol{\epsilon}_{n-k} \quad (n=0,\pm1,\pm2,\dots). \tag{59}$$

Then, from (58),

$$\boldsymbol{v}_n-\boldsymbol{v}_n^*=G^{n-p}\left(\boldsymbol{v}_{\mathrm{p}}-\boldsymbol{v}_p^*\right) \quad (n=p+1,p+2,\dots), \tag{60}$$

holds due to $\boldsymbol{v}_p^*=\sum_{k=0}^{\infty}G^k\boldsymbol{\epsilon}_{p-k}$. Let the autocovariance function of $\boldsymbol{v}_n^*$ be

$$\gamma_v(h)=E(v_{n+h}^*v_n^*) \quad (h=0,1,\dots).$$

From the ergodic theorem,

$$\sum_{n=p+1}^{m}v_{n+h}^*v_n^*/(m-p)\to\gamma_v(h) \quad P^*-a.s.,$$

and

$$\sum_{n=p+1}^{m} \boldsymbol{v}_{n+h}^{*\prime} / (m-p) = \left(\sum_{n=p+1}^{m} v_{n+h-i}^{*} v_{n-j}^{*} \right)_{i,j} / (m-p) \to \boldsymbol{\Gamma}_{v}^{h} \quad P^{*} - a.s.,$$

where $i, j = 0,1,\ldots, p-2$, and $\boldsymbol{\Gamma}_{v}^{h} = \left(\gamma_{v}(|h+j-i|)\right)_{i,j}$. Using (60),

$$\begin{aligned} & \frac{1}{m-p} \sum_{n=p+1}^{m} v_{n+h} v_{n}^{\prime} \\ = \;& \frac{1}{m-p} \sum_{n=p+1}^{m} \boldsymbol{v}_{n+h}^{*\prime} + \frac{1}{m-p} \sum_{n=p+1}^{m} \left(G^{n-p+h}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)\left(G^{n-p}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)^{\prime} \\ + \;& \frac{1}{m-p} \sum_{n=p+1}^{m} \boldsymbol{v}_{n+h}^{*} \left(G^{n-p}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)^{\prime} + \frac{1}{m-p} \sum_{n=p+1}^{m} \left(G^{n-p+h}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)^{\prime}. \end{aligned} \tag{61}$$

By the Cayley–Hamilton theorem,

$$G^{n} = \psi_{1} G^{n-1} + \psi_{2} G^{n-1} + \cdots + \psi_{p-1} \boldsymbol{I}, \tag{62}$$

which implies that each element $\left(G^{n}\right)_{i,j}$ of G^{n} satisfies the difference equation with the characteristic polynomial $x^{n} = \psi_{1} x^{n-1} + \psi_{2} x^{n-1} + \cdots + \psi_{p-1}$. Thus, each element $\left(G^{n}\right)_{i,j}$ of G^{n} can be represented as a linear combination of ${}_{n}C_{k} \alpha_{i}^{n-k}, (k = 0,1,\ldots, p-1)$ and is bounded by $Kn^{p-2}\xi^{n}$ with some constant $K > 0$ and some $\xi \in (0,1)$ satisfying $\max_{i<p} |\alpha_{i}| < \xi$. Therefore, as $n \to \infty$,

$$\left(G^{n-p+h}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)\left(G^{n-p}\left(\boldsymbol{v}_{\mathrm{p}} - \boldsymbol{v}_{p}^{*}\right)\right)^{\prime} \to \boldsymbol{0} \quad P^{*} - a.s..$$

The Cesàro mean indicates that the second term of (61) P^{*} almost surely converges to **0**. The last two terms of the right side of (61) P^{*} almost surely converge to **0**, since there exists $n \geq \exists n_{0}$ such that $Kn^{p-2}|\xi|^{n} \leq 1$.

Then, we have (36) from

$$\sum_{n=p+1}^{m} \boldsymbol{v}_{n+h} \boldsymbol{v}_{n}^{\prime} / (m-p) \to \boldsymbol{\Gamma}_{v}^{h} \quad P - a.s..$$

From the FCLT with respect to ergodic, stationary martingale differences (Billingsley, 1999; Theorem 18.3),

$$\sum_{n=p+1}^{\lfloor \sqrt{c}t \rfloor} \boldsymbol{v}_{n-1}^{*} \epsilon_{n} / \left(c^{1/4} \sigma \sqrt{\gamma_{v}(0)}\right) \Rightarrow \left(W_{i}(t)\right)_{i=1,\ldots p-1}.$$

Using (60), we have

$$
\begin{aligned}
\frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{ct}\rfloor}\boldsymbol{v}_{n-1}\epsilon_n &= \frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{ct}\rfloor}\boldsymbol{v}^*_{n-1}\epsilon_n \\
&+ \frac{1}{c^{1/4}\sigma\sqrt{\gamma_v(0)}}\sum_{n=p+1}^{\lfloor\sqrt{ct}\rfloor}\epsilon_n\boldsymbol{G}^{n-1-p}\left(\boldsymbol{v}_{\mathrm{p}}-\boldsymbol{v}^*_p\right).
\end{aligned}
\tag{63}
$$

For $|\alpha|<1$, define $M_m=\sum_{n=p+1}^{m}{}_nC_k\alpha^{n-k}\epsilon_n$, $(k=0,1,\ldots,p-1)$. Then each term of $\sum_{n=p+1}^{m}\epsilon_n G^n$ can be rewritten as a linear combination of such M_m's with $\alpha=\alpha_1,\ldots,\alpha_{p-1}$. Due to $\sum_{n=p+1}^{\infty}\left({}_nC_k\alpha^{n-k}\right)^2<\infty$, M_m is a uniformly integrable L^2 martingale and $M_m\to M_\infty$ *a.s.*. There exists m_0 such that $|M_m|<|M_\infty|+1$ for any $m\geq m_0$. As $c\to\infty$

$$
\max_{t\leq t_0}\left|M_{\lfloor\sqrt{ct}\rfloor}\right|/c^{1/4}\leq\max_{t_0\leq m_0}|M_m|/c^{1/4}+\left(|M_\infty|+1\right)/c^{1/4}\to 0 \quad a.s.,
$$

which implies that the second term of the right side of (63) converges to $\mathbf{0}$ in $D[0,\infty)$.

Lemma 8. *Suppose that for some random sequence* w_n, $n=1,2,\ldots$ *and* $\theta\in\mathbb{R}$ $\sum_{n=1}^{m}w_n/m\to\theta$ *a.s. Let* $I:[0,\infty)\to[0,\infty)$ *be* $I(t)=\theta t$ *and*

$$
I_c(t)=\sum_{n=1}^{\lfloor\sqrt{ct}\rfloor}w_n/\sqrt{c}. \tag{64}
$$

Then for any $t_0>0$ *as* $c\uparrow\infty$,

$$
\sup_{t\leq t_0}|I_c(t)-\theta t|\to 0 \quad a.s.
$$

Proof. Fix $\omega\in\left\{\sum_{n=1}^{m}w_n/m\to\theta\right\}$ and for any $t_0>0$ and $\varepsilon>0$ find m_0 so that for any $m\geq m_0$, $\left|\sum_{n=1}^{m}w_n/m-\theta\right|<\varepsilon/2t_0$. Then, for large enough $c>0$,

$$
\begin{aligned}
&\sup_{t\leq t_0}|I_c(t)-\theta t| \\
=\;& \sup_{t\leq t_0}\left|\sum_{n=1}^{\lfloor\sqrt{ct}\rfloor}(w_n-\theta)/\sqrt{c}+\left(\lfloor\sqrt{ct}\rfloor-\sqrt{ct}\right)\theta/\sqrt{c}\right| \\
\leq\;& \max_{m\leq m_0}\left|\sum_{n=1}^{m}(w_n-\theta)/\sqrt{c}\right|\vee\sup_{m_0\leq\sqrt{ct}\leq\sqrt{ct_0}}\left|t\sum_{n=1}^{\lfloor\sqrt{ct}\rfloor}(w_n-\theta)/\left(\sqrt{ct}\right)\right|+|\theta|/\sqrt{c} \\
\leq\;& \varepsilon/4+\varepsilon t_0/2t_0+\varepsilon/4=\varepsilon.
\end{aligned}
\tag{65}
$$

The following lemma reforms x_n and Δx_n as linear combinations of near-unit-root AR(1) process Z_n in (27), a strongly stationary process y_n, and a nearly strongly stationary process $y_{c,n}$ defined in (66).

Lemma 9. *Let*

$$y_n = \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\psi_i\right)v_{n-k+1} \quad \text{and} \quad y_{c,n} = \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\frac{\psi_i}{\alpha_c^{i-k+1}}\right)v_{n-k+1}. \tag{66}$$

Then the near-unit-root $AR(p)$ process x_n in (6) can be represented as

$$x_n = \left(z_n - y_{c,n}\right)/\,\Psi(\alpha_c^{-1}), \tag{67}$$

and the key relations for the proofs of asymptotic properties are obtained:

$$\Delta y_n = \epsilon_n - \Psi(1)v_n, \tag{68}$$

$$\Delta x_n = \left(\frac{\delta}{\sqrt{c}}z_{n-1} + \Psi(1)v_n + \Delta y_n - \Delta y_{c,n}\right)/\,\Psi(\alpha_c^{-1}), \tag{69}$$

$$y_n - y_{c,n} \to 0 \quad a.s. \quad \text{as} \quad c\to\infty, \quad \text{for} \quad \text{any} \quad n. \tag{70}$$

Proof. Since $x_{n-1} = x_n/\alpha_c - v_n/\alpha_c$ from (27), we can obtain by induction, for $i = 1,\ldots,p-1$,

$$x_{n-i} = \frac{x_n}{\alpha_c^i} - \sum_{k=1}^{i}\frac{v_{n-k+1}}{\alpha_c^{i-k+1}}.$$

Since $z_n = x_n - \sum_{i=1}^{p-1}\psi_i x_{n-i}$, we have

$$\begin{aligned} z_n &= x_n - \sum_{i=1}^{p-1}\psi_i\left(\frac{x_n}{\alpha_c^i} - \sum_{k=1}^{i}\frac{v_{n-k+1}}{\alpha_c^{i-k+1}}\right) \\ &= \left(1 - \sum_{i=1}^{p-1}\frac{\psi_i}{\alpha_c^i}\right)x_n + \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\frac{\psi_i}{\alpha_c^{i-k+1}}\right)v_{n-k+1} \\ &= \Psi(\alpha_c^{-1})x_n + y_{c,n}, \end{aligned} \tag{71}$$

which implies (67). As to Δy_n,

$$\begin{aligned} \Delta y_n &= \sum_{k=0}^{p-2}\sum_{i=k+1}^{p-1}\psi_i v_{n-k} - \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\psi_i v_{n-k} = -\sum_{k=1}^{p-1}\psi_k v_{n-k} + \sum_{i=1}^{p-1}\psi_i v_n \\ &= \left(\Psi(L) - \Psi(1)\right)v_n = \epsilon_n - \Psi(1)v_n. \end{aligned}$$

Taking the difference of (67), we have

$$\begin{aligned}\Delta x_n &= \left(\Delta z_n - \Delta y_{c,n}\right)/\Psi(\alpha_c^{-1}) = \left(\frac{\delta}{\sqrt{c}} z_{n-1} + \epsilon_n - \Delta y_{c,n}\right)/\Psi\left(\alpha_c^{-1}\right) \\ &= \left(\frac{\delta}{\sqrt{c}} z_{n-1} + \Psi(1)v_n + \Delta y_n - \Delta y_{c,n}\right)/\Psi\left(\alpha_c^{-1}\right).\end{aligned} \tag{72}$$

Here we use (28) and (68) for the last two equations. Since $\alpha_c \to 1$, as $c \to \infty$,

$$y_n - y_{c,n} = \sum_{k=1}^{p-1}\left(\sum_{i=k}^{p-1}\psi_i\left(1-1/\alpha_c^{i-k+1}\right)\right)v_{n-k+1} \to 0 \quad a.s.$$

Define for $i, j = 1,\ldots, p-1$,

$$F_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}^2/\left(c\sigma^2\right), \qquad J_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta z_n/\left(\sqrt{c}\sigma^2\right), \tag{73}$$

$$I_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \epsilon_n^2/\sqrt{c}, \qquad I_{ij,c}(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-i}v_{n-j}/\sqrt{c}, \tag{74}$$

$$H_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_n/\sqrt{c}, \qquad H_{i,c}(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \epsilon_n v_{n-i}/\sqrt{c}, \tag{75}$$

$$W_{i,c}(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-i}\epsilon_n/\left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right). \tag{76}$$

The convergence with respect to (Z_c, F_c, J_c) in Lemma 8 and Lemma 7, we obtain, in the sense of $D[0,\infty)$,

$$\begin{aligned}&\begin{pmatrix} Z_c & F_c & J_c & W_{i,c}\end{pmatrix} \Rightarrow \begin{pmatrix} Z & F & J & W_i\end{pmatrix}, \\ &\begin{pmatrix} I_c & I_{ij,c} & H_c & H_{i,c}\end{pmatrix} \to \begin{pmatrix} I & I_{ij} & 0 & 0\end{pmatrix} \quad a.s.,\end{aligned}$$

where Z is in (33), and

$$\begin{aligned}F(t) &= \int_0^t Z_u^2 du, & J(t) &= \int_0^t Z_u dZ_u, \\ I(t) &= \sigma^2 t, & I_{ij}(t) &= \gamma_v(|i-j|)t.\end{aligned}$$

Applying Skorohod's theorem and changing $(\Omega, \mathcal{F}, P)$ to a completed probability space in which, as $c \to \infty$, in the sense of space $D[0,\infty)$,

$$\begin{aligned}&\begin{pmatrix} Z_c & F_c & J_c & W_{i,c} & I_c & I_{ij,c} & H_c & H_{i,c}\end{pmatrix} \\ \to &\begin{pmatrix} Z & F & J & W_i & I & I_{ij} & 0 & 0\end{pmatrix} \quad a.s.\end{aligned} \tag{77}$$

where

$$F_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}^2 / \left(c\sigma^2\right), \quad J_c(t) = \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta z_n / \left(c\sigma^2\right). \tag{78}$$

Next lemma proves the convergence of the first component in (38).

Lemma 10. *As* $c \to \infty$,

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 / \left(c\sigma^2\right) \to F(t) / \Psi^2(1) = \int_0^t Z^2(u)du / \Psi^2(1)$$

in the sense of space $D[0,\infty)$.

Proof. We obtain the first convergence of (38). We use (67) and write

$$\begin{aligned} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}^2 / c\sigma^2 &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\left(z_{n-1} - y_{c,n-1}\right) / \Psi(\alpha_c^{-1})\right)^2 / c\sigma^2 \\ &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(z_{n-1}^2 - 2z_{n-1}y_{c,n-1} + y_{c,n-1}^2\right) / \left(\Psi^2(\alpha_c^{-1})c\sigma^2\right). \end{aligned} \tag{79}$$

For the right-hand side of the last equation, we prove as $c \to \infty$,

$$\begin{aligned} &\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{c,n-1}^2 / c \to 0, \quad \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(y_n - y_{c,n}\right)^2 / \sqrt{c} \to 0, \\ &\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\Delta y_n - \Delta y_{c,n}\right)^2 / \sqrt{c} \to 0, \end{aligned} \tag{80}$$

uniformly in $t \in [0, m]$ for any $m > 0$ a.s. As $c \to \infty$, we obtain

$$\begin{aligned} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} y_{c,n-1}^2 / c &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\sum_{k=1}^{p-1}\sum_{i=k}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} v_{n-k} \right)^2 / c \\ &= \frac{1}{\sqrt{c}} \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} \frac{\psi_j}{\alpha_c^{j-l+1}} \left(\frac{1}{\sqrt{c}} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} v_{n-k}v_{n-l} \right) \\ &= \frac{1}{\sqrt{c}} \sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \frac{\psi_i}{\alpha_c^{i-k+1}} \frac{\psi_j}{\alpha_c^{j-l+1}} I_{kl,c}(t) \to 0, \end{aligned}$$

since

$$\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1} \psi_i\psi_j / (\alpha_c^{i-k+1}\alpha_c^{j-l+1}) = O(1).$$

For $\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_n-y_{c,n}\right)^2/\sqrt{c}\to 0$, as $c\to\infty, 1/\alpha_c^{i-k+1}\to 0$. We have

$$
\begin{aligned}
&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(y_n-y_{c,n}\right)^2\\
=\;&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\psi_i\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)v_{n-k+1}\right)^2\\
=\;&\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1}\psi_i\psi_j\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)\left(1-\frac{1}{\alpha_c^{j-l+1}}\right)\left(\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}v_{n-k}v_{n-l}\right)\\
=\;&\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\sum_{l=1}^{p-1}\sum_{j=l}^{p-1}\psi_i\psi_j\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)\left(1-\frac{1}{\alpha_c^{j-l+1}}\right)I_{kl,c}(t)\to 0
\end{aligned}
$$

uniformly in $t\in[0,m]$ for any $m>0$ a.s. The other convergence holds from

$$
\begin{aligned}
\left(\Delta y_n-\Delta y_{c,n}\right)^2 &= \left(y_n-y_{c,n}-\left(y_{n-1}-y_{c,n-1}\right)\right)^2\\
&\le 2\left(y_n-y_{c,n}\right)^2+2\left(y_{n-1}-y_{c,n-1}\right)^2.
\end{aligned}
$$

Using the Cauchy–Schwarz inequality and (77), (80), we can obtain

$$
\left|\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}y_{c,n-1}/c\right|\le\sqrt{\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}^2/c}\sqrt{\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}y_{c,n-1}^2/c}\to 0.\quad a.s.,
$$

which shows Lemma 10.

Next lemma proves the convergence of the third component in (38).

Lemma 11. $\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\Delta x_{n-i}\Delta x_{n-j}/\sqrt{c}\right)_{i,j}\to\left(\gamma_v\left(|i-j|\right)t\right)_{i,j}\quad a.s.$

Proof. Using (72), we have

$$
\begin{aligned}
&\frac{1}{\sqrt{c}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\Delta x_{n-i}\Delta x_{n-j}\\
=\;&\frac{1}{\sqrt{c}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\frac{\delta}{\sqrt{c}}z_{n-i-1}+\Psi(1)v_{n-i}+\Delta y_{n-i}-\Delta y_{c,n-i}\right)\\
&\times\left(\frac{\delta}{\sqrt{c}}z_{n-j-1}+\Psi(1)v_{n-j}+\Delta y_{n-j}-\Delta y_{c,n-j}\right).
\end{aligned}
$$

Using the ergodic theorem, we have

$$\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Psi^2(1)v_{n-i}v_{n-j} / \left(\sqrt{c}\Psi^2(\alpha_c^{-1})\right) = \Psi^2(1)I_{kl,c}(t) / \left(\Psi^2(\alpha_c^{-1})\right) \to \gamma_v\left(|i-j|\right)t$$

uniformly in $t \in [0,m]$ for any $m > 0$ a.s. The Cauchy–Schwarz inequality and (80) help for the evaluation of the other terms.

Next lemma proves the convergence of the forth component in (38).

Lemma 12. $\sum_{n=p+1}^{\sqrt{c}t} x_{n-1}\epsilon_n / \left(\sqrt{c}\sigma^2\right) \to \int_0^t Z_s dW_s / \Psi(1)$.

Using (67), $\left(I_c, W_{i,c}\right) \to \left(I, W_i\right)$ in (77) and (80), we have

$$\begin{aligned}
\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} x_{n-1}\epsilon_n / \sqrt{c} &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(z_{n-1} - y_{n-1} + \left(y_{n-1} - y_{c,n-1}\right)\right)\epsilon_n / \left(\sqrt{c}\Psi(\alpha_c^{-1})\right) \\
&= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\epsilon_n / \left(\sqrt{c}\Psi(\alpha_c^{-1})\right) + o_p(1).
\end{aligned}$$

Using (77), we have

$$\begin{aligned}
\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\epsilon_n / \left(\sqrt{c}\sigma^2\right) &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\left(\Delta z_n - \delta z_{n-1} / \sqrt{c}\right) / \left(\sqrt{c}\sigma^2\right) \\
&= J_c(t) - \delta F_c(t) \to \int_0^t Z_s dZ_s - \delta \int_0^t Z_s^2 ds = \int_0^t Z_s dW_s.
\end{aligned} \tag{81}$$

Next lemma proves the convergence of the fifth component in (38).

Lemma 13. *As* $c \to \infty$,

$$\left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n / \left(c^{1/4}\sqrt{\gamma_v(0)}\sigma\right)\right)_i \to \left(W_i(t)\right)_i. \tag{82}$$

Proof. By (72), we have

$$\begin{aligned}
&\frac{1}{c^{1/4}\sigma^2} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \Delta x_{n-i}\epsilon_n \\
&= \frac{1}{c^{1/4}\sigma^2\Psi(\alpha_c^{-1})} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\frac{\delta}{\sqrt{c}} z_{n-i-1} + \Psi(1)v_{n-i} + \Delta y_{n-i} - \Delta y_{c,n-i}\right)\epsilon_n.
\end{aligned}$$

By (28), $\Delta z_n = \delta z_{n-1} / \sqrt{c} + \epsilon_n$, $J_c \to J$ in (77) and (81) yield

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(\Delta z_n)^2/\sqrt{c} &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\delta^2 z_{n-1}^2/c+2\delta z_{n-1}\epsilon_n/\sqrt{c}+\epsilon_n^2\right)/\sqrt{c} \\ &\to I(t)=\sigma^2 t\end{aligned}. \quad (83)$$

(81) in Lemma 12 and (83) imply, as $c\to\infty$,

$$\frac{1}{\sqrt{c}c^{1/4}\sigma^2}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(\delta z_{n-1-i})\,\epsilon_n=\frac{1}{\sqrt{c}c^{1/4}\sigma^2}\delta\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(z_{n-1}-\sum_{k=1}^{i}\Delta z_{n-k}\right)\epsilon_n\to 0$$

from the Cauchy–Schwarz inequality. Since $W_{c,i}\to W_i$ in (77), and $a_c\to 1$,

$$\frac{1}{c^{1/4}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(y_{n-i}-y_{c,n-i})\,\epsilon_n=\sum_{k=1}^{p-1}\sum_{i=k}^{p-1}\psi_i\left(1-\frac{1}{\alpha_c^{i-k+1}}\right)\frac{1}{c^{1/4}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}v_{n-i+k+1}\,\epsilon_n\to 0,$$

we have (82).

Next lemma proves the convergence of the second component in (38).

Lemma 14. *As* $c\to\infty, \left(\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}x_{n-1}\Delta x_{n-i}/c^{3/4}\sigma^2\right)_i\to\mathbf{0}$.

Proof. Using (67) we have

$$\begin{aligned}&\frac{1}{c^{3/4}}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}x_{n-1}\Delta x_{n-i}\\
=\;&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(z_{n-1}-y_{c,n-1})(\Delta z_{n-i}-\Delta y_{c,n-i})\\
=\;&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(z_{n-1}-y_{n-1}+(y_{n-1}-y_{c,n-1})\right)\\
&\times\left(\Delta z_{n-i}-\Delta y_{n-i}+(\Delta y_{n-i}-\Delta y_{c,n-i})\right)\\
=\;&\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}\Delta z_{n-i}-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}\Delta y_{n-i}\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}z_{n-1}(\Delta y_{n-i}-\Delta y_{c,n-i})+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}y_{n-1}\Delta z_{n-i}\\
&-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}y_{n-1}\Delta y_{n-i}+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}y_{n-1}(\Delta y_{n-i}-\Delta y_{c,n-i})\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(y_{n-i}-y_{c,n-i})\Delta z_{n-i}-\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(y_{n-i}-y_{c,n-i})\Delta y_{n-i}\\
&+\frac{1}{c^{3/4}}\frac{1}{\Psi^2(\alpha_c^{-1})}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}(y_{n-i}-y_{c,n-i})(\Delta y_{n-i}-\Delta y_{c,n-i}).\end{aligned} \quad (84)$$

Except for the first two terms, the terms in the last equality converge to 0 by (80) and (83), and $I_{ij,c} \to I_{ij}$. As for the first two terms, since $z_{n-1} = \sum_{k=1}^{i} \Delta z_{n-k} + z_{n-i-1}$, we have

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta z_{n-i} / c^{3/4} &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\left(\sum_{k=1}^{i}\Delta z_{n-k} + z_{n-i-1}\right)\Delta z_{n-i} / c^{3/4} \\ &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor}\sum_{k=1}^{i}\Delta z_{n-k}\Delta z_{n-i} / c^{3/4} + \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-i-1}\Delta z_{n-i} / c^{3/4}.\end{aligned}$$

As $c \to \infty$, the second term in the last equation converges to 0 uniformly in $t \in [0, m]$ for any $m > 0$ a.s., since

$$\sum_{n=1}^{\lfloor\sqrt{c}t\rfloor} z_{n-i-1}\Delta z_{n-i} / \left(\sqrt{c}\sigma^2\right) = J_c(t) + o_p(1) \to \int_0^t Z_u dZ_u \tag{85}$$

uniformly in $t \in [0, m]$ for any $m > 0$ a.s.

Next, as $c \to \infty$,

$$\begin{aligned}\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\Delta y_{n-i} / c^{3/4} &= \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} z_{n-1}\left(y_{n-i} - y_{n-i-1}\right) / c^{3/4} \\ &= -\sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor-1}\Delta z_n y_{n-i} / c^{3/4} - z_p y_{p-i} / c^{3/4} + z_{\lfloor\sqrt{c}t\rfloor-1} y_{\lfloor\sqrt{c}t\rfloor-i} / c^{3/4} \\ &\to 0.\end{aligned}$$

Using the Cauchy–Schwarz inequality and (83), we find that the first term converges to 0 uniformly in t. The third term in the last equation converges to 0, since $z_{\lfloor\sqrt{c}t\rfloor-1} / c^{1/4}\sigma \to Z(t)$ and $v_{\lfloor\sqrt{c}t\rfloor-i} / \sqrt{c} = H_c\left(t - i/\sqrt{c}\right) - H_c\left(t - (i-1)/\sqrt{c}\right) \to 0$.

6.2 Proof of Theorem 2

Now, we conclude the proof of the main theorem.

Proof. Let $\tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} = \left(x_{n-1}, x_{n-2}, \cdots, x_{n-p}\right)'$, $\boldsymbol{\psi}^* = \left(0, \psi^t\right)'$ and $\hat{\boldsymbol{\psi}}^*_{\mathrm{m}} = \left(0, \hat{\boldsymbol{\psi}}^{\mathrm{t}}_{\mathrm{m}}\right)'$, according to (17), (25),

$$\boldsymbol{\psi}^* = A_I \boldsymbol{\phi}^{\mathrm{c}} \quad \text{and} \quad \hat{\boldsymbol{\psi}}^*_{\mathrm{m}} = \hat{A}_{I,m}\hat{\boldsymbol{\phi}}^{\mathrm{c}}_{\mathrm{m}}, \tag{86}$$

where $A_I = \begin{pmatrix} 0 & 0 \\ 0 & A_c^{-1} \end{pmatrix}$, $\hat{A}_{I,m} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{A}_{c,m}^{-1} \end{pmatrix}$ and

$$\hat{A}_{c,m}^{-1} = \begin{pmatrix} 1/\hat{\alpha}_{c,m} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{2} & \cdots & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-2} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-1} \\ 0 & 1/\hat{\alpha}_{c,m} & \cdots & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-3} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{p-2} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1/\hat{\alpha}_{c,m} & \left(1-\hat{\alpha}_{c,m}\right)/\hat{\alpha}_{c,m}^{2} \\ 0 & 0 & \cdots & 0 & 1/\hat{\alpha}_{c,m} \end{pmatrix}.$$

Using (27),

$$\begin{aligned} \hat{\Psi}_m(L)x_{n-1} &= \Psi(L)x_{n-1} + \left(\hat{\Psi}_m(L) - \Psi(L)\right)x_{n-1} \\ &= z_{n-1} + \left(\hat{\boldsymbol{\psi}}_{\boldsymbol{m}}^{*} - \boldsymbol{\psi}^{*}\right)' \tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} \\ &= z_{n-1} + \left(\hat{A}_{I,m}\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\boldsymbol{c}} - A_I \boldsymbol{\phi}^{c}\right)' \tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} \\ &= z_{n-1} + \left(\hat{A}_{I,m}\left(\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\boldsymbol{c}} - \boldsymbol{\phi}^{c}\right) + o(1)\right)' \tilde{\boldsymbol{x}}_{\boldsymbol{n-1}}. \end{aligned}$$

The LSE

$$\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\boldsymbol{c}} - \boldsymbol{\phi}^{c} = \boldsymbol{D}_{\boldsymbol{c}}^{-1}\boldsymbol{\Xi}_{\boldsymbol{c},\boldsymbol{m}}^{-1}\boldsymbol{D}_{\boldsymbol{c}}^{-1}X_m'\boldsymbol{\epsilon}_{\boldsymbol{m}},$$

where $\boldsymbol{\Xi}_{\boldsymbol{c},\boldsymbol{m}} = \boldsymbol{D}_{\boldsymbol{c}}^{-1}X_m'X_m\boldsymbol{D}_{\boldsymbol{c}}^{-1}$. Then the observed Fisher information becomes

$$\begin{aligned} \sum_{n=p+1}^{m}\left(\hat{\Psi}_m(L)x_{n-1}\right)^2 &= \sum_{n=p+1}^{m} z_{n-1}^2 + 2\left(\hat{A}_{I,m}\left(\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\mathrm{c}} - \boldsymbol{\phi}^{\mathrm{c}}\right) + o(1)\right)' \sum_{n=p+1}^{m} \tilde{\boldsymbol{x}}_{\boldsymbol{n-1}} z_{n-1} \\ &+ \left(\hat{A}_{I,m}\left(\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\mathrm{c}} - \boldsymbol{\phi}^{\mathrm{c}}\right) + o(1)\right)' \sum_{n=p+1}^{m} \tilde{\boldsymbol{x}}_{\boldsymbol{n-1}}\tilde{\boldsymbol{x}}_{n-1}' \qquad (87) \\ &\times \left(\hat{A}_{I,m}\left(\hat{\boldsymbol{\phi}}_{\boldsymbol{m}}^{\mathrm{c}} - \boldsymbol{\phi}^{\mathrm{c}}\right) + o(1)\right), \qquad (88) \end{aligned}$$

and the estimator of σ^2 is

$$s_m^2 = \left(\boldsymbol{\epsilon}_{\boldsymbol{m}}'\boldsymbol{\epsilon}_{\boldsymbol{m}} - \boldsymbol{\epsilon}_{\boldsymbol{m}}X_m\boldsymbol{D}_{\boldsymbol{c}}^{-1}\boldsymbol{\Xi}_{\boldsymbol{c},\boldsymbol{m}}^{-1}\boldsymbol{D}_{\boldsymbol{c}}^{-1}X_m'\boldsymbol{\epsilon}_{\boldsymbol{m}}\right)'/(m-p).$$

As $c \to \infty$, the main terms of

$$\boldsymbol{\Xi}_{\boldsymbol{c},\lfloor\sqrt{\boldsymbol{c}t}\rfloor} \to \begin{pmatrix} F(t)/\Psi^2(1) & \mathbf{0} \\ \mathbf{0} & t\boldsymbol{\Gamma}_{\boldsymbol{v}} \end{pmatrix},$$

and

$$\boldsymbol{D}_{\boldsymbol{c}}^{-1}X_{\lfloor\sqrt{\boldsymbol{c}t}\rfloor}'\boldsymbol{\epsilon}_{\lfloor\sqrt{\boldsymbol{c}t}\rfloor} \to \begin{pmatrix} \left(J(t) - \delta F(t)\right)/\Psi(1) \\ \boldsymbol{W}^{(1)}(t) \end{pmatrix}.$$

Stopping time $\hat{\tau}_c$ defined in (39) can be represented as

$$\frac{\hat{\tau}_c}{\sqrt{c}} = \inf\left\{t : \frac{\lfloor\sqrt{c}t\rfloor - p}{\sqrt{c}} \frac{1}{c\sigma^2} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\hat{\Psi}_{\lfloor\sqrt{c}t\rfloor}(L)x_{n-1}\right)^2 \geq \frac{\lfloor\sqrt{c}t\rfloor - p}{\sqrt{c}} \frac{s^2_{\lfloor\sqrt{c}t\rfloor}}{\sigma^2}\right\}.$$

As $c \to 0$,

$$\left|\boldsymbol{\Xi}_{c,\lfloor\sqrt{ct}\rfloor}\right| \to F(t)t\left|\boldsymbol{\Gamma}_v\right| / \Psi^2(1).$$

Since the right side close to 0, as $t \downarrow 0$. Multiplying both sides of the inequality in stopping time $\hat{\tau}_c$ by $\left|\boldsymbol{\Xi}_{c,\lfloor\sqrt{ct}\rfloor}\right|^2$, as $c \to \infty$, the right side

$$\begin{aligned} & \frac{\lfloor\sqrt{c}t\rfloor - p}{\sqrt{c}} \left|\boldsymbol{\Xi}_{c,\lfloor\sqrt{ct}\rfloor}\right|^2 \frac{1}{c\sigma^2} \sum_{n=p+1}^{\lfloor\sqrt{c}t\rfloor} \left(\hat{\Psi}_{\lfloor\sqrt{c}t\rfloor}(L)x_{n-1}\right)^2 \\ \to \quad & t\left(F(t)t\left|\boldsymbol{\Gamma}_v\right| / \Psi^2(1)\right)^2 F(t) \end{aligned}$$

and the left side

$$\frac{\lfloor\sqrt{c}t\rfloor - p}{\sqrt{c}} \left|\boldsymbol{\Xi}_{c,\lfloor\sqrt{ct}\rfloor}\right|^2 \frac{s^2_{\lfloor\sqrt{c}t\rfloor}}{\sigma^2} \to t\left(F(t)t\left|\boldsymbol{\Gamma}_{\mathrm{v}}\right| / \Psi^2(1)\right)^2.$$

We have

$$\begin{aligned} \hat{\tau}_c / \sqrt{c} \quad & \to \quad \inf\left\{t : t\left(F(t)t\left|\boldsymbol{\Gamma}_{\mathrm{v}}\right| / \Psi^2(1)\right)^2 F(t) = t\left(F(t)t\left|\boldsymbol{\Gamma}_{\mathrm{v}}\right| / \Psi^2(1)\right)^2\right\}, \\ & = \quad \inf\{t : F(t) = 1\} = U_1 \end{aligned}$$

where U_1 is defined in (41).

As for $\hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c} / \hat{\Psi}_{\hat{\tau}_c}(1) \Rightarrow \delta + B_1$, by the fact $\int_0^{U_1} Z_u^2 du = \langle M \rangle_{U_1} = 1$ and using DDS theorem $\int_0^{U_1} Z_u dW_u = M_{U_1} = B_1$, we have

$$\begin{aligned} \hat{\delta}_{\hat{\tau}_c} = \sqrt{c}\hat{\phi}^c_{1,\hat{\tau}_c} / \hat{\Psi}_{\hat{\tau}_c}(1) \quad & = \quad \frac{\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\epsilon_n / \sqrt{c}}{\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}^2 / c} \frac{1}{\hat{\Psi}_{\hat{\tau}_c}(1)} \\ & \to \quad \frac{J(\hat{\tau}_c / \sqrt{c})}{F(\hat{\tau}_c / \sqrt{c}) / \Psi^2(1)} \frac{1}{\Psi^2(1)} \\ & = \quad \frac{\delta + \int_0^{U_1} Z_u dW_u}{\int_0^{U_1} Z_u^2 du} \quad \text{a.s.} \\ & = \quad \delta + B_1. \end{aligned} \tag{89}$$

Finally we obtain the representation of the stopping time U_1 by using the Bessel process with a drift δ. The inverse function theorem gives $dU_s / ds = 1/Z_{U_s}^2$. By Ito's formula,

$$Z_u^2 = 2\int_0^u Z_t dZ_t + u = 2\delta\int_0^u Z_t^2 dt + 2\int_0^u Z_t dW_t + u. \tag{90}$$

Letting $u = U_s$, we have

$$Z_{U_s}^2 = 2\delta\int_0^{U_s} Z_t^2 dt + 2\int_0^{U_s} Z_t dW_t + U_s = 2\delta s + 2B_s + U_s.$$

Thus

$$dU_s / ds = 1/\left(2\delta s + 2B_s + U_s\right).$$

Put $\rho_s = Z_{U_s}^2 / 2 = (2\delta s + 2B_s + U_s)/2$, then we have

$$d\rho_s = \left(\delta + \frac{1}{4\rho_s}\right)ds + dB_s. \tag{91}$$

This indicates that ρ_s is the Bessel process of dimension 3/2 with a drift δ and a initial value $\rho_0 = 0$. Then, we have

$$U_t = \int_0^t dU_s = \int_0^t \frac{1}{2\delta s + 2B_s + U_s} ds = \int_0^t \frac{1}{2\rho_s} ds.$$

(45) in Theorem 2 is shown as follows. Let $\boldsymbol{D}_{\hat{\tau}_c}$ be the $p \times p$ matrix $\operatorname{diag}\left(\sqrt{c}, \sqrt{\hat{\tau}_c}, \ldots, \sqrt{\hat{\tau}_c}\right)$. Multiplying the Normal Equation by $\boldsymbol{D}_{\hat{\tau}_c}^{-1}$, we have

$$\boldsymbol{D}_{\hat{\tau}_c}^{-1} X_{\hat{\tau}_c}' X_{\hat{\tau}_c} \boldsymbol{D}_{\hat{\tau}_c}^{-1} \boldsymbol{D}_{\hat{\tau}_c} \left(\hat{\phi}_{\hat{\tau}_c}^c - \phi^c\right) = \boldsymbol{D}_{\hat{\tau}_c}^{-1} X_{\hat{\tau}_c}' \epsilon_{\hat{\tau}_c}. \tag{92}$$

Letting $t = \hat{\tau}_c / \sqrt{c}$ in Proposition 1, as $c \to \infty$, the convergence of $\boldsymbol{D}_{\hat{\tau}_c}^{-1} X_{\hat{\tau}_c}' X_{\hat{\tau}_c} \boldsymbol{D}_{\hat{\tau}_c}^{-1}$ is

$$\begin{pmatrix} \sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}^2 / c & \left(\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\Delta x_{n-j} / \sqrt{c\hat{\tau}_c}\right)_j \\ \left(\sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\Delta x_{n-i} / \sqrt{c\hat{\tau}_c}\right)_i' & \left(\sum_{n=p+1}^{\hat{\tau}_c} \Delta x_{n-i}\Delta x_{n-j} / \hat{\tau}_c\right)_{i,j} \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} \sigma^2 \int_0^{U_1} Z^2(u)du / \Psi(1)^2 & \mathbf{0}' \\ \mathbf{0} & \left(\sigma^2\gamma_v\left(|i-j|\right)\right)_{i,j} \end{pmatrix}, \tag{93}$$

and the convergence of $\boldsymbol{D}_{\hat{\tau}_c}^{-1} X_{\hat{\tau}_c}' \boldsymbol{\epsilon}_{\hat{\tau}_c}$ is

$$\begin{pmatrix} \sum_{n=p+1}^{\hat{\tau}_c} x_{n-1}\epsilon_n / \sqrt{c} \\ \left(\sum_{n=p+1}^{\hat{\tau}_c} \Delta x_{n-i}\epsilon_n / \sqrt{\hat{\tau}_c} \right)_i' \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma^2 \int_0^{U_1} Z(u) dW(u) / \Psi(1) \\ \sqrt{\gamma_v(0)} \sigma \boldsymbol{W}^{(1)}(1) \end{pmatrix}. \tag{94}$$

where $i, j = 1, \ldots, p-1$. (45) could be computed through (92).

Finally, we prove $s_{\hat{\tau}_c}^2 \rightarrow_p \sigma^2$. Setting $m = \hat{\tau}_c$ and substituting $\boldsymbol{\Delta x_m} = \boldsymbol{X_m} \boldsymbol{\phi}^{\mathrm{c}} + \boldsymbol{\epsilon_m}$ into (24), we have

$$\begin{aligned} s_{\hat{\tau}_c}^2 &= \left(\boldsymbol{\Delta x}_{\hat{\tau}_c} - X_{\hat{\tau}_c} \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right)' \left(\boldsymbol{\Delta} x_{\hat{\tau}_c} - X_{\hat{\tau}_c} \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right) / (\hat{\tau}_c - p) \\ &= \left(\boldsymbol{X}_{\hat{\tau}_c} \left(\boldsymbol{\phi}^c - \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right) + \boldsymbol{\epsilon}_{\hat{\tau}_c} \right)' \left(\boldsymbol{X}_{\hat{\tau}_c} \left(\boldsymbol{\phi}^c - \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right) + \boldsymbol{\epsilon}_{\hat{\tau}_c} \right) / (\hat{\tau}_{\boldsymbol{c}} - p) \\ &= \left\{ \left(\boldsymbol{\phi}^c - \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right)' \boldsymbol{D}_{\hat{\tau}_c} D_{\hat{\tau}_c}^{-1} \boldsymbol{X}_{\hat{\tau}_c}' \boldsymbol{X}_{\hat{\tau}_c} \boldsymbol{D}_{\hat{\tau}_c}^{-1} \boldsymbol{D}_{\hat{\tau}_c} \left(\boldsymbol{\phi}^c - \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right) \right. \\ &\quad \left. + 2 \left(\boldsymbol{\phi}^c - \hat{\boldsymbol{\phi}}_{\hat{\tau}_c}^{\boldsymbol{c}} \right) D_{\hat{\tau}_c} D_{\hat{\tau}_c}^{-1} X_{\hat{\tau}_c}' \boldsymbol{\epsilon}_{\hat{\tau}_c} + \boldsymbol{\epsilon}_{\hat{\tau}_c}' \boldsymbol{\epsilon}_{\hat{\tau}_c} \right\} / (\hat{\tau}_{\boldsymbol{c}} - p) \\ &\rightarrow_p \sigma^2. \end{aligned}$$

6.3 Proof of Theorem 3

Proof. Under null hypothesis, according to (40) $M_u = \int_0^u W_s dW_s$. Using a time-change $v = \langle M \rangle_u = \int_0^u W_s^2 ds$ and $dv = W_u^2 du$,

$$\begin{aligned} &\int_0^\infty e^{-\gamma v} P\{U_v \le t, 2\rho_v \le y\} dv = E\left[\int_0^\infty e^{-\gamma \langle M \rangle_u} 1\left\{ U_{\langle M \rangle_u} \le t, W_{U_{\langle M \rangle_u}}^2 \le y \right\} W_u^2 du \right] \\ = \quad &\int_0^t E\left[e^{-\gamma \langle M \rangle_u} 1\{W_u^2 \le y\} W_u^2 \right] du. \end{aligned}$$

By the fundamental theorem of calculus

$$\begin{aligned} &\int_0^\infty e^{-\gamma v} \frac{\partial^2}{\partial y \partial t} P\{U_v \le t, 2\rho_v \le y\} dv = \frac{\partial}{\partial y} E\left[e^{-\gamma \langle M \rangle_t} 1\{W_t^2 \le y\} W_t^2 \right] \\ = \quad &\int_0^\infty e^{-\gamma v} f_{\langle M \rangle_t, W_t^2}(v, y) y d\nu, \end{aligned}$$

where $f_{\langle M \rangle_t, W_t^2}(v, y)$ is the density function of $\left(\langle M \rangle_t, W_t^2 \right)$. Hence,

$$\frac{\partial^2}{\partial y \partial t} P\{U_v \le t, 2\rho_v \le y\} = f_{\langle M \rangle_t, W_t^2}(v, y) y.$$

According to the theory of Bessel Bridge in Pitman and Yor (1982)

$$E\left[\exp\left(-\frac{b^2}{2}\int_0^u W_s^2 ds \mid W_u^2 = y\right)\right]$$

$$= \lim_{x\to 0}\frac{bu}{\sinh(bu)}\exp\left(\frac{y}{2u}(1-bu\coth(bu))\right)\frac{I_{-1/2}\left(\frac{b\sqrt{xy}}{\sinh(bu)}\right)}{I_{-1/2}\left(\frac{\sqrt{xy}}{u}\right)} \tag{95}$$

$$= \left(\frac{bu}{\sinh(bu)}\right)^{1/2}\exp\left(\frac{y}{2u}(1-bu\coth(bu))\right), \tag{96}$$

where I_ν is the modified Bessel function defined as $I_\nu(x)=\sum_{k=0}^{\infty}(x/2)^{2k+\nu}/(k!\Gamma(k+\nu+1))$ for $\nu \geq -1$. Since W_u^2 has the marginal density

$$f_{W_u^2}(y) = y^{-1/2}e^{-y/(2u)}/\left(\Gamma(1/2)(2u)^{1/2}\right), \tag{97}$$

then

$$\int_0^\infty e^{-\gamma\nu} f_{\langle M\rangle_u, W_u^2}(\nu, y) y d\nu = \frac{y^{1/2}}{\sqrt{2\pi}}\left(\frac{\sqrt{2\gamma}}{\sinh\left(u\sqrt{2\gamma}\right)}\right)^{1/2}\exp\left(-\frac{y}{2}\sqrt{2\gamma}\coth\left(u\sqrt{2\gamma}\right)\right).$$

Since $2\rho_v = 2B_v + U_v$, one can obtain (46) from the expression of the es function in Borodin and Salminen (2002).

6.4 Proof of Theorems 5 and 6

Proof. Assuming $\gamma > 0, \alpha > 0, t > 0$, and integrating (96) with (97), we have

$$E_{H_0}\left[e^{-\gamma<M>_t-\alpha W_t^2}\right] = \int_0^\infty E_{H_0}\left[e^{-\gamma<M>_t} \mid W_t^2 = y\right]e^{-\alpha y} f_{W_t^2}(y)dy$$

$$= \frac{2^{1/4}\gamma^{1/4}}{\sqrt{\sqrt{2\gamma}\cosh\left(t\sqrt{2\gamma}\right)+2\alpha\sinh\left(t\sqrt{2\gamma}\right)}}.$$

Using the time change $t = U_v$ and the above expression

$$\begin{aligned}
&\int_0^\infty e^{-\gamma v} E_{H_0}\left[\exp\left(-\alpha 2\rho_v - \beta U_v\right)/2\rho_v\right]dv \\
=\ &\int_0^\infty e^{-\beta t} E_{H_0}\left[e^{-\gamma \langle M \rangle_t - \alpha W_t^2}\right]dt \qquad (98)\\
=\ &\int_0^\infty e^{-\beta t} \frac{2^{1/4}\gamma^{1/4}}{\sqrt{\sqrt{2\gamma}\cosh\left(t\sqrt{2\gamma}\right) + 2\alpha\sinh\left(t\sqrt{2\gamma}\right)}}dt \\
=\ &\int_0^1 \frac{s^{-3/4+\beta/(2\sqrt{2\gamma})}}{2^{3/4}\gamma^{1/4}\sqrt{2\alpha(1-s)+\sqrt{2\gamma}(s+1)}}ds \equiv \int_0^1 H(s,\alpha,\beta,\gamma)\,ds. \qquad (99)
\end{aligned}$$

Here we made a variable change $t = -\log(s)/\left(2\sqrt{2\gamma}\right)$ in the third equation. Since

$$\frac{\partial^{k+l} H}{\partial\alpha^k\,\partial\beta^l}(s,0,0,\gamma) = \frac{2^{-\frac{k}{2}-\frac{3l}{2}-1}\left(-\frac{1}{2}\right)^{(k)}(1-s)^k(1+s)^{-k-\frac{1}{2}}\log^l(s)}{s^{3/4}}\gamma^{-\frac{1}{2}(k+l+1)}.$$

and

$$\mathfrak{L}_\gamma^{-1}\left(\gamma^{-(j+1)/2}\right)(v) = v^{(j-1)/2}/\Gamma\left((j+1)/2\right),$$

$\mathfrak{L}_\gamma^{-1}\left(\partial^{k+l} H(s,0,0,\gamma)/\partial\alpha^k\,\partial\beta^l\right)(v)$ can be written as $J(t,v,k,l)$ in (53). Use Girsanov's theorem to obtain (54).

This page intentionally left blank

PART II

NONLINEARITY

This page intentionally left blank

CHAPTER 5

FUNCTIONAL-COEFFICIENT COINTEGRATING REGRESSION WITH ENDOGENEITY

Han-Ying Liang[a], Yu Shen[a] and Qiying Wang[b]

[a]*School of Mathematical Sciences, Tongji University, Shanghai, P.R. China*
[b]*School of Mathematics and Statistics, The University of Sydney, Sydney, NSW, Australia*

ABSTRACT

Joon Y. Park is one of the pioneers in developing nonlinear cointegrating regression. Since his initial work with Phillips (Park & Phillips, 2001) in the area, the past two decades have witnessed a surge of interest in modeling nonlinear nonstationarity in macroeconomic and financial time series, including parametric, nonparametric and semiparametric specifications of such models. These developments have provided a framework of econometric estimation and inference for a wide class of nonlinear, nonstationary relationships. In honor of Joon Y. Park, this chapter contributes to this area by exploring nonparametric estimation of functional-coefficient cointegrating regression models where the structural equation errors are serially dependent and the regressor is endogenous. The self-normalized local kernel and local linear estimators are shown to be asymptotic normal and to be pivotal upon an estimation of co-variances. Our new results improve those of Cai et al. (2009) and open up inference by conventional nonparametric method to a wide class of potentially nonlinear cointegrated relations.

Keywords: Cointegration; functional-coefficient model; nonstationary time series; endogeneity; kernel estimation; local linear estimation

JEL classifications: C14; C22

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 157–186

Published under exclusive licence by Emerald Publishing Limited
ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A005

1. INTRODUCTION

Linear cointegrating regression was suggested by Granger (1981) and Engle and Granger (1987) and has attracted extensive researches in both theory and empirical applications. The specification in linear structure is convenient for practical work and software packages have many standard routines for dealing with such a system, encouraging extensive usage of the methods. While common in applications, the linear structure is often too restrictive and linear cointegration models are often rejected by the data even when there is a clear long-run relationship in the series (see, for instance, Park & Phillips, 1988; Saikkonen, 1995; Terasvirta et al., 2011).

To overcome such deficiencies, various nonlinear cointegrating regression models have been suggested in past two decades. Among many other contributors, we refer to Park and Phillips (2001), Chang et al. (2001), Chang and Park (2003), Chan and Wang (2015) and Wang (2021) for parametric nonlinear cointegrating regression; Wang and Phillips (2009a, 2009b, 2016), Kim and Kim (2012), Gao and Phillips (2013a), Duffy (2016, 2020), Dong and Linton (2018) and Wang et al. (2021) for nonparametric and semiparametric approaches that can cope with the unknown functional form of the response in a nonstationary time series setting; together with the references cited therein. Nonlinear cointegrating regression with functional-coefficients was introduced in Cai et al. (2009) and Xiao (2009), where the authors suggested a model of the form:

$$y_t = x_t^T \beta_0(z_t) + \epsilon_t, \tag{1.1}$$

where y_t, z_t and ϵ_t are all scalars, $x_t = (x_{t1}, \cdots, x_{td})^T$ is of dimension d, $\beta_0(\cdot)$ is a $d \times 1$ vector of unknown smooth function defined on $\mathbb{R}$ and A^T denotes the transpose of a vector or a matrix A. Extension of the model (1.1) to more general formulations can be found in Gao and Phillips (2013b), Li et al. (2017), Hirukawa and Sakudo (2018) and Tu and Wang (2019, 2022). Also see Phillips and Wang (2022a, 2022b) for recent developments.

Model (1.1) allows cointegrating relationships that vary or evolve smoothly over time. This framework seems particularly useful in empirical applications where there may be structural evolution in a relationship over time. Practical examples in forecasting regional long-run energy and electricity demand can be found in Chang et al. (2016, 2021). Asymptotic theory of estimation and inference for model (1.1) and more general related models has been established in the literature. Technical difficulties, however, have confined much of the asymptotic theory to the case of sequential exogeneity where a martingale structure is usually assumed in the models (see, for instance, Cai et al., 2009; Gao & Phillips, 2013a, 2013b; Hirukawa & Sakudo, 2018; Li et al., 2017; Tu & Wang, 2019, 2022; Xiao, 2009). Exogeneity is a natural starting point for a pure cointegrated system and provides some useful insight into the properties of various estimates of nonlinear long-run linkages between the system variables. But the assumption is restrictive, especially in a cointegrated framework where the explanatory variables may be expected to be temporally and contemporaneously correlated.[1] Exogeneity therefore limits potential applications as well as removing a central technical difficulty in the development of the asymptotics.

The aim of this chapter is to remove the exogeneity restriction on the nonstationary regressor. Generalizing earlier models of Cai et al. (2009), Xiao (2009) and others, our framework allows for a wider class of regressors and temporal dependence properties within the system, particularly, we may have $E(\epsilon_t \mid x_t, z_t) \neq 0$, thereby introducing the endogeneity in model (1.1). Another contribution of this chapter is to address the technical difficulties. Unlike the papers cited above, our methodology in investigating the asymptotics builds up the techniques recently developed in Wang and Phillips (2009b, 2016), enabling our assumptions to be neat and our proofs to be quite straightforward.

The rest of this chapter is organized as follows. In Section 2, we investigate the asymptotics for local kernel and local linear nonparametric estimators of $\beta_0(\cdot)$ in model (1.1). This chapter considers two different situations:

1. x_t is nonstationary and z_t is stationary; and
2. x_t is stationary and z_t is nonstationary.

In both situations, we allow for $E(\epsilon_t \mid x_t, z_t) \neq 0$, thereby introducing the endogeneity in model (1.1) and providing an essential extension to previous works in the related fields. It should be mentioned that model (1.1) has been extensively investigated in literature in case that both x_t and z_t are stationary. We only refer to Cai et al. (2000), Fan and Zhang (2008) and citations therein. It is of great interest to consider the asymptotics for the nonparametric estimators of $\beta_0(\cdot)$ in model (1.1) when both x_t and z_t are $I(1)$ processes, but there are technical challenges at the moment. See Remark 8 for more details. We conclude in Section 3. The proofs of main results are given in Appendix A. The proofs of some auxiliary results are collected in Appendix B.

Throughout the chapter, we make use of the following notation: for $a = (a_{ij}), 1 \leq i \leq m, 1 \leq j \leq k, \| a \| = \sum_{i=1}^{m} \sum_{j=1}^{k} |a_{ij}|$.

2. MAIN RESULTS

The local kernel estimator of $\beta_0(z)$ in model (1.1) is given by

$$
\begin{aligned}
\hat{\beta}_n(z) &= \arg\min_{\beta} \sum_{t=1}^{n} \left(y_t - x_t^T \beta \right)^2 K\left(\frac{z_t - z}{h} \right) \\
&= \left[\sum_{t=1}^{n} x_t x_t^T K\left(\frac{z_t - z}{h} \right) \right]^{-1} \sum_{t=1}^{n} x_t y_t K\left(\frac{z_t - z}{h} \right),
\end{aligned}
$$

where $K(x)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. When one of the regressors x_t and z_t is nonstationary, the limit behavior of $\hat{\beta}_n(z)$ has been investigated in some special situations, notably where the error process ϵ_t is a martingale difference sequence and there is no contemporaneous correlation between (x_t, z_t) and ϵ_t (see Cai et al., 2009; Gao & Phillips, 2013a, 2013b; Li et al., 2017; Xiao, 2009, for instance).

This work provides more general results with advantages for empirical applications. Our assumptions permit dependence between the error process ϵ_t and the regressors x_t or/and z_t. These relaxations of the conditions in previous works are particularly important in nonlinear cointegrated systems because finite time horizon independence between the regressor and the equation error will often be restrictive in practice.

We further consider the local linear estimator $\hat{\beta}_L(z)$ of $\beta_0(z)$ (e.g., Fan & Gijbels, 1996) defined by

$$\begin{pmatrix} \hat{\beta}_L(z) \\ \hat{\beta}_L'(z) \end{pmatrix} = \arg\min_{\beta,\beta_1} \sum_{t=1}^{n} \{y_t - x_t^T[\beta + \beta_1(z_t - z)]\}^2 K\left(\frac{z_t - z}{h}\right).$$

Namely, we have

$$\hat{\beta}_L(z) = \left[\sum_{t=1}^{n} w_t x_t x_t^T K\left(\frac{z_t - z}{h}\right)\right]^{-1} \sum_{t=1}^{n} w_t x_t y_t K\left(\frac{z_t - z}{h}\right),$$

where $w_t = V_{n2} - (z_t - z)V_{n1}$ and $V_{nj} = \sum_{i=1}^{n} x_t x_t^T K\left(\frac{z_t - z}{h}\right)(z_t - z)^j$ for $j = 0, 1$ and 2.

The asymptotics of $\hat{\beta}_n(z)$ and $\hat{\beta}_L(z)$ will be investigated in two different cases mentioned in the Introduction section. Since the conditions set on x_t, z_t and ϵ_t are quite different, we consider their theoretical results in Sections 2.1 and 2.2 separately. In Section 2.3, we discuss an extension of the model to multivariate settings.

2.1. Model With Nonstationary x_t and Stationary z_t

This section makes use of the following assumptions in the asymptotic development.

A1

(i) $\{z_t, \epsilon_t, \eta_t\}_{t\geq 1}$ (where $\eta_t = x_t - x_{t-1}$) is a strict stationary α-mixing process of $d + 2$ dimension with $E\eta_t = 0$ and mixing coefficients $\alpha(n) = O(n^{-\gamma})$, where $\gamma > 0$ is specified later.
(ii) $E(\epsilon_1 \mid z_1) = 0$, $E(|\epsilon_1|^3 \mid z_1 = z)$ is bounded and (z_1, ϵ_1) has a joint density function $p(x, y)$ so that $p(x, y)$ is continuous in a neighbourhood of z.
(iii) z_1 has a density function $g(x)$ which is continuous in a neighbourhood of z.
(iv) $\Omega = \lim_{n\to\infty} \frac{1}{n} E x_n x_n^T > 0$ and $E\|\eta_1\|^3 < \infty$.

A2

(i) $K(x)$ is a nonnegative real function having a compact support and $\int_{-\infty}^{\infty} K(x)dx = 1$;

(ii) $\int_{-\infty}^{\infty} xK(x)dx = 0$·

A3 In a neighbourhood of z, for some $\nu > 0$,

(i) $\| \beta_0(y+z) - \beta_0(z) - \beta'_0(z)y \| \leq C_z \mid y \mid^{1+\nu}$.

(ii) $\| \beta_0(y+z) - \beta_0(z) - \beta'_0(z)y - \frac{1}{2}\beta''(z)y^2 \| \leq C_z \mid y \mid^{2+\nu}$,

where C_z is a constant depending only on z.

Conditions **A2** and **A3** are standard in literature (see, for instance, Cai et al., 2000, 2009). The smoothness condition on $\beta_0(x)$ in **A3** (ii) is stronger than that of **A3** (i), which is required to provide a better bias term in the local linear estimator $\hat{\beta}_L(z)$. Under **A1** (i), we have $x_t = \sum_{j=1}^{t} \eta_j$, that is, x_t is a standard *I*(1) process. It is possible to allow for the x_t to be a nearly *I*(1) process. Such an extension is omitted since it will involve complicated calculations. As in the situation where both x_t and z_t are stationary, the conditional mean $E(\epsilon_1 \mid z_1) = 0$ in **A1** (ii) is necessary to establish the consistency for both estimators $\hat{\beta}_n(z)$ and $\hat{\beta}_L(z)$. However, under **A1**, we may have $E(\epsilon_t \mid z_t, x_t) \neq 0$, which introduces endogeneity in the model. This differs from the previous work (e.g., Cai et al., 2009; Xiao, 2009) where the model is often assumed to have a martingale structure. The other conditions in **A1** are standard. It should be mentioned that, when the bandwidth parameter h converges to 0 satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, the condition **A1** with $\gamma \geq \max\{21/2, 6/\delta\}$, together with **A2** (i), implies that

$$\left(\frac{x_{[nt]}}{\sqrt{n}}, \frac{1}{\sqrt{nh}} \sum_{t=1}^{[nt]} K[(z_t - z)/h]\epsilon_t \right) \Rightarrow \left(B_t, \sigma_z B_{1t} \right), \tag{2.1}$$

on $D_{R^2}[0,1]$, where

$$\sigma_z^2 = E(\epsilon_1^2 \mid z_1 = z) \int_{-\infty}^{\infty} K^2(x)dx,$$

$B_1 = \{B_{1t}\}_{t \geq 0}$ is a standard Brownian motion independent of $B = \{B_t\}_{t \geq 0}$, and B is a *d*-dimensional Brownian motion with covariance matrix $\Omega = \lim_{n\to\infty} \frac{1}{n} Ex_n x_n^T$.

Result (2.1) is vital to establish the asymptotics of $\hat{\beta}_n(z)$ and $\hat{\beta}_L(z)$ in our technical development and the condition that $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$ is essential to enable the independence between two Brownian motions B_1 and B. To establish (2.1), there is a tradeoff condition (i.e., $\gamma \geq \max\{21/2, 6/\delta\}$) between the mixing coefficient in **A1** (i) and the convergence rate for the bandwidth parameter h. It is natural for such a joint convergence although it might be slightly stronger than necessary. For a proof of (2.1), see Lemma B.1 in Appendix B.

Let I_d be a *d*-dimensional identity matrix and $z \in R$ be a fixed constant. The next is our first result.

Theorem 2.1. *Under **A1**, **A2** (i) and **A3** (i), for any h satisfying* $nh^{3/2} = O(1)$ *and* $nh^{1+\delta} \to \infty$ *for some* $\delta > 0$, *if* γ, *which is defined in **A1** (i), satisfies that* $\gamma > \max\{21/2, 6/\delta\}$, *then*

$$\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h]\right)^{1/2} \left(\hat{\beta}_n(z) - \beta_0(z) - c_1\beta_0{}'(z)h\right) \to_D \sigma_z \mathbb{N}, \tag{2.2}$$

where $c_1 = \int_{-\infty}^{\infty} xK(x)dx$ *and* $\mathbb{N} \sim N(0, I_d)$ *is a standard d-dimensional normal vector.*

Remark 1. From the Proof of Theorem 2.1, we have also established the following result:

$$nh^{1/2}\left(\hat{\beta}_n(z) - \beta_0(z) - c_1\beta_0{}'(z)h\right) \to_D \tau_1\left(\int_0^1 B_s B_s^T ds\right)^{-1/2} \mathbb{N}, \tag{2.3}$$

where $\tau_1^2 = g^{-1}(z)\sigma_z^2$ and $\mathbb{N}$ is independent of $B = \{B_s\}_{s\geq 0}$. As expected in nonparametric cointegrating regression, due to the nonstationarity of the regressor x_t, the convergence rate $n\sqrt{h}$ in (2.3) is faster than $\sqrt{nh}$ comparing to the conventional functional-coefficient estimators in stationary time series regression (e.g., Cai et al., 2000).

Remark 2. In applications, one may choose $c_1 = 0$ (take $K(x)$ to be a symmetric kernel, for instance) or the bandwidth h satisfying $nh^{3/2} = o(1)$ so that the term $c_1\beta_0{}'(z)h$ disappears. Consequently, the self-normalized limit (2.2) is pivotal and well-suited to inference and confidence interval construction upon estimation of $E(\epsilon_1^2 \mid z_1 = z)$, which can be constructed by

$$\hat{\sigma}_z^2 = \frac{\sum_{t=1}^{n}\left[y_t - x_t^T\hat{\beta}_n(z_t)\right]^2 K[(z_t - z)/h]}{\sum_{t=1}^{n} K[(z_t - z)/h]}$$

Remark 3. Results (2.2) and (2.3) provide a first-order bias $c_1\beta_0{}'(z)h$. Surprisingly, it is unrealistic to add a deterministic higher order bias term into the result even if we have more smoothness conditions on $\beta_0(z)$. To see this claim, let $\tilde{K}(x) = xK(x)$, $c_1 = \int_{-\infty}^{\infty} xK(x)dx = 0$ and

$$\Lambda_n = \frac{h^{1/2}}{n}\sum_{t=1}^{n} x_t x_t^T \tilde{K}[(z_t - z)/h].$$

In order to add a deterministic bias term having a order $O(h^2)$ in result (2.3), from the Proof of Theorem 2.1 in Section A.1, we have to show that the bandwidth condition that $nh^{3/2} = O(1)$ can be reduced to $nh^{5/2} = O(1)$ and, as $nh^{5/2} = O(1)$,

$$\Lambda_n - c_0 nh^{5/2} = o_P(1), \tag{2.4}$$

for some constant c_0 (c_0 is allowed to be zero). This seems to be impossible except when $\tilde{K}(x) \equiv 0$. Indeed, by letting $d = 1$ and $x_t = \sum_{j=1}^{t} u_j$, where $u_t \sim$ i.i.d. $N(0, 1)$ and x_t is independent of $z_t \sim$ i.i.d. $N(0, 1)$, it is readily seen that

$$E\Lambda_n^2 \geq E\{\tilde{K}[(z_1 - z)/h] - E\tilde{K}[(z_1 - z)/h]\}^2 \frac{h}{n^2}\sum_{t=1}^{n} Ex_t^4$$
$$= [1+o(1)]\int_{-\infty}^{\infty} \tilde{K}^2(x)dx\, nh^2,$$

that is, $\sqrt{n}h / \Lambda_n = O_P(1)$, indicating that (2.4) is impossible whenever $nh^{5/2} = O(1)$. The similar phenomena has also be noticed in Sun and Li (2011).

It is possible to reduce the bias in local linear estimator $\hat{\beta}_L(z)$, as indicated in the following theorem.

Theorem 2.2. *Under **A1–A2** and **A3** (ii), for any h satisfying* $nh^{5/2} = O(1)$ *and* $nh^{1+\delta} \to \infty$ *for some* $\delta > 0$, *if* γ, *which is defined in **A1** (i), satisfies that* $\gamma > \max\{21/2, 6/\delta\}$, *then*

$$\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h]\right)^{1/2} \left(\hat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) \to_D \sigma_z \mathbb{N}, \tag{2.5}$$

where $c_2 = \dfrac{1}{2}\displaystyle\int_{-\infty}^{\infty} x^2 K(x)dx$.

Remark 4. As in Remark 2, the self-normalized limit (2.5) is pivotal upon estimation of $E(\epsilon_1^2 \mid z_1 = z)$. Theorem 2.2 also indicates that the local linear estimator is always better in reducing the bias when z_t is stationary in a functional-coefficient cointegrating regression model. In a related paper, under more restrictive conditions (in particular, without consideration of endogeneity), Theorem 2.1 of Cai et al. (2009) established a similar version of (2.3):

$$nh^{1/2}\left(\hat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) \to_D \tau_1\left(\int_0^1 B_s B_s^T ds\right)^{-1/2}\mathbb{N}, \tag{2.6}$$

where τ_1 is given in Theorem 2.1.

2.2. Model With Stationary x_t and Nonstationary z_t

In this section, let $\eta_i \equiv (\nu_i, \eta_{1i}, ..., \eta_{mi})^T, i \in Z, m \geq 1$ be a sequence of i.i.d. random vectors with $E\eta_0 = 0$, $E\left(\eta_0\eta_0^T\right) = \Sigma$ and $E\|\eta_0\|^4 < \infty$. We further make use of the following assumptions in the asymptotic development.

A4

(i) $\xi_j, j \geq 1$, is a linear process defined by $\xi_j = \sum_{k=0}^{\infty} \phi_k \nu_{j-k}$, where the coefficients $\phi_k, k \geq 0$, satisfy one of the following conditions:
   (a) **LM.** $\phi_k \sim k^{-\mu}\rho(k)$ where $1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞.
   (b) **SM.** $\sum_{k=0}^{\infty}|\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty}\phi_k \neq 0$;

(ii) $z_k = (1 - c/n)z_{k-1} + \xi_k$, where $z_0 = 0$ and $c \geq 0$ is a constant;

(iv) $E\nu_1^2 = 1$ and $\lim_{t|\to\infty} |t|^\lambda |Ee^{it\nu_1}| < \infty$ for some $\lambda > 0$.

A5

(i) $\begin{pmatrix} \epsilon_j \\ x_j \end{pmatrix} = \sum_{k=0}^{\infty} \psi_k \eta_{j-k}$, *where the coefficient matrix,*

$$\psi_k = \begin{pmatrix} \psi_k^{(1)} \\ \psi_k^{(2)} \\ \cdot \\ \psi_k^{(d+1)} \end{pmatrix} \quad \text{with} \quad \psi_k^{(s)} = (\psi_{k,s1}, \psi_{k,s2}, \ldots, \psi_{k,s(m+1)}),$$

satisfies $\sum_{k=0}^{\infty} k^{1/4} \| \psi_k^{(s)} \| < \infty$, for each $1 \le s \le d+1$;
(ii) $E(x_1 \epsilon_1) = 0$ and $Ex_1 x_1^T > 0$, that is, $Ex_1 x_1^T$ is a positive-definite matrix.

A6 In addition to **A2**, $\int_{\infty}^{\infty} |\hat{K}(x)| \, dx < \infty$, *where* $\hat{K}(x) = \int_{-\infty}^{\infty} e^{ixt} K(t) dt.$

Assumption **A4** (i) allows for short (under **SM**) and long (under **LM**) memory innovations ξ_j driving the (near) integrated regressor z_t given in Assumption **A4** (ii). As noticed in Wang and Phillips (2016), this setting is quite general for empirical applications. Set $d_n^2 = \mathbb{E}\left(\sum_{k=1}^{n} \xi_k\right)^2$, $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^{\infty} x^{-\mu}(x+1)^{-\mu} dx$ and denote by $W_\beta(t)$ a fractional Brownian motion with Hurst parameter $0 < \beta < 1$. It is well-known that

$$d_n^2 \sim \begin{cases} c_\mu n^{3-2\mu} \rho^2(n), & \text{under } \mathbf{LM}, \\ \phi^2 n, & \text{under } \mathbf{SM}, \end{cases}$$

and on $D[0,1]$ the following weak convergence applies (e.g., Chapter 2.1.3 of Wang (2015))

$$z_{\lfloor nt \rfloor} / d_n \Rightarrow Z(t) := \psi(t) + c \int_0^t e^{-c(t-s)} \psi(s) ds, \tag{2.7}$$

where

$$\psi(t) = \begin{cases} W_{3/2-\mu}(t), & \text{under } \mathbf{LM} \\ W(t), & \text{under } \mathbf{SM} \end{cases}$$

and $W = W_{1/2}$ is Brownian motion. Furthermore, the limit process $Z(t)$ has continuous local time process $L_Z(t,s)$ [2] with dual (time and space) parameters (t, s) in $[0,\infty) \times \mathbb{R}$. The characteristic function condition $\lim_{|t|\to\infty} |t|^\lambda | Ee^{it\nu_1} | < \infty$ for some $\lambda > 0$ is not necessary for the establishment of (2.7), but it is required for the convergence to local time in Lemma A.2 in Appendix 1. These notations are used throughout the rest of this chapter without further explanation.

Assumption **A5** (i) allows (ϵ_t, x_t) to be cross correlated with z_s for all $s \le t$, thereby inducing endogeneity and giving the structural model more natural temporal dependence properties than those used in previous works (e.g., Cai et al., 2009; Gao & Phillips, 2013b). We may have $\text{cov}(\epsilon_t, z_t) \neq 0$ under Assumption **A5** (i), which differs from the previous work where the model is often assumed to form a martingale difference sequence structure. In the latter case, $E(\epsilon_t \mid x_t, z_t) = 0$. Assumption **A5** (ii) is necessary to establish the consistency for both estimators $\hat{\beta}_n(z)$ and $\hat{\beta}_L(z)$. These quantities are well-defined due to $E \mid \eta_0 \mid^4 < \infty$. We further have $E \parallel x_1 x_1^T \parallel e_1^2 < \infty$, which is required in the following main result. Assumption **A6**, which is the same as in Wang and Phillips (2009b), is quite weak, and are easily verified for various kernels $K(x)$.

Let z be a fixed constant in R. We have the following main result in this section.

Theorem 2.3. *Under **A4–A6** and **A3** (ii), for any h satisfying $nh^5 / d_n = O(1)$ and $nh / d_n \to \infty$, we have*

$$\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h] \right)^{1/2} \left(\hat{\beta}_n(z) - \beta_0(z) - c_2 \beta_0''(z) h^2 \right) \to_D \sigma \mathbb{N}, \qquad (2.8)$$

where $c_2 = \frac{1}{2} \int_{-\infty}^{\infty} x^2 K(x) dx$, $\sigma^2 = [E x_1 x_1^T]^{-1} E(\epsilon_1^2 x_1 x_1^T) \int_{-\infty}^{\infty} K^2(x) dx$ and $\mathbb{N} \sim N(0, I_d)$ is a standard d-dimensional normal vector. Result (2.8) also holds if we replace $\hat{\beta}_n(z)$ by $\hat{\beta}_L(z)$.

Remark 5. In comparison with Theorem 2.2 where the result is derived under stationary z_t, (2.8) has a similar structure but with different co-variance σ^2, indicating the limit distributions of $\hat{\beta}_n(z)$, also for $\hat{\beta}_L(z)$, is not mutually independent. As in Theorem 2.2, the self-normalized limit (2.8) is pivotal upon estimation of σ^2, which can be constructed by

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} x_t x_t^T \left[y_t - x_t^T \hat{\beta}_n(z_t) \right]^2 K^2[(z_t - z)/h]}{\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h]}.$$

We may establish (result also holds if we replace $\hat{\beta}_n(z)$ by $\hat{\beta}_L(z)$)

$$\left(\frac{nh}{d_n} \right)^{1/2} \left(\hat{\beta}_n(z) - \beta_0(z) - c_2 \beta_0''(z) h^2 \right) \to_D \tau_2 L_Z^{-1/2}(1,0) \mathbb{N}, \qquad (2.9)$$

where $\tau_2^2 = [E x_1 x_1^T]^{-1} \sigma^2$ and $\mathbb{N}$ is independent of $L_Z(1,0)$. Note that (2.9) is quite different from (2.3) or (2.6), indicating that quite different techniques are used in establishing the results. Result (2.9) has a slow convergence rate due to the fact that, in the nonstationary case, the amount of time spent by the process z_t around any particular spatial point z is n/d_n rather than n so that the corresponding convergence rate in such a regression is now $\sqrt{nh/d_n}$. This was first explained in Wang and Phillips (2009a, 2009b). Furthermore, unlike Theorem 2.2, the bias reducing advantage of the local linear nonparametric estimator is lost under point-wise estimation as first noticed in Wang and Phillips (2011). In contrast to point-wise estimation, the local linear nonparametric estimator does have superior performance characteristics to the Nadaraya–Watson estimator in terms of uniform asymptotics over wide domains (Chan & Wang, 2014; Duffy, 2017).

Remark 6. Let $x_{1t} = x_t + A_0$, where A_0 is a constant vector. Note that

$$Ex_{11}\epsilon_1 = Ex_1\epsilon_1 + A_0 E\epsilon_1 = 0.$$

A routine modification of Theorem 2.3 yields that result (2.8) still holds if we replace x_t by x_{1t} and σ^2 by σ_1^2 defined by

$$\sigma_1^2 = (A_0 A_0^T + Ex_1 x_1^T)^{-1} E\left[\epsilon_1^2 (x_1 + A_0)(x_1 + A_0)^T\right] \int_{-\infty}^{\infty} K^2(x)dx.$$

This fact indicates that Theorem 2.3 provides a natural extension of Wang and Phillips (2009b, 2016) to a functional-coefficient cointegrating regression model. As noted in Wang and Phillips (2009b), there is no inverse problem in structural models of nonlinear cointegration of the form (1.1) where the regressor z_t is an endogenously generated integrated process, avoiding the need for instrumentation and completely eliminating ill-posed functional equation inversions. As a consequence, Theorem 2.3 has important implications for applications.

2.3. Multivariate Extension

In economic applications, it is important to consider multivariate extension of model (1.1), that is, to consider the model having the form:

$$y_t = x_t^T \beta_0(z_t, w_t) + \epsilon_t, \tag{2.10}$$

where y_t, z_t and ϵ_t are all scalars, $x_t = (x_{t1}, \cdots, x_{td})^T$ and $w_t = (w_{t1}, ..., w_{td_1})$ are of dimension d and d_1, respectively, and $\beta_0(\cdot, \cdots)$ is a $d \times 1$ vector of unknown smooth functions defined on $\mathbb{R}^{1+d_1}$. As in the one-dimension situation, the local kernel estimator $\hat{\beta}_0(\cdot,\cdot)$ of $\beta_0(\cdot,\cdot)$ can be similarly defined by

$$\hat{\beta}_0(z,w) = \frac{\sum_{t=1}^n x_t y_t K[(z_t - z)/h] \Pi_{j=1}^{d_1} L_j[(w_{tj} - w_j)/h_j]}{\sum_{t=1}^n x_t x_t^T K[(z_t - z)/h] \Pi_{j=1}^{d_1} L_j[(w_{tj} - w_j)/h_j]}, \tag{2.11}$$

where $K(x), L_j(x)$ are nonnegative kernel functions and the bandwidth $h, h_j \equiv h_{jn} \to 0$ for $j = 1, ..., d_1$.

If x_t is nonstationary and both z_t and w_t are stationary, asymptotics of $\hat{\beta}_0(z,w)$ can be obtained by using similar arguments as in Section 2.1 under some regular settings and hence the details are omitted. We next consider the situation that x_t is stationary and z_t is an $I(1)$ process. As noticed in Section 5.1.5 of Wang (2015), to enable $\hat{\beta}_0(z,w)$ being a consistent estimator, it is also essential to assume that w_t is stationary. We further assume $d_1 = 1$ for the sake of notation convenience. The extension to $d_1 \geq 2$ is straightforward.

To investigate the asymptotics of $\hat{\beta}_0(z,w)$, as in Section 2.2, let $\eta_i \equiv (\nu_i, \eta_{1i}, ..., \eta_{mi})^T, i \in Z, m \geq 1$ is a sequence of i.i.d. random vectors with

$E\eta_0 = 0$, $E\left(\eta_0\eta_0^T\right) = \Sigma$ and $E \| \eta_0 \|^4 < \infty$. We also make use of the following assumptions.

A7

(i) The kernels $K(x)$ and $L_1(x)$ have a common compact support satisfying $\int_{-\infty}^{\infty} K(x)dx = \int_{-\infty}^{\infty} L_1(x)dx = 1$ and $\int_{-\infty}^{\infty} |\hat{K}(t)|\, dt < \infty$, where $\hat{K}(t) = \int_{-\infty}^{\infty} e^{itx} K(x)dx$;
(ii) When (x, y) is in a compact set, we have

$$|\, \beta_0(x+\delta_1, y+\delta_2) - \beta_0(x,y)\,| \le C(|\,\delta_1\,| + |\,\delta_2\,|), \tag{2.12}$$

where $C > 0$ is an absolute constant, whenever δ_1 and δ_2 are sufficiently small.

A8

(i) z_t is defined as in ***A4*** (ii) and ***A4*** holds.
(ii) $x_t = (x_{t1}, \cdots, x_{td})^T$, where $x_{ti} = \Gamma_i(\eta_t, \ldots, \eta_{t-m_0+1})$ for some $m_0 > 0$, where $\Gamma_i(.), 1 \le i \le d$, are real measurable functions of its contents.
(iii) $w_t = \Gamma_0(\eta_t, \ldots, \eta_{t-m_0+1})$ for some $m_0 > 0$, where $\Gamma_0(.)$ is a real measurable function of its contents.
(iv) For any fixed w and each $1 \le i \le d$, $E(|\, x_{ti}\,|^{4+\delta}|\, w_t = w) < \infty$ with $t = m_0$ for some $\delta > 0$.
(v) For any fixed w and each $1 \le i, j \le d$, (x_{ti}, x_{tj}, w_t) and (x_{ti}, w_t) have joint density functions $p_{ij}(x, y, z)$ and $p_j(x, z)$, respectively, that are continuous in a neighbourhood of w.
(vi) For any fixed w, $D_w = \left(d_{ij}(w)\right)_{1 \le i, j \le d}$ is a positive-definite matrix, where $d_{ij}(w) = E(x_{m_0 i} x_{m_0 j} \,|\, w_{m_0} = w)$.

A9 $\{\epsilon_i, \mathcal{F}_{i+1}\}_{i \ge 1}$, where $\mathcal{F}_i = \sigma(\eta_i, \eta_{i-1}, \ldots)$, is a martingale difference sequence such that, as $i \to \infty$, $E(\epsilon_i^2 \,|\, \mathcal{F}_{i-1}) \to_{\text{a.s.}} \sigma^2 > 0$, and, as $A \to \infty$,

$$\sup_{i \ge 1} E\left[\epsilon_i^2 \mathrm{I}(|\,\epsilon_i\,| \ge A) \,|\, \mathcal{F}_{i-1}\right] = o_{\mathrm{P}}(1).$$

Theorem 2.4. *Under Assumptions **A7–A9**, for any h and h_1 satisfying $nhh_1/d_n \to \infty$ and $(h+h_1)^2\, nhh_1 / d_n \to 0$, we have*

$$D_n^{1/2}\left[\hat{\beta}_0(z,w) - \beta_0(z,w)\right] \to_D \tau\mathbb{N}, \tag{2.13}$$

where $D_n = \sum_{t=1}^n x_t x_t^T K[(z_t - z)/h] L_1[(w_t - w)/h_1]$ *and* $\tau^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(x)dx \int_{-\infty}^{\infty} L_1^2(x)dx$, $\mathbb{N} \sim N(0, I_d)$ *is a standard d-dimensional normal vector.*

Remark 7. Under Assumption ***A9***, we have $E(\epsilon_t \,|\, x_t, z_t, w_t) = 0$. Unlike the asymptotics developed in Theorem 2.3, such a martingale difference restriction seems hard to be reduced in the multivariate extension to model (1.1) due to the technical reasons. This also indicates that a complicated calculation will be involved in investigating the asymptotics of $\hat{\beta}_0(z,w)$, even when the regressor w_t in model (2.10) is stationary.

Remark 8. Theorem 2.4 provides an extension of Theorem 5.7 in Wang (2015) to a functional-coefficient cointegrating regression model. In a related paper, Gao and Phillips (2013b) (also see Sun et al., 2013) investigated model (2.10) in the case that both x_t and z_t are $I(1)$ processes under some similar conditions. Their main theorems made use of a result established by Phillips (2009), where the independence between x_t and (z_t, w_t) is essentially required. In terms of possible empirical applications, it is of interest to remove these restrictions, in particular, to establish the asymptotics without imposing $E(\epsilon_t \mid x_t, z_t, w_t) = 0$ as those given in Theorems 2.1–2.3.

3. CONCLUSION

This chapter studies nonparametric estimation for functional-coefficient cointegrating regression models of the form (1.1) in two different situations: (1) x_t is nonstationary and z_t is stationary; and (2) x_t is stationary and z_t is nonstationary. Both self-normalized local kernel and local linear estimators are shown to be asymptotic normal and to be pivotal upon an estimation of co-variances, indicating that the model (1.1) may be estimated by kernel regression just as in the case that both x_t and z_t are stationary. Importantly, and in contrast to stationary nonparametric regression, our asymptotic results allow for endogenous regressor in the models, namely, we assume $E(\epsilon_t \mid x_t, z_t) \neq 0$ in (1.1). As explained in Wang and Phillips (2009b), the reason for this robustness to endogeneity in the regressor is that nonstationary regressors such as unit root processes have a wandering character that assists in tracing out the true regression function. These structural models differ from various previous works and open up some interesting possibilities for functional-coefficient regression in empirical research with integrated processes. In terms of many possible empirical applications, some extensions of the ideas presented here to other useful models involving nonlinear functions of integrated processes seems to be interesting. In particular, partial linear cointegration models (e.g., Gao & Phillips, 2013a) may be treated in a similar way to (1.1), but there are difficulties for multiple nonstationary regression models, due to the nonrecurrence of the limit processes in high dimensions (*cf*. Park & Phillips, 2001). It will also be of interest in exploring the functional-coefficient cointegration models by the use of instrumental variables in the present nonstationary context. We plan to report some of these extensions in later work. Finally, in both situations described above, we suggest that it is always better to use the local linear nonparametric estimator with symmetric kernel rather than the Nadaraya–Waston estimator in empirical applications, in terms of the bias deduction and uniform asymptotics over wide domains. Furthermore, to ensure our theoretical results work, a unit root pretest on x_t (or z_t) is essentially necessary.

NOTES

1. To give an example, let us consider the demand for money. It has often been suggested that endogeneity exists in this typical cointgerated system, where the Central Bank supplies base money on demand at its prevailing interest rate, and broad money is created by the banking system (see, for instance, King, 1994). A survey paper on the literature related to the demand for money across a range of industrial and developing countries can be found in Sriram (2000).

2. The local time process $L_G(t,s)$ of a stochastic process $G(x)$ is defined by (e.g., Chapter 2 of Wang, 2015).

$$L_G(t,s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t I\{|G(r) - s| \le \epsilon\} dr.$$

ACKNOWLEDGMENTS

The authors thank the co-editor J. I. Miller and the referee for constructive and insightful comments which lead to significant improvement of this chapter. Liang acknowledges research support from the National Natural Science Foundation of China (12071348). Wang acknowledges research support from the Australian Research Council.

REFERENCES

Aldous, D. J. (1989). Stopping times and tightness, II. *Annals of Probability*, *17*, 586–595.

Cai, Z., Fan, J., & Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Heart Association*, *95*, 941–956.

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, *148*, 101–113.

Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., & Park, J. (2016). Disentangling temporal patterns in elasticities: A functional coefficient panel analysis of electricity demand. *Energy Economics*, *60*, 232–243.

Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., & Park, J. (2021). Forecasting regional long-run energy demand: A functional coefficient panel approach. *Energy Economics*, *96*, 105–117.

Chang, Y., & Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, *114*(1), 73–106.

Chang, Y., Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, *4*, 1–36.

Chan, N., & Wang, Q. (2014). Uniform convergence for nonparametric estimators with non-stationary data. *Econometric Theory*, *30*, 1110–1133.

Chan, N., & Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, *185*(1), 182–195.

Davidson, J. (1994). *Stochastic limit theory*. Oxford University Press

Dong, C., & Linton, O. (2018). Additive nonparametric models with time variables and both stationary and nonstationary regressors. *Econometric Theory*, *34*(4), 754–789.

Duffy, J. A. (2016). A uniform law for convergence to the local times of linear fractional stable motions. *The Annals of Applied Probability*, *25*(1), 45–72.

Duffy, J. A. (2017). Uniform convergence rates over maximal domains in structural nonparametric cointegrating regression. *Econometric Theory*, *33*, 1387–1417.

Duffy, J. A. (2020). Asymptotic theory for kernel estimators under moderate deviations from a unit root, with an application to the asymptotic size of nonparametric tests. *Econometric Theory*, *36*, 559–582.

Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica*, *55*, 251–276.

Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.

Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, *1*, 179–195.

Gao, J., & Phillips, P. C. B. (2013a). Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*, *176*, 59–79.

Gao, J., & Phillips, P. C. B. (2013b). *Functional coefficient nonstationary regression*. Cowles Fundation Discussion Paper No. 1911. USA: Yale University.

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*, 121–130.

Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press.

Hirukawa, M., & Sakudo, M. (2018). Functional-coefficient cointegration models in the presence of deterministic trends. *Econometric Review*, *37*(5), 507–533.

Kim, C. S., & Kim, I. M. (2012). Partial parametric estimation for nonstationary nonlinear regressions. *Journal of Econometrics*, *167*, 448–457.

King, M. (1994). The transmission mechanism of monetary policy. *Bank of England Quarterly Bulletin*, 261–267. August.

Li, K., Li, D., Liang, Z., & Hsiao, C. (2017). Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Review*, *36*, 354–369.

Liang, H.-Y., Phillips, P. C. B., Wang, H., & Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications. *Econometric Theory*, *32*, 1349–1375.

McLeish, D. L. (1975) A maximal inequality and dependent strong laws. *Annals of Probability*, *3*, 829–839.

Park, J. Y., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*, 468–497.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, *69*, 117–161.

Phillips, P. C. B. (2009). Local limit theorem and spurious regressions. *Econometric Theory*, *15*, 269–298.

Phillips, P. C. B., & Wang, Y. (2022a). Functional coefficient panel modeling with communal smoothing covariates. *Journal of Econometrics*, *227*, 371–407.

Phillips, P. C. B., & Wang, Y. (2022b). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Econometric Theory*, *25*, 1466–1497.

Saikkonen, P. (1995). Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory*, *11*(5), 888–911.

Sriram, S. S. (2000). A survey of recent empirical money demand studies. *IMF Staff Papers*, *47*, 334–365.

Sun, Y., & Li, Q. (2011). Data-driven bandwidth selection for nonstationary semiparametric models. *Journal of Business and Economic Statistics*, *29*, 541–551.

Sun, Y., Cai, Z, & Li, Q. (2013). Semiparametric functional coefficient models with integrated covariaates. *Econometric Theory*, *29*, 659–672.

Terasvirta, T., Tjostheim, D., & Granger, C. W. J. (2011). *Modelling nonlinear economic time series*. Oxford University Press.

Tu, Y., & Wang, Y. (2019). Functional coefficient cointegration models subject to time–varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, *81*(6), 1401–1423.

Tu, Y., & Wang, Y. (2022). Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics*, *229*, 396–421.

Wang, Q. (2014). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, *30*, 509–535.

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression*. World Scientific.

Wang, Q. (2021). Least squares estimation for nonlinear regression models with heteroscedasticity, *Econometric Theory*, *37*(6), 1267–1289.

Wang, Q., & Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, *25*, 710–738.

Wang, Q., & Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression, *Econometrica*, *25*, 1901–1948.

Wang, Q., & Phillips, P. C. B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory*, *27*, 235–259.

Wang, Q., & Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, *32*, 359–401.

Wang, Q., Phillips, P. C. B., & Kasparis, I. (2021). Latent variable nonparametric cointegrating regression. *Econometric Theory*, *37*, 138–168.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, *152*, 81–92.

APPENDICES

A. PROOFS OF MAIN RESULTS

Since the methodology is different, the Proofs of Theorems 2.1–2.4 will be given in Sections A.1–A.3, respectively.

A1.1. Proofs of Theorems 2.1 and 2.2

We start with the some preliminaries. Write $x_{nt} = x_t/\sqrt{n}$, $K_j(x) = x^j K(x)$ and $\mu_j = \int_{-\infty}^{\infty} K_j(x)dx$ for $j \geq 0$. Other notation is the same as in previous sections except mentioned explicitly. Furthermore, we always assume **A1** and **A2** (i) hold in the following lemmas.

Lemma A.1. *(a) For any* $0 \leq \alpha \leq 3$, *we have*

$$EK[(z_1 - z)/h](1 + |\epsilon_1|^{\alpha}) \leq C_z h, \tag{A.1}$$

where C_z *is a constant depending only on z; (b) As* $h \to 0$, *we have*

$$h^{-1} EK_j\big[(z_1 - z)/h\big] = g(z)\mu_j + o(1), \quad j = 0,1,2. \tag{A.2}$$

The Proof of Lemma A.1 is routine, and hence the details are omitted. In the next lemma, suppose that $H(x)$ and $H_1(x)$ are locally bounded real functions on R^d and $H_1(x)$ satisfies the local Lipschitz condition, that is, for any $\| x \| + \| y \| \leq K$,

$$\big| H_1(x) - H_1(y) \big| \leq C_K \| x - y \|, \tag{A.3}$$

where C_K is a constant depending only on K.

Lemma A.2.

(i) *For any real function* $A_n(x, y)$,
(a) *we have*

$$\frac{1}{n}\sum_{t=1}^{n} H(x_{nt}) A_n(z_t, \epsilon_t) = O_P\big(E\,|\,A_n(z_1, \epsilon_1)\,|\big); \tag{A.4}$$

(b) *If* $EA_n(z_1, \epsilon_1) = 0$ *for each* $n \geq 1$, *then for any* $\alpha > 0$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} H_1(x_{nt}) A_n(z_t, \epsilon_t) = O_P\big\{\big[E\,|\,A_n(z_1, \epsilon_1)\,|^{2+\alpha}\big]^{1/(2+\alpha)}\big\}. \tag{A.5}$$

(ii) *For any* $h \to 0$ *satisfying* $nh^{1+\delta} \to \infty$ *for some* $\delta > 0$, *if* $\gamma \geq \max\{21/2, 6/\delta\}$ *where* γ *is given in **A1** (i), then*

$$\begin{gathered}\left\{\frac{1}{n}\sum_{t=1}^{n} H(x_{nt}), \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} H_1(x_{nt}) K[(z_t - z)/h]\epsilon_t\right\}, \\ \to_D \left\{\int_0^1 H(B_s)ds, a_1\Big(\int_0^1 H_1^2(B_s)ds\Big)^{1/2} N\right\}\end{gathered} \tag{A.6}$$

where $a_1^2 = g(z)\sigma_z^2$, *N is a standard normal variate independent of* B_s.

The Proof of Lemma A.2 will be given in Appendix B. Note that $|K_j(x)| \le CK(x)$, where $C > 0$ is an absolute constant, as $K(x)$ has a compact support. Result (A.4), together with (A.1), implies that, as $h \to 0$,

$$\frac{1}{nh}\sum_{t=1}^{n} \| x_{nt}x_{nt}^T \| | K_j\big[(z_t - z)/h\big]| = O_P(1), \quad j = 0,1,2. \tag{A.7}$$

Similarly, by using (A.2) and (A.5) with $A_n(z_t, \epsilon_t) = K_j\big[(z_t - z)/h\big] - EK_j\big[(z_t - z)/h\big]$, we have

$$\begin{aligned}
\Delta_{nj} &:= \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T K_j\big[(z_t - z)/h\big] \\
&= \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\, EK_j\big[(z_t - z)/h\big] + \frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\big[K_j\big[(z_t - z)/h\big] - EK_j\big[(z_t - z)/h\big]\big] \\
&= [g(z)\mu_j + o(1)]\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + O_P\Big(\big(nh^{1+\alpha/(2+\alpha)}\big)^{-1/2}\Big) \\
&= g(z)\mu_j \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1), \quad j = 0,1,2
\end{aligned} \tag{A.8}$$

by taking α sufficiently small so that $nh^{1+\alpha/(2+\alpha)} \ge nh^{1+\delta} \to \infty$. Furthermore, it follows from (A.6) that, for any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$,

$$\begin{aligned}
&\left\{\left(\frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T, \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t\right\}, \\
&\to_D \left\{\int_0^1 B_s B_s^T ds,\, a_1\Big(\int_0^1 B_s B_s^T ds\Big)^{1/2}\mathbb{N}\right\}
\end{aligned} \tag{A.9}$$

where $\mathbb{N} \sim N(0, I_d)$ is a d-dimensional normal vector independent of B_s with covariance I_d.

We are now ready to prove the main results.

Proof of Theorem 2.1. We may write

$$n\sqrt{h}\Big(\hat{\beta}_n(z) - \beta_0(z) - c_1\beta_0{}'(z)h\Big) = \Delta_{n0}^{-1}\Big(S_n + R_{n1} + R_{n2}\Big), \tag{A.10}$$

where

$$\begin{aligned}
S_n &= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K(z_t - z)\epsilon_t, \\
R_{n1} &= h^{-1/2}\sum_{t=1}^{n} x_{nt}x_{nt}^T K[(z_t - z)/h][\beta_0(z_t) - \beta_0(z) - \beta_0{}'(z)(z_t - z)], \\
R_{n2} &= h^{1/2}\beta_0{}'(z)\sum_{t=1}^{n} x_{nt}x_{nt}^T\Big(K_1[(z_t - z)/h] - c_1K[(z_t - z)/h]\Big).
\end{aligned}$$

A3 (i) and (A.7) imply that, for some $\nu > 0$,

$$\| R_{n1} \| \le C(1+|z|^{\beta})h^{1/2+\nu}\sum_{t=1}^{n}\| x_{nt}x_{nt}^{T} \| K[(z_t - z)/h] = O_P(nh^{3/2+\nu}) = o_P(1).$$

Write $A_n(z_t, \epsilon_t) = K_1[(z_t - z)/h] - c_1 K[(z_t - z)/h]$. Lemma A.1 implies that $h^{-1}EA_n(z_1,\epsilon_1) = o(1)$ and $E|A_n(z_1,\epsilon_1)|^{2+\alpha} = O(h)$. It is readily seen from (A.5) that

$$\| R_{n2} \| = o_P(1)nh^{3/2} + O_P(1)\sqrt{n}h^{1/2+1/(2+\alpha)} = o_P(1)$$

whenever $nh^{3/2} = O(1)$. Taking these estimates into (A.10), we get

$$\begin{aligned}
&\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h]\right)^{1/2} \left(\hat{\beta}_n(z) - \beta_0(z) - c_1\beta_0'(z)h\right) \\
&\quad = \Delta_{n0}^{-1/2}\left[S_n + o_P(1)\right] \to_D \sigma_z \mathbb{N},
\end{aligned}$$

due to (A.8)–(A.9) and the continuous mapping theorem. Theorem 2.1 is now proved.

Proof of Theorem 2.2. Similarly to the proof of $\hat{\beta}_n(z)$, we may write

$$n\sqrt{h}\left(\hat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) = \Delta_n^{-1}\left(P_n + T_{n1} + \beta_0''(z)nh^{5/2}T_{n2}\right), \tag{A.11}$$

where, by letting $v_t = K[(z_t - z)/h][\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z) - \frac{1}{2}\beta_0''(z)(z_t - z)^2]$,

$$\begin{aligned}
\Delta_n &= \frac{1}{nh}\sum_{t=1}^{n} w_t x_{nt}x_{nt}^T K[(z_t - z)/h], \\
P_n &= \frac{1}{\sqrt{nh}}\sum_{t=1}^{n} w_t x_{nt} K[(z_t - z)/h]\epsilon_t, \\
T_{n1} &= \frac{1}{\sqrt{h}}\sum_{t=1}^{n} w_t x_{nt}x_{nt}^T v_t, \\
T_{n2} &= \frac{1}{nh}\sum_{t=1}^{n} w_t x_{nt}x_{nt}^T \left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\right\},
\end{aligned}$$

where we have used the fact:

$$\sum_{t=1}^{n} w_t x_{nt}x_{nt}^T K[(z_t - z)/h](z_t - z) = 0. \tag{A.12}$$

Note that, as $h \to 0$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$,

$$n^{-2}h^{1-j}V_{nj} = \Delta_{nj} = g(z)\mu_j \frac{1}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1), \quad j = 0,1,2,$$

by (A.8). It is readily seen from (A.8) and Lemma A.2 that, by recalling $\mu_1 = 0$ and $\mu_0 = 1$,

$$
\begin{aligned}
\Delta_n &= V_{n2}\Delta_{n0} - hV_{n1}\Delta_{n1} = V_{n2}\left[\frac{g(z)}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1)\right];\\
P_n &= V_{n2}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t - hV_{n1}\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K_1[(z_t - z)/h]\epsilon_t\\
&= V_{n2}\left[\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t + o_P(1)\right];\\
\|T_{n1}\| &\le Ch^{3/2+\delta}\sum_{t=1}^{n}|w_t|\,\| x_{nt}x_{nt}^T\|K[(z_t - z)/h]\\
&\le Ch^{3/2+\delta}\left(|V_{n2}|\sum_{t=1}^{n}\| x_{nt}x_{nt}^T\|K[(z_t - z)/h] + h|V_{n1}|\sum_{t=1}^{n}\| x_{nt}x_{nt}^T\|K_1[(z_t - z)/h]\right)\\
&= O_P(nh^{5/2+\delta})V_{n2};\\
T_{n2} &= V_{n2}\frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2K[(z_t - z)/h]\right\}\\
&\quad -hV_{n1}\frac{1}{nh}\sum_{t=1}^{n} x_{nt}x_{nt}^T\left\{\frac{1}{2}K_3[(z_t - z)/h] - c_2K_1[(z_t - z)/h]\right\}\\
&= o_P(1)V_{n2}.
\end{aligned}
$$

Taking these facts into (A.11), we obtain

$$
\begin{aligned}
&\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t - z)/h]\right)^{1/2}(\hat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2)\\
= &\left[\frac{g(z)}{n}\sum_{t=1}^{n} x_{nt}x_{nt}^T + o_P(1)\right]^{-1/2}\\
&\left[\frac{1}{\sqrt{nh}}\sum_{t=1}^{n} x_{nt}K[(z_t - z)/h]\epsilon_t + o_P(1)nh^{5/2}\right]\\
\to_D &\quad \sigma_z\mathbb{N}
\end{aligned}
$$

as $nh^{5/2} = O(1)$ and $nh^{1+\delta} \to \infty$ for some $\delta > 0$, due to (A.9) and the continuous mapping theorem. Theorem 2.2 is now proved.

A.2 Proof of Theorem 2.3

As in Section A.1, let $K_j(x) = x^j K(x)$ and $\mu_j = \int_{-\infty}^{\infty} K_j(x)dx$ for $j \ge 0$. Let

$$u_k = \sum_{l,m=0}^{\infty} \varphi_l^T \eta_{k-l} \eta_{k-m}^T \tilde{\varphi}_m, \tag{A.13}$$

where coefficient constants φ_l and $\tilde{\varphi}_m$ are the $d+1$ dimensional vectors satisfying $\sum_{l=0}^{\infty} l^{1/4} \|\varphi_l\| < \infty$ and $\sum_{m=0}^{\infty} m^{1/4} \|\tilde{\varphi}_m\| < \infty$. We start with the following lemmas. Under the conditions **A4** and **A6**, the Proof of Lemma A.3 is similar to Lemma 2.2 and Theorem 3.16 of Wang (2015). An outline will be given in Appendix B. For a Proof of Lemma A.4, we refer to Theorem 3.18 of Wang (2015) (see, also, Wang & Phillips, 2011).

Lemma A.3. *Let z be a fixed constant. For any $1 \le s,t \le d+1$ and any h satisfying $h \log n \to 0$ and $nh/d_n \to \infty$, we have*

$$\sum_{k=1}^{n} (1+|u_k|) K_j\left(\frac{z_k - z}{h}\right) = O_P\left(nh/d_n\right); \tag{A.14}$$

$$\sum_{k=1}^{n} \left(u_k - Eu_k\right) K_j\left(\frac{z_k - z}{h}\right) = O_P\left(\left(\frac{nh}{d_n}\right)\right)^{1/2} \sum_{l,m=0}^{\infty} \left(l^{1/4} m^{1/4} \|\varphi_l\| \|\tilde{\varphi}_m\|\right), \tag{A.15}$$

$j = 0$, 1, 2, and

$$\left\{\frac{d_n}{nh} \sum_{t=1}^{n} K\left(\frac{z_t - z}{h}\right), \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} \left(u_k - Eu_k\right) K\left(\frac{z_t - z}{h}\right)\right\}, \tag{A.16}$$
$$\to_D \left\{L_Z(1,0), a_2 L_Z^{1/2}(1,0) N\right\}$$

where $a_2^2 = E(u_1^2) \int_{-\infty}^{\infty} K^2(t)dt$, and N is standard normal variate independent of $L_Z(1,0)$;

Lemma A.4. *Let $f(x)$ be a real function having a compact support. If $\int_{-\infty}^{\infty} f(x)dx = 0$, then*

$$\sum_{k=1}^{n} f\left(\frac{z_k - z}{h}\right) = O_P\left(\left(nh/d_n\right)^{1/2}\right), \tag{A.17}$$

for any h satisfying $nh/d_n \to \infty$.

Since, due to Assumption **A5**, each element of $x_t x_t^T$ and $x_t \epsilon_t$ can be representated as u_k for some specified φ_l and $\tilde{\varphi}_m$, it follows from Lemma A.3 that

$$
\begin{aligned}
D_{nj} := & \frac{d_n}{nh}\sum_{t=1}^{n} x_t x_t^T K_j\left(\frac{z_t - z}{h}\right) \\
= & \frac{d_n}{nh}\sum_{t=1}^{n} E\left(x_t x_t^T\right) K_j\left(\frac{z_t - z}{h}\right) + \frac{d_n}{nh}\sum_{t=1}^{n}\left[x_t x_t^T - E\left(x_t x_t^T\right)\right] K_j\left(\frac{z_t - z}{h}\right), \\
= & E\left(x_1 x_1^T\right)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\left(\frac{z_t - z}{h}\right) + O_P\left(\left(\frac{d_n}{nh}\right)^{1/2}\right) \\
= & E\left(x_1 x_1^T\right)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\left(\frac{z_t - z}{h}\right) + o_P(1), \quad j = 0,1,2
\end{aligned}
\tag{A.18}
$$

as $nh/d_n \to \infty$. Furthermore, due to $E(x_1\epsilon_1) = 0$, it follows from (A.16) and the continuous mapping theorem that

$$
\begin{aligned}
& \left\{\frac{d_n}{nh}\sum_{t=1}^{n} K\left(\frac{z_t - z}{h}\right), \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} x_t \epsilon_t K\left(\frac{z_t - z}{h}\right)\right\}, \\
& \to_D \left\{L_Z(1,0), a_2 L_Z^{1/2}(1,0)\mathbb{N}\right\}
\end{aligned}
\tag{A.19}
$$

where $a_2^2 = E(\epsilon_1^2 x_1 x_1^T)\int_{-\infty}^{\infty} K^2(x)dx$, and $\mathbb{N} \sim N(0, I_d)$ is a d-dimensional normal vector independent of $L_Z(1,0)$.

We are now ready to prove Theorems 2.3. By letting $v_t = K\left(\frac{z_t - z}{h}\right)$ $\left[\beta_0(z_t) - \beta_0(z) - \beta_0'(z)(z_t - z) - \frac{1}{2}\beta_0''(z)(z_t - z)^2\right]$, we may write

$$
\left(\frac{nh}{d_n}\right)^{1/2}\left(\hat{\beta}_n(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) = D_{n0}^{-1}\left(S_n + R_{n1} + \beta_0'(z)R_{n2} + \beta_0''(z)R_{n3}\right), \tag{A.20}
$$

where

$$
\begin{aligned}
S_n &= \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} x_t \epsilon_t K\left(\frac{z_t - z}{h}\right), \\
R_{n1} &= \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} x_t x_t^T v_t, \\
R_{n2} &= h\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} x_t x_t^T K_1\left(\frac{z_t - z}{h}\right), \\
R_{n3} &= h^2\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n} x_t x_t^T \left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\right\}.
\end{aligned}
$$

From **A3** (ii) and (A.14), as $nh^5/d_n = O(1)$ we have

$$\|R_{n1}\| \le Ch^{2+\eta}\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\|x_t x_t^T\| K\left(\frac{z_t - z}{h}\right) = O_P\left(\left(\frac{nh}{d_n}\right)^{1/2} h^{2+\eta}\right) = o_P(1). \quad \text{(A.21)}$$

From (A.15) with $j = 1$ and (A.17) with $f(x) = K_1(x)$, as $h \to 0$ we have

$$\begin{aligned} R_{n2} &= h\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\left[x_t x_t^T - E\left(x_t x_t^T\right)\right]K_1\left(\frac{z_t - z}{h}\right) \\ &\quad + h\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}E\left(x_t x_t^T\right)K_1\left(\frac{z_t - z}{h}\right). \qquad \text{(A.22)} \\ &= O_P(h) = o_P(1) \end{aligned}$$

Note that $\int_{-\infty}^{\infty}\left[\frac{1}{2}K_2(x) - c_2 K(x)\right]dx = 0$. It follows from (A.15) and (A.17) again that, as $nh^5/d_n = O(1)$,

$$\begin{aligned} R_{n3} &= h^2\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\left[x_t x_t^T - E\left(x_t x_t^T\right)\right]\left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\right\} \\ &\quad + \left(\frac{nh^5}{d_n}\right)^{1/2}\frac{d_n}{nh}\sum_{t=1}^{n}E\left(x_t x_t^T\right)\left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\right\} \qquad \text{(A.23)} \\ &= O_P(h^2) + O_p\left(\left(nh^5/d_n\right)^{1/2}\right)o_P(1) = o_P(1) \end{aligned}$$

Combining (A.18) and (A.20)–(A.23), we obtain

$$\begin{aligned} &\left(\sum_{t=1}^{n}x_t x_t^T K[(z_t - z)/h]\right)^{1/2}\left(\hat{\beta}_n(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) \\ &= \left[E\left(x_1 x_1^T\right)\frac{d_n}{nh}\sum_{t=1}^{n}K\left(\frac{z_t - z}{h}\right) + o_P(1)\right]^{-1/2}\left[\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}x_t\epsilon_t K\left(\frac{z_t - z}{h}\right) + o_P(1)\right] \to_D \sigma\mathbb{N}, \end{aligned}$$

due to (A.19) and the continuous mapping theorem. This proves (2.8).

We next prove that (2.8) still holds if $\hat{\beta}_n(z)$ is replaced by $\hat{\beta}_L(z)$. In fact, as in the Proof of Theorem 2.2, we may write

$$\left(\frac{nh}{d_n}\right)^{1/2}\left(\hat{\beta}_L(z) - \beta_0(z) - c_2\beta_0''(z)h^2\right) = D_n^{-1}\left(P_n + T_{n1} + \beta_0''(z)T_{n2}\right), \quad \text{(A.24)}$$

by virtue of (A.12), where

$$D_n = \frac{d_n}{nh}\sum_{t=1}^{n} w_t x_t x_t^T K\left(\frac{z_t - z}{h}\right),$$

$$P_n = \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} w_t x_t \epsilon_t K\left(\frac{z_t - z}{h}\right),$$

$$T_{n1} = \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} w_t x_t x_t^T v_t,$$

$$T_{n2} = h^2\left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} w_t x_t x_t^T \left\{\frac{1}{2}K_2[(z_t - z)/h] - c_2 K[(z_t - z)/h]\right\}.$$

Noting that from (A.18), it can be obtained

$$\frac{d_n}{nh} h^{-j} V_{nj} = D_{nj} = E\left(x_1 x_1^T\right)\frac{d_n}{nh}\sum_{t=1}^{n} K_j\left(\frac{z_t - z}{h}\right) + o_P(1).$$

Since $\mu_1 = 0$ and $\mu_0 = 1$, from Lemmas A.3 and A.4 we have

$$\begin{aligned}
D_n &= V_{n2} D_{n0} - h V_n 1 D_{n1} = V_{n2}[D_{n0} + o_P(1)]; \\
P_n &= V_{n2}\left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} x_t \epsilon_t K\left(\frac{z_t - z}{h}\right) - h V_{n1}\left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} x_t \epsilon_t K_1\left(\frac{z_t - z}{h}\right) \\
&= V_{n2}\left[\left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} x_t \epsilon_t K\left(\frac{z_t - z}{h}\right) + o_P(1)\right]; \\
\|T_{n1}\| &\le V_{n2}\|R_{n1}\| + h\,|V_{n1}|\sum_{t=1}^{n} x_t x_t^T\,|v_t(z_t - z)| \qquad . \\
&= O_P\left[\left(\frac{nh}{d_n}\right)^{1/2} h^{2+\eta}\right] V_{n2}; \\
T_{n2} &= V_{n2} R_{n3} - V_{n1} h^3 \left(\frac{d_n}{nh}\right)^{1/2} \sum_{t=1}^{n} x_t x_t^T \left\{\frac{1}{2}K_3[(z_t - z)/h] - c_2 K_1[(z_t - z)/h]\right\} \\
&= o_P(1) V_{n2}
\end{aligned}$$

Taking these facts into (A.24), the claim follows from

$$\begin{aligned}
&\left(\sum_{t=1}^{n} x_t x_t^T K[(z_t-z)/h]\right)^{1/2}\left(\hat{\beta}_L(z)-\beta_0(z)-c_2\beta_0''(z)h^2\right)\\
=\ &\left[E\left(x_1x_1^T\right)\frac{d_n}{nh}\sum_{t=1}^{n}K\left(\frac{z_t-z}{h}\right)+o_P(1)\right]^{-1/2}\left[\left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}x_t\epsilon_t K\left(\frac{z_t-z}{h}\right)+o_P(1)\right],\\
\to_D\ &\sigma\mathbb{N}
\end{aligned}$$

due to (A.19) and the continuous mapping theorem. Theorem 2.3 is now proved. ■

A.3 Proof of Theorem 2.4

Let $V_t = x_t K[(z_t-z)/h]L_1[(w_t-w)/h_1]$. We may write

$$D_n^{1/2}\left[\hat{\beta}_0(z,w)-\beta_0(z,w)\right]=D_n^{-1/2}S_n+R_n. \tag{A.25}$$

where $S_n=\sum_{t=1}^{n}\epsilon_t V_t$ and, by (2.12),

$$\begin{aligned}
\|R_n\| &= \sum_{t=1}^{n}\left|\beta_0(z_t,w_t)-\beta_0(z,w)\right|\left\|D_n^{-1/2}V_t\right\|\\
&\leq C(|h_1|+|h_2|)\sum_{t=1}^{n}\left\|D_n^{-1/2}V_t\right\|
\end{aligned}.$$

By the continuous mapping theorem, result (2.13) will follow if we prove

$$\|R_n\|=o_P(1), \tag{A.26}$$

and for any $A=(A_1,\ldots,A_d)^T\in R^d$,

$$\begin{aligned}
&\left\{\frac{d_n}{nhh_1}A^T D_n A,\ \left(\frac{d_n}{nhh_1}\right)^{1/2}A^T S_n\right\}\\
\to_D\ &\left\{(A^T D_w A)L_Z(1,0),\ \tau^{1/2}(A^T D_w A)NL_Z^{1/2}(1,0)\right\}
\end{aligned}, \tag{A.27}$$

where $L_Z(1,0)$ is given as in Section 2.2 and $N\sim N(0,1)$ is independent of $L_Z(1,0)$. We start with some preliminaries. Set $\Delta_t=\sum_{k,j=1}^{d}A_kA_jx_{tk}x_{tj}L_1[(w_t-w)/h_1]$. Since, by **A8**(d) and some standard arguments,

$$\begin{aligned}
Ex_{ti}x_{tj}L_1^{\gamma}[(w_t-w)/h_1] &= h_1E(x_{ti}x_{tj}\mid w_t=w)\int_{\Omega}L_1^{\gamma}(x)dx+o(h_1),\\
E|x_{ti}|^{\beta}L_1^{\gamma}[(w_t-w)/h_1] &= h_1E(|x_{ti}|^{\beta}\mid w_t=w)\int_{\Omega}L_1^{\gamma}(x)dx+o(h_1)=O(h_1),
\end{aligned}$$

for any $\gamma > 0, 0 \le \beta \le 4+\delta$ and uniformly for all $t \ge m_0$, we have $E\Delta_t = h_1[A^T DA + o(1)]$ and $E\Delta_t^2 = O(h_1)$. Now it follows from Lemma 2.2 (ii) of Wang (2015) that, for any $A = (A_1, \ldots, A_d)^T \in R^d$,

$$\begin{aligned}\frac{d_n}{nhh_1} A^T D_n A &= \frac{d_n}{nhh_1}\sum_{t=1}^n K[(z_t - z)/h]E\Delta_t + \frac{d_n}{nhh_1}\sum_{t=1}^n K[(z_t - z)/h](\Delta_t - E\Delta_t) \\ &= \left[A^T D_w A + o(1)\right]\frac{d_n}{nh}\sum_{t=1}^n K[(z_t - z)/h] + O_P\left[\left(\frac{d_n}{nhh_1}\right)^{1/2}\right], \\ &= \left[A^T D_w A + o(1)\right]\frac{d_n}{nh}\sum_{t=1}^n K[(z_t - z)/h] + o_P(1)\end{aligned} \tag{A.28}$$

due to $nhh_1 / d_n \to \infty$. Similarly, we have

$$\begin{aligned}\frac{d_n}{nhh_1}\sum_{t=1}^n \left(A^T V_t\right)^2 \;&=\; \left[A^T D_w A + o(1)\right]\int_\Omega L_1^2(x)dx \frac{d_n}{nh}\sum_{t=1}^n K^2[(z_t - z)/h] \\ &\quad + o_P(1)\end{aligned}, \tag{A.29}$$

and

$$\begin{aligned}\sum_{t=1}^n \left(\| V_t \| + \| V_t \|^{2+\delta/2}\right) &= \sum_{t=1}^n \left(\| x_t \| + \| x_t \|^{2+\delta}\right) K[(z_t - z)/h] L_1[(w_t - w)/h_1] \\ &\le O(h_1)\sum_{t=1}^n K[(z_t - z)/h] + O_P\left[(nhh_1/d_n)^{1/2}\right], \\ &= O_P(nhh_1/d_n)\end{aligned} \tag{A.30}$$

where we have used the fact that $\sum_{t=1}^n K[(z_t - z)/h] = O_P(nh/d_n)$. By virtue of Theorem 2.21 of Wang (2015), results (A.28)–(A.29) imply that

$$\begin{aligned}&\left\{\frac{\sum_{j=1}^{[nt]}\nu_j}{\sqrt{n}}, \frac{\sum_{j=1}^{[nt]}\nu_{-j}}{\sqrt{n}}, \frac{d_n}{nhh_1} A^T D_n A, \frac{d_n}{nhh_1}\sum_{t=1}^n (A^T V_t)^2\right\}, \\ \Rightarrow \;& \left\{B_t, B_{-t}, (A^T D_w A) L_Z(1,0), \tau_1 (A^T D_w A) L_Z(1,0)\right\}\end{aligned} \tag{A.31}$$

on $D_{R^4}[0,\infty)$, where $B = \{B_t\}_{t\in R}$ is a standard Brown motion and

$$\tau_1 = \int_\Omega L_1^2(x)dx \int_\Omega K^2(x)dx.$$

We are now ready to prove (A.26) and (A.27). By noting that D_w is positive definite, it is readily seen from (A.28) that $D_n^{-1} = O_P(d_n / nhh_1)$. This, together with (A.30), yields that

$$\| R_n \| = (|h| + |h_1|) O_P\left[(d_n / nhh_1)^{1/2}\right]\sum_{t=1}^n \|V_t\| = O_P\left[(|h| + |h_1|)(nhh_1/d_n)^{1/2}\right] = o_P(1),$$

implying (A.26).

To prove (A.27), write $u_{nt} = \left(\frac{d_n}{nhh_1}\right)^{1/2} A^T V_t$, namely, we have $\left(\frac{d_n}{nhh_1}\right)^{1/2} A^T S_n = \sum_{t=1}^{n} \epsilon_t u_{nt}$. By using (A.28), routine calculations show that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} |u_{nt}| = o_P(1)$

and

$$\max_{1\le t\le n} |u_{nt}| \le \left(\frac{d_n}{nhh_1}\right)^{1+\delta/4} \sum_{t=1}^{n} |A^T V_t|^{2+\delta/2} = o_P(1).$$

Now, by recalling **A8** and (A.31), (A.27) follows from Wang's extended martingale limit theorem, for example, Wang (2014) or Theorem 3.14 of Wang (2015). The Proof of Theorem 2.4 is complete.

B. PROOFS OF AUXILIARY RESULTS

Throughout this section, we denote an absolute positive constant by C, which may be different at each appearance. A sequence $\{\xi_k, k \ge 1\}$ is said to be α mixing if the α mixing coefficient

$$\alpha(n) := \sup_{k\ge 1} \sup\left\{ | P(A\cap B) - P(A)P(B) | : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^{k} \right\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m$ denoted the $\sigma-$ algebra generated by $\xi_l, \ldots, \xi_m$ with $l \le m$.

The following results for the moment properties of α-mixing sequence are well-known (e.g., McLeish, 1975 or Hall & Heyde, 1980, p. 278), which will be used in the proofs of other results.

Suppose $X \in \mathcal{F}_k^{\infty}$ and $Y \in \mathcal{F}_{-\infty}^{i}$, where $k > i$. Then,

(a) for any $1 \le p \le r \le \infty$,

$$\left\| E(X \mid \mathcal{F}_{-\infty}^{i}) - EX \right\|_p \le 2(2^{1/p} + 1)\{\alpha(k-i)\}^{1/p-1/r} \|X\|_r; \tag{B.1}$$

(b) for any $p, q > 1$, $p^{-1} + q^{-1} < 1$,

$$| EXY - EXEY | \le 8\|X\|_p \|Y\|_q \{\alpha(k-i)\}^{1-p^{-1}-q^{-1}}. \tag{B.2}$$

Lemma B.1. *For any $h \to 0$ satisfying $nh^{1+\delta} \to \infty$ for some $\delta > 0$, the condition **A1** with $\gamma \ge \max\{21/2, 6/\delta\}$, together with **A2** (i), implies (2.1).*

Proof. Write $A_k = K\left[(z_k - z)/h\right]\epsilon_k$, $W_{nk} = A_k/\sqrt{nh}$ and $R_n(t) = \sum_{k=1}^{[nt]} W_{nk}$. It is well-known (see, e.g., Davidson, 1994) that

$$\frac{x_{[nt]}}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \eta_k \Rightarrow B_t,$$

on $D_R[0,1]$, namely, $\left\{x_{[nt]}/\sqrt{n}\right\}_{n\geq 1}$ is tight. As a consequence, to prove (2.1), it suffices to show that

(i) the finite dimensional distributions of $\left(x_{[nt]}/\sqrt{n}, R_n(t)\right)$ converges to those of $(B_t, \sigma_z B_{1t})$;
(ii) $\{R_n(t)\}_{n\geq 1}$ is tight.

The proof of the finite dimensional convergence is of somewhat standard. See, for instance, Cai et al. (2000) with some routine modification. The independence between B_t and B_{1t} comes from the fact that, for any $0 < t \leq 1$, the covariance of $x_{[nt]}/\sqrt{n}$ and $R_n(t)$ converges to zero in probability. Indeed, by using (B.2) and Lemma A.1, we have

$$\begin{aligned}
\left|\mathrm{Cov}\left(x_{[nt]}/\sqrt{n}, R_n(t)\right)\right| &\leq \frac{1}{n\sqrt{h}} \sum_{k=1}^{n} E\,|\,\eta_k A_k\,| + \frac{2}{n\sqrt{h}} \sum_{k=1}^{n} \sum_{j=0}^{n-k} |E(\eta_k A_{k+j})\,| \\
&\leq 8h^{-1/2} (E\,|\,A_1\,|^{7/4})^{4/7} (E\,|\,\eta_1\,|^3)^{1/3} \sum_{j=0}^{\infty} j^{-2\gamma/21} \qquad , \\
&\leq C h^{1/14} \to 0
\end{aligned}$$

for any $0 < t \leq 1$ and $\gamma > 21/2$. Simple calculations by using similar arguments (see, e.g., Lemma A1 (c) of Cai et al., 2000) also yield that

$$\sup_{n\geq 1} E R_n^2(t) \leq C t, \tag{B.3}$$

indicating that $\{R_n(t)\}_{n\geq 1}$, for any $0 < t \leq 1$, is uniformly integrable. This fact will be used later.

We next prove the tightness. To this end, let $\mathcal{F}_k = \sigma(z_i, \epsilon_i; i \leq k)$,

$$\beta_{nk} = \sum_{i=1}^{\infty} E(W_{n,i+k} \,|\, \mathcal{F}_k), \quad w_{nk} = \sum_{i=0}^{\infty} \left[E(W_{n,i+k} \,|\, \mathcal{F}_k) - E(W_{n,i+k} \,|\, \mathcal{F}_{k-1})\right]$$

It is well-known that β_{nk} and w_{nk} are well-defined and $W_{nk} = w_{nk} + \beta_{n,k-1} - \beta_{nk}$. Since $R_n(t) = \sum_{k=1}^{[nt]} w_{nk} + \beta_{n,[nt]}$, the tightness of $R_n(t)$ will follow if we prove that $\sum_{k=1}^{[nt]} w_{nk}$ is tight and

$$E \max_{1\leq k\leq n} |\,\beta_{nk}\,| = o_P(1). \tag{B.4}$$

Note that $\{w_{nk}, \mathcal{F}_k\}$ forms a sequence of martingale differences and the finite dimensional distribution converges to a joint normal distribution. To prove $\sum_{k=1}^{[nt]} w_{nk}$ is tight, it suffice to show that, for any $t > 0$, $\sum_{k=1}^{[nt]} w_{nk}$ is uniformly integrable (see, e.g., Proposition 1.2 of Aldous, 1989), which follows from (B.3), (B.4) and the fact $R_n(t) = \sum_{k=1}^{[nt]} w_{nk} + \beta_{n,[nt]}$ again.

It remains to prove (B.4). Note that $EA_1 = 0$ and $E|A_1|^r \le C_z h$ for any $1 \le r \le 3$ by (A.1). Standard arguments by using (B.1), together with $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 0$, show that, for any $1 \le p < 3$ and $0 < \alpha \le 3 - p$,

$$\left(E|E(A_{i+k}|\mathcal{F}_i)|^p\right)^{1/p} \le C\alpha(k)^{\alpha/p(p+\alpha)}\left(E|A_1|^{p+\alpha}\right)^{1/(p+\alpha)}$$

and

$$(E|\beta_{ni}|^p)^{1/p} \le (nh)^{-1/2}\sum_{k=1}^{\infty}\left(E|E(A_{i+k}|\mathcal{F}_i)|^p\right)^{1/p} \le C(nh)^{-1/2}h^{1/(p+\alpha)}\sum_{k=1}^{\infty}k^{-\gamma\alpha/p(p+\alpha)}. \quad \text{(B.5)}$$

This implies that, for any $2 < p < 3$, $0 < \alpha \le 3 - p$ and $\gamma\alpha/p(p+\alpha) > 1$,

$$E\max_{1\le k\le n}|\beta_{nk}| \le \left[\sum_{i=1}^{n}E|\beta_{ni}|^p\right]^{1/p} \le C\left(nh^{1+\alpha/[(p+\alpha)(p/2-1)]}\right)^{(1-p/2)/p}.$$

We have now established (B.4) by taking $\gamma > 6/\delta$, and α sufficiently small so that

$$nh^{1+\alpha/[(p+\alpha)(p/2-1)]} \ge nh^{1+\delta} \to \infty.$$

The Proof of Lemma B.1 is now complete.

Proof of Lemma A.2. We only prove (A.6). Due to the local boundedness of $H(x)$, (A.4) is obvious. The proof of (A.5) follows from similar arguments as in the proof of (A.6). We omit the details.

As in the Proof of Lemma B.1 for (2.1), let $A_i = K[(z_i - z)/h]\epsilon_i$, $\mathcal{F}_t = \sigma(\eta_{i+1}, z_i, \epsilon_i, 0 < i \le t)$, and $\mathcal{F}_s = \sigma(\phi, \Omega)$ be the trivial σ-field for $s \le 0$. By writing

$$u_i = \sum_{k=1}^{\infty}E(A_{i+k}|\mathcal{F}_i) \quad \text{and} \quad u_i = \sum_{k=0}^{\infty}\left[E(A_{i+k}|\mathcal{F}_i) - E(A_{i+k}|\mathcal{F}_{i-1})\right],$$

it is readily seen that $\{v_i, \mathcal{F}_i\}_{i\ge1}$ forms a sequence of martingale differences and, as in Liang et al. (2016),

$$\begin{aligned}
\sum_{k=1}^{n}H_1(x_{nk})A_k &= \sum_{k=1}^{n}H_1(x_{nk})(v_k + u_{k-1} - u_k) \\
&= \sum_{k=1}^{n}H_1(x_{nk})v_k + \sum_{k=1}^{n}H_1\left[H_1(x_{n,k+1}) - H_1(x_{n,k})\right]u_k - H_1(x_{n,n+1})u_n \qquad \text{(B.6)} \\
&= \sum_{k=1}^{n}H_1(x_{nk})v_k + R(n), \quad \text{say,}
\end{aligned}$$

where we recall $x_{n,t} = x_{nt} = x_t/\sqrt{n}$. As in the proof of (B.4), we have

$$\max_{1\le i\le n}|A_i - v_i| \le 2\max_{1\le i\le n}|u_i| = o_P\left(\sqrt{nh}\right) \quad \text{(B.7)}$$

This, together with (A.4) and Lemma B.1 (i.e., (2.1) holds), implies that

$$\left(x_{n,[nt]}, \frac{1}{\sqrt{nh}}\sum_{k=1}^{[nt]} v_k\right) = \left\{\frac{x_{[nt]}}{\sqrt{n}}, \frac{1}{\sqrt{nh}}\sum_{t=1}^{[nt]} K[(z_t - z)/h]\epsilon_t\right\} + o_P(1) \\ \Rightarrow \left\{B_t, \sigma_z B_{1t}\right\},$$

on $D_{R^2}[0,1]$. Now, by recalling that B_{1t} is independent of B_t, standard argument on the convergence to stochastic integrals yields that

$$\left(\frac{1}{n}\sum_{t=1}^{n} H(x_{nt}), \frac{1}{\sqrt{nh}}\sum_{k=1}^{n} H_1(x_{nk})v_k\right) \to_D \left\{\int_0^1 H(B_s)ds, \sigma_z\left(\int_0^1 H_1^2(B_s)ds\right)^{1/2} N\right\},$$

where $N \sim N(0,1)$ independent of B_t. Taking this estimation into (B.6), (A.6) will follow if we prove

$$|R(n)| = o_P\left(\sqrt{nh}\right). \tag{B.8}$$

To this end, write $\Omega_K = \{x_{ni} : \max_{1\le i\le n+1} |x_{ni}| \le K\}$. Note that (B.5) implies $E|u_j|^p \le Ch^{p/(p+\alpha)}$ for any $\alpha > 0, 1 \le p \le 3$ and for any $j \ge 1$. It follows from (A.3) and $E\|\eta_1\|^3 < \infty$ that

$$\begin{aligned} E\big[|R(n)|I(\Omega_K)\big] &\le C_K\left(\sum_{k=1}^{n} E\big(\|x_{n,k+1} - x_{n,k}\||u_k|\big) + E|u_n|\right) \\ &\le \frac{C_K}{\sqrt{n}}\sum_{k=1}^{n} E(\|\eta_k\||u_k|) + o(1) \\ &\le C_K\sqrt{n}\big(E\|\eta_1\|^3\big)^{1/3}(E|u_1|^{3/2})^{2/3} + o(1) \\ &\le C_K\sqrt{n}h^{2/(3+2\alpha)} + o(1) = o\left(\sqrt{nh}\right) \end{aligned},$$

by taking $\alpha < 1/2$. This implies that $R(n) = o_P\left(\sqrt{nh}\right)$ due to $P(\Omega_K) \to 1$ as $K \to \infty$.

The Proof of Lemma A.2 is complete.

Proof of Lemma A.3. We only provide an outline. Results (A.14) and (A.17) follow from (2.94) and Theorem 3.18 of Wang (2015), respectively. By using similar arguments as in proof of (2.96) in Wang (2015), for any $l, m \ge 0$, we have

$$\begin{aligned} &E\left(\sum_{k=1}^{n}\left[\eta_{k-l}\eta_{k-m}^T - E\left(\eta_{k-l}\eta_{k-m}^T\right)\right] K[(z_k - z)/h]\right)^2 \\ \le\ & C(1 + \max\{l^{1/2}, m^{1/2}\} + h\log n)\left[E\left(\eta_1\eta_1^T\right) + E\left(\eta_1\eta_2^T\right)\right]nh/d_n \end{aligned},$$

This, together with Hólder's inequality, yields that

$$
\begin{aligned}
& E\left|\sum_{k=1}^{n}(u_k - Eu_k)K\left(\frac{z_k - z}{h}\right)\right| \\
\leq\; & \sum_{l,m=0}^{\infty}\|\varphi_l\|\;\|\tilde{\varphi}_m\|E\left|\sum_{k=1}^{n}\left[\eta_{k-l}\eta_{k-m}^T - E\left(\eta_{k-l}\eta_{k-m}^T\right)\right]K[(z_k - z)/h]\right|, \\
=\; & O\left[\left(nh/d_n\right)^{1/2}\right]\sum_{l,m=0}^{\infty}\left(l^{1/4}m^{1/4}\|\varphi_l\|\;\|\tilde{\varphi}_m\|\right)
\end{aligned}
$$

implying (A.15). To see (A.16), let $u_{kM} = \sum_{l,m=0}^{M}\varphi_l^T\eta_{k-l}\eta_{k-m}^T\tilde{\varphi}_m$ and $\overline{u}_k = u_k - Eu_k$, $\overline{u}_{kM} = u_{kM} - Eu_{kM}$. For any $M \geq 1$, (3.8) of Wang and Phillips (2009b) implies that

$$
\begin{aligned}
&\left\{\frac{d_n}{nh}\sum_{t=1}^{n}K\left(\frac{z_t - z}{h}\right), \left(\frac{d_n}{nh}\right)^{1/2}\sum_{t=1}^{n}\overline{u}_{kM}K\left(\frac{z_t - z}{h}\right)\right\}, \\
&\to_D \left\{L_Z(1,0), a_{M2}L_Z^{1/2}(1,0)N\right\}
\end{aligned}
\tag{B.9}
$$

where $a_{M2}^2 = E(u_{kM}^2)\int_{-\infty}^{\infty}K^2(t)dt$. Since $a_{M2}^2 \to a_2^2$ as $M \to \infty$, (A.16) follows easily from (B.9) and the fact:

$$
\begin{aligned}
&\left(\frac{d_n}{nh}\right)^{1/2}\sum_{k=1}^{n}\left(\overline{u}_k - \overline{u}_{kM}\right)K\left(\frac{z_t - z}{h}\right) \\
&= O_P(1)\left(\sum_{l=M,m=0}^{\infty} + \sum_{l=0,m=M}^{\infty} l^{1/4}m^{1/4}\|\varphi_l\|\;\|\tilde{\varphi}_m\| = o_P(1)\right)
\end{aligned},
$$

as $n \to \infty$ first and then $M \to \infty$. The Proof of Lemma A.3 is now complete.

CHAPTER 6

A SPECIFICATION TEST BASED ON CONVOLUTION-TYPE DISTRIBUTION FUNCTION ESTIMATES FOR NON-LINEAR AUTOREGRESSIVE PROCESSES

Kun Ho Kim[a], Hira L. Koul[b] and Jiwoong Kim[c]

[a]*Concordia University, Montreal, QC, Canada*
[b]*Michigan State University, East Lansing, MI, USA*
[c]*University of South Florida, Tampa, FL, USA*

ABSTRACT

This chapter proposes a test for a parametric specification of the autoregressive function of a given stationary autoregressive time series. This test is based on the integrated square difference between the empirical distribution function estimate and a convolution-type distribution function estimate of the stationary distribution function obtained from the autoregressive residuals. Some asymptotic properties of the proposed convolution-type distribution function estimate are studied when the model's innovation density is unknown. These properties are in turn used to derive the asymptotic null distribution of the proposed test statistic. We also discuss some finite sample properties of the test statistic based on the block bootstrap methodology. A simulation study shows that the proposed test competes favorably with some existing tests in terms of the empirical level and power.

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 187–206

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A006

Keywords: Nonlinear autoregressive processes; integrated squared difference of the two d.f.'s; empirical and convolution d.f. estimators; block bootstrap; asymptotic power; empirical level and power;

JEL classifications: C12; C13; C14

1. INTRODUCTION

Consider the autoregressive model framework in time series:

$$X_i = m(X_{i-1}) + \varepsilon_i, \quad i \in \mathbb{Z} := \{0, \pm 1, \cdots\}, \tag{1.1}$$

where X_i is a real valued stationary and ergodic process, m is a measurable function defined on the real line $\mathbb{R}$ to $\mathbb{R}$ and $\varepsilon, \varepsilon_i$ are independent and identically distributed (i.i.d.) random variables (r.v.'s) having $E\varepsilon \equiv 0, 0 < E\varepsilon^2 < \infty$, and X_{i-1} is independent of ε_i, for each $i \in \mathbb{Z}$. Hence, $m(x) = E(X_1 \mid X_0 = x)$ is the conditional mean function. The framework (1.1) has been a popular benchmark in econometrics due to its simple but effective representation of time series in many applications.

Let q be a known positive integer, $\Theta \subset \mathbb{R}^q$, and $\{m_\vartheta(x), \vartheta \in \Theta, x \in \mathbb{R}\}$ a family of parametric functions. The problem of interest here is to test the hypothesis:

$$H_0 : m(x) = m_\theta(x), \quad \text{for some } \theta \in \Theta \quad \text{and for all } x \in \mathbb{R} \tag{1.2}$$

based on the data $X_i, 0 \le i \le n$.

Various tests for testing H_0 have been proposed in the time-series literature. Aït-Sahalia (1996) proposed a parametric specification test by comparing the non-parametric kernel density estimate of the marginal density of X_0 with its closed-form density estimate under the parametric form. Unfortunately, the test proposed in Aït-Sahalia (1996) is hampered by the slow consistency rate of the kernel density estimates, which is closely examined by Pritsker (1998). In an attempt to overcome this shortcoming, Corradi and Swanson (2005) propose a test that utilizes the empirical d.f. of X_i's and the closed-form d.f. estimate under the parametric forms of the mean function because, unlike the density estimate, the d.f. estimates achieve the root-n-consistency rate. However, the critical values of their test cannot be tabulated even for the large samples because it is the limiting distribution involves a functional of a Gaussian process with unknown covariance function. They employ the block bootstrap procedures (Künsch, 1989; Lahiri, 1992; 2003) that can capture the dependence structure of the process to perform their proposed test. Aït-Sahalia and Park (2012) analyze the asymptotic behavior of the specification test of Aït-Sahalia (1996) for the stationary density of a diffusion process, but when the diffusion is not stationary. They consider integrated and explosive processes, as well as nearly integrated ones in the spirit of the local to unity analysis in classical unit root theory. This chapter finds that the behavior of the test predicted by the asymptotic distribution under an integrated process provides a better approximation to the finite sample distribution of the test than that predicted by the asymptotic distribution under strict stationarity.

Despite the innovative nature of their ideas, the tests in Aït-Sahalia (1996) and Corradi and Swanson (2005) are *not applicable* if the closed-form density and closed-form d.f. of the process are not available. This significantly reduces the applicability of these tests because the closed-form density and d.f. are typically *unavailable* for many prominent linear and non-linear time-series models. To address this issue, Kim et al. (2015) introduce a convolution-type *density* estimate that is used to test for the framework (1.1), regardless of whether or not the closed-form density of X_i is available. They propose a test based on the maximal deviation of this convolution density estimate from the traditional kernel density estimate. However, because of the well-known slow consistency rate of the kernel density estimates, Kim et al. (2015) test leads to size distortion and a low power of the test for moderate sample sizes.

One way to address the issue with the Kim et al. (2015) test is to construct tests based on d.f.'s. More precisely, for any r.v. ξ, let F_ξ, f_ξ denote its d.f. and density, respectively. Let X, ε denote copies of X_0, ε_1, respectively, and let $Z := m(X)$. Because of the independence between X and ε in (1.1), we propose a test based on a statistic that compares an estimate for the d.f. F_X based on the convolution between F_Z and F_ε to the traditional empirical d.f. of $X_i, 1 \le n$. Since both the convolution and empirical d.f. estimates enjoy the $n^{1/2}$-consistency rate, which is faster than that of the kernel density estimate, our test based on these d.f. estimates is expected to perform better than that of the Kim et al. (2015) test in terms of preserving the level of significance in finite sample applications.

The organization of this chapter is the following: Section 2 introduces the convolution d.f. estimate of F_X and the test statistic based on the integrated square difference between this estimate and the empirical d.f. estimate of F_X. Section 3 provides the assumptions required for the asymptotics, technical lemmas, and the main theoretic result of the chapter. Throughout the chapter, we assume that the innovation density in *unknown*. Section 4 derives the asymptotic null distribution of the test statistic of Section 2. The proposed test is compared to competing benchmark tests in the literature. Section 5 provides a simulation study that investigates the finite sample properties of the proposed statistic based on the block bootstrap methodology. Section 6 concludes the chapter.

2. TEST STATISTICS

In this section, we combine the ideas of Corradi and Swanson (2005) and Kim et al. (2015) to propose a test statistic based on the empirical d.f. estimate and the convolution-type d.f. estimate of F_X. Throughout the rest of the chapter, $Z = m_\theta(X), Z_i = m_\theta(X_i)$, where θ is the true parameter value for which H_0 holds.

Let K be a density kernel, $G(y) = \int_{-\infty}^{y} K(x)dx$, $y \in \mathbb{R}$ and $b \equiv b_n$ be a bandwidth sequence. For $x \in \mathbb{R}, \vartheta \in \Theta$, define

$$F_n(x,\vartheta) := \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{x - X_i + m_\vartheta(X_{i-1})}{b}\right), \qquad f_n(x,\vartheta) := \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - m_\vartheta(X_{i-1})}{b}\right).$$

When the null hypothesis H_0 is true, then specifying the true parameter θ among all possible ϑ's will lead to $X_i - m_\vartheta(X_{i-1}) = X_i - m_\theta(X_{i-1}) = \varepsilon_i$. Consequently, we have

$$F_n(x,\theta) = \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{x-\varepsilon_i}{b}\right), \qquad f_n(x,\theta) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - m_\theta\left(X_{i-1}\right)}{b}\right).$$

A natural estimate of the d.f. F_X is the empirical d.f. defined by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \le x\right), \qquad x \in \mathbb{R}. \tag{2.1}$$

Alternately, we can estimate F_X in the following way. Let $\hat{\theta}$ be a $n^{1/2}$-consistent estimator of θ, under H_0. Let $\hat{\varepsilon}_i = X_i - m_{\hat{\theta}}(X_{i-1})$ and

$$\hat{F}_\varepsilon(x) := F_n(x,\hat{\theta}) \equiv \frac{1}{n}\sum_{i=1}^{n} G\left(\frac{x-\hat{\varepsilon}_i}{b}\right).$$

Next, suppose $Z := m_\theta(X)$ has Lebesgue density denoted by f_Z. Then, under H_0, an estimate of f_Z is given by

$$\hat{f}_Z(x) := f_n(x,\hat{\theta}) \equiv \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - m_{\hat{\theta}}\left(X_{i-1}\right)}{b}\right).$$

Given the independence between X_{i-1} and ε_i, the d.f. $F_X(x) = \int F_\varepsilon(x-z) f_Z(z)dz$ can also be estimated by the convolution d.f.

$$\hat{F}_c(x) = \int \hat{F}_\varepsilon(x-z)\hat{f}_Z(z)dz. \tag{2.2}$$

Because $X_i, i \in \mathbb{Z}$ is stationary and ergodic, by the Ergodic Theorem, the estimator $\hat{F}_n$ is uniformly consistent for the true d.f. F_X, regardless of whether H_0 is true or not, while, typically under H_0, $\hat{F}_c(x)$ will be uniformly consistent for F_X. Hence a suitably centered and scaled difference between the two d.f.'s $\hat{F}_n$ and $\hat{F}_c$ can be used as a test statistic for testing H_0. Let ψ be a density on $\mathbb{R}$. In this chapter, we propose a class of tests, one corresponding to each ψ, based on the statistics

$$\mathcal{V}_n := \int V_n^2(y)\psi(y)dy, \qquad V_n(y) := n^{1/2}\left(\hat{F}_n(y) - \hat{F}_c(y)\right), \quad y \in \mathbb{R}. \tag{2.3}$$

We first derive the asymptotic null distribution of $\mathcal{V}_n$ and then illustrate how the bootstrap method can be used to implement the test for moderate sample sizes.

3. ASSUMPTIONS AND SOME PRELIMINARIES

In this section, we shall describe the assumptions and some preliminary results needed for deriving the asymptotic null distribution of V_n. In the sequel, all limits are taken as $n \to \infty$, unless mentioned otherwise. For any Euclidean vector x and any smooth function ℓ from $\mathbb{R}$ to $\mathbb{R}$, let $\|x\|$ denote the Euclidean norm of x, $\|\ell\|_\infty := \sup_{x\in\mathbb{R}} |\ell(x)|$, and $\dot{\ell}$ and $\ddot{\ell}$ denote the first and second derivatives of ℓ, respectively. We are now ready to state the needed assumptions.

Assumption K. The kernel K is a Lebesgue density on $[-1,1]$, vanishes off $(-1,1)$, $\int uK(u)du = 0$ and K is differentiable, having a bounded derivative $\dot{K}$.

Assumption M. The process $X_i, i \in \mathbb{Z}$ is strictly stationary, ergodic, and strongly mixing with the mixing coefficient sequence α_i satisfying

$$\sum_{i=1}^{\infty} i^\delta \alpha_i < \infty, \quad \exists \delta > 0. \tag{3.1}$$

Assumption F1. F_ε has Lebesgue density f_ε that is bounded and twice differentiable with the two bounded derivatives $\dot{f}_\varepsilon, \ddot{f}_\varepsilon$.

Assumption F2. The r.v. Z takes values in a bounded set $\mathcal{B}$ and F_Z has Lebesgue density f_Z that is bounded from below on $\mathcal{B}$ and twice differentiable with the two derivatives $\dot{f}_Z, \ddot{f}_Z$ satisfying $\int |\dot{f}_Z(z)|\, dz + \int |\ddot{f}_Z(z)|\, dz < \infty, \|\ddot{f}_Z\|_\infty < \infty$.

Assumption T1. The estimator $\hat{\theta}$ of θ satisfying the following condition.

$$n^{1/2}\left\|\hat{\theta} - \theta\right\| = O_p(1), \qquad \text{under } H_0. \tag{3.2}$$

Assumption T2. The estimator $\hat{\theta}$ is asymptotically linear, that is,

$$n^{1/2}\left(\hat{\theta} - \theta\right) = n^{-1/2}\sum_{i=1}^{n} \varphi(\varepsilon_i, X_{i-1}, \theta) + o_p(1), \quad \text{under } H_0, \tag{3.3}$$

where $\varphi_i := \varphi(\varepsilon_i, X_{i-1}, \theta), i \geq 1$ are p-vectors of stationary and ergodic r.v.'s with $E\varphi_1 = 0, E\|\varphi_1\|^2 < \infty$ and such that $\lim_n n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(\varphi_i\varphi_j'\right)$ exists.

Assumption M1. There exists a vector of q functions $\dot{m}_\theta(x), \theta \in \Theta, x \in \mathbb{R}$ such that for every $0 < k < \infty$,

$$\max_{1\leq i\leq n, n^{1/2}\|\vartheta-\theta\|\leq k} n^{1/2} \left| m_\vartheta(X_{i-1}) - m_\theta(X_{i-1}) - (\vartheta - \theta)'\dot{m}_\theta(X_{i-1}) \right| = o_p(1), \tag{3.4}$$

$$E\|\dot{m}_\theta(X_0)\|^2 < \infty. \tag{3.5}$$

Assumption M2. $\sup_{x\neq x'} |m_\theta(x) - m_\theta(x')| / |x - x'| < 1$ and $E\left[|\epsilon_i|^p\right] < \infty$, for some $p > 0$.

Assumption b. $b \to 0, n^{1/2}b^3 \to 0$ and $n^{1/2}b^2 \to \infty$.

Assumption K implies that K is bounded. It is satisfied by many popular kernels such as Parzen, Epanechnikov and uniform kernels. Assumption M and (1.1) imply that the process $Z_i := m_\theta(X_i), i \in \mathbb{Z}$ is strongly mixing with the mixing sequence α_i, under H_0. Assumptions F1 and F2 are satisfied by numerous distributions. Assumptions T1 and T2 are important, intermediate assumptions which turn out to be useful for obtaining an approximation of the test statistic and its Gaussian process limit. Under some smoothness conditions on m_θ, it is satisfied by a class of M-estimators, see, for example, Koul (2002, Ch. 8). Clearly, Assumption T2 implies Assumption T1. Assumption M1 is a minimal smoothness assumption on the null model under which the asymptotic results of this chapter are obtained. It is obviously satisfied by $m_\theta(x) = \theta' h(x)$, provided $E\|h(X)\|^2 < \infty$, where $h(x) = (h_1(x), \cdots, h_p(x))'$ is a vector of p real valued functions. Assumption M2 represents a contraction condition. It ensures that the process $X_i, i \in \mathbb{Z}$ is stationary and ergodic, cf. Tong (1990). The Assumption b on the bandwidth sequence implies that $n^{1/2}b \to \infty$. It is in particular satisfied by any $b \propto n^{-1/5}$.

The following lemmas are useful in assessing the asymptotic behavior of $\hat{F}_n$ and $\hat{F}_c$ and in deriving the asymptotic null distribution of $\mathcal{V}_n$. Their proofs are deferred to Section 7.

Lemma 3.1. *Suppose Assumptions K, T1, M1, and b hold. Then the followings hold for every θ for which H_0 holds.*

$$\sup_{x\in\mathbb{R}} \left| F_n(x,\hat{\theta}) - F_n(x,\theta) \right| = O_p\left(\left(n^{1/2}b\right)\right)^{-1}, \tag{3.6}$$

$$\sup_{x\in\mathbb{R}} \left| F_n(x,\theta) - EF_n(x,\theta) \right| = O_p\left(n^{-1/2}\right), \tag{3.7}$$

$$\sup_{x\in\mathbb{R}} \left| F_n(x,\hat{\theta}) - F_\varepsilon(x) \right| = O_p\left(\left(n^{1/2}b\right)^{-1} + n^{-1/2}\right) + O\left(b^2\right) = O_p\left(\left(n^{1/2}b\right)^{-1}\right). \tag{3.8}$$

Lemma 3.2. *Suppose Assumptions K, M, F2, and b hold. Then, the following results hold.*

$$\int \left| f_n(z,\theta) - Ef_n(z,\theta) \right| dz = O_p\left(\left(n^{1/2}b\right)\right)^{-1}\Big), \tag{3.9}$$

$$\int \left| f_n(z,\hat{\theta}) - f_Z(z) \right| dz = O_p\left(\left(n^{1/2}b\right)^{-1}\right) + O_p\left(n^{-1/2}\right) + O_p\left(b^2\right) = O_p\left(\left(n^{1/2}b\right)^{-1}\right). \tag{3.10}$$

Lemma 3.3. *Assume f_ε, f_Z are twice differentiable with the second-order derivatives $\ddot{f}_\varepsilon, \ddot{f}_Z$ bounded. Then the followings hold*:

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n \left[\dot{f}_\varepsilon(x - m_{\hat{\theta}}(X_{i-1})) - \dot{f}_\varepsilon(x - m_{\theta_0}(X_{i-1})) \right] \right| = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{3.11}$$

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n \left[\dot{f}_Z(x - \hat{\varepsilon}_i) - \dot{f}_Z(x - \varepsilon_i) \right] \right| = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{3.12}$$

Lemma 3.4. *Assume f_ε, f_Z are differentiable with $E\left|\dot{f}_\varepsilon(x-Z)\right| + E\left|\dot{f}_Z(x-\varepsilon)\right| < \infty$, for all $x \in \mathbb{R}$. Then*

$$E\left\{\dot{f}_\varepsilon(x-Z) + \dot{f}_Z(x-\varepsilon)\right\} = 0, \qquad \forall x \in \mathbb{R}. \tag{3.13}$$

Moreover,

$$\sup_x \left| \frac{1}{n}\sum_{i=1}^n \left[\dot{f}_\varepsilon\left(x - m_{\hat\theta}\left(X_{i-1}\right)\right) + \dot{f}_Z\left(x - \hat\varepsilon_i\right)\right] \right| = O_P\left(n^{-1/2}\right). \tag{3.14}$$

Next, let

$$\begin{aligned}
&\hat{S}_{n1}(x) := n^{-1}\sum_{i=1}^n F_\varepsilon\left(x - m_{\hat\theta}(X_{i-1})\right), \quad S_{n1}(x) := n^{-1}\sum_{i=1}^n F_\varepsilon\left(x - m_\theta(X_{i-1})\right), \\
&\hat{S}_{n2}(x) := n^{-1}\sum_{i=1}^n F_Z\left(x - \hat\varepsilon_i\right), \qquad S_{n2}(x) := n^{-1}\sum_{i=1}^n F_Z\left(x - \varepsilon_i\right), \quad x \in \mathbb{R}.
\end{aligned} \tag{3.15}$$

The following proposition describes the asymptotic uniform linearity of $\hat{S}_{nj}, j = 1,2$ in the standardized estimate $n^{1/2}(\hat\theta - \theta)$. Its proof is given in Section 7.

Proposition 3.1. *Under the Assumptions T1, F1, F2, and M1, the followings hold.*

$$\sup_{x\in\mathbb{R}} \left| n^{1/2}\left(\hat{S}_{n1}(x) - S_{n1}(x)\right) - n^{1/2}\left(\hat\theta - \theta\right)' E\left[\dot{m}_\theta(X) f_\varepsilon(x - m_\theta(X)\right]\right| = o_p(1). \tag{3.26}$$

$$\sup_{x\in\mathbb{R}} \left| n^{1/2}\left(\hat{S}_{n2}(x) - S_{n2}(x)\right) - n^{1/2}\left(\hat\theta - \theta\right)' E\left[\dot{m}_\theta(X)\right] f_Z(x)\right| = o_p(1). \tag{3.17}$$

Let

$$\begin{aligned}
\mu(x) &:= E\left[\dot{m}_\theta(X)\left\{f_Z(x) + f_\varepsilon\left(x - Z\right)\right\}\right], \\
\hat{S}_n(x) &:= \hat{S}_{n1}(x) + \hat{S}_{n2}(x), \quad S_n(x) := S_{n1}(x) + S_{n2}(x).
\end{aligned}$$

Proposition 3.2. *Under the Assumptions K, T1, F1, F2, and M2, the following holds*:

$$\sup_{x\in\mathbb{R}} \left| \int n^{1/2}\hat{S}_n(x - ub)K(u)du - n^{1/2}S_n(x) - n^{1/2}\Delta_n'\mu(x)\right| = o_p(1). \tag{3.18}$$

Proof. By Proposition 3.1,

$$\sup_{x\in\mathbb{R}} \left| \int \left[n^{1/2}\hat{S}_n(x - ub) - n^{1/2}S_n(x - ub) - n^{1/2}\Delta_n'\mu(x - ub)\right]K(u)du\right| = o_p(1).$$

By the uniform continuity of f_ε, the continuity of f_Z, and the assumptions that $E\|\dot{m}_\theta(X)\| < \infty$ and $b \to 0$, we have $\sup_{x\in\mathbb{R}, |u|\le 1} |\mu(x - ub) - \mu(x)| \to 0$.

Observe that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\Big|\int n^{1/2}S_n(x-ub)K(u)du-n^{1/2}S_n(x)\Big|\\
&=(b^2/2)\sup_{x\in\mathbb{R}}\Big|\int n^{-1/2}\sum_{i=1}^{n}\big\{\dot f_\varepsilon(x-Z_{i-1}-\eta ub)+\dot f_Z(x-\varepsilon_i-\eta ub)\big\}u^2K(u)du\Big|\\
&\le b^2\sup_{y\in\mathbb{R}}\Big|n^{-1/2}\sum_{i=1}^{n}\big\{\dot f_\varepsilon(y-Z_{i-1})+\dot f_Z(y-\varepsilon_i)\big\}\Big|=O_p(b^2)=o_p(1),
\end{aligned}$$

where the first equality follows from the Taylor's expansion and $\nu_1=0$, and the inequality holds true due to $\int u^2K(u)du\le 1$ while the last two equalities follow from (3.14) of Lemma 3.4 and $b\to 0$ of Assumption b, respectively.

The stationarity of $\{X_i, i\in\mathbb{Z}\}$, the independence of X_{i-1} and ε_i for each $i\in\mathbb{Z}$, integration by parts and the change of variable formula yield that for all $x\in\mathbb{R}$,

$$ES_{n1}(x)=\int F_\varepsilon(x-y)f_Z(y)dy=\int F_Z(x-z)f_\varepsilon(z)dz=ES_{n2}(x)=F_X(x). \quad (3.19)$$

Next, recall the definition of φ_i from the Assumption T2 and let

$$\begin{aligned}
\tilde T_n(x) &:= n^{1/2}\big(S_n(x)-2F_X(x)\big)+n^{1/2}\Delta_n'\mu(x)\\
Y_i(x) &:= F_\varepsilon\big(x-m_\theta(X_{i-1})\big)+F_Z(x-\varepsilon_i)-2F_X(x)+\mu(x)'\varphi_i, \quad i\ge 1,\\
T_n(x) &:= n^{-1}\sum_{i=1}^{n}Y_i(x), \quad x\in\mathbb{R}.
\end{aligned}$$

The following proposition describes an approximation of $n^{1/2}\big(\hat F_c-F_X\big)$. In the sequel, for any two stochastic processes $U_{n1}(x), U_{n2}(x)$, the statement $U_{n1}(x)=U_{n2}(x)+u_p(1)$ means that $\sup_x|U_{n1}(x)-U_{n2}(x)|=o_p(1)$.

Proposition 3.3. *Under the Assumptions K, F1, T2, F2, M2, and H_0,*

$$\sup_{x\in\mathbb{R}}n^{-1/2}\Big|\tilde T_n(x)-T_n(x)\Big|=o_p(1), \quad (3.20)$$

$$\sup_x\Big|n^{1/2}\big(\hat F_c(x)-F_X(x)\big)-T_n(x)\Big|=o_p(1). \quad (3.21)$$

Proof. The proof of (3.20) follows immediately from Assumption T2 and (3.3).

Proof of (3.21). Recall $\hat f_Z(z)\equiv f_n(z,\hat\theta)$ and $\hat F_\varepsilon(x)\equiv F_n(x,\hat\theta)$. Let

$$\begin{aligned}
\mathcal{Z}_n(x) &:= \int\big[\hat F_\varepsilon(x-z)-F_\varepsilon(x-z)\big]\big[\hat f_Z(z)-f_Z(z)\big]dz\\
&= \int\big[F_n\big(x-z,\hat\theta\big)-F_\varepsilon(x-z)\big]\big[f_n\big(z,\hat\theta\big)-f_Z(z)\big]dz \quad .
\end{aligned}$$

By (3.8), (3.10), and $n^{1/2}b^2 \to \infty$ (guaranteed by Assumption b), we have:

$$\begin{aligned}\sup_x n^{1/2}\left|\mathcal{Z}_n(x)\right| &\le \sup_x n^{1/2}\left|F_n\left(x,\hat{\theta}\right)-F_\varepsilon(x)\right|\int\left|f_n\left(z,\hat{\theta}\right)-f_Z(z)\right|dz \\ &= n^{1/2}O_p\left((n^{1/2}b)^{-1}\right)O_p\left((n^{1/2}b)^{-1}\right)=O_p\left((n^{1/2}b^2)^{-1}\right)=o_p(1)\end{aligned}. \quad (3.22)$$

We also have the decomposition

$$F_c(x)-F_X(x)=\int F_\varepsilon(x-z)\hat{f}_Z(z)dz+\int \hat{F}_\varepsilon(x-z)f_Z(z)dz-2F_X(x)+\mathcal{Z}_n(x). \quad (3.23)$$

By the change of variable formula,

$$\begin{aligned}\int F_\varepsilon(x-z)\hat{f}_Z(z)dz &= \int \frac{1}{nb}\sum_{i=1}^{n}K\left(\frac{z-m_{\hat{\theta}}\left(X_{i-1}\right)}{b}\right)F_\varepsilon(x-z)dz, \\ &= \int \frac{1}{n}\sum_{i=1}^{n}F_\varepsilon(x-bu-m_{\hat{\theta}}(X_{i-1}))K(u)du \\ &= \int \hat{S}_{n1}(x-bu)K(u)du.\end{aligned}$$

Similarly, integration by parts and the change of variable formula yield

$$\begin{aligned}&\int \hat{F}_\varepsilon(x-z)f_Z(z)dz \\ &=\frac{1}{n}\sum_{i=1}^{n}\int G\left(\frac{x-z-\hat{\varepsilon}_i}{b}\right)dF_Z(z)=\frac{1}{nb}\sum_{i=1}^{n}\int F_Z(z)K\left(\frac{x-z-\hat{\varepsilon}_i}{b}\right)dz \\ &=\frac{1}{n}\sum_{i=1}^{n}\int F_Z\left(x-ub-\hat{\varepsilon}_i\right)K(u)du=\int \hat{S}_{n2}(x-bu)K(u)du\end{aligned}.$$

Thus, by Proposition 3.2, (3.20), (3.22), and (3.23),

$$\begin{aligned}n^{1/2}\left(F_c(x)-F_X(x)\right) &= \int n^{1/2}\left(\hat{S}_n(x-bu)-2F_X(x)\right)K(u)du+n^{1/2}\mathcal{Z}_n(x) \\ &= n^{1/2}\left(S_n(x)-2F_X(x)\right)+n^{1/2}\Delta_n'\mu(x)+u_p(1), \\ &= \tilde{T}_n(x)+o_p(1)=T_n(x)+u_p(1)\end{aligned} \quad (3.24)$$

thereby completing the proof of (3.21) and of the lemma.

4. MAIN RESULT

In this section, we shall describe the asymptotic null distribution of the test statistic $\mathcal{V}_n$ of (2.3).

Theorem 4.1. *Under the Assumptions (1.1), K, F1, T2, F2, M2, and H_0, for any density ψ on $\mathbb{R}$,*

$$\mathcal{V}_n = \int V_n^2(x)\psi(x)dx \Rightarrow \mathcal{V} := \int \mathcal{G}^2(x)\psi(x)dx, \tag{4.1}$$

where $\mathcal{G}(\cdot)$ is a mean zero Gaussian process with the covariance function $\tilde{\mathcal{C}}(x,y)$ given at (4.3) below.

Proof. Rewrite

$$V_n(x) = n^{1/2}\left(\hat{F}_n(x) - F_X(x)\right) - n^{1/2}\left(\hat{F}_c(x) - F_X(x)\right). \tag{4.2}$$

Recall the definition of φ_i from the Assumption T2 and let

$$\begin{aligned} \tilde{Y}_i(x) &:= I\left(X_i \le x\right) - F_X(x) \\ &\quad -\left\{F_\varepsilon\left(x - m_\theta(X_{i-1}\right) + F_Z(x-\varepsilon_i) - 2F_X(x) + \mu(x)'\varphi_i\right\}, \quad i \ge 1, \\ \tilde{V}_n(x) &:= n^{-1/2}\sum_{i=1}^n \tilde{Y}_i(x), \qquad \tilde{\mathcal{C}}(x,y) := \lim_n \mathrm{Cov}\left(\tilde{V}_n(x), \tilde{V}_n(y)\right). \end{aligned} \tag{4.3}$$

Let $\mathcal{G}(x), x \in \mathbb{R}$ be a continuous Gaussian process with mean zero and the covariance function $\mathrm{Cov}\left(\mathcal{G}(x), \mathcal{G}(y)\right) \equiv \tilde{\mathcal{C}}(x,y)$.

Note that $\tilde{V}_n(x) \equiv n^{1/2}\left(\hat{F}_n(x) - F_X(x)\right) - T_n(x)$. By (3.21) of Proposition 3.3,

$$\sup_x \left| V_n(x) - \tilde{V}_n(x) \right| = \sup_x \left| n^{1/2}\left(\hat{F}_c(x) - F_X(x)\right) - T_n(x) \right| = o_p(1).$$

The r.v. $n^{-1/2}\tilde{V}_n(x)$ is an average of stationary and ergodic r.v.'s $\tilde{Y}_i(x)$ where, in view of (3.19) and Assumption T2, $E\tilde{Y}_0(x) \equiv 0$. Hence, by Theorem 2 in Wu and Shao (2004), $\tilde{\mathcal{C}}(x,y)$ exists and is finite for all $x, y \in \mathbb{R}$ and $\tilde{\mathcal{C}}(x,x) > 0$ for all $x \in \mathbb{R}$. Moreover, by Theorem 3 of the same paper, $\tilde{V}_n(x) \to_D N\left(0, \mathcal{C}(x,x)\right)$ r.v., for every $x \in \mathbb{R}$. By the Cramér–Wold device, all finite dimensional distributions of $\tilde{V}_n$ converge weakly to those of $\mathcal{G}$.

To complete the proof of (4.1), it remains to prove the tightness of the process $\tilde{V}_n(x), x \in \mathbb{R}$ in the uniform metric. For this purpose, recall the definition of S_{nj}, j = 1,2 from (3.15) and with $U_n := n^{-1/2}\sum_{i=1}^n \varphi_i$, rewrite

$$\begin{aligned} \tilde{V}_n(x) &= n^{1/2}\left(\hat{F}_n(x) - F_X(x)\right) - n^{1/2}\left(S_{n1}(x) - ES_{n1}(x)\right) \\ &\qquad - n^{1/2}\left(S_{n2}(x) - ES_{n2}(x)\right) - \mu(x)'U_n \\ &= T_{n1}(x) - T_{n2}(x) - T_{n3}(x) - \mu(x)'U_n, \qquad \text{say.} \end{aligned}$$

The processes $T_{n1}(x), T_{n2}, x \in \mathbb{R}$ are empirical processes of bounded functions of a Markov chain. Their tightness in the uniform metric follows from Levental (1989). Similarly, $T_{n3}(x)$, $x \in \mathbb{R}$ is an empirical process of bounded functions i.i.d. r.v.'s and its tightness follows from van der Vaart and Wellner (1996). These facts together with the uniform continuity of $\mu(x)$ and $\|U_n\| = O_p(1)$, guaranteed by

Assumption T2, imply the tightness of the process $\tilde{V}_n(x), x \in \mathbb{R}$, thereby completing the proof of (4.1) and of the theorem.

Remark 4.1. Similarly, we can formulate a test statistic based on density estimation

$$v_{n,b}(x) := \sqrt{nb}\Big(\hat{f}_k(x) - \hat{f}_c(x)\Big) \tag{4.4}$$

where $\hat{f}_k(x)$ is a kernel density estimate for the marginal density $f_X(x)$ of X, while $\hat{f}_c(x)$ is the convolution density estimate of $f_X(x)$. Kim and Wu (2007) proved that under H_0, $v_{n,b}(\cdot) \Rightarrow N\Big(0,\ f_X(\cdot)\int K^2(u)du\Big)$. Hence,

$$\mathcal{K}_n := \int v_{n,b}^2(u)\psi(u)du \Rightarrow \int \mathcal{Z}^2(u)\psi(u)du, \quad \text{for all densities } \psi \text{ on } \mathbb{R}, \tag{4.5}$$

where $\mathcal{Z}$ is a continuous Gaussian process. Note, that the consistency rate in (4.2) is faster than that in (4.4), given the order of the bandwidth b. Thus, the test based on (4.2) is expected to perform better than the test based on (4.4) in finite sample applications for moderate sample sizes. This is evidenced in the following simulation study.

5. SIMULATION STUDIES

5.1. Setup

This section reports findings of a finite sample study that compares two tests for the comparison purpose: our proposed test and the benchmark test proposed by Kim et al. (2015). In this simulation study, test statistics considered are $\mathcal{V}_n$ in (4.1) and $\mathcal{K}_n$ in (4.5), respectively, with ψ equal to the uniform density of (0, 10). These tests are similar in that both tests exploit the fact that the deviation from the null hypothesis will lead to the significant discrepancy between the convolution estimate and classical kernel density estimates. On the other hand, the quintessential difference between the two tests lies in the fact that the proposed test based on $\mathcal{V}_n$ employs the d.f.'s while the density functions are used in the benchmark test based on $\mathcal{K}_n$. In this section, we demonstrate that our proposed test outperforms the benchmark test when testing various hypotheses for the chosen sample sizes and alternatives.

Consider the following two sets of hypotheses

$$\begin{aligned}
&H_{01}: X_i = 0.5\,|X_{i-1}| + \varepsilon_i, &&H_{a1}: X_i = 0.5\,|X_{i-1}| + \varepsilon_i\sqrt{0.8(1+X_{i-1}^2)},\\
&H_{02}: X_i = 0.1 + 0.5X_{i-1} + \varepsilon_i, &&H_{a2}: X_i = 0.1 + 0.5X_{i-1} + 0.8\mathrm{e}^{-|X_{i-1}|} + \varepsilon_i,
\end{aligned}$$

where $\varepsilon_i, 1 \le i \le n$ is a random sample from $N(0, 1)$ distribution. The null model H_{01} is the threshold autoregressive (TAR) model (Tong, 1990) of order one, while the alternative H_{a1} is the TAR-ARCH(1,1) model (Zako ï an, 1994). The purpose of testing hypotheses H_{01} versus H_{a1} is to see whether our test can differentiate the conditional homoscedastic process (i.e., TAR) from the conditional heteroscedastic one (i.e., TAR-ARCH). The hypotheses H_{02} and H_{a2} simply test whether the autoregressive function is linear or non-linear function of the lag variable.

For T = 200, 400, 600, we generate $\{X_i : i = 1,2,...,T\}$ according to the given null hypothesis with the initial value $X_0 = 0.8$. When we compare the performance of the proposed test with that of the benchmark test, empirical level and power will be used to assess the performance. For the benchmark test, we emulate Monte Carlo simulation as described in Kim et al. (2015) while we employ block-wise bootstrap method proposed by Künsch (1989) and Liu and Singh (1992) for the proposed test. We first determine the size of the block l_B, so that the number of blocks, n_B, is T / l_B. Naik-Nimbalakar and Rajarshi (1994) showed that the weak convergence of block-wise bootstrapped empirical process depends on the order of the l_B. They obtained desired results when $l_B = O(n^{1/2} - \epsilon)$, with $0 < \epsilon < \frac{1}{2}$. Motivated by their work, we chose $l_B = \{8,10,16,20\}$ in this finite sample study. We found that the proposed test displays the optimal result when l_B = 10, for all T. Once we determine the value of l_B, we construct a block: we draw any uniform random number between 1 and $T - l_B + 1$, say k, and choose l_B consecutive observations, $X_{k+1},...,X_{k+l_B}$. We repeat constructing a block n_B times, combine these n_B blocks all together, and obtain re-sampled observations, $X_1^*,...,X_T^*$. For the calculation of the statistics, we use uniform kernel function $K(u) := 2^{-1} I(|u| \leq 1)$. Therefore,

$$G(u) = \int_{-\infty}^{u} K(x)dx = \begin{cases} 0, & u < -1; \\ \dfrac{u+1}{2}, & -1 \leq u < 1; \\ 1, & u \geq 1. \end{cases}$$

For $\hat{\theta}$, we use least squares estimator. Define $h(t) := (-t^2 + 2c_i t) / 8b$ where $c_i := x + b - \hat{\varepsilon}_i$. Then $\hat{F}_c$ in (2.2) can be rewritten as

$$\hat{F}_c(x) = \frac{1}{n^2 b} \sum_{i=1}^{n} \sum_{j=1}^{n} \int G\left(\frac{x - t - \hat{\varepsilon}_i}{b}\right) K\left(\frac{x - m_{\hat{\theta}}\left(X_{j-1}\right)}{b}\right) dt = \frac{1}{n^2 b} \sum_{i=1}^{n} \sum_{j=1}^{n} IF_{ij}(x),$$

where, with $Z_{ij} := m_{\hat{\theta}}\left(X_{j-1}\right) + \hat{\varepsilon}_i$,

$$IF_{ij}(x) = \begin{cases} 0, & x < Z_{ij} - 2b, \\ h(x - \hat{\varepsilon}_i + b) - h(m_{\hat{\theta}}(X_{j-1}) - b), & Z_{ij} - 2b \leq x < Z_{ij}, \\ bG((x - Z_{ij} - b) / b) + h(m_{\hat{\theta}}(X_{j-1}) + b) & \\ \qquad\qquad - h(x - \hat{\varepsilon}_i - b), & Z_{ij} \leq x < Z_{ij} + 2b, \\ b, & \text{otherwise} \end{cases}$$

Consequently, the great deal of simplification in the computation of $V_n(x)$ in (2.3) follows directly.

Define the bootstrap test statistic:

$$\mathcal{V}_n^* = \int_0^{10} \left(V_n^*(x) - V_n(x)\right)^2 dx \tag{5.1}$$

where $V_n^*(x)$ denotes the counterpart of $V_n(x)$, which is obtained from re-sampled observations. We repeat block-wise bootstrap B_{Iter} times, obtain $\mathcal{V}_n^*$'s, and calculate $100(1-\alpha)$ percentiles, $q_{1-\alpha}^*$, obtained from these values of $\mathcal{V}_n^*$'s. As various l_B's are tried, so are B_{Iter}'s. Our findings show that empirical levels approaches more closely to the suggested significance level α as B_{Iter} increases (see, e.g., Table 1). After $q_{1-\alpha}^*$ is obtained, we reject H_0 if $\mathcal{V}_n > q_{1-\alpha}^*$. As a final step, we repeat this procedure 1,000 times, count the number of rejections, and obtain empirical levels and powers by dividing it by 1,000.

5.2. *Selection of B_{Iter}, b, and l_B*

In the simulation study, $b = \{0.05, 0.1, 0.15, 0.2\}$ are tried for the bandwidth. Since the choice of b does not affect the powers and levels much, we only report the result corresponding to $b = 0.1$. Tables 1 and 2 report empirical levels of the proposed test for H_{01} and H_{02} corresponding to various sizes of blocks and numbers of bootstrap iterations. As shown in the tables, we obtain the optimal results at $(l_B, B_{Iter}) = (10, 200)$ and $(l_B, B_{Iter}) = (20, 160)$ for H_{01} and H_{02}, respectively. Therefore, we use these optimal values in the following simulation studies. One point worth noting here is that the empirical levels of H_{01} show good convergence to corresponding α's while the same conclusion does not hold for H_{02}. For H_{01}, the empirical levels corresponding to $(l_B, B_{Iter}) = (10, 200)$ are 0.045 and 0.011 which are quite close to $\alpha = 0.05$ and 0.01, respectively. The counterparts of H_{02} corresponding to $(l_B, B_{Iter}) = (20, 160)$ are 0.094 and 0.037. In view of this, we conjecture that the proposed test for H_{02} might not be as efficient as it is for H_{01}: we will discuss more about this in the next section.

Table 1. Levels of H_{01} When B_{Iter} and l_B Vary With T Being Fixed at 400.

	$l_B = 8$					$l_B = 10$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.079	0.072	0.063	0.059	0.055	0.074	0.064	0.053	0.047	0.045
0.01	0.050	0.025	0.016	0.014	0.016	0.029	0.013	0.012	0.014	0.011
	$l_B = 16$					$l_B = 20$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.064	0.057	0.055	0.054	0.052	0.097	0.076	0.075	0.067	0.065
0.01	0.025	0.015	0.015	0.012	0.013	0.058	0.029	0.020	0.019	0.016

Table 2. Levels of H_{02} When B_{Iter} and l_B Vary With T Being Fixed at 400.

	$l_B = 8$					$l_B = 10$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.134	0.123	0.120	0.144	0.114	0.137	0.129	0.117	0.092	0.124
0.01	0.078	0.057	0.052	0.065	0.042	0.076	0.059	0.047	0.044	0.047
	$l_B = 16$					$l_B = 20$				
α	$B_{Iter} = 40$	80	120	160	200	$B_{Iter} = 40$	80	120	160	200
0.05	0.131	0.104	0.104	0.107	0.102	0.112	0.103	0.123	0.094	0.094
0.01	0.064	0.045	0.038	0.037	0.045	0.051	0.045	0.055	0.030	0.037

5.3. Tests $\mathcal{V}_n$ and $\mathcal{K}_n$ for H_{01} and H_{02}

This section compares the proposed test $\mathcal{V}_n$ and the benchmark test $\mathcal{K}_n$ for H_{01} and H_{02}. Tables 3 and 4 report their empirical levels and powers when $T = 200, 400, 600$ and $\alpha = 0.01, 0.05$. To begin with, consider the result for H_{01}. It is hard to tell which test is superior in terms of the level. However, there is no room for argument in terms of the power: the proposed test dominates the benchmark test. When $T = 200$, the differences in the power between two tests are approximately 0.3 and 0.17 for $\alpha = 0.05$ and 0.01, respectively. For other T's (400 and 600), the differences, however, decreases to approximately 0.19 and 0.15 for $\alpha = 0.05$ and 0.01, respectively. Note that the benchmark test does not obtain the power larger than 0.8 even when T reaches 600: 0.747 and 0.692 for $\alpha = 0.05$ and 0.01, respectively. On the contrary, the proposed test with $\alpha = 0.05$ accomplishes the power larger than 0.9 when $T = 400$.

Next, we proceed to analyze the result for H_{02} which is reported in Table 4. A quick glance at the table reveals that the proposed and benchmark tests for H_{02}

Table 3. The Proposed $\mathcal{V}_n$ and Benchmark $\mathcal{K}_n$ Tests for H_{01}.

	α	$T = 200$		$T = 400$		$T = 600$	
		$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$
Level	0.05	0.041	0.052	0.045	0.050	0.058	0.054
	0.01	0.013	0.007	0.013	0.010	0.012	0.007
Power	0.05	0.658	0.343	0.901	0.710	0.933	0.747
	0.01	0.387	0.215	0.714	0.561	0.854	0.692

Table 4. The Proposed ($\mathcal{V}_n$) and Benchmark ($\mathcal{K}_n$) Tests for H_{02}.

	α	$T = 200$		$T = 400$		$T = 600$	
		$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$	$\mathcal{V}_n$	$\mathcal{K}_n$
Level	0.05	0.134	0.023	0.096	0.013	0.091	0.013
	0.01	0.065	0.003	0.032	0.002	0.031	0.004
Power	0.05	0.424	0.045	0.617	0.120	0.702	0.233
	0.01	0.260	0.015	0.433	0.072	0.498	0.109

display poor performances (especially, power) for all α and T when compared with the result for H_{01}. In the case of empirical level, both tests fail to show a sign of any convergence to α, as T increases. For example, the empirical level of the proposed test closest to $\alpha = 0.01$ is 0.031 when $T = 600$, which is still far away from 0.01.

While the empirical power of both tests for H_{02} is much smaller than worse that for H_{01}, there exists a stark difference between the two tests. The extent of decrease in the power displayed by the benchmark test $\mathcal{K}_n$ is extremely large, compared with that of the proposed test $\mathcal{V}_n$. For example, the empirical power of the $\mathcal{V}_n$ test corresponding to $T = 600$ and $\alpha = 0.05$ decreased from 0.933 to 0.702 while the power of the $\mathcal{K}_n$ test plummets from 0.747 to 0.233. Even though both tests show consistency, that is, display an increase in the power as T increases, the benchmark test always yields a power below 0.3 while the power of the proposed test starts with 0.424 and rises to 0.702 as T changes from 400 to 600. In summary, the proposed test for H_{02} does not retain the same efficiency as that for H_{01} anymore, but it still outperforms the benchmark test.

This result accords closely with the fact that both tests assume the independence between the error term and the lagged X_i's. The presence of the dependence between them as in the alternative H_{1a} will lead to the instant rejection of the null H_{10} and beget a larger statistical power than would otherwise be the case. Unlike H_{1a}, the alternative H_{2a} does not violate the assumption of the independence, and, hence, it is plausible for both tests to return smaller powers. Based on all findings in this section, it is undoubtedly the case that the proposed test is indeed superior to the benchmark test.

6. CONCLUSION

In this chapter, we propose a test for a parametric specification of the autoregressive function. This test is based on the integrated square difference between the empirical d.f. and the convolution d.f. estimates of the stationary d.f. We prove the weak convergence of a suitably standardized convolution d.f. estimate process to a Gaussian process. This in turn is used to derive the asymptotic null distribution of the proposed test statistic. The consistency rate of this test statistic is $n^{1/2}$. The block bootstrap approach is employed to run the proposed test under a Monte Carlo setting. As expected, the simulation study shows the superiority of the proposed test over the benchmark test of Kim et al. (2015) based on density estimates in term of the empirical level and power.

This chapter can be extended in the following directions. First, the framework considered in the current study can be extended to cover continuous-time processes. The continuous-time processes have played an important role in economic time-series analysis, due to their usefulness in modeling financial data of high frequency. Some of the recent studies focus on the inference of potentially non-stationary continuous-time processes, see, for example, Aït-Sahalia and Park (2012, 2016). The methodology suggested in the current study can be extended to this end. Secondly, a theoretical justification of the bootstrap methodology employed in the simulation study needs to be provided. This study employs the bootstrap to handle the issue of non-pivotal asymptotic distribution for the proposed test statistic. A formal justification of the approach will make the simulation result more convincing.

7. PROOFS

This section contains the proofs of Lemmas 3.1–3.4 and Proposition 3.1. Before proceeding to the proof, we shall discuss some implications of the Assumption M1, which are often used in the proofs below. Let

$$\begin{aligned} &d_i := m_{\hat\theta}(X_{i-1}) - m_\theta(X_{i-1}), \qquad \Delta_n := \hat\theta - \theta, \qquad D_n := \max_{1\le i\le n} |d_i|, \\ &\delta_i := d_i - \Delta_n{}'\dot m_\theta(X_{i-1}), \quad 1 \le i \le n. \end{aligned} \tag{7.1}$$

By (3.5), stationarity, ergodicity, and the Ergodic Theorem imply that

$$\max_{1\le i\le n} n^{-1/2}\|\dot m_\theta(X_{i-1})\| = o_p(1), \qquad n^{-1}\sum_{i=1}^n \dot m_\theta(X_{i-1}) = E\dot m_\theta(X_0) + o_p(1).$$

Together with these facts, Assumptions (3.2) and (3.4), in turn, yield the following results.

$$\begin{aligned} &\max_{1\le i\le n} |\Delta_n'\dot m_\theta(X_{i-1})| \le n^{1/2}\|\Delta_n\| \max_{1\le i\le n} n^{-1/2}\|\dot m_\theta(X_{i-1})\| = o_p(1), \quad D_n = o_p(1) \\ &\max_{1\le i\le n} \big|\delta_i\big| = \max_{1\le i\le n}\big| m_{\hat\theta}(X_{i-1}) - m_\theta(X_{i-1}) - \Delta_n'\dot m_\theta(X_{i-1})\big| = o_p(n^{-1/2}), \\ &n^{-1/2}\sum_{i=1}^n d_i = n^{1/2}\Delta_n' n^{-1}\sum_{i=1}^n \dot m_\theta(X_{i-1}) + o_p(1) = n^{1/2}\Delta_n' E\dot m_\theta(X_0) + o_p(1) = O_p(1), \\ &n^{-1/2}\sum_{i=1}^n |d_i| \le n^{-1/2}\sum_{i=1}^n \big| d_i - \Delta_n'\dot m_\theta(X_{i-1})\big| + n^{1/2}\|\Delta_n\| n^{-1}\sum_{i=1}^n \|\dot m_\theta(X_{i-1})\| \\ &\qquad = O_p(1) \end{aligned} \tag{7.2}$$

In the proofs below, $\sup_x \equiv \sup_{x\in\mathbb{R}}$, unless mentioned otherwise. We also use the notation $\nu_j := \int u^j K(u)du$, $j = 1,2$. Under Assumption K, $\nu_1 = 0$ and $\nu_2 \le 1$.

Proof of Lemma 3.1. Because $\kappa := \|K\|_\infty < \infty$,

$$\big|G(u) - G(v)\big| = \Big|\int_v^u K(y)dy\Big| \le |u - v|\kappa, \quad \forall u, v \in \mathbb{R}.$$

Hence

$$\begin{aligned} \sup_x \big| F_n(x,\hat\theta) - F_n(x,\theta)\big| &\le \sup_x \frac{1}{n}\sum_{i=1}^n \left| G\left(\frac{x-\hat\varepsilon_i}{b}\right) - G\left(\frac{x-\varepsilon_i}{b}\right)\right| \\ &\le \frac{\kappa}{n}\sum_{i=1}^n \left|\frac{\varepsilon_i - \hat\varepsilon_i}{b}\right| = \frac{1}{nb}\sum_{i=1}^n |d_i| = O_p\big((n^{1/2}b)^{-1}\big), \end{aligned}$$

by (7.2). This completes the proof of (3.6).

Proof of (3.7). The stochastic process $n^{1/2}\big(F_n(x,\theta)-EF_n(x,\theta)\big), x\in\mathbb{R}$ is a mean zero empirical process of bounded functions of i.i.d. r.v.s' ε_i, which are known to converge weakly to a Gaussian process in the uniform metric, see, for example, van der Vaart and Wellner (1996). This fact in turn implies (3.7).

Proof of (3.8). The left hand side of (3.8) is bounded from the above by

$$\begin{aligned}\sup_x\big|\big(F_n(x,\hat\theta)-F_n(x,\theta)\big|+\sup_x\big|F_n(x,\theta)-EF_n(x,\theta)\big|\\ +\sup_x\big|EF_n(x,\theta)-F_\varepsilon(x)\big|.\end{aligned}\tag{7.3}$$

Consider the third term in this bound. The integration by parts, the change of variable, $\nu_1=0$, $\dot f_\varepsilon$ being bounded and the Taylor expansion yields that for some $0<\eta<1$,

$$\begin{aligned}EF_n(x,\theta)-F_\varepsilon(x)&=\int\big[F_\varepsilon(x-bu)-F_\varepsilon(x)\big]K(u)du\\ &=\frac{b^2}{2}\int\dot f_\varepsilon(x-\eta bu)K(u)du,\\ \sup_x\big|EF_n(x,\theta)-F_\varepsilon(x)\big|&\le\big\|\dot f_\varepsilon\big\|_\infty b^2.\end{aligned}$$

This bound together with the assumption $n^{1/2}b^3\to 0$, guaranteed by Assumption b, (3.7), (3.8) and (7.3) completes the proof of (3.8), and that of the lemma.

Proof of Lemma 3.2. Assumption M and (1.1) imply that, under H_0, the sequence $Z_i\equiv m_\theta(X_i), i\in\mathbb{Z}$ is stationary and ergodic α-mixing with the mixing sequence α_i satisfying (3.1). This fact and the compactness of the support of f_Z make Theorem 4.2.2 (iii) of Györfi et al. (1989) applicable, which yields that the expected value of the left hand side of (3.9) is of the order $O\big((n^{1/2}b)^{-1}\big)$, which together with the Markov inequality implies (3.9). Note that this result does not need the smoothness of f_Z and K.

To prove (3.10), let

$$A_n:=\int\big|f_n(z,\theta)-Ef_n(z,\theta)\big|dz,\qquad B_n:=\int\big|Ef_n(z,\theta)-f_Z(z)\big|dz.$$

By the triangle inequality and (3.9),

$$\begin{aligned}\int\big|f_n(z,\hat\theta)-f_Z(z)\big|dz&\le\int\big|f_n(z,\hat\theta)-f_n(z,\theta)\big|dz+A_n+B_n\\ &=\int\big|f_n(z,\hat\theta)-f_n(z,\theta)\big|dz+O_p\big((n^{1/2}b)^{-1}\big)+B_n\end{aligned}.\tag{7.4}$$

To obtain a bound on B_n, let $\gamma(z,\eta):=\int u^2K(u)\ddot f_Z(z-\eta bu)du$, $z\in\mathbb{R}, 0<\eta<1$. With $Z_i=m_\theta(X_i)$, rewrite

$$f_n(z,\theta) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{z - Z_{i-1}}{b}\right).$$

Because $\nu_1 = 0$, by the Taylor expansion, for some $0 < \eta < 1$,

$$\begin{aligned} EK\left(\frac{z-Z}{b}\right) &= b\int K(u) f_Z(z - ub)du \\ &= b\int K(u)\left[f_Z(z) - ub\dot{f}_Z(z) + \frac{u^2b^2}{2}\ddot{f}_Z(z - \eta bu)\right]du \\ &= bf_Z(z) + \frac{b^3}{2}\gamma(z,\eta). \end{aligned}$$

Hence, by stationarity,

$$Ef_n(z,\theta) = \frac{1}{b}EK\left(\frac{z-Z}{b}\right) = f_Z(z) + \frac{b^2}{2}\gamma(z,\eta).$$

Let $J_{2Z} := \int \left|\ddot{f}_Z(z)\right| dz$. By Assumption F2, $J_{2Z} < \infty$. By the Fubini Theorem,

$$\int |\gamma(z,\eta)|\, dz = \int\int \left|\ddot{f}_Z(z - \eta bu)\right| dz\, u^2 K(u)du = \nu_2 J_{2Z}, \quad \forall n \geq 1.$$

Hence

$$B_n = \int \left|Ef_n(z,\theta) - f_Z(z)\right| dx = \frac{b^2}{2}\int |\gamma(z,\eta)|\, dx = \frac{b^2}{2}\nu_2 J_{2Z}, \quad \forall n \geq 1. \quad (7.5)$$

Next, K having bounded derivative implies that

$$\begin{aligned} &\int \left|f_n\left(z,\hat{\theta}\right) - f_n\left(z,\theta\right)\right| dz \\ &\leq \frac{1}{nb}\sum_{i=1}^{n}\int\left|K\left(\frac{z - m_{\hat{\theta}}(X_{i-1})}{b}\right) - K\left(\frac{z - m_\theta(X_{i-1})}{b}\right)\right| dz \\ &= \frac{1}{n}\sum_{i=1}^{n}\int\left|K(u) - K\left(u + \frac{d_i}{b}\right)\right| du \leq \frac{\|\dot{K}\|_\infty}{n}\sum_{i=1}^{n}|d_i| = O_p\left(n^{-1/2}\right), \end{aligned}$$

by (7.2). This fact together with (7.5) and (7.4) completes the proof of the lemma.

Proof of Lemma 3.3. Because $\ddot{f}_\varepsilon$ is bounded, by the mean value theorem, the left hand side of (3.11) is bounded from the above by $\left\|\ddot{f}_\varepsilon\right\|_\infty n^{-1}\sum_{i=1}^{n}|d_i| = O_p(n^{-1/2})$, where the equality directly follows from (7.2). Similarly, the left hand side of (3.12) is bounded from the above by $\left\|\dot{f}_Z\right\|_\infty n^{-1}\sum_{i=1}^{n}|\hat{\varepsilon}_i - \varepsilon_i| = n^{-1}\sum_{i=1}^{n}|d_i| = O_p(n^{-1/2})$.

Proof of Lemma 3.4. The integration by parts and the change of variables yield that

$$E\left[\dot{f}_Z(x-\varepsilon)\right] = \int \dot{f}_Z(x-v) f_\varepsilon(v) dv = -f_Z(x-v) f_\varepsilon(v) \Big|_{v=-\infty}^{v=\infty} + \int f_Z(x-v) \dot{f}_\varepsilon(v) dv$$
$$= -\int f_Z(z) \dot{f}_\varepsilon(x-z) dz = -E\dot{f}_\varepsilon(x-Z), \quad \forall x \in \mathbb{R}.$$

By Lemmas 3.3 and (3.13), the left hand side of (3.14) is bounded from the above by

$$\frac{1}{n}\left|\sum_{i=1}^{n}\left[\dot{f}_\varepsilon\left(x-m_\theta\left(X_{i-1}\right)\right) - E\dot{f}_\varepsilon\left(x-m_\theta(X)\right) + \dot{f}_Z\left(x-\varepsilon_i\right) - E\dot{f}_Z\left(x-\varepsilon\right)\right]\right| + O_p\left(n^{-1/2}\right)$$
$$= O_p\left(n^{-1/2}\right),$$

by the Ergodic Theorem.

Proof of Proposition 3.1. Under Assumption F1, $\dot{f}_\varepsilon$ is bounded, which in turn implies that f_ε is uniformly continuous. Hence, by (7.2),

$$\sup_{x\in\mathbb{R}} n^{-1/2}\left|\sum_{i=1}^{n}\left\{F_\varepsilon\left(x-m_{\hat{\theta}}(X_{i-1})\right) - F_\varepsilon\left(x-m_\theta(X_{i-1})\right) - d_i f_\varepsilon\left(x-m_\theta(X_{i-1})\right)\right\}\right|$$
$$= \sup_{x\in\mathbb{R}} n^{-1/2}\left|\sum_{i=1}^{n}\int_0^{d_i}\left[f_\varepsilon\left(x-m_\theta(X_{i-1})-s\right) - f_\varepsilon(x-m_\theta(X_{i-1}))\right]ds\right|. \tag{7.6}$$
$$\leq n^{-1/2}\sum_{i=1}^{n}|d_i| \sup_{|y-z|\leq D_n}\left|f_\varepsilon(y)-f_\varepsilon(z)\right| = o_p(1)$$

Next, (7.2) and f_ε being bounded, guaranteed by Assumption F1, readily imply

$$\sup_{x\in\mathbb{R}}\left|n^{-1/2}\sum_{i=1}^{n}\delta_i f_\varepsilon\left(x-m_\theta(X_{i-1})\right)\right| \leq n^{1/2}\max_{1\leq i\leq n}\left|\delta_i\right|\sup_{y\in\mathbb{R}} f_\varepsilon(y) = o_p(1). \tag{7.7}$$

Thus, by (7.6) and (7.7),

$$\begin{aligned} n^{1/2}(\hat{S}_{n1}(x) - S_{n1}(x)) &= n^{1/2}\Delta_n' n^{-1}\sum_{i=1}^{n}\dot{m}_\theta(X_{i-1}) f_\varepsilon(x-m_\theta(X_{i-1})) + o_p(1) \\ &= n^{1/2}\Delta_n' E\left(\dot{m}_\theta(X) f_\varepsilon(x-m_\theta(X))\right) + o_p(1) \end{aligned},$$

by the Ergodic Theorem, thereby proving (3.16). The proof of (3.17) is exactly similar.

REFERENCES

Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, *9*, 385–426.

Aït-Sahalia, Y., & Park, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *Journal of Econometrics*, *169*, 279–292.

Aït-Sahalia, Y., & Park, J. Y. (2016). Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *Journal of Econometrics*, *192*, 119–138.

Corradi, V., & Swanson, N. R. (2005). Bootstrap specification tests for diffusion processes. *Journal of Econometrics*, *124*, 117–148.

Györfi, L., Härdle, W., Sards, P., & Vieu, P. (1989). *Nonparametrie curve estimation from time series*. Springer-Verlag.

Kim, K. H., & Wu, W. B. (2007). Density estimation for nonlinear time series. Manuscript.

Kim, K. H., Zhang, T., & Wu, W. B. (2015). Parametric specification test for nonlinear autoregressive models. *Econometric Theory*, *31*, 1078–1101.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. 2nd Edition. *Lecture Notes Series in Statistics* (p. 166). New York, NY: Springer.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, *17*, 1217–1241.

Lahiri, S. N. (1992). Edgeworth correction by "moving block" bootstrap for stationary and nonstationary data. In *Exploring the limits of bootstrap*. Wiley Series in Probability and Mathematical Statistics (pp. 183–214). Wiley-Interscience Publisher.

Lahiri, S. N. (2003). *Resampling methods for dependent data*. Springer Series in Statistics. Springer-Verlag.

Levental, S. (1989). A Uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability*, *2*(3), 171–187.

Liu, R. Y., & Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In R. Lepage & L. Billard (Eds.), *Exploring the limits of bootstrap* (pp. 225–248). Wiley.

Naik-Nimbalkar, U. V., & Rajarshi, M. B. (1994). Validity of block-wise bootstrap for empirical processes with stationary observations. *Annals of*, 22, 980–994.

Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies*, *11*, 449–487.

Tong, H. (1990). *Nonlinear time series: A dynamical system approach*. Oxford University Press.

van der Vaart, A. W., & Wellner, J. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer-Verlag.

Wu, W. B., & Shao, X. F. (2004). Limit theorems for iterated random functions. *Journal of Applied Probability*, *41*, 425–436.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931–955.

CHAPTER 7

TRANSFORMATION MODELS WITH COINTEGRATED AND DETERMINISTICALLY TRENDING REGRESSORS

Yingqian Lin[a] and Yundong Tu[b]

[a]*School of Economics, Shanghai University of Finance and Economics, Shanghai, China*
[b]*Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, China*

ABSTRACT

This chapter develops an asymptotic theory for a general transformation model with a time trend, stationary regressors, and unit root nonstationary regressors. This model extends that of Han (1987) to incorporate time trend and nonstationary regressors. When the transformation is specified as an identity function, the model reduces to the conventional cointegrating regression, possibly with a time trend and other stationary regressors, which has been studied in Phillips and Durlauf (1986) and Park and Phillips (1988, 1989). The limiting distributions of the extremum estimator of the transformation parameter and the plug-in estimators of other model parameters are found to critically depend upon the transformation function and the order of the time trend. Simulations demonstrate that the estimators perform well in finite samples.

Keywords: Cointegration; extremum estimation; nonlinear model; time trend; transformation model; unit root
JEL classifications: C13; C22; C51

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 207–232

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A007

1. INTRODUCTION

The concept of cointegration has been proved important since the seminal work of Granger (1981) and Engle and Granger (1987), with numerous developments witnessed in both theoretical and empirical analysis during the past three decades. The generalization of this concept to nonlinear cointegration dates back to Granger (1991), and gives rise to a wide range of studies in economics and related fields. As put forward by Park (2006), the nonlinear cointegration models provide more flexibilities and allow for much more diverse types of relationships among integrated processes. Well-known applications of nonlinear cointegrations include the modeling of money demand functions, as demonstrated by Bae et al. (2006), Bae and De Jong (2007), and Kim and Kim (2012), and the environmental Kuznets curve as illustrated by Chan and Wang (2015), Wang et al. (2018), Lin et al. (2020), among others. The readers are referred to Granger and Teräsvirta (1993) and Teräsvirta et al. (2010) for more empirical illustrations.

The statistical foundation of the nonlinear cointegration theory has been built up by Professor Joon Park and his co-authors in a sequence of influential papers. In particular, Park and Phillips (1999, 2001) developed a general asymptotic theory of nonlinear regressions with integrated time series, which has been generalized later by Chan and Wang (2015). Park and Phillips (2000) studied the binary choice models with integrated regressors. Chang et al. (2001) extended the model of Park and Phillips (2001) to the multivariate case by accommodating a time trend and stationary regressors, as well as multiple I(1) regressors. Chang and Park (2003) considered the nonlinear index models driven by integrated processes. Chang et al. (2012) considered nonstationary logistic regression. In addition, Chang and Park (2011) and Chan and Wang (2015) considered the endogeneity in nonstationary nonlinear regressions. More recently, Lin and Tu (2021b) investigated transformed linear cointegration models with multiple unit root processes.

Besides the parametric nonlinear cointegration, the nonstationary semi/nonparametric models become increasingly popular in econometrics, due to their flexibility to characterize nonlinear cointegrated relationships. For instance, Phillips and Park (1998), Park and Hahn (1999), and Juhl (2005) developed early nonparametric asymptotic analyses with nonstationary data; Karlsen et al. (2007), Wang and Phillips (2009a, 2009b, 2016), Wang (2015), and Linton and Wang (2016) studied nonparametric cointegration models with kernel estimation method; Cai et al. (2009), Xiao (2009), Gao and Phillips (2013), Hirukawa and Sakudo (2018), and Tu and Wang (2019, 2020) considered functional coefficient cointegration models; Dong et al. (2016) considered the (partially) single index integrated model; Phillips et al. (2017) considered smooth structural changes in cointegration models; Dong and Linton (2018) proposed an additive nonparametric regression with time variable, nonstationary and stationary variables, while Dong et al. (2021) proposed a weighted sieve estimator for the more general nonparametric setting. Lin et al. (2020) proposed a double-nonlinear cointegration model in which a monotonic transformation of the dependent variable and a nonparametric transformation of a unit root regressor are cointegrated.

In addition, a variety of specification tests have been developed in nonlinear cointegration models. See, for example, Wang and Phillips (2012), Wang et al. (2018), and Dong and Gao (2018) for specification tests in cointegrations with a univariate nonstationary regressor, Kasparis and Phillips (2012) for dynamic misspecification tests, and Kasparis et al. (2015) for inferences in nonparametric predictive regressions, Dong et al. (2017) and Tu et al. (2021) for specification tests in cointegrations with stationary covariates, among others. Moreover, Phillips (2009) and Tu and Wang (2022) studied spurious regressions in nonparametric regression and functional coefficient regressions with integrated processes.

This chapter contributes to the above growing literature by investigating a new transformation nonlinear cointegration model, where a monotonic nonlinear transformation of the dependent variable is cointegrated with the multivariate unit root regressors, the time trend, and the stationary regressors. Such a transformation model extends the model of Han (1987) to the case with a time trend and nonstationary regressors. In the special case in which the transformation function becomes identity, the proposed model degenerates to the conventional linear cointegration model (possibly with stationary regressors and time trend) studied by Phillips and Durlauf (1986), Park and Phillips (1988, 1989), etc. Compared to Lin et al. (2020), the current model incorporates time trends, multivariate I(1) processes, and stationary regressors, though these components are linearly related with the monotonic transformation of the dependent variable.

Transformation models have been important tools to analyze economic and financial data. Since Box and Cox (1964) and Bickel and Doksum (1981), a large body of literature has been developed. See, for example, Han (1987) and Abrevaya (1999) for rank estimation of the transformation model; Breiman and Friedman (1985) and Wang and Ruppert (1995) for transform-both-sides models; Chen (2002) and Horowitz (1996) for $\sqrt{n}$ -consistent semiparametric estimators of a regression model with an unknown transformation of the dependent variable; Fan and Fine (2013) for linear transformation models with parametric covariate transformation; Chiappori et al. (2015) for identification and estimation of nonparametric transformations; and Lewbel et al. (2015) for a specification test for nonparametric transformation models. More recently, Florens and Sokullu (2017), Vanhems and Van Keilegom (2019), and Lin and Tu (2021a) studied semiparametric transformation models in the presence of endogeneity. For more references on this literature, see Lin and Tu (2021a), Lin et al. (2020), and references therein.

This chapter first presents an estimation strategy for the proposed model. An extremum estimator of the transformation parameter is proposed via the loss function that measures the relative variation of the regression residual compared to the variation in the transformed dependent variable (Breiman & Friedman, 1985; Lin et al., 2020). The plug-in estimator for rest parameters in the linear component is then obtained. Second, asymptotic distributions for the extremum estimator and the plug-in estimators are then established under a set of regularity conditions. In particular, the limiting distribution of the transformation parameter estimator is nonstandard, with the rate of convergence depending on the model parameters, the properties of the transformation, and time trend. For unit root and time trend regressors, the slope estimators converge at order

n and $\sqrt{n}\kappa_{nd}$ (the order of time trend), respectively, and have nonstandard distributions that involve functionals of Brownian motions. The estimators for the slope parameters before stationary regressors are shown to be $\sqrt{n}$-consistent and asymptotically normal. The derivations build upon Park and Phillips (2001), Chan and Wang (2015), and Hu et al. (2021), which considered nonlinear parametric regressions with univariate I(1) regressor, and Chang et al. (2001) which studied a nonlinear additive parametric model that accommodates all three types of regressors as in the proposed model. Finally, numerical studies illustrate the merit of our proposed estimators. Simulation results show that the biases of the proposed estimators are small, and their variances decay to zero as the sample size increases. The sampling behavior of the t ratios associated with the estimators largely corroborates with our theoretical results, and this finding is robust to various choices of parameters.

The rest of this chapter is organized as follows. Section 2 introduces the model and the estimators, whose asymptotic properties are presented in Section 3. Section 4 reports some simulation results. Section 5 concludes the chapter. The proof of the main theorem is contained in the Appendix, while additional technical details and simulation results are relegated to the Online Supplementary Document.

Notations. Throughout the chapter, convergence in probability and convergence in distribution are denoted as $\xrightarrow{p}$ and $\Rightarrow$, respectively; and $\boldsymbol{A}^\top$ refers to the transpose of the matrix $\boldsymbol{A}$.

2. MODEL AND ESTIMATION

The transformation model of interest is given by

$$\begin{aligned}\Lambda(y_t,\boldsymbol{\beta}_0) &= \boldsymbol{w}_t^\top\boldsymbol{\theta}_0 + u_t \\ &= \boldsymbol{x}_t^\top\boldsymbol{\theta}_1^0 + \boldsymbol{z}_t^\top\boldsymbol{\theta}_2^0 + \boldsymbol{d}_t^\top\boldsymbol{\theta}_3^0 + u_t,\end{aligned} \tag{2.1}$$

for $t = 1,2,\ldots,n$, where $\Lambda:\mathbb{R}\times\mathbb{R}^{\ell_0}\to\mathbb{R}$ is a known strictly increasing function, $\boldsymbol{x}_t$ is an ℓ_1-dimensional integrated process of order one, and $\boldsymbol{z}_t$ is an ℓ_2-dimensional stationary variable, $\boldsymbol{d}_t$ is an ℓ_3-dimensional deterministic sequence, u_t is the stationary error, and $\boldsymbol{\vartheta}_0 = (\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top})^\top$ is the unknown true parameter vector. This model extends the transformation model of Han (1987) to the cases with time trend and nonstationary regressors. When Λ is specialized to an identity function, (2.1) reduces to the conventional linear cointegrating regression, possibly with a time trend and other stationary regressors, which has been developed in the earlier work of Phillips and Durlauf (1986) and Park and Phillips (1988, 1989). Compared to Lin et al. (2020), this model can allow for multivariate unit root regressors as in Lin and Tu (2021b), but also accommodates a time trend component, which leads to much more complication in the resulting limiting theory. In the specification (2.1), the integrated processes, the deterministic and stationary regressors are assumed to be additively separable. The assumption of additive separability here is not essential, but significantly simplifies the subsequent theoretical development.

The estimators of the unknown parameters are defined sequentially. First, for given $\boldsymbol{\beta}, \boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\theta}_3^\top)^\top$ is estimated by the least squares method, that is,

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\beta}) = (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{\Lambda}(\boldsymbol{\beta}), \tag{2.2}$$

where $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^\top$, $\boldsymbol{w}_t = (\boldsymbol{x}_t^\top, \boldsymbol{z}_t^\top, \boldsymbol{d}_t^\top)^\top$, $\boldsymbol{\Lambda}(\boldsymbol{\beta}) = (\Lambda(y_1, \boldsymbol{\beta}), \ldots, \Lambda(y_n, \boldsymbol{\beta}))^\top$. Second, for fixed $\boldsymbol{\beta}$, define the loss function

$$L_n(\boldsymbol{\beta}) = \frac{\sum_{t=1}^n \left[\Lambda(y_t, \boldsymbol{\beta}) - \boldsymbol{w}_t^\top \hat{\boldsymbol{\theta}}(\boldsymbol{\beta})\right]^2}{\sum_{t=1}^n \Lambda(y_t, \boldsymbol{\beta})^2}. \tag{2.3}$$

Then, the extremum estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ is obtained by minimizing $L_n(\boldsymbol{\beta})$ over $\boldsymbol{\beta} \in \Theta_0$, that is, $\hat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta} \in \Theta_0} L_n(\boldsymbol{\beta})$. Consequently, a plug-in estimator for $\boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top}\right)^\top$ is defined as $\hat{\boldsymbol{\theta}}_n = (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{\Lambda}(\hat{\boldsymbol{\beta}}_n)$.

The loss function in (2.3) has a normalizing denominator, which is different from that used for the standard nonlinear regressions (e.g., Chan & Wang, 2015; Park & Phillips, 2001). Since there are unknown parameters on the both sides of (2.1), the direct least squares estimation (i.e., without the normalizing denominator) for (2.1) tends to choose $\boldsymbol{\vartheta}$ such that Λ has little variation. On the other hand, minimizing the loss function (2.3) is equivalent to minimizing the fraction of variance in $\Lambda(y, \boldsymbol{\beta})$ not explained by $\boldsymbol{\omega}_t^\top \hat{\boldsymbol{\theta}}(\boldsymbol{\beta})$. Furthermore, the normalization excludes the trivial specification $\Lambda(y, \beta) = \beta y$ or y / β, under which the loss function in (2.3) is invariant to β. See Breiman and Friedman (1985), Lin et al. (2020), and Lin and Tu (2021b, 2020) for a similar objective function used in estimating transformation models and related discussions.

3. ASYMPTOTIC THEORY

3.1. Assumptions

To present the main theorem, the following assumptions are needed.

Assumption 1.

(a) *Define* $\zeta_t = (u_t, \varepsilon_{t+1}', \eta_{t+1}')'$ *and the filtration* $\mathcal{F}_{nt} = \sigma\{(\zeta_s)_{-\infty}^t\}$, *that is, the* σ*-field generated by* $(\zeta_s)_{s \le t}$. $\{\zeta_t, \mathcal{F}_{nt}\}$ *is a stationary and ergodic martingale difference sequence with* $E(\zeta_t \zeta_t' \mid \mathcal{F}_{n,t-1}) = \Sigma$ *and* $\sup_{t \ge 1} E\left(\|\zeta_t\|^r \left|\mathcal{F}_{n,t-1}\right| < \infty\right)$ *for some* $r > 4$.

(b) *Let* $\Delta \boldsymbol{x}_t = \boldsymbol{v}_t = \boldsymbol{\varphi}(L) \boldsymbol{\varepsilon}_t = \sum_{i=0}^\infty \varphi_i \boldsymbol{\varepsilon}_{t-i}$ *and* $\boldsymbol{z}_t = \phi(L) \boldsymbol{\eta}_t = \sum_{i=0}^\infty \phi_i \boldsymbol{\eta}_{t-i}$, *with* $\varphi_0 = I_{\ell_1}, \phi_0 = I_{\ell_2}$. *Furthermore,* $\varphi(1) = \sum_{i=0}^\infty \varphi_i$ *is nonsingular,* $\sum_{k=0}^\infty k \|\varphi_k\| < \infty$, *and* $\sum_{k=0}^\infty k^{1/2} \|\phi_k\| < \infty$.

(c) *There exists a nonsingular sequence of normalizing matrices* $\boldsymbol{\kappa}_{nd}$ *such that if* $\boldsymbol{d}_n(r) = \boldsymbol{\kappa}_{nd}^{-1}\boldsymbol{d}_{[nr]}$ *on* $[0,1]$, *then* $\sup_{n\geq 1}\sup_{0\leq r\leq 1}\|\boldsymbol{d}_n(r)\| \leq \infty$, *and* $\boldsymbol{d}_n \xrightarrow{L^2} \boldsymbol{d}$ *for some* $\boldsymbol{d} \in L^2[0,1]$ *such that* $\int_0^1 \boldsymbol{d}(r)\boldsymbol{d}(r)^\top dr > 0$. *The order of the first component of* $\boldsymbol{d}_t$, *denoted as* κ_{nd1}, *is largest among the components of* $\boldsymbol{\kappa}_{nd}$.

Assumption 1 (a) and (b) stipulate that the regressor $\boldsymbol{x}_t$ is an integrated process generated by a linear process $\boldsymbol{v}_t$, which has the martingale difference sequence $\{\boldsymbol{\varepsilon}_j, -\infty < j < \infty\}$ as building blocks, $\boldsymbol{z}_t$ is stationary, ergodic, and could be correlated with $\boldsymbol{x}_t$. In addition, the regressors $\boldsymbol{x}_t$ and $\boldsymbol{z}_t$ are predetermined, that is, $E(\boldsymbol{x}_t \mid \mathcal{F}_{n,t-1}) = \boldsymbol{x}_t$ and $E(\boldsymbol{z}_t \mid \mathcal{F}_{n,t-1}) = \boldsymbol{z}_t$. The nonsingularity of $\varphi(1)$ implies that there is no cointegrating relationship among the component time series in $\boldsymbol{x}_t$. See, for example, Phillips and Solo (1992) for more discussions on these conditions. The conditions in (c) are general enough to allow for deterministic regressors such as constant and time polynomials, possibly with breaks, which are commonly used in time-series analyses (see Park, 1992, for the asymptotics of integrated processes with such time trends). The convergence here is quite weak, as in most cases of practical interest we will have uniform convergence $\|\boldsymbol{d}_n - \boldsymbol{d}\| \to 0$.

For u_t and $\boldsymbol{v}_t$, we define stochastic processes

$$U_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} u_t, \qquad \text{and} \qquad \boldsymbol{V}_n(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} \boldsymbol{v}_t, \tag{3.1}$$

on $[0,1]$, where $\lfloor s\rfloor$ denotes the largest integer not exceeding s. The process $(U_n, \boldsymbol{V}_n^\top)^\top$ is defined in $D[0,1]^{1+\ell_1}$, where $D[0,1]$ is the space of cadlag functions on [0,1]. Under Assumption 1, an invariance principle holds for $\left(U_n, \boldsymbol{V}_n^\top\right)^\top$, that is, we have as $n \to \infty$, $\left(U_n, \boldsymbol{V}_n^\top\right)^\top \Rightarrow \left(U, \boldsymbol{V}^\top\right)^\top$, where $(U, \boldsymbol{V}^\top)^\top$ is $(1+\ell_1)$-dimensional vector Brownian motion, as shown in Phillips and Solo (1992).

The martingale difference assumption on the regression errors in Assumption 1 (a) is standard in nonlinear nonstationary time-series regression, such as Park and Phillips (2000, 2001), Chang et al. (2001), Dong et al. (2016), among others, and it can be relaxed like Chang and Park (2011), Chan and Wang (2015), and Hu et al. (2021). However, a generalization of our theory allowing for correlated errors would involve substantial level of complexity and is not attempted here. However, it does not seem overly restrictive at this point to assume the absence of serial correlation in the errors, especially given our flexible nonlinear specification of the transformation function and the inclusion of stationary regressors in the model.

The following assumption is standard in the extremum estimation theory.

Assumption 2. $\Theta = \Theta_0 \times \Theta_1 \times \Theta_2 \times \Theta_3$, *where* $\Theta_0 \subset \mathbb{R}^{\ell_0}, \Theta_1 \subset \mathbb{R}^{\ell_1}, \Theta_2 \subset \mathbb{R}^{\ell_2}$ *and* $\Theta_3 \subset \mathbb{R}^{\ell_3}$ *are convex and compact. And the* ℓ*-dimensional true parameter vector* $\boldsymbol{\vartheta}_0 = \left(\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_1^{0\top}, \boldsymbol{\theta}_2^{0\top}, \boldsymbol{\theta}_3^{0\top}\right)^\top$ *is an interior point of* Θ.

For the ease of presentation, we define $\boldsymbol{\xi}(x, \boldsymbol{\beta}) = \dot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x, \boldsymbol{\beta}_0), \boldsymbol{\beta})$, $\dot{\boldsymbol{\xi}}(x, \boldsymbol{\beta}) = \ddot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x, \boldsymbol{\beta}_0), \boldsymbol{\beta})$, $\ddot{\boldsymbol{\xi}}(x, \boldsymbol{\beta}) = \dddot{\boldsymbol{\Lambda}}(\Lambda^{-1}(x, \boldsymbol{\beta}_0), \boldsymbol{\beta})$, where $\Lambda^{-1}(x, \boldsymbol{\beta})$ is the inverse of $\Lambda(x, \boldsymbol{\beta})$ with respect to x, $\dot{\boldsymbol{\Lambda}}(x, \boldsymbol{\beta}) = \left(\partial\Lambda(x, \boldsymbol{\beta})/\partial\beta_i\right)_{\ell_0\times 1}$, $\ddot{\boldsymbol{\Lambda}}(x, \boldsymbol{\beta}) = \left(\partial^2\Lambda(x, \boldsymbol{\beta})/\partial\beta_i\partial\beta_j\right)_{\ell_0^2\times 1}$, $\dddot{\boldsymbol{\Lambda}}(x, \boldsymbol{\beta}) = \left(\partial^3\Lambda(x, \boldsymbol{\beta})/\partial\beta_i\partial\beta_j\partial\beta_k\right)_{\ell_0^3\times 1}$ are vectors, arrange by the lexicographic ordering of their indices.

Assumption 3.

(a) $\Lambda(\cdot,\boldsymbol{\beta})$ *is strictly increasing for any given* $\boldsymbol{\beta}$ *and is supposed to be three times continuously differentiable with respect to* $\boldsymbol{\beta}$. $\boldsymbol{\xi}(\cdot,\boldsymbol{\beta})$ *is an H-regular function with asymptotic order* $\boldsymbol{\kappa}_{\xi}(\cdot,\boldsymbol{\beta})$, *and limiting homogeneous function* $\boldsymbol{h}_{\xi}(\cdot,\boldsymbol{\beta})$.

(b) *Define a neighborhood of* $\boldsymbol{\beta}_0$ *by* $N(\varepsilon,\omega_1,\omega_2)=\left\{\boldsymbol{\beta}:\left\|\boldsymbol{\kappa}_{\xi}(\omega_1,\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\|\le\omega_2^{-1+\varepsilon}\right\}$ *for given* $\varepsilon>0$. *For any given* $\bar{s}>0$, *there exists* $\varepsilon>0$ *such that as* $\omega_1,\omega_2\to\infty$,

$$\left\|(\boldsymbol{\kappa}_{\xi}\otimes\boldsymbol{\kappa}_{\xi})^{-1}(\omega_1,\boldsymbol{\beta}_0)\left(\sup_{|s|\le\bar{s}}\dot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta}_0)\right)\right\|\to 0, \tag{3.2}$$

$$\omega_2^{-1+\varepsilon}\left\|\boldsymbol{\kappa}_{\xi}^{-1}(\omega_1,\boldsymbol{\beta}_0)\left(\sup_{|s|\le\bar{s}}\sup_{\beta\in N(\varepsilon,\omega_1,\omega_2)}\boldsymbol{\xi}(\omega_1 s,\boldsymbol{\beta})\right)\right\|\to 0, \tag{3.3}$$

$$\omega_2^{-1+\varepsilon}\left\|(\boldsymbol{\kappa}_{\xi}\otimes\boldsymbol{\kappa}_{\xi})^{-1}(\omega_1,\boldsymbol{\beta}_0)\left(\sup_{|s|\le\bar{s}}\sup_{\beta\in N(\varepsilon,\omega_1,\omega_2)}\dot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta})\right)\right\|\to 0, \tag{3.4}$$

$$\omega_2^{-1+\varepsilon}\left\|(\boldsymbol{\kappa}_{\xi}\otimes\boldsymbol{\kappa}_{\xi}\otimes\boldsymbol{\kappa}_{\xi})^{-1}(\omega_1,\boldsymbol{\beta}_0)\left(\sup_{|s|\le\bar{s}}\sup_{\beta\in N(\varepsilon,\omega_1,\omega_2)}\ddot{\boldsymbol{\xi}}(\omega_1 s,\boldsymbol{\beta})\right)\right\|\to 0, \tag{3.5}$$

where $\boldsymbol{\kappa}_{\xi}$ *are defined in Assumption 3 (a).*

The strictly increasing property of Λ in Assumption 3 (a) is commonly imposed for identification in transformation models. Assumption 3 (a) further stipulates that the function $\boldsymbol{\xi}$ is H-regular (see Definition B.2 in the Supplementary Document), and $\boldsymbol{\kappa}_{\xi},\boldsymbol{h}_{\xi}$ may depend on $\boldsymbol{\beta}_0$. For theoretical derivation, we can also consider the integrable function class as in Park and Phillips (2001), and the corresponding limiting theory can be obtained following Park and Phillips (2000), Chang and Park (2003), and Dong et al. (2016), etc. However, it is very hard to find a transformation function Λ such that the corresponding composite function $\boldsymbol{\xi}$ is integrable. Thus, we do not consider the integrable function class from the practical point of view and leave it as a future work. Assumption 3 (b) is similar to Assumption (b) of Theorem 5.3 in Park and Phillips (2001), and is required to prove a uniform convergence result. It holds for many H-regular functions used in nonlinear analysis. In addition, we may replace (3.3)–(3.5) with stronger, yet easier to verify conditions. See Park and Phillips (2001) for more related details.

3.2. Distribution Theory

The following gives the asymptotic distributions for the estimators $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\theta}}_n$.

Theorem 1. *Let Assumptions 1–3 hold. Assume that for all* $\delta > 0$,

$$\int_{|s|\le\delta} \boldsymbol{h}_{\xi,0}(s)\boldsymbol{h}_{\xi,0}^{\top}(s)ds > 0. \tag{3.6}$$

Then the following assertions hold as $n \to \infty$.

(a) *If* $\kappa_{nd1} / \sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$,

$$\begin{aligned}
&\sqrt{n}\boldsymbol{\kappa}_{\xi}\left(\sqrt{n}, \boldsymbol{\beta}_0\right)\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \Rightarrow \boldsymbol{\mathcal{B}}_1 \\
&\quad \equiv -\Big[\int_0^1 \tau_3^2(r)dr \int_0^1 \boldsymbol{h}_{\xi,0}(\tau_3(r))\boldsymbol{h}_{\xi,0}^{\top}(\tau_3(r))dr\Big]^{-1} \\
&\quad \times \Big\{\Big(\int_0^1 \tau_3^2(r)dr\Big)\Big[\int_0^1 \boldsymbol{h}_{\xi,0}(\tau_3(r))dU(r) + \sigma_u^2 \int_0^1 \boldsymbol{h}_{\xi,0}^{x}(\tau_3(r))dr\Big] \\
&\quad -\sigma_u^2[\textstyle\int_0^1 \tau_1(r)\boldsymbol{h}_{\xi,0}(\tau_3(r))dr + \pi\int_0^1 \tau_2(r)\boldsymbol{h}_{\xi,0}(\tau_3(r))dr]\Big\},
\end{aligned}$$

and $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \Rightarrow \boldsymbol{\Delta}_J^{-1}\boldsymbol{\Delta}_S + \begin{pmatrix} \boldsymbol{\mathcal{A}}_1^{\top} & \boldsymbol{0}^{\top} & \boldsymbol{\mathcal{A}}_2^{\top} \end{pmatrix}^{\top}$;

(b) *if* $\sqrt{n} / \kappa_{nd1} \to 0$,

$$\begin{aligned}
&\sqrt{n}\boldsymbol{\kappa}_{\xi}(\kappa_{nd1}, \boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \Rightarrow \boldsymbol{\mathcal{B}}_2 \\
&\quad \equiv -\Big(\int_0^1 \boldsymbol{h}_{\xi,0}(\tau_2(r))\boldsymbol{h}_{\xi,0}^{\top}(\tau_2(r))dr\Big)^{-1} \int_0^1 \boldsymbol{h}_{\xi,0}(\tau_2(r))dU(r),
\end{aligned}$$

and $\boldsymbol{D}_n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \Rightarrow \boldsymbol{\Delta}_J^{-1}\boldsymbol{\Delta}_S + \begin{pmatrix} \boldsymbol{\mathcal{A}}_3^{\top} & \boldsymbol{0}^{\top} & \boldsymbol{\mathcal{A}}_4^{\top} \end{pmatrix}^{\top}$,

where

$$\mathrm{D}_n = \begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 & 0 \\ 0 & \sqrt{n}\boldsymbol{I}_{\ell_2} & 0 \\ 0 & 0 & \sqrt{n}\boldsymbol{\kappa}_{nd} \end{pmatrix}, \boldsymbol{\Delta}_S = \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ N(0, \boldsymbol{\Omega}_{zu}) \\ \int_0^1 \boldsymbol{d}(r)dU(r) \end{pmatrix},$$

$$\Delta_J = \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}^{\top}(r)dr & 0 & \int_0^1 \boldsymbol{V}(r)\boldsymbol{d}^{\top}(r)dr \\ 0 & E\left[\boldsymbol{z}_t\boldsymbol{z}_t^{\top}\right] & 0 \\ \int_0^1 \boldsymbol{d}(r)\boldsymbol{V}^{\top}(r)dr & 0 & \int_0^1 \boldsymbol{d}(r)\boldsymbol{d}^{\top}(r)dr \end{pmatrix},$$

$\tau_1(r) = \boldsymbol{V}(r)^\top \boldsymbol{\theta}_1^0, \tau_2(r) = d_1(r)\theta_{31}^0, \qquad \tau_3(r) = \tau_1(r) + \pi\tau_2(r), \qquad \sigma_u^2 = \mathrm{var}(u_t),$

$\boldsymbol{h}_{\xi,0}(x) = \boldsymbol{h}_\xi(x, \beta_0), \ \boldsymbol{h}_{\xi,0}^x(x) = \partial \boldsymbol{h}_{\xi,0}(x)/\partial x, \ \boldsymbol{\Omega}_{zu} = \lim_{n\to\infty} n^{-1}\mathrm{var}\left(\sum_{t=1}^n z_t u_t\right),$

$$\mathcal{A}_1 = -\left(\int_0^1 \boldsymbol{V}(r)\boldsymbol{h}_{\xi,0}^\top(\tau_3(r))dr\right) \times \mathcal{B}_1, \mathcal{A}_2 = -\left(\int_0^1 \boldsymbol{d}(r)\boldsymbol{h}_{\xi,0}^\top(\tau_3(r))dr\right) \times \mathcal{B}_1,$$

$$\mathcal{A}_3 = -\left(\int_0^1 \boldsymbol{V}(r)\boldsymbol{h}_{\xi,0}^\top(\tau_2(r))dr\right) \times \mathcal{B}_2, \quad \text{and} \quad \mathcal{A}_4 = -\left(\int_0^1 \boldsymbol{d}(r)\boldsymbol{h}_{\xi,0}^\top(\tau_2(r))dr\right) \times \mathcal{B}_2.$$

Here, $d_1(r)$ and θ_{31}^0 are the first elements of $\boldsymbol{d}(r)$ and θ_3^0, respectively.

Remark 3.1. *The condition (3.6) is introduced for identification purpose and is similar to that of Park and Phillips (2001, Theorem 5.2) and Uematsu (2019, Theorem 3.1). Detailed discussions of these conditions are given in Park and Phillips (2001). Note that the limiting distributions for nonstationary parameters ($\boldsymbol{\beta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3$) in the two cases are nonstandard and likely to have an asymptotic bias. However, how to construct the bias-corrected estimator in this setup remains a challenging issue, and is beyond the scope of this study. See Chang et al. (2001) and Chan and Wang (2015) for similar issues.*

Remark 3.2. *The rate of convergence and limiting distribution of $\hat{\boldsymbol{\beta}}_n$ are nonstandard and depend on $\boldsymbol{\beta}_0$, Λ and the relatively size of κ_{nd1} and $\sqrt{n}$. When $\kappa_{nd1}/\sqrt{n} \to 0$, $\hat{\boldsymbol{\beta}}_n$ converges at rate $\sqrt{n}\boldsymbol{\kappa}_\xi\left(\sqrt{n}, \beta_0\right)$; when $\sqrt{n}/\kappa_{nd1} \to 0$, the convergence rate becomes $\sqrt{n}\boldsymbol{\kappa}_\xi(\kappa_{nd1}, \boldsymbol{\beta}_0)$. The result is similar to that in* Lin et al. (2020), *which also found that the limiting distribution of the transformation parameter is nonstandard with the rate of convergence depending on the property of the unknown transformations. In addition, the convergence rates of the estimators for the coefficients of the stationary, unit root, and deterministic regressors are given, respectively, by $\sqrt{n}$, n, and $\sqrt{n}\boldsymbol{\kappa}_{nd}$, as in standard regressions. The limiting distribution of the estimators of the coefficients before the stationary regressors is normal. However, the asymptotic distributions of the estimators of time trend or unit root coefficients are generally non-Gaussian. As shown in Theorem 1, the limiting distribution for the estimators of time trend or unit root coefficients contains two parts. The first part depends on $\boldsymbol{V}(r)$ and U(r), and is Gaussian as long as $\boldsymbol{V}$ and U are asymptotically independent; the second part accounts for the estimation effect induced by the transformation parameters. Such a finding is brand new in the literature and has not been discovered in Chang et al. (2001), Lin et al. (2020), Lin and Tu (2021b), etc. The limiting distributions are thus not centered, and the usual chi-squared approach to inference is not possible. The critical values of the usual tests are dependent upon nuisance parameters. See Park and Phillips (2001), Chang and Park (2003), etc., for similar discussions. While in some specific cases, these limiting distributions will degenerate to Gaussian. For example, when the transformation Λ is the Box–Cox function, and $\boldsymbol{V}(r)$ and U(r) are asymptotically independent, the limiting distribution for the estimators of the time trend and unit root coefficients is* *mixed normal and the conventional inference can be applied. See Examples 3.1–3.2 and Section 4 for details.*

Remark 3.3. *The limiting distribution of $\hat{\boldsymbol{\theta}}_{2n}$ is not affected by $\hat{\boldsymbol{\beta}}_n$, and is identical to that of the least squares estimator in $\Lambda(y_t,\boldsymbol{\beta}_0)=\boldsymbol{z}_t^\top\boldsymbol{\theta}_2^0+u_t$, that is, the stationary regressors are asymptotically orthogonal to the integrated regressors and the deterministic trends, as found in Chang et al. (2001). However, the asymptotic distributions of $\hat{\boldsymbol{\theta}}_{1n}$ and $\hat{\boldsymbol{\theta}}_{3n}$ both depend on that of $\hat{\boldsymbol{\beta}}_n$, thus affect each other. This finding is different from the findings in Chang et al. (2001) and Kim and Kim (2012) because of the presence of transformation function Λ.*

Example 3.1. *As an illustrative example, we look at the regression with the constant deterministic component $d_t = 1$. Consider the Box–Cox transformation (Box & Cox, 1964) function $\Lambda(y,\beta)=(y^\beta-1)/\beta,\beta\neq 0$, it is easy to show that $\xi(y,\beta)=(\beta\beta_0)^{-1}(\beta_0 y+1)^{\beta/\beta_0}\ln(\beta_0 y+1)-\beta^{-2}(\beta_0 y+1)^{\beta/\beta_0}+\beta^{-2}$, the homogeneous function is $h_\xi(y,\beta)=\dfrac{1}{\beta\beta_0}(\beta_0 y)^{\beta/\beta_0}$, and the asymptotic order is $\kappa_\xi(\lambda,\beta)=\lambda^{\beta/\beta_0}\ln(\lambda)$. Then, by Theorem 1, we have*

$$n\ln\left(\sqrt{n}\right)\left(\hat{\beta}_n/\beta_0-1\right)\Rightarrow \mathcal{B}_1 \quad = \quad -\Big(\int_0^1\tau_1^2(r)dr\Big)^{-1}\int_0^1\tau_1(r)dU(r), \tag{3.7}$$

$$\begin{aligned}\begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0\\ 0 & \sqrt{n}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\theta}}_{n1}-\boldsymbol{\theta}_1^0\\ \hat{\theta}_{n3}-\theta_3^0\end{pmatrix} &\Rightarrow \begin{pmatrix}\int_0^1\boldsymbol{V}(r)\boldsymbol{V}^\top(r)dr & \int_0^1\boldsymbol{V}(r)dr\\ \int_0^1\boldsymbol{V}^\top(r)dr & 1\end{pmatrix}^{-1}\\ &\quad \begin{pmatrix}\int_0^1\boldsymbol{V}(r)dU(r)\\ \int_0^1 dU(r)\end{pmatrix}\\ &\quad -\begin{pmatrix}\Big(\int_0^1\boldsymbol{V}(r)\tau_1(r)dr\Big)\times\mathcal{B}_1\\ \Big(\int_0^1\tau_1(r)dr\Big)\times\mathcal{B}_1\end{pmatrix},\end{aligned} \tag{3.8}$$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}_{n2}}-\boldsymbol{\theta}_2^0\right) \quad\Rightarrow\quad E[\boldsymbol{z}_t\boldsymbol{z}_t^\top]^{-1}N(0,\Omega_{zu}). \tag{3.9}$$

In addition, we consider the monotonic transformation function $\Lambda(y,\beta)=e^{\beta y}$. Similarly, by a simple calculation, we have $\xi=\dfrac{1}{\beta_0}y^{\beta/\beta_0}\ln(y)$, with the homogeneous function $h_\xi(y,\beta)=\dfrac{1}{\beta_0}y^{\beta/\beta_0}$ and the asymptotic order $\kappa_\xi(\lambda,\beta)=\lambda^{\beta/\beta_0}\ln(\lambda)$. The limiting results for model

$$e^{\beta y_t}=\boldsymbol{x}_t^\top\boldsymbol{\theta}_1^0+\boldsymbol{z}_t^\top\boldsymbol{\theta}_2^0+\boldsymbol{d}_t^\top\boldsymbol{\theta}_3^0+u_t,$$

can be obtained following Theorem 1, which are the same as (3.7)–(3.9).

Example 3.2. *Consider the model with the linear deterministic component $d_t = t$ and the Box–Cox transformation $\Lambda(y,\beta) = (y^{\beta}-1)/\beta$, $\beta \neq 0$. By Theorem 1, we have*

$$n^{3/2}\ln(n)\left(\hat{\beta}_n/\beta_0-1\right) \Rightarrow \mathcal{B}_2 = -\Big(\int_0^1 \tau_2^2(r)dr\Big)^{-1}\int_0^1 \tau_2(r)dU(r), \tag{3.10}$$

$$\begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 \\ 0 & n^{3/2}\end{pmatrix}\begin{pmatrix}\boldsymbol{\theta}_{n1}-\boldsymbol{\theta}_1^0 \\ \hat{\theta}_{n3}-\theta_3^0\end{pmatrix} \Rightarrow \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}^{\top}(r)^{-1}dr & \int_0^1 \boldsymbol{V}(r)dr \\ \int_0^1 \boldsymbol{V}^{\top}(r)dU(r) & 1 \end{pmatrix}^{-1}$$

$$\begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ \int_0^1 dU(r) \end{pmatrix}.$$

$$-\begin{pmatrix} \Big(\int_0^1 \boldsymbol{V}(r)\tau_2(r)dr\Big)\times\mathcal{B}_2 \\ \Big(\int_0^1 \tau_2(r)dr\Big)\times\mathcal{B}_2 \end{pmatrix}, \tag{3.11}$$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}_{n2}}-\boldsymbol{\theta}_2^0\right) \Rightarrow E\left[\boldsymbol{z}_t\boldsymbol{z}_t^{\top}\right]^{-1}N(0,\Omega_{zu}). \tag{3.12}$$

For the exponential transformation $\Lambda(y,\beta) = e^{\beta y}$, by Theorem 1, we also have (3.10)–(3.12) hold.

The finite sample performance of the estimators in the above two examples shall be investigated in the following section via Monte Carlo simulations.

4. SIMULATION

This section investigates the finite sample performance of the proposed estimators. To this end, consider the model

$$\Lambda(y_t,\beta_0) = \boldsymbol{x}_t^{\top}\boldsymbol{\theta}_1^0 + z_t^{\top}\boldsymbol{\theta}_2^0 + \theta_3^0 d_t + u_t, \qquad t = 1,2,\ldots,n,$$

where $d_t \in \{1, t\}$, $\boldsymbol{\theta}_1^0 = (1.5,1)^{\top}$, $\boldsymbol{\theta}_2^0 = (0.2,0.4)^{\top}$, and $\theta_3^0 = 2$. Let $\boldsymbol{x}_t = (x_{1t},x_{2t})^{\top}$ be generated by $\boldsymbol{x}_t = \boldsymbol{x}_{t-1}+\boldsymbol{v}_t$, where $\left(\boldsymbol{v}_t^{\top},\boldsymbol{z}_t^{\top}\right)^{\top}$ is a bivariate normal vector with zero mean and covariance matrix $\{1,0,\rho,0;0,1,0,\rho;\rho,0,0.5,0;0,\rho,0,0.5\}$, for $\rho = 0, 0.3, 0.7$, and the error u_t is $N(0,0.5^2)$. The transformation function Λ is set as the Box–Cox transformation, $\Lambda(y,\beta_0) = (y^{\beta_0}-1)/\beta_0$ (M1) and $\Lambda(y,\beta) = e^{\beta y}$ (M2) for $\beta_0 = 0.8, 1, 1.2$. In our simulations, we draw samples of sizes n = 250 and 500 to obtain the proposed estimators and their t ratios. The bias, standard deviation (SD) and root of mean squared errors (RMSEs) of each estimator are calculated over 1, 000 replications. Kernel densities of the t-ratios are computed using the standard normal kernel with the rule-of-thumb bandwidth. Due to space limitation, we only report some selected results. Other results are similar and are contained in the Supplementary Document.

The findings are summarized as follows. First, the finite sample performances of the estimators are quite close to what would be expected from the limit theory. As can be seen from Tables 1 and 2, the biases of the estimators are close to zero, and the RMSE of the estimators are small and decrease as the sample size increases in all cases. This confirms that our estimators are accurate in all scenarios. The realized ratios of the RMSEs for $n = 250$ to those for $n = 500$ are close to the theoretical counterparts as shown in Tables 3 and 4, which are consistent with the convergence rates obtained in Examples 3.1–3.2. In particular, the estimator for transformation parameter $\hat{\beta}_n$ converges the fastest among all the estimators, for both cases with $d_t = 1$ or $d_t = t$. When $d_t = 1$, the estimators in stationary and time trend parts $\left(\hat{\boldsymbol{\theta}}_{2n} \text{ and } \hat{\theta}_{3n}\right)$ converge at the same rate, which is slower than that of the estimators for the coefficients before unit root processes $\hat{\boldsymbol{\theta}}_{1n}$. When $d_t = t$, $\hat{\theta}_{3n}$ converges faster than $\hat{\boldsymbol{\theta}}_{1n}$ does, and the latter converges much faster than $\hat{\boldsymbol{\theta}}_{2n}$ does. Second, the sampling behavior of the t-ratios of the estimators largely corroborates with our theoretical results in Section 3. The asymptotic theory demonstrated in Examples 3.1–3.2 shows that the limiting distributions for all estimators are standard (either normal or mixed normal). It has been verified in Fig. 1 that the kernel density curves of t-ratios are symmetric and centered around zero, which are approximating the limiting standard normal density reasonably well. These findings are robust to various specifications of Λ, β_0, ρ, d_t, and sample size n. Note that in other simulation settings, the limiting distributions for the estimators of the transformation parameters, unit root and time trend coefficients may not centered, and the associated t-ratios will have different performances, which should be discussed case by case.

Table 1. Bias, SD, and RMSE ($\times 10^3$) for $\hat{\beta}_n$ and $\hat{\theta}_n$ with $\rho = 0.3$, M1.

n		250						500					
β_0		$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$	$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$
							$d_t = 1$						
0.8	Bias	−0.04	−0.12	−0.27	1.43	−1.41	1.33	−0.02	−0.15	0.00	0.71	−1.25	1.58
	SD	0.34	7.41	7.12	45.9	44.1	116	0.14	3.91	3.84	31.4	30.4	89.0
	RMSE	0.34	7.41	7.12	45.9	44.1	116	0.14	3.91	3.84	31.4	30.5	89.0
1	Bias	−0.02	−0.37	0.09	−1.55	−1.34	1.01	−0.01	−0.06	−0.10	−0.84	1.70	1.07
	SD	0.38	7.24	7.74	45.6	46.4	121	0.17	3.84	3.77	31.7	32.8	87.5
	RMSE	0.38	7.25	7.74	45.7	46.4	121	0.17	3.84	3.77	31.7	32.8	87.5
1.2	Bias	−0.02	−0.02	−0.15	−1.08	−1.40	−0.87	−0.02	0.09	−0.21	−0.99	−0.01	0.56
	SD	0.41	7.43	7.20	44.4	46.5	126	0.19	3.77	3.95	32.3	32.1	86.9
	RMSE	0.41	7.43	7.20	44.4	46.6	126	0.19	3.77	3.96	32.3	32.1	86.9
							$d_t = t$						
0.8	Bias	0.00	−0.10	0.08	−0.28	−1.77	0.01	0.00	−0.08	−0.02	−1.39	−1.07	0.00
	SD	0.02	7.17	7.37	47.2	44.8	0.72	0.01	3.74	3.81	31.5	31.5	0.25
	RMSE	0.02	7.17	7.37	47.2	44.8	0.72	0.01	3.74	3.81	31.5	31.5	0.25
1	Bias	0.00	−0.23	−0.09	−0.50	2.06	0.02	0.00	0.02	−0.10	0.27	−1.15	0.00
	SD	0.02	7.43	7.56	46.1	44.8	0.71	0.01	3.60	3.52	30.1	32.3	0.26
	RMSE	0.02	7.43	7.57	46.1	44.8	0.71	0.01	3.60	3.52	30.1	32.3	0.26
1.2	Bias	0.00	−0.36	0.20	−2.28	0.76	0.00	0.00	0.05	−0.25	−0.39	0.66	0.01
	SD	0.02	7.00	7.15	44.5	44.8	0.72	0.01	3.72	3.62	32.1	30.5	0.25
	RMSE	0.02	7.01	7.15	44.6	44.8	0.72	0.01	3.72	3.63	32.1	30.5	0.25

Table 2. Bias, SD, and RMSE ($\times 10^3$) for $\hat{\beta}_n$ and $\hat{\theta}_n$ with $\rho = 0$, M2.

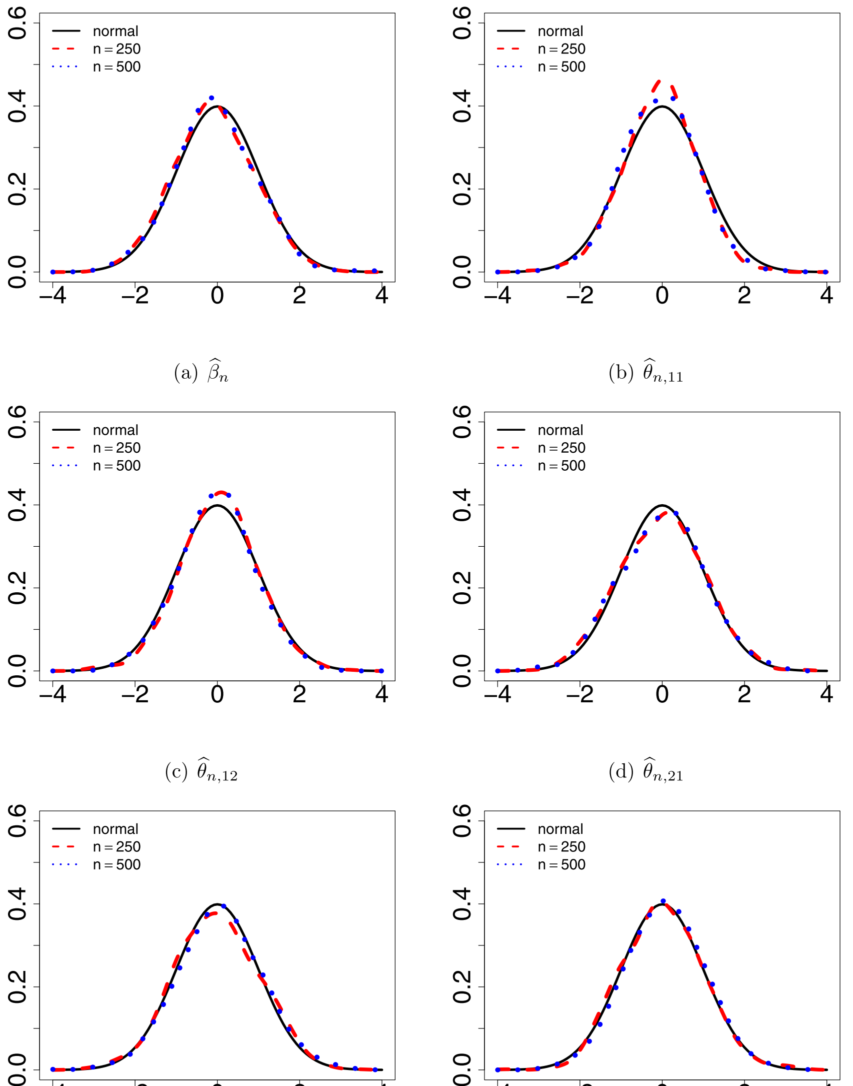
n		250						500					
β_0		$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$	$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$
							$d_t = 1$						
0.8	Bias	−0.01	−0.26	−0.22	2.45	1.03	2.28	0.00	0.21	−0.04	0.29	0.04	−4.96
	SD	0.23	7.41	7.78	44.6	43.0	130	0.11	3.78	3.81	31.9	32.9	91
	RMSE	0.23	7.42	7.78	44.7	43.0	130	0.11	3.78	3.81	31.9	32.9	91
1	Bias	−0.03	−0.27	−0.01	0.63	1.23	−0.24	−0.01	0.00	−0.04	−0.04	0.69	−0.28
	SD	0.28	6.92	7.00	44.0	46.0	119	0.13	3.68	3.83	32.6	30.9	87
	RMSE	0.28	6.93	7.00	44.0	46.1	119	0.13	3.68	3.83	32.6	30.9	87
1.2	Bias	−0.02	−0.20	−0.10	−0.76	−1.53	0.65	−0.01	0.00	−0.22	0.59	1.25	3.03
	SD	0.36	7.66	7.54	44.2	46.2	120	0.16	3.90	3.67	32.4	31.6	86
	RMSE	0.36	7.67	7.54	44.3	46.2	120	0.17	3.90	3.67	32.4	31.6	86
							$d_t = t$						
0.8	Bias	0.00	0.01	0.15	−0.30	−3.49	0.00	0.00	0.00	0.03	1.49	1.57	0.00
	SD	0.01	7.14	7.24	47.1	46.1	0.68	0.00	3.45	3.63	30.8	32.1	0.24
	RMSE	0.01	7.14	7.24	47.1	46.3	0.68	0.00	3.45	3.63	30.9	32.2	0.24
1	Bias	0.00	0.16	−0.31	−1.77	−1.21	0.02	0.00	0.22	−0.09	−0.91	−1.69	−0.01
	SD	0.02	7.05	7.12	46.3	44.5	0.65	0.01	3.51	3.54	31.6	33.3	0.24
	RMSE	0.02	7.05	7.13	46.4	44.5	0.66	0.01	3.51	3.54	31.6	33.4	0.24
2	Bias	0.00	0.26	−0.20	0.84	−0.45	0.01	0.00	0.06	0.01	−2.81	1.18	−0.01
	SD	0.02	7.15	7.03	45.4	45.2	0.66	0.01	3.72	3.58	30.7	31.6	0.26
	RMSE	0.02	7.15	7.03	45.4	45.2	0.66	0.01	3.72	3.58	30.9	31.6	0.26

Table 3. Ratios of the RMSEs for $n = 250$ to those for $n = 500$ with $\rho = 0.3$, M1.

		β_0	$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.385	1.895	1.857	1.463	1.449	1.300
		1	2.193	1.885	2.053	1.439	1.413	1.379
		1.2	2.146	1.967	1.818	1.375	1.450	1.450
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.138	1.917	1.936	1.499	1.425	2.846
		1	3.093	2.065	2.150	1.533	1.387	2.701
		1.2	3.100	1.882	1.970	1.388	1.470	2.909

Table 4. Ratios of the RMSEs for $n = 250$ to those for $n = 500$ with $\rho = 0$, M2.

		β_0	$\hat{\beta}_n$	$\hat{\theta}_{n,11}$	$\hat{\theta}_{n,12}$	$\hat{\theta}_{n,21}$	$\hat{\theta}_{n,22}$	$\hat{\theta}_{n,3}$
$d_t = 1$	Theoretical		2.215	2	2	1.414	1.414	1.414
	Realized	0.8	2.177	1.961	2.044	1.400	1.310	1.425
		1	2.110	1.881	1.826	1.349	1.491	1.358
		1.2	2.195	1.968	2.052	1.364	1.460	1.391
$d_t = t$	Theoretical		3.183	2	2	1.414	1.414	2.828
	Realized	0.8	3.135	2.067	1.995	1.527	1.439	2.847
		1	3.095	2.009	2.013	1.468	1.334	2.723
		1.2	2.959	1.920	1.964	1.473	1.429	2.494



(a) $\widehat{\beta}_n$ (b) $\widehat{\theta}_{n,11}$

(c) $\widehat{\theta}_{n,12}$ (d) $\widehat{\theta}_{n,21}$

(e) $\widehat{\theta}_{n,22}$ (f) $\widehat{\theta}_{n,3}$

Fig. 1. Kernel Density of t-ratios with $d_t = 1, \beta_0 = 1, \rho = 0$, M1. (a) $\hat{\beta}_n$. (b) $\hat{\theta}_{n,11}$. (c) $\hat{\theta}_{n,12}$. (d) $\hat{\theta}_{n,21}$. (e) $\hat{\theta}_{n,22}$. (f) $\hat{\theta}_{n,3}$.

In addition, we compute the percentages of rejection for t-ratios to test the null hypothesis $H_0 : \varsigma = \varsigma_0 + j$ ($\varsigma = \beta, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}, \theta_3$), at significance level $\alpha = 0.01, 0.05, 0.10$, where j is taken from a grid formed between -0.20 and 0.20 with an equal increment 0.05. Form the figures in the Supplementary Document, we conclude that the sizes of the t-ratios ($j = 0$) are close to the nominal level α for various combinations of Λ, ρ, β, and α. Furthermore, the power of t-ratios ($j \neq 0$) for $\varsigma = \beta, \theta_{11}, \theta_{12}$ is close to 1 when $\varsigma \neq \varsigma_0$ even with sample size 250, and that for θ_{21} and θ_{22} becomes larger as the absolute value of j increases and grows with the sample size n. For θ_3, the performance of power of t-ratios is affected by the choice of d_t. When $d_t = 1$, the power increases as $|j|$ and n become larger; when $d_t = t$, the power is close to 1 for various combinations of n and j. Overall, the t-ratios enjoy nice finite sample performance.

5. CONCLUSION

This chapter considers a transformation model with a time trend, stationary regressors, as well as multiple I(1) regressors, which is a hybrid of a transformation model and a conventional cointegration model. Estimation of the unknown quantities is investigated and an asymptotic theory of the proposed estimators is established. Numerical results demonstrate the nice performance of the estimators and corroborate the limiting results.

There are several possible directions to extend the chapter.

1. The right-hand side of our model is linear in parameters, which can be extended to the general nonlinear setting, such as the index model of Park and Phillips (2000) and Chang and Park (2003), the additive model of Chang et al. (2001), or the (partially linear) single index model of Dong et al. (2016).
2. Endogeneity could be incorporated in the current setting and worths consideration in the future work. Some further development following Chang and Park (2011) and Chan and Wang (2015) could be made.
3. Specification test for the parametric form of the transformation and the test for the existence of such cointegration relationship are still underdeveloped. These issues are technically involved and are left as future research.

ACKNOWLEDGMENTS

Lin acknowledges the support of the Fundamental Research Funds for the Central Universities (China) (2021110462), the Chenguang Program of Shanghai Education Development Foundation and Shanghai Municipal Education Commission (Grant 21CGA43), the National Natural Science Foundation of China (Grant 72203141). Tu (corresponding author) would like to thank support from National Natural Science Foundation of China (Grant 72073002, 12026607, 92046021), the Center for Statistical Science at Peking University, and Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education.

REFERENCES

Abrevaya, J. (1999). Rank estimation of a transformation model with observed truncation. *Econometrics Journal*, *2*(2), 292–305.

Bae, Y., & De Jong, R. M. (2007). Money demand function estimation by nonlinear cointegration. *Journal of Applied Econometrics*, *22*(4), 767–793.

Bae, Y., Kakkar, V., & Ogaki, M. (2006). Money demand in Japan and nonlinear cointegration. *Journal of Money, Credit, and Banking*, *38*(6), 1659–1667.

Bickel, P. J., & Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, *76*(374), 296–311.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–243.

Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, *80*(391), 580–598.

Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, *148*(2), 101–113.

Chan, N., & Wang, Q. (2014). Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory*, *30*(5), 1110–1133.

Chan, N., & Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, *185*(1), 182–195.

Chang, Y., Jiang, B., & Park, J. Y. (2012). Nonstationary regression with logistic transition. *The Econometrics Journal*, *15*(2), 255–287.

Chang, Y., & Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, *114*(1), 73–106.

Chang, Y., & Park, J. Y. (2011). Endogeneity in nonlinear regressions with integrated time series. *Econometric Reviews*, *30*(1), 51–87.

Chang, Y., Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal*, *4*(1), 1–36.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, *70*(4), 1683–1697.

Chiappori, P.-A., Komunjer, I., & Kristensen, D. (2015). Nonparametric identification and estimation of transformation models. *Journal of Econometrics*, *188*(1), 22–39.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, *55*(2), 251–276.

Dong, C., & Gao, J. (2018). Specification testing in structural nonparametric cointegration. *Econometric Theory*, *34*(4), 754–789.

Dong, C., Gao, J., & Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *Annals of Statistics*, *44*(1), 425–453.

Dong, C., Gao, J., Tjøstheim, D., & Yin, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics*, *200*(1), 104–117.

Dong, C., & Linton, O. (2018). Additive nonparametric models with time variable and both stationary and nonstationary regressors. *Journal of Econometrics*, *207*(1), 212–236.

Dong, C., Linton, O., & Peng, B. (2021). A weighted sieve estimator for nonparametric time series models with nonstationary variables. *Journal of Econometrics*, *222*(2), 909–932.

Fan, C., & Fine, J. P. (2013). Linear transformation model with parametric covariate transformations. *Journal of the American Statistical Association*, *108*(502), 701–712.

Florens, J.-P., & Sokullu, S. (2017). Nonparametric estimation of semiparametric transformation models. *Econometric Theory*, *33*(4), 839–873.

Gao, J., & Phillips, P. C. B. (2013). *Functional coefficient nonstationary regression with non- and semi-parametric cointegration*. Working Paper, Monash University.

Granger, C. (1991). Some recent generalisations of cointegration and the analysis of longrun relationships. In R. Engle & C. Granger (Eds.), *Long-run economic relationships* (pp. 277–287). Oxford University Press.

Granger, C., & Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford University Press.

Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*(1), 121–130.

Han, A. K. (1987). A non-parametric analysis of transformations. *Journal of Econometrics*, *35*(2–3), 191–209.

Hirukawa, M., & Sakudo, M. (2018). Functional-coefficient cointegration models in the presence of deterministic trends. *Econometric Reviews*, *37*(5), 507–533.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, *64*(1), 103–137.

Hu, Z., Phillips, P. C. B., & Wang, Q. (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory*. *37*(6), 1173–1213. doi:10.1017/S0266466620000560.

Juhl, T. (2005). Functional-coefficient models under unit root behaviour. *Econometrics Journal*, *8*(2), 197–213.

Karlsen, H. A., Myklebust, T., & Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *The Annals of Statistics*, *35*(1), 252–299.

Kasparis, I., Andreou, E., & Phillips, P. C. (2015). Nonparametric predictive regression. *Journal of Econometrics*, *185*(2), 468–494.

Kasparis, I., & Phillips, P. C. B. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*, *168*(2), 270–284.

Kim, C. S., & Kim, I. M. (2012). Partial parametric estimation for nonstationary nonlinear regressions. *Journal of Econometrics*, *167*(2), 448–457.

Lewbel, A., Lu, X., & Su, L. (2015). Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, *184*(1), 81–96.

Lin, Y., & Tu, Y. (2020). Sieve extremum estimation of a semiparametric transformation model. *Economics Letters*, *189*, 109020.

Lin, Y., & Tu, Y. (2021a). *Identification and estimation of a semiparametric single index transformation model*. Working Paper, Peking University.

Lin, Y., & Tu, Y. (2021b). On transformed linear cointegration models. *Economics Letters*, *198*, 109686.

Lin, Y., Tu, Y., & Yao, Q. (2020). Estimation for double-nonlinear cointegration. *Journal of Econometrics*, *216*(1), 175–191.

Linton, O., & Wang, Q. (2016). Nonparametric transformation regression with nonstationary data. *Econometric Theory*, *32*(1), 1–29.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica*, *60*(1), 119–143.

Park, J. Y. (2006). Nonstationary nonlinearity: An outlook for new opportunities. In D. Corbae, S. N. Durlauf, & B. E. Hansen (Eds.), *Econometric theory and practice*: *Frontiers of analysis and applied research* (pp. 178–211). Cambridge University Press.

Park, J. Y., & Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, *15*(5), 664–703.

Park, J. Y., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*(3), 468–497.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*(1), 95–131.

Park, J. Y., & Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, *15*(3), 269–298.

Park, J. Y., & Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, *68*(5), 1249–1280.

Park, J. Y., & Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica*, *69*(1), 117–161.

Phillips, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory*, *25*(6), 1466–1497.

Phillips, P. C. B., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, *53*(4), 473–495.

Phillips, P. C. B., Li, D., & Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, *196*(1), 180–195.

Phillips, P. C. B., & Park, J. Y. (1998). Nonstationary density estimation and Kernel autoregression. Unpublished Manuscript.

Phillips, P. C. B., & Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, *20*(2), 971–1001.

Teräsvirta, T., Tjøstheim, D., & Granger, C. W. J. (2010). *Modelling nonlinear economic time series*. Oxford University Press.

Tu, Y., Liang, H.-Y., & Wang, Q. (2021). Nonparametric inference for quantile cointegrations with stationary covariates. *Journal of Econometrics*, *230*(2), 453–482.

Tu, Y., & Wang, Y. (2019). Functional coefficient cointegration models subject to time-varying volatility with an application to the purchasing power parity. *Oxford Bulletin of Economics and Statistics*, *81*(6), 1401–1423.

Tu, Y., & Wang, Y. (2020). Adaptive estimation of heteroskedastic functional-coefficient regressions with an application to fiscal policy evaluation on asset markets. *Econometric Reviews*, *39*(3), 299–318.

Tu, Y., & Wang, Y. (2022). Spurious functional-coefficient regression models and robust inference with marginal integration. *Journal of Econometrics*, *229*(2), 396–421.

Uematsu, Y. (2019). Nonstationary nonlinear quantile regression. *Econometric Reviews*, *38*(4), 386–416.

Vanhems, A., & Van Keilegom, I. (2019). Estimation of a semiparametric transformation model in the presence of endogeneity. *Econometric Theory*, *35*(1), 73–110.

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression*. World Scientific.

Wang, Q., & Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, *25*(3), 710–738.

Wang, Q., & Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, *77*(6), 1901–1948.

Wang, Q., & Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, *40*(2), 727–758.

Wang, Q., & Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory*, *32*(2), 359–401.

Wang, N., & Ruppert, D. (1995). Nonparametric estimation of the transformation in the transform-both-sides regression model. *Journal of the American Statistical Association*, *90*(430), 522–534.

Wang, Q., Wu, D., & Zhu, K. (2018). Model checks for nonlinear cointegrating regression. *Journal of Econometrics*, *207*(2), 261–284.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, *152*(2), 81–92.

APPENDIX

The proof of Theorem 1 is given below, while all lemmas used in this proof are given in the Supplementary Document.

Proof of the Main Theorem

Proof of Theorem 1. We have two cases, that is, Case (a), if $\kappa_{nd1}/\sqrt{n} \to \pi \in \{0, \mathbb{R}^+\}$; Case (b), if $\sqrt{n}/\kappa_{nd1} \to 0$. Because the arguments used to prove the two cases are similar, we present the proofs for Case (a) below, with the detailed proofs for Case (b) omitted. This proof contains four parts. We first investigate the limiting distribution of $\hat{\boldsymbol{\beta}}_n$. Specifically, Part I gives the score and hessian, Part II establishes their joint asymptotics, and Part III includes a detailed proof for the limit distribution of $\hat{\boldsymbol{\beta}}_n$. Then Part IV proves that of $\hat{\boldsymbol{\theta}}_n$.

Part I – The Loss Function. Since n^2 does not rely on $\boldsymbol{\beta}$, then minimizing $L_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is equivalent to minimizing

$$\tilde{L}_n(\boldsymbol{\beta}) = \frac{n^2}{2} \frac{\sum_{t=1}^{n}\left[\Lambda(y_t, \boldsymbol{\beta}) - \boldsymbol{w}_t' \boldsymbol{\theta}_0\right]^2}{\sum_{t=1}^{n}\Lambda(y_t, \boldsymbol{\beta})^2}. \tag{A.1}$$

Therefore, the score function is

$$S_n(\boldsymbol{\beta}) = n^2 A_{n1}^{-2}(\boldsymbol{\beta})[A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}) - A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})], \tag{A.2}$$

and the hessian matrix is

$$\begin{aligned} J_n(\boldsymbol{\beta}) &= n^2 A_{n1}^{-3}(\boldsymbol{\beta})\big\{A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n6}(\boldsymbol{\beta})] - A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta}) + A_{n7}(\boldsymbol{\beta})] \\ &\quad - 2A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})A_{n4}^{\top}(\boldsymbol{\beta}) + A_{n4}(\boldsymbol{\beta})A_{n3}^{\top}(\boldsymbol{\beta})] + 4A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})A_{n3}^{\top}(\boldsymbol{\beta})\big\} \\ &\equiv n^2 J_{n1}(\boldsymbol{\beta}) / J_{n2}(\boldsymbol{\beta}), \end{aligned} \tag{A.3}$$

where the definitions of J_{n1}, J_{n2} should be obvious.

Part II – The Score and the Hessian. In this case, $\bar{\boldsymbol{\kappa}}_{n\xi,0} = \boldsymbol{\kappa}_{n\xi,0}$. Let $\boldsymbol{D}_n = \sqrt{n}\boldsymbol{\kappa}_{n\xi,0}$. By Lemma C.2, we have

$$
\begin{aligned}
\boldsymbol{D}_n^{-1}S_n(\boldsymbol{\beta}_0) &= \left(n^{-2}A_{n1}(\boldsymbol{\beta}_0)\right)^{-2}\left[n^{-2}A_{n1}(\boldsymbol{\beta}_0)\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}\right. \\
&\quad \left. A_{n4}(\boldsymbol{\beta}_0) - n^{-1}A_{n2}(\boldsymbol{\beta}_0)\,(n^{3/2}\boldsymbol{\kappa}_{n\xi,0})^{-1}A_{n3}(\boldsymbol{\beta}_0)\right] \\
&\Rightarrow \left(\int_0^1 \tau_3^2(r)dr\right)^{-2}\left\{\left(\int_0^1 \tau_3^2(r)dr\right)\left[\int_0^1 \boldsymbol{h}_{\xi,0}\right.\right. \\
&\quad \left.(\tau_3(r))dU(r) + \sigma_u^2 \int_0^1 \boldsymbol{h}_{\xi,0}^x(\tau_3(r))dr\right] \\
&\quad \left. - \ \sigma_u^2 \int_0^1 \tau_3(r)\boldsymbol{h}_{\xi,0}(\tau_3(r))dr\right\}, \\
\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top} &= (n^{-2}A_{n1}(\boldsymbol{\beta}_0))^{-1}\cdot\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}A_{n5}(\boldsymbol{\beta}_0) \\
&\quad \left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1} + o_p(1) \\
&\Rightarrow \left(\int_0^1 \tau_3^2(r)dr\right)^{-1}\int_0^1 \boldsymbol{h}_{\xi,0}(\tau_3(r))\boldsymbol{h}_{\xi,0}^{\top}(\tau_3(r))dr \qquad \text{(A.4)}
\end{aligned}
$$

And the joint convergence of S_n and J_n follows from the joint convergence results in Lemma C.2.

Part III – Detailed proof for the limit of $\hat{\boldsymbol{\beta}}_n$. We shall use Theorem A.1 in Hu et al. (2021) to complete our proof. To apply this theorem, we need to show two conditions, that is,

1. the smallest eigenvalue of $\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive; and
2. $\sup_{\boldsymbol{\beta}\in N_n}\left\|\boldsymbol{C}_n^{-1}[J_n(\boldsymbol{\beta}) - J_n(\boldsymbol{\beta}_0)]\boldsymbol{C}_n^{-1}\right\| = o_p(1),$ (A.5)

where $\boldsymbol{C}_n = n^{-\rho}\boldsymbol{D}_n$ for $0<\rho<\varepsilon/6$ with ε defined in Assumption 3, $N_n = \{\boldsymbol{\beta} : \left\|C_n(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\| \le 1\}$.

(1) We first show that $\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive definite. Assume that there exists a nonzero vector $\check{c}$ such that $\check{c}^T\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}\check{c} \le 0$. Then, by *Part II*, we have

$$
\begin{aligned}
\check{c}^{\top}\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}\check{c} &= \frac{\check{c}^{\top}\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}A_{n5}(\boldsymbol{\beta}_0)\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}\check{c}}{n^{-2}A_{n1}(\boldsymbol{\beta}_0)} + o_p(1) \\
&\Rightarrow \left(\int_0^1 \tau_3^2(r)dr\right)^{-1}\bullet\ \check{c}^{\top}\int_0^1 \boldsymbol{h}_{\xi,0}(\tau_3(r))\boldsymbol{h}_{\xi,0}^{\top}(\tau_3(r))dr\,\check{c} > 0
\end{aligned}
.
$$

Then, $\boldsymbol{D}_n^{-1}J_n(\boldsymbol{\beta}_0)\boldsymbol{D}_n^{-\top}$ is positive definite with the smallest eigenvalue larger than 0.

(2) Next, we show (A.5). To do so, first write that

$$
\begin{aligned}
& n^{2\rho-1}\boldsymbol{\kappa}_{n\xi,0}^{-1}[J_n(\boldsymbol{\beta})-J_n(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\xi,0}^{-\top} \\
&= n^{2\rho+1}\boldsymbol{\kappa}_{n\xi,0}^{-1}\left\{\frac{J_{n2}(\boldsymbol{\beta})-J_{n2}(\boldsymbol{\beta}_0)}{J_{n1}(\boldsymbol{\beta}_0)}-\frac{J_{n2}(\boldsymbol{\beta}_0)}{J_{n1}^2(\boldsymbol{\beta}_0)}(J_{n1}(\boldsymbol{\beta})-J_{n1}(\boldsymbol{\beta}_0))\right\}\boldsymbol{\kappa}_{n\xi,0}^{-\top}\frac{J_{n1}(\boldsymbol{\beta}_0)}{J_{n,1}(\boldsymbol{\beta})} \ . \\
&= [n^{-6}J_{n1}(\boldsymbol{\beta}_0)]^{-1}\left\{n^{2\rho-5}\boldsymbol{\kappa}_{n\xi,0}^{-1}\left[J_{n,2}(\boldsymbol{\beta})-J_{n2}(\boldsymbol{\beta}_0)\right]\boldsymbol{\kappa}_{n\xi,0}^{-\top}\right. \\
&\left. -\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}J_{n,11}(\boldsymbol{\beta}_0)\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-\top}\bullet\ n^{2\rho-6}\left[J_{n1}(\boldsymbol{\beta})-J_{n1}(\boldsymbol{\beta}_0)\right]\right\}\ \bullet\ \frac{J_{n1}(\boldsymbol{\beta}_0)}{J_{n1}(\boldsymbol{\beta})}
\end{aligned}
$$

Since by Lemma C.2 (a1), Lemma C.3 (i) and (A.4), we have

$$
\begin{aligned}
& n^{-6}J_{n1}(\boldsymbol{\beta}_0) = O_p(1), \\
& \left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-1}J_n(\boldsymbol{\beta}_0)\left(\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\right)^{-\top} = O_p(1).
\end{aligned}
$$

To show (A.5), it suffices to show

$$
\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\xi,0}^{-1}[J_{n2}(\boldsymbol{\beta})-J_{n2}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\xi,0}^{-\top}\right\| = o_p(1), \tag{A.6}
$$

$$
\sup_{\beta\in N_n}\left\|n^{2\rho-6}[J_{n1}(\boldsymbol{\beta})-J_{n1}(\boldsymbol{\beta}_0)]\right\| = o_p(1), \tag{A.7}
$$

$$
\sup_{\beta\in N_n}\left\|J_{n1}(\boldsymbol{\beta}_0)/J_{n1}(\boldsymbol{\beta})\right\| = O_p(1). \tag{A.8}
$$

Consider (A.6), write

$$
\begin{aligned}
& J_{n2}(\boldsymbol{\beta})-J_{n2}(\boldsymbol{\beta}_0) \\
&= \left\{A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})+A_{n6}(\boldsymbol{\beta})]-A_{n1}^2(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)]\right\} \\
&-\left\{A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})+A_{n7}(\boldsymbol{\beta})]-A_{n1}(\boldsymbol{\beta}_0)A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\right\} \\
&-2\begin{Bmatrix}A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})A_{n4}^\top(\boldsymbol{\beta})+A_{n4}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta})]-A_{n1}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0)+ \\ A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)]\end{Bmatrix} \\
&+4\left\{A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})A_{n3}^\top(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)A_{n3}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)\right\} \\
&\equiv \Upsilon_{112}^a-\Upsilon_{112}^b-2\Upsilon_{112}^c+4\Upsilon_{112}^d,
\end{aligned}
$$

where the definitions of $\Upsilon_{112}^a-\Upsilon_{112}^d$ should be obvious. For the first term,

$$
\begin{aligned}
\Upsilon_{112}^a &= A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)]+A_{n1}^2(\boldsymbol{\beta})[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)] \\
&+ [A_{n1}^2(\boldsymbol{\beta})-A_{n1}^2(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)].
\end{aligned}
$$

By Lemmas C.2 and C.3, we have

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}^2(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$\le \sup_{\beta\in N_n}\left\|n^{-4}A_{n1}^2(\boldsymbol{\beta})\right\|\cdot\sup_{\beta\in N_n}\left\|n^{2\rho-1}\boldsymbol{\kappa}_{n0\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1),$$

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}^2(\boldsymbol{\beta})[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$\le \sup_{\beta\in N_n}\left\|n^{-4}A_{n1}^2(\boldsymbol{\beta})\right\|\cdot\sup_{\beta\in N_n}\left\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n6}(\boldsymbol{\beta})-A_{n6}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1),$$

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\left[A_{n1}^2(\boldsymbol{\beta})-A_{n1}^2(\boldsymbol{\beta}_0)\right]\left[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)\right]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$\le \sup_{\beta\in N_n}\left\|n^{2\rho-4}\left[A_{n1}^2(\boldsymbol{\beta})-A_{n1}^2(\boldsymbol{\beta}_0)\right]\right\|\cdot\left\|\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}\left[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)\right]\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-\top}\right\|$$
$$\le \sup_{\beta\in N_n}\left\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\right\|\cdot\sup_{\beta\in N_n}\left\|n^{-2}[A_{n1}(\boldsymbol{\beta})+A_{n1}(\boldsymbol{\beta}_0)]\right\|$$
$$\times\left\|\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n6}(\boldsymbol{\beta}_0)]\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-\top}\right\|=o_p(1).$$

Thus,

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{a}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1). \tag{A.9}$$

For the second term,

$$\begin{aligned}\Upsilon_{112}^{b} &= A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\\ &+ A_{n1}(\boldsymbol{\beta})[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\\ &+ [A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)].\end{aligned}$$

Similarly, by Lemmas C.2 and C.3, we have

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n2}(\boldsymbol{\beta})[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$\le n^{-1/2}\cdot\sup_{\beta\in N_n}\left\|n^{-2}A_{n1}(\boldsymbol{\beta})\right\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{-1}A_{n2}(\boldsymbol{\beta})\right\|$$
$$\times\sup_{\beta\in N_n}\left\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta})-A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta})-A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1),$$

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)][A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$= n^{-1/2}\cdot\sup_{\beta\in N_n}\left\|n^{-2}A_{n1}(\boldsymbol{\beta})\right\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]\right\| \quad .$$
$$\times\left\|n^{-3/2}\boldsymbol{\beta}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1),$$

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]A_{n2}(\boldsymbol{\beta}_0)[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|$$
$$\le n^{-1/2}\cdot\sup_{\beta\in N_n}\left\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\right\|\cdot\left\|n^{-1}A_{n2}(\boldsymbol{\beta}_0)\right\|$$
$$\times\left\|n^{-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n5}(\boldsymbol{\beta}_0)+A_{n7}(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1).$$

Thus,

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{b}\,\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1). \tag{A.10}$$

For the third term,

$$\begin{aligned}\Upsilon_{112}^{c} &= A_{n1}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta})-A_{n4}(\boldsymbol{\beta}_0)]^\top + A_{n1}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta})-A_{n4}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta})^\top \\ &+ A_{n1}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]A_{n4}^\top(\boldsymbol{\beta}_0)+A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]^\top \\ &+ [A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)][A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0)+A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)].\end{aligned}$$

By Lemmas C.2 and C.3, we have

$$\begin{aligned}&\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n4}(\boldsymbol{\beta})-A_{n4}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\| \\ &\quad\le n^{-1/2}\cdot\sup_{\beta\in N_n}\left\|n^{-2}A_{n1}(\boldsymbol{\beta})\right\|\cdot\sup_{\boldsymbol{\beta}\in N_n}\left\|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta})\right\| \\ &\qquad\times\sup_{\beta\in N_n}\left\|n^{2\rho-1}\boldsymbol{\kappa}_{n\boldsymbol{\beta},0}^{-1}[A_{n4}(\boldsymbol{\beta})-A_{n4}(\boldsymbol{\beta}_0)]\right\|=o_p(1), \\ &\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}A_{n1}(\boldsymbol{\beta})A_{n4}(\boldsymbol{\beta}_0)[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\| \\ &\quad\le n^{-1}\ \cdot\ \sup_{\beta\in N_n}\left\|n^{-2}A_{n1}(\boldsymbol{\beta})\right\|\ \cdot\ \left\|\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}A_{n4}(\boldsymbol{\beta}_0)\right\| \\ &\qquad\times\sup_{\beta\in N_n}\left\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]\right\|=o_p(1), \\ &\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)][A_{n3}(\boldsymbol{\beta}_0)A_{n4}^\top(\boldsymbol{\beta}_0)+A_{n4}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)]\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\| \\ &\quad\le n^{-1}\cdot\sup_{\boldsymbol{\beta}\in N_n}\left\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\right\| \\ &\qquad\times\left\|(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0})^{-1}A_{n3}(\boldsymbol{\beta}_0)\left[\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}A_{n4}(\boldsymbol{\beta}_0)^\top\right]\right. \\ &\qquad\left.+\left(\sqrt{n}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}A_{n4}(\boldsymbol{\beta}_0)\left[\left(n^{3/2}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}\right)^{-1}A_{n3}(\boldsymbol{\beta}_0)\right]^\top\right\|=o_p(1)\end{aligned}$$

Thus,

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-1}\Upsilon_{112}^{c}\,\boldsymbol{\kappa}_{n\boldsymbol{\xi},0}^{-\top}\right\|=o_p(1). \tag{A.11}$$

For the last term,

$$\begin{aligned}\Upsilon_{112}^{d} &= A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]^\top + A_{n2}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})- \\ &\quad A_{n3}(\boldsymbol{\beta}_0)]A_{n3}^\top(\boldsymbol{\beta}_0) \\ &+ [A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0).\end{aligned}$$

By Lemmas C.2 and C.3, we have

$$\begin{aligned}
&\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\xi,0}^{-1}A_{n2}(\boldsymbol{\beta})A_{n3}(\boldsymbol{\beta})[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)]^\top\boldsymbol{\kappa}_{n\xi,0}^{-\top}\right\| \\
&\quad = n^{-1}\cdot\sup_{\beta\in N_n}\left\|n^{-1}A_{n2}(\boldsymbol{\beta})\right\|\cdot\sup_{\beta\in N_n}\left\|(n^{3/2}\boldsymbol{\kappa}_{n\xi,0})^{-1}A_{n3}(\boldsymbol{\beta})\right\| \\
&\quad\quad\times\sup_{\beta\in N_n}\left\|n^{2\rho-3/2}\boldsymbol{\kappa}_{n\xi,0}^{-1}\left[A_{n3}(\boldsymbol{\beta})-A_{n3}(\boldsymbol{\beta}_0)\right]\right\| = o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\xi,0}^{-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]A_{n3}(\boldsymbol{\beta}_0)A_{n3}^\top(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\xi,0}^{-\top}\right\| \\
&\quad = n^{-1}\cdot\sup_{\beta\in N_n}\left\|n^{2\rho-1}[A_{n2}(\boldsymbol{\beta})-A_{n2}(\boldsymbol{\beta}_0)]\right\|\cdot\left\|(n^{3/2}\boldsymbol{\kappa}_{n\xi,0})^{-1}A_{n3}(\boldsymbol{\beta}_0)\right\|^2 = o_p(1).
\end{aligned}$$

Thus,

$$\sup_{\beta\in N_n}\left\|n^{2\rho-5}\boldsymbol{\kappa}_{n\xi,0}^{-1}\Upsilon_{112}^{d}\boldsymbol{\kappa}_{n\xi,0}^{-\top}\right\| = o_p(1). \tag{A.12}$$

Combining (A.9)–(A.12) gives rise to (A.6).

Next consider (A.7). By Lemma C.2 and Lemma C.3,

$$\begin{aligned}
&\sup_{\beta\in N_n}\left\|n^{2\rho-6}[J_{n1}(\boldsymbol{\beta})-J_{n1}(\boldsymbol{\beta}_0)]\right\| \\
&\le \sup_{\beta\in N_n}\left\|n^{2\rho-2}[A_{n1}(\boldsymbol{\beta})-A_{n1}(\boldsymbol{\beta}_0)]\right\|\cdot\sup_{\beta\in N_n}\left\|n^{-4}\left[A_{n1}^2(\boldsymbol{\beta})+A_{n1}(\boldsymbol{\beta})A_{n1}(\boldsymbol{\beta}_0)+A_{n1}^2(\boldsymbol{\beta}_0)\right]\right\| \\
&= o_p(1),
\end{aligned}$$

showing (A.7). By Lemma C.3, using arguments similar to Theorem 2.2 in Chan and Wang (2014), (A.8) is established.

We have shown in *Part II* the convergence of $S_n(\boldsymbol{\beta}_0)$ and $J_n(\boldsymbol{\beta}_0)$. Thus, by Theorem A.1 in Hu et al. (2021), there exists a sequence of estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ such that $S_n\left(\hat{\boldsymbol{\beta}}_n\right)=0$ and the limiting distribution follows.

Part IV – Proof for the limit of $\hat{\boldsymbol{\theta}}_n$. Finally, we show the limiting distribution of $\hat{\boldsymbol{\theta}}_n$. Define $\tilde{\boldsymbol{D}}_n = \begin{pmatrix} n\boldsymbol{I}_{\ell_1} & 0 & 0 \\ 0 & \sqrt{n}\boldsymbol{I}_{\ell_2} & 0 \\ 0 & 0 & \sqrt{n}\boldsymbol{\kappa}_{\mathrm{nd}} \end{pmatrix}$. Note that

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_n &= \left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top\boldsymbol{\Lambda}\left(\hat{\boldsymbol{\beta}}_n\right)\\
&= \left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)+\left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top\left[\boldsymbol{\Lambda}\left(\hat{\boldsymbol{\beta}}_n\right)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)\right]\\
&= \boldsymbol{\theta}_0+\left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top u+\left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top\left[\boldsymbol{\Lambda}\left(\hat{\boldsymbol{\beta}}_n\right)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)\right],\\
\tilde{\boldsymbol{D}}_n\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\right) &= \tilde{\boldsymbol{D}}_n\left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top u+\tilde{\boldsymbol{D}}_n\left(\boldsymbol{W}^\top\boldsymbol{W}\right)^{-1}\boldsymbol{W}^\top\left[\boldsymbol{\Lambda}\left(\hat{\boldsymbol{\beta}}_n\right)-\boldsymbol{\Lambda}(\boldsymbol{\beta}_0)\right].
\end{aligned}$$

First, we have

$$\tilde{\boldsymbol{D}}_n^{-1}\boldsymbol{W}^\top\boldsymbol{W}\tilde{\boldsymbol{D}}_n^{-1} \Rightarrow \begin{pmatrix} \int_0^1 \boldsymbol{V}(r)\boldsymbol{V}'(r)dr & 0 & \int_0^1 \boldsymbol{V}(r)\boldsymbol{d}'(r)dr \\ 0 & E[\boldsymbol{z}_t\,\boldsymbol{z}_t{}'] & 0 \\ \int_0^1 \boldsymbol{d}(r)\boldsymbol{V}(r)'dr & 0 & \int_0^1 \boldsymbol{d}(r)\boldsymbol{d}(r)'dr \end{pmatrix} \equiv \Delta_V,$$

by Lemma 5 in Chang et al. (2001). And

$$\tilde{\boldsymbol{D}}_n(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^T u = \Delta_V^{-1}\left(1+o_p(1)\right)\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t u_t \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t u_t \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t u_t \end{pmatrix} \Rightarrow \Delta_V^{-1}\begin{pmatrix} \int_0^1 \boldsymbol{V}(r)dU(r) \\ N(0,\boldsymbol{\Omega}_{zu}) \\ \int_0^1 \boldsymbol{d}(r)dU(r) \end{pmatrix},$$

where $\boldsymbol{\Omega}_{zu} = \lim_{n\to\infty} n^{-1}\mathrm{var}\left(\sum_{t=1}^n \boldsymbol{z}_t u_t\right)$. In addition, using the mean value theorem, we have.

$$\begin{aligned}
\tilde{\boldsymbol{D}}_n(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top[\boldsymbol{\Lambda}\left(\hat{\boldsymbol{\beta}}_n\right)-\Lambda(\boldsymbol{\beta}_0)] &= \boldsymbol{\Delta}_V^{-1}(1+o_p(1))\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\left[\Lambda_t\left(\hat{\boldsymbol{\beta}}_n\right)-\Lambda_t(\boldsymbol{\beta}_0)\right] \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\left[\Lambda_t\left(\hat{\boldsymbol{\beta}}_n\right)-\Lambda_t(\boldsymbol{\beta}_0)\right] \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\left[\Lambda_t\left(\hat{\boldsymbol{\beta}}_n\right)-\Lambda_t(\boldsymbol{\beta}_0)\right] \end{pmatrix}\\
&= \Delta_V^{-1}(1+o_p(1))\begin{pmatrix} \frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \\ \frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^\top(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \end{pmatrix}.
\end{aligned}$$

Using arguments similar to Lemma C.2, tt follows from Lemma C.2 and Theorem 1 that

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)\\
&\quad=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\left[\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_n)-\dot{\boldsymbol{\Lambda}}_t(\boldsymbol{\beta}_0)\right]^{\top}\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)\\
&\quad=\frac{1}{n}\sum_{t=1}^{n}\mathrm{x}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)=\frac{1}{n^{3/2}}\sum_{t=1}^{n}\boldsymbol{x}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\xi,0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\\
&\qquad\quad\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)\\
&\quad\Rightarrow-\int_0^1\boldsymbol{V}(r)\boldsymbol{h}_{\xi,0}^{\top}(\tau_3(r))dr\cdot\boldsymbol{\mathcal{B}}_1,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathrm{z}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_t\left[\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)-\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\right](\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathrm{z}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)=\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{z}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\xi,0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\\
&\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)\xrightarrow{p}0,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)\\
&\quad=\frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+\frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{d}_t\left[\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_n)-\dot{\boldsymbol{\Lambda}}_t^{\top}(\hat{\boldsymbol{\beta}}_0)\right]\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)\\
&\quad=\frac{\boldsymbol{\kappa}_{nd}^{-1}}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)=\frac{\boldsymbol{\kappa}_{nd}^{-1}}{n}\sum_{t=1}^{n}\boldsymbol{d}_t\dot{\boldsymbol{\Lambda}}_t^{\top}(\boldsymbol{\beta}_0)\boldsymbol{\kappa}_{n\xi,0}^{-1}\cdot\sqrt{n}\boldsymbol{\kappa}_{n\xi,0}\\
&\left(\hat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)+o_p(1)\\
&\quad\Rightarrow-\int_0^1\boldsymbol{d}(r)\boldsymbol{h}_{\xi,0}^{\top}(\tau_3(r))dr\cdot\boldsymbol{\mathcal{B}}_1,
\end{aligned}
$$

where $\boldsymbol{\mathcal{B}}_1$ is defined in Theorem 1. Combining the above pieces, with an application of Lemma C.2, completes the proof.

CHAPTER 8

MINIMAX RISK IN ESTIMATING KINK THRESHOLD AND TESTING CONTINUITY

Javier Hidalgo[a], Heejun Lee[b], Jungyoon Lee[c], and Myung Hwan Seo[d]

[a]*London School of Economics, London, UK*
[b]*Brown University, Providence, RI, USA*
[c]*Royal Holloway, University London, London, UK*
[d]*Seoul National University, Seoul, Korea*

ABSTRACT

The authors derive a risk lower bound in estimating the threshold parameter without knowing whether the threshold regression model is continuous or not. The bound goes to zero as the sample size n *grows only at the cube-root rate. Motivated by this finding, the authors develop a continuity test for the threshold regression model and a bootstrap to compute its* p*-values. The validity of the bootstrap is established, and its finite-sample property is explored through Monte Carlo simulations.*

Keywords: Continuity test; kink; risk lower bound; unknown threshold; bootstrap; minimax

JEL classifications: C12; C13; C24

1. INTRODUCTION

The threshold model has been widely used to model the non-linearity of time series. For instance, threshold autoregressive (TAR) model is one of the earliest

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 233–259

Published under exclusive licence by Emerald Publishing Limited
ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A008

regime switching models. In its simplest form, it is assumed that there are two regimes. The regime is determined depending on the realization of the threshold variable and the threshold level. See Tong (1990) for a review. Hansen (2000) has extended it to the regression with more Economics application and Hansen and Seo (2002) and Seo (2006) to the threshold cointegration. Park and Shintani (2016) and Seo (2008) examined testing issues surrounding threshold effect and unit root. Chang et al. (2017) proposed an interesting generalization by introducing a latent factor threshold variable, while Lee et al. (2021) extended it further by estimating the factors from an external big data set.

Once one accepts the hypothesis of a threshold effect in the regression function via any of the available tests, see, for example, Hansen (1996), Lee et al. (2011), among others, one is then interested in deciding whether the "segmented" regression model is a model with a discontinuity (jump) or a model with a kink, since the so-called break/threshold tests are unable to discriminate between the two models. A powerful reason to test for a kink comes from the statistical inferential point of view. As we discuss in Section 2, the kink design can be represented by a set of restrictions on the parameter space of the threshold regression model. Thus, the parameters can be consistently estimated by the unrestricted least squares estimator. Unlike in the linear regression model, where one can make valid inferences based on the unconstrained estimation without knowing if the constraint holds, inferences in our context have very different statistical properties under the kink design when using the unrestricted estimation. More specifically, Hidalgo et al. (2019) shows that the rate of convergence of the estimate of the threshold point via unrestricted least squares method is $n^{1/3}$ if there is a kink, which is in contrast to $n^{1/2}$-rate when the (true) constraint of a kink is employed in its estimation (Chan & Tsay, 1998; Feder, 1975). If there is not a kink but a jump, then the unrestricted estimate converges in n-rate (Chan, 1993), which is also a ℓ_1-minimax rate (Korostelev, 1987).

On the other hand, we may focus on the fact that the worst-case convergence rate of unrestricted estimate slows down to $n^{1/3}$ if the type of threshold is unknown compared with the situation in which the type of threshold is known so that it is used in the estimation. We show in Section 3 that the cube-root convergence rate cannot be improved in terms of ℓ_1-risk if the model does not specify the type of threshold. We extend these results to the diminishing threshold model, where the threshold degenerates in polynomial order. The diminishing threshold was introduced by Hansen (2000), and it can be understood as an asymptotic approximation of a small threshold. By allowing the diminishing threshold, we investigate how the size of the threshold affects the performance of estimators. Also, we develop a test valid under both fixed and diminishing threshold effect.

The main contribution of this chapter is to develop a testing procedure to distinguish between jump and kink designs. Hansen (2017) considers inference under the kink design and mentioned

> one could imagine testing the assumption of continuity within the threshold model class. This is a difficult problem, one to which we are unaware of a solution, and therefore is not pursued in this paper.

We propose a test statistic that is based on the quasi-likelihood ratio and develop its asymptotic distribution. The difficulty stems from the degeneracy of the hessian matrix of the expected pseudo-Gaussian likelihood function under the null of continuity. The test is not asymptotically pivotal since it involves multiple restrictions related to the continuity and conditional heteroscedasticity, and a bootstrap method is proposed in Section 4 to estimate p-values of the test.

We then present the results of a Monte Carlo experiment in Section 5, which reports a good finite-sample performance of our bootstrap procedures for the continuity test. In our empirical application in Section 6, we employ our test of continuity on the long-span time-series data of US real GDP growth and debt-to-GDP ratio data used in Hansen (2017) which had fitted the kink model. Our test of continuity rejects the null of continuity, and we present the estimated jump model. We also consider data from Sweden and find substantial variations across countries not only in the values of parameter estimates but also in the results of tests on the presence of threshold effect and continuity.

2. MODEL AND ASSUMPTIONS

We consider a threshold/segmented regression model

$$Y_i = X_i'\beta + X_i'\delta\mathbb{I}\{Q_i > \tau\} + U_i, \tag{1}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function, Y_i is dependent variable and X_i is a d-dimensional vector of regressors. The parameter τ represents a change/break-point or threshold, taking values in a compact parameter space $\mathbb{T}$ which lies in the interior of the domain of the threshold variable Q_i. In addition, we assume that $\delta \neq 0$, which implies that the model has a threshold effect.

As mentioned before, we consider the case where the conditional expectation of Y_i given the regressor X_i is allowed to be either continuous, that is, to have a kink, or discontinuous, that is, to have a jump. We let the threshold variable Q_i be an element of the covariate vector X_i since otherwise, it would not be possible for the regression function to be continuous. We shall decompose the d-dimensional parameters and regressors as follows:

$$X_i = \left(1, X_{i2}', Q_i\right)'; \quad \delta = \left(\delta_1, \delta_{i2}', \delta_3\right)', \tag{2}$$

where δ is partitioned to match the dimensionality of X_i and X_{i2} is a $(d-2)$-dimensional vector. Also we shall abbreviate $\mathbb{I}_i(\tau) = \mathbb{I}\{Q_i > \tau\}$ and $X_i(\tau) = \left(X_i', X_i'\mathbb{I}_i(\tau)\right)'$, so that we can write (1) as

$$Y_i = X_i'\beta + \delta_1\mathbb{I}_i(\tau) + X_{i2}'\delta_2\mathbb{I}_i(\tau) + \delta_3 Q_i\mathbb{I}_i(\tau) + U_i \tag{3}$$

$$= X_i(\tau)'\alpha + U_i, \quad \text{where} \quad \alpha = (\beta', \delta')'. \tag{4}$$

Notation. Before stating some regularity assumptions on the model, we introduce some extra notations. Let $f(\cdot)$ denote the density function of Q_i and $\sigma^2(\tau) = E\left(U_i^2 \mid Q_i = \tau\right)$, the conditional variance function of the error term, while $\sigma^2 = E(U_i^2)$ denotes the unconditional variance. Denote $d \times d$ matrices $D(\tau) = E\left(X_i X_i' \mid Q_i = \tau\right), V(\tau) = E\left(X_i X_i' U_i^2 \mid Q_i = \tau\right)$ and let $D = D(\tau_0)$ and $V = V(\tau_0)$. As usual the "0" subscript on a parameter indicates its true unknown value. Finally, let $M = E(\mathbf{X}_i \mathbf{X}_i')$ and $\Omega = E(\mathbf{X}_i \mathbf{X}_i' U_i^2)$ with $\mathbf{X}_i = X_i(\tau_0)$.

We shall now introduce some regularity conditions.

Assumption 1. *Let $\{X_i, U_i\}_{i\in\mathbb{Z}}$ be a strictly stationary, ergodic sequence of random variables such that their ρ-mixing coefficients satisfy $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$ and $E(U_i \mid \mathcal{F}_{i-1}) = 0$, where $\mathcal{F}_i$ is the filtration up to time i. Furthermore, $M, \Omega > 0$, $E\| X_i \|^4 < \infty$, $E\| X_i U_i \|^4 < \infty$ and $E\| U_i \|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption 2. *The functions $f(\tau), V(\tau)$ and $D(\tau)$ are continuous at $\tau = \tau_0$. For all $\tau \in \mathbb{T}$, the functions $f(\tau)$, $E\left(X_i X_i' \mathbb{I}\{Q_i \le \tau\}\right)$ and $\mathrm{Var}(X_{i2} \mid Q_i = \tau)$ are positive and continuous, and the functions $f(\tau)$, $E\left(| X_i |^4 | Q_i = \tau\right)$ and $E\left(| X_i U_i |^4 | Q_i = \tau\right)$ are bounded by some $C < \infty$.*

These are similar to those in Hansen (2000). Note that the SETAR model of Tong (1990) satisfies Assumption 1. The condition for the conditional moment $\mathrm{Var}(X_{i2} \mid Q_i = \tau)$ is written in terms of X_{i2} as the other elements in X_i are fixed given $Q_i = \tau$. While we allow conditional heteroscedasticity of a general form, Assumption 2 requires continuity of the conditional variance function $\sigma^2(\cdot)$ at τ_0. We need to estimate the conditional variance via non-parametric methods.

We shall emphasize that the model (1) encompasses both the kink and jump models. The kink model is characterized by the continuity restriction:

Assumption C. $\delta_{30} \neq 0$ *and*

$$\delta_{10} + \delta_{30}\tau_0 = 0; \quad \delta_{20} = 0. \tag{5}$$

Note that we require δ_{30} to be non-zero to identify τ_0. Under (5), we observe that (3) becomes

$$Y_i = X_i'\beta_0 + \delta_{30}(Q_i - \tau_0)\mathbb{I}_i(\tau_0) + U_i. \tag{6}$$

For the sake of completeness, we define the jump threshold:

Assumption J. $\delta_0 \neq 0$ *and*

$$\delta_0{}' D \delta_0 > 0. \tag{7}$$

In the following sections, we allow for the threshold effect δ_0 to converge to zero at a polynomial rate, as in Hansen (2000). Specifically, $\delta_0 = d_0 \cdot n^{-\varphi}$ where $\varphi \geq 0$ and d_0 is fixed over n. We call the case where $\varphi = 0$ a fixed threshold and the case where $\varphi > 0$ a diminishing threshold.

3. ESTIMATORS AND RISK BOUND

This section elaborates on how the continuity restriction affects the estimation of the threshold location τ_0. As mentioned before, when the continuity restriction is not employed in the estimation, the rate of convergence is either n if there is a jump or $n^{1/3}$ if there is a kink, which means that the worst-case performance of the unrestricted estimator is $n^{1/3}$ under the situation that the type of threshold is unknown. A generalized result that includes the diminishing threshold effect is presented in Proposition 1. One may pursue to propose an estimation procedure that outperforms the unrestricted estimator with respect to the worst-case convergence rate. However, it is impossible to overcome the cube-root rate in ℓ_1-minimax sense if the information about the threshold type is unavailable, as we show in Proposition 2.

3.1. Estimators

We choose the residual sum of squares as the objective function. Denote parameters by $\theta = (\alpha', \tau)' \in \mathbb{R}^{2d+1}$ and denote the objective function by $\mathbb{S}_n$ where

$$\mathbb{S}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - X_i'(\tau)\alpha\right)^2. \tag{8}$$

If the continuity restriction is not imposed on the true parameter θ_0, then it can be estimated by minimizing the objective function, that is,

$$\hat{\theta} = \left(\hat{\alpha}', \hat{\tau}\right)' = \underset{\theta \in \Theta}{\operatorname{argmin}}\ \mathbb{S}_n(\theta), \tag{9}$$

where $\Theta = \Lambda \times \mathbb{T}$ is a compact set in $\mathbb{R}^{2d+1}$. Following convention, we let $\hat{\tau}$ be an element of $\{Q_i\}$.

On the other hand, we can minimize (8) among the elements of Θ that satisfy constraints in (5), yielding the constrained least squares estimator (CLSE):

$$\tilde{\theta} = \left(\tilde{\alpha}', \tilde{\tau}\right)' := \underset{\theta \in \Theta : \delta_1 + \delta_3 \tau = 0; \delta_2 = 0}{\operatorname{argmin}}\ \mathbb{S}_n(\theta). \tag{10}$$

Since criterion is not smooth, we compute the unconstrained least squares estimator (LSE) as a two-step algorithm. Since the criterion $\mathbb{S}_n$ is in fact a step function along τ with jumps at each Q_i, we may first discretize the parameter space of threshold $\mathbb{T}$ as $\mathbb{T}_n = \mathbb{T} \cap \{Q_1, \ldots, Q_n\}$ to find $\hat{\tau}$. Then, find $\hat{\alpha}(\tau)$ which minimizes the sum of squared errors for each τ:

$$\hat{\alpha}(\tau) = \underset{\alpha \in \Lambda}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left(\mathrm{Y}_i - \mathrm{X}_i'(\tau)\alpha \right)^2 \tag{11}$$

Finally, we define the LSE $\hat{\tau}$ as the minimizer of the sum of squared errors:

$$\hat{\tau} = \underset{\tau \in \mathbb{T}_n}{\operatorname{argmin}} \hat{\mathbb{S}}_n(\tau), \tag{12}$$

where

$$\hat{\mathbb{S}}_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - X_i'(\tau)\hat{\alpha}(\tau) \right)^2. \tag{13}$$

Then, our estimator of α is $\hat{\alpha} = \hat{\alpha}(\hat{\tau})$. The CLSE can be obtained similarly.

Suppose that $\delta_0 = d_0 \cdot n^{-\varphi}$, for $0 \le \varphi < 1/2$ and a non-zero vector $d_0 \in \mathbb{R}^d$. When $\varphi = 0$, the rate of convergence is n if there is a jump (Chan, 1993), whereas it is only $n^{1/3}$ if there is a kink and the restriction is not used in the estimation (Hidalgo et al., 2019). If the restriction is used in the estimation, the rate of convergence is $n^{1/2}$. On the other hand, when $\varphi > 0$, the rate is $n^{1-2\varphi}$ when there is a jump (Hansen, 2000), and we shall show that when there is a kink, the rate of convergence becomes $n^{(1-2\varphi)/3}$ if the restriction is not used in the estimation.

Proposition 1. *Let Assumptions C, 1, and 2 hold. If* $\delta_0 = d_0 \cdot n^{-\varphi}$ *for some* $0 \le \varphi < 1/2$ *and* $d_0 \neq 0$, *we have that*

$$\hat{\alpha} - \alpha_0 = O_p\left(n^{-1/2}\right) \text{ and } \hat{\tau} - \tau_0 = O_p\left(n^{(2\varphi-1)/3}\right).$$

This proposition generalizes the rate of convergence of the LSE $\hat{\theta}$ under the fixed threshold assumption explored in Hidalgo et al. (2019) to encompass the diminishing threshold.

3.2. Risk Bound

In this section, we shall develop an ℓ_1-minimax lower bound in estimating the threshold τ. The ℓ_1-risk of an estimator $\hat{\tau}$ for τ is defined as

$$\mathcal{R}_n(\hat{\tau}\,; \alpha, \tau, \mathbb{Q}_n) = E\left(|\hat{\tau} - \tau|\right), \tag{14}$$

where the expectation depends on α_0, τ_0, and $\mathbb{Q}_n$, the joint distribution of $\{X_i, U_i\}_{i=1}^n$. Let $\mathcal{P}(n, \kappa, \underline{\sigma}^2, \overline{f})$ denote the class of joint distributions of $\{Y_i, X_i\}_{i=1}^n$ such that $Y_i = X_i(\tau)'\alpha + U_i$ for all $i \in \mathbb{Z}, |\delta_3| \ge \kappa$, and $f(\tau) \le \overline{f}, \sigma^2(\tau) \ge \underline{\sigma}^2$ for all $\tau \in \mathbb{T}$. We evaluate the performance of an estimator based on the most adverse choice of the distribution $\mathbb{P}_n \in \mathcal{P}(n, \kappa, \underline{\sigma}^2, \overline{f})$, namely,

$$\sup_{\mathbb{P}_n \in \mathcal{P}(n,\kappa,\underline{\sigma}^2,\overline{f})} \mathcal{R}_n(\hat{\tau}\,;\alpha(\mathbb{P}_n),\tau(\mathbb{P}_n),\mathbb{Q}_n(\mathbb{P}_n)), \tag{15}$$

where $\alpha(\mathbb{P}_n),\tau(\mathbb{P}_n)$, and $\mathbb{Q}_n(\mathbb{P}_n)$ make the joint distribution of $\{Y_i, X_i\}_{i=1}^n$ equal to $\mathbb{P}_n$. We will show that the worst-case risk (15) of any estimator cannot tend to zero faster than the cube-root rate by providing a lower bound for the ℓ_1-minimax risk.

Our lower bound is valid even for a restrictive subclass of $\mathcal{P}(n,\kappa,\underline{\sigma}^2,\overline{f})$ induced by Assumption 1, that is,

Assumption L. *Let $\{X_i, U_i\}_{i\in\mathbb{Z}}$ be a sequence of independent and identically distributed (i.i.d.) random vectors. Assume that U_i follows $\mathcal{N}(0,\sigma^2(Q_i))$ given X_i.*

Even if we assume that the δ_2 is known to be zero, the cube-root lower bound cannot be improved. Let η be the diameter of $\mathbb{T}$, that is, $\eta = \sup_{\tau_1,\tau_2\in\mathbb{T}} |\tau_2 - \tau_1|$. For notational convenience, we focus on $\mathbb{T}\subset(0,1)$ since for any interval (a, b), there exists a trivial affine transformation to (0, 1). Let $\kappa = \kappa_0 n^{-\varphi}$. If $\varphi > 0$, it represents the diminishing threshold effect. Then the minimax risk is lower bounded as follows:

Proposition 2. *Assume that $\mathbb{T}$ is a closed interval in (0, 1) and $\delta_2 = 0$. Under Assumption L, we have that*

$$\inf_{\hat{\tau}} \sup_{\mathbb{P}_n \in \mathcal{P}(n,\kappa,\underline{\sigma}^2,\overline{f})} E_{\mathbb{P}_n}\left(|\hat{\tau}-\tau(\mathbb{P}_n)|\right) \geq \begin{cases} \dfrac{\underline{\sigma}^{2/3}}{3\overline{f}^{1/3}\kappa_0^{2/3}} n^{(2\varphi-1)/3} & \text{if } n^{(1-2\varphi)} \geq \dfrac{3\underline{\sigma}^2}{\overline{f}\kappa^2\eta^3} \\ \dfrac{1}{4}\eta & \text{if } n^{(1-2\varphi)} < \dfrac{3\underline{\sigma}^2}{\overline{f}\kappa^2\eta^3} \end{cases},$$

for $0 \leq \varphi < 1/2$, where the infimum is taken over all estimators $\hat{\tau}$ of τ.

Note that there are reasonable relationships between the constant factor multiplied to $n^{-1/3}$ and nuisance parameters in Proposition 2. When the noise $\underline{\sigma}^2$ is a large constant or the minimal slope change κ is small, the estimation of τ becomes harder.

Thus far, in this section, we considered one of the simplest forms of the threshold model except that it includes both the jump and kink threshold. Therefore, the major complexity that causes the slow decay rate of the minimax risk lies in the fact that it is unknown whether the regression function is continuous or not. From this observation, we can see that there would be little gain from searching for an estimator with better accuracy without knowing the continuity of regression function, which motivates the test for the continuity.

Remark 1. *We derived the risk lower bound under Assumption **L** instead of Assumption 1. As mentioned earlier, Assumption **L** is more restrictive than Assumption 1. In some sense, Assumption **L** is a favorable scenario of Assumption 1. Since the worst-case performance under the favorable scenario cannot be better than that under the general scenario, the risk bound in Proposition 2 is also valid* *under Assumption 1. The implication of Proposition 2 is that the minimax risk cannot tend to zero faster than the cube-root rate even under the favorable scenario if the type of the threshold is unknown.*

4. TESTING CONTINUITY

This section considers testing of the continuity restriction, stated formally as

$$H_0 : \delta_{10} + \delta_{30}\tau_0 = 0 \quad \text{and} \quad \delta_{20} = 0, \tag{16}$$

along with an auxiliary condition of $\delta_{30} \neq 0$ to ensure the identification of the threshold point τ_0.

The alternative hypothesis is its negation

$$H_1 : \delta_{10} + \delta_{30}\tau_0 \neq 0 \quad \text{and/or} \quad \delta_{20} \neq 0. \tag{17}$$

Provided that $\operatorname{Var}[X_{i2} \mid Q_i = \tau_0] > 0$, the H_1 yields that

$$E\left[\left(\delta_{10} + \delta_{30}\tau_0 + X_{i2}'\delta_{20}\right)^2 \mid Q_i = \tau_0\right] > 0,$$

which implies that the regression function has a jump (non-zero change) at $Q_i = \tau_0$ with positive probability. As mentioned in the previous section, we develop a test valid for both fixed and diminishing threshold. In order to obtain such a test, we extend the earlier results of Hidalgo et al. (2019) about the fixed threshold to the diminishing threshold.

4.1. Continuity Test

To develop the test, we first need to derive the asymptotic distributions of the LSE $\hat{\theta}$ and CLSE $\tilde{\theta}$ under Assumption **C**. Feder (1975) and later Chan and Tsay (1998) or Hansen (2017) have already established the asymptotic normality of $\tilde{\theta}$ with the standard squared root consistency. Thus, we only need to examine the asymptotic properties of $\hat{\theta}$. We present the asymptotic distribution of $\hat{\theta}$ under the null.

Theorem 1. *Let Assumptions **C**, 1, and 2 hold, and $B_1(\cdot), B_2(\cdot)$ be two independent standard Brownian motions. Define $W(g) := B_1(-g)\mathbb{I}\{g<0\} + B_2(g)\mathbb{I}\{g>0\}$. Let $\delta_0 = d_0 \cdot n^{-\varphi}$. If $0 \leq \varphi < 1/2$,*

$$n^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}\left(0, M^{-1}\Omega M^{-1}\right)$$
$$n^{(1-2\varphi)/3}(\hat{\tau} - \tau_0) \xrightarrow{d} \underset{g\in\mathbb{R}}{\operatorname{argmax}}\left(2d_{30}\sqrt{\frac{\sigma^2(\tau_0)f(\tau_0)}{3}}W\left(g^3\right) + \frac{d_{30}^2}{3}f(\tau_0)\,|\,g\,|^3\right),$$

where the two limit distributions are independent of each other.

This result is an extension of Theorem 1 in Hidalgo et al. (2019) where only the fixed threshold case, $\varphi = 0$, is considered.

4.2. Test Statistic

Our testing problem is non-standard. First, the score-type test is not straightforward due to the non-differentiability of the criterion function $\mathbb{S}_n(\theta)$ with respect to τ. Second, the unconstrained estimators $\hat{\tau}$ and $\hat{\delta}$ converge at different rates to different family of probability distribution functions making the construction of a Wald-type test non-obvious. Thus, we consider a quasi-likelihood ratio statistic, which compares the constrained sum of squared residuals with the unconstrained one, that is,

$$T_n = n\frac{\tilde{\mathbb{S}}_n - \hat{\mathbb{S}}_n}{\hat{\mathbb{S}}_n} \tag{18}$$

where $\hat{\mathbb{S}}_n = \mathbb{S}_n\left(\hat{\theta}\right)$ and $\tilde{\mathbb{S}}_n = \mathbb{S}_n\left(\tilde{\theta}\right)$.

Deriving the asymptotic distribution of T_n is also non-standard due to the lack of expansion of the criterion function $\mathbb{S}_n(\theta)$ with respect to τ. Therefore, we employ the approach developed by Lee et al. (2011), which reformulates the statistic as a continuous functional of a stochastic process over an expanded domain. In particular, denote by I_d and $0_{a\times b}$ the identity matrix of dimension d and the matrix of zeros of dimension $a\times b$, respectively, and let

$$R = \begin{pmatrix} I_d & 0_{d\times d} \\ 0_{1\times d} & -\tau_0 : 0_{1\times(d-2)} : 1 \\ 0_{1\times d} & -\beta_{30} - \delta_{30} : 0_{1\times(d-1)} \end{pmatrix}.$$

Define a Gaussian process

$$\begin{aligned} \mathbb{K}(h,g,\ell) &= \ell' E\left(R\mathbf{X}_i\mathbf{X}_i' R'\right)\ell + h' E\left(\mathbf{X}_i\mathbf{X}_i'\right)h - 2\left(\ell' R + h'\right)B \\ &+ \left(2d_{30}\sqrt{\frac{\sigma^2(\tau_0)f(\tau_0)}{3}}W\left(g^3\right) + \frac{d_{30}^2}{3}f(\tau_0)\,|\,g\,|^3\right), \end{aligned}$$

where B follows $\mathcal{N}(0,\Omega)$ and is independent of the Gaussian process W that was introduced in Theorem 1.

Theorem 2. *Under Assumptions 1*, 2, and C *with* $\delta_{30} = d_{30}\cdot n^{-\varphi}$, $0 \le \varphi < 1/2$, *and* $d_{30} \ne 0$,

$$T_n \xrightarrow{d} \left(\min_{\ell: g=0,h=0} \mathbb{K}(h,g,\ell) - \min_{h: g=0,\ell=0} \mathbb{K}(h,g,\ell) - \min_{g: h=0,\ell=0} \mathbb{K}(h,g,\ell)\right) / \sigma^2.$$

It is worthwhile to note that δ_{30} is allowed to degenerate at the rate $n^{-\varphi}$ as well as stay fixed when $\varphi = 0$. We allow non-zero ϕ to examine the property of the continuity test when δ_{30} is small, and thus the identification of τ_0 is relatively weak. Along with the Monte Carlo experiments reported in Section 5, this theorem provides support for the good finite-sample performance of our continuity test based on the statistic T_n even when δ_{30} is small.

Next, we remark on the auxiliary assumption that $\delta_{30} \neq 0$. Recall that the discussion following (5) that without $\delta_{30} \neq 0$ the continuity restriction (5) implies that $\delta_0 = 0$ and thus, the null model is not a model with a kink but a linear regression model, a consequence being that τ_0 is unidentifiable as well. Indeed, testing that $\delta_0 = 0$ is a classic non-standard testing problem, also known as Davies' problem, where the null hypothesis induces a loss of identification. It has been studied intensively in the literature as in, for example, Hansen (1996) and Lee et al. (2011) to cite a few. Our testing problem is different from this Davies' problem and does not involve a loss of identification. Another related testing problem is the testing of the jump hypothesis against more general transition functions like Kim and Seo (2017).

Next, we establish the consistency of the test. Since $\hat{\mathbb{S}}_n \xrightarrow{p} EU_i^2$ while $\tilde{\mathbb{S}}_n \xrightarrow{p} EU_i^2 + c$ for some $c > 0$, which is due to the rank conditions in Assumptions 1 and 2, T_n diverges to $+\infty$ under the alternative. Formally,

Theorem 3. Under Assumptions 1 and 2 and the alternative H_1 (17),

$$P\{T_n > c\} \to 1,$$

for any $c < \infty$.

As the limiting distribution of T_n is not pivotal as it depends on the multiple restrictions and conditional heteroskedasticity, it is not practically useful to derive an explicit expression of its limit distribution, and hence we do not pursue it here. Instead, to compute its critical values, we proceed by examining a valid bootstrap to estimate the *p*-values of the test statistic.

4.3. Bootstrapping Continuity Test

This section provides a bootstrap procedure for the test of continuity based on the T_n statistic. We shall mention that the bootstrap-based test inversion confidence interval for the unknown threshold parameter τ_0 is developed in Hidalgo et al. (2019). We proceed as follows:

Algorithm 1: Bootstrapping the Continuity Test

STEP 1. Obtain both LSE $\hat{\theta} = \left(\hat{\alpha}', \hat{\tau}\right)'$ and CLSE $\tilde{\theta} = \left(\tilde{\alpha}', \tilde{\tau}\right)'$ of $\theta_0 = \left(\alpha_0', \tau_0\right)'$ as given in (11), (12), and (10), and compute the least squares residuals

$$\hat{U}_i = Y_i - X_i\left(\hat{\tau}\right)' \hat{\alpha}, \quad i=1,\ldots,n.$$

STEP 2. Generate $\{\eta_i\}_{i=1}^{n}$ as *i.i.d.* zero mean random variables with unit variance and finite fourth moments, and compute

$$Y_i^* = X_i(\tilde{\tau})' \tilde{\alpha} + \hat{U}_i \eta_i, \quad i=1,...,n.$$

STEP 3. Using $\{Y_i^*\}_{i=1}^{n}$ and $\{X_i\}_{i=1}^{n}$, construct the bootstrap statistic T_n^* as in (18) of Section 4. Specifically,

$$T_n^* = n\frac{\tilde{\mathbb{S}}_n^* - \hat{\mathbb{S}}_n^*}{\hat{\mathbb{S}}_n^*},$$

where $$\hat{\mathbb{S}}_n^* = \min_{\theta} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^* - X_i(\tau)'\alpha(\tau)\right)^2,$$

$$\tilde{\mathbb{S}}_n^* = \min_{\theta:\delta_1+\delta_3\tau_0=0;\delta_2=0} \frac{1}{n}\sum_{i=1}^{n}\left(Y_i^* - X_i(\tau)'\alpha(\tau)\right)^2.$$

STEP 4. Compute the bootstrap p-value, p^* by repeating *STEPS 2 and 3 B* times and obtain the proportion of times that T_n^* exceeds the sample statistic T_n given in (18).

The validity of this procedure is given in the following theorem. As usual, the superscript "*" indicates the bootstrap quantities and convergences of bootstrap statistics conditional on the original data. The notation $\xrightarrow{d^*}$ *in Probability* signifies the convergence in probability of the random distribution functions of the bootstrap statistics in terms of the uniform metric.

Theorem 4. *Suppose Assumptions 1 and 2 hold. Then, under Assumption* **C**,

$$T_n^* \xrightarrow{d^*} T \text{ in Probability},$$

where T denotes the limit variable in Theorem 2.

5. MONTE CARLO EXPERIMENT

As in Hidalgo et al. (2019, Section 5), our simulation is based on the following three specifications:

$$\begin{aligned}
A: &\quad Y_i = 2 + 3Q_i + \delta Q_i \mathbb{I}\{Q_i > \tau_0\} + U_i, \quad \tau_0 = E(Q_i) = 0,\\
B: &\quad Y_i = 2 + 3Q_i + \delta Q_i \mathbb{I}\{Q_i > \tau_0\} + U_i, \quad \tau_0 = E(Q_i) = 2,\\
C: &\quad Y_i = 2 + 3X_i + \delta X_i \mathbb{I}\{Q_i > \tau_0\} + U_i, \quad \tau_0 = E(Q_i) = 2.
\end{aligned}$$

Settings B and C are jump models considered in Hansen (2000, Section 4.2), and setting A represents the kink case. However, our data generating process differs from Hansen (2000) in that we assume the conditional heteroscedasticity in U_i such that $U_i = |Q_i| e_i$ where $\{e_i\}_{i\geq 1}$ and $\{Q_i\}_{i\geq 1}$ were generated as mutually independent and *i.i.d.* normal random variables with unit variance. This leads to conditional heteroscedasticity of the form $E(U_i^2 | Q_i) = Q_i^2$, in contrast to Hansen (2000) where U_i was generated from $\mathcal{N}(0,1)$. In A, we generated X_i as *i.i.d.* draws from $\mathcal{N}(2,1)$, independent of $\{U_i\}_{i\geq 1}$ and $\{Q_i\}_{i\geq 1}$. For the grid $\mathbb{T}_n$ used in estimation of τ_0, we discard 10% of extreme values of realized Q_i and use $n/2$ number of equidistant points.

We investigate finite-sample performance of the bootstrap-based test of continuity proposed in Section 4. Results are based on 10,000 iterations, with one bootstrap per iteration, using the warp-speed method of Giacomini et al. (2013). Table 1 presents Monte Carlo size results of the test for nominal size $s = 0.1, 0.05, 0.01$. We first try two settings for δ that are in line with conditions of Theorem 2: for columns 2–4 in rows 3–6, δ is fixed at 2,[1] while δ is shrinking in n for columns 5–7, with $\delta = n^{-1/4}\sqrt{10}/4 = 0.25, 0.1988, 0.1672, 0.1406$ for $n = 100, 250, 500, 1{,}000$, respectively. $\delta = 0.25$ was the smallest δ used in Hansen (2000) and by letting it diminish further for n = 250, 500, 1,000, we hope to investigate size performance of our test for very small δ. The results show satisfactory size performance for both cases, with the fixed δ case producing better size results, as expected. It is reassuring that the size performance is satisfactory for δ as small as 0.1406 in the diminishing δ case. We have also tried $\delta = n^{-1/2}\sqrt{10}/4 = 0.0791, 0.05, 0.0354, 0.025$ for $n = 100, 250, 500, 1{,}000$ and δ = 0. These settings are outside the scope of the chapter, but obtaining some informal evidence of what happens in such cases is nonetheless of interest, and the results are reported in rows 9–12 of Table 1.

Table 1. Monte Carlo Size of Test of Continuity, Setting A.

Δ	2			$n^{-1/4}\sqrt{10}/4$		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.1195	0.0737	0.0204	0.152	0.0843	0.0177
250	0.0832	0.0477	0.0122	0.1404	0.0775	0.0162
500	0.0897	0.0408	0.0076	0.1318	0.0684	0.0135
1,000	0.105	0.0491	0.0109	0.1312	0.0662	0.0135
δ	$n^{-1/2}\sqrt{10}/4$			0		
$n\backslash s$	0.1	0.05	0.01	0.1	0.05	0.01
100	0.1508	0.0867	0.0165	0.1485	0.0837	0.0177
250	0.1347	0.072	0.0144	0.1394	0.0745	0.0147
500	0.1263	0.0653	0.0147	0.1237	0.0633	0.0141
1,000	0.1444	0.0749	0.0154	0.14	0.0737	0.0164

The size results are somewhat worse than in the earlier two cases for larger $n =$ 1,000, but still they are satisfactory, with some over-sizing, not in excess of half the nominal size.

Tables 2 and 3 report Monte Carlo power results for the test of continuity for the nominal size of test s in jump settings B and C, respectively. Power results naturally are affected by the size of δ, and four sets of δ have been tried. For the first three sets, we use values tried in Hansen (2000), $\delta = 0.25, 0.5, 1$, for $n = 100$ and let it diminish according to Assumption **J** with $\varphi = 1/4$. For the fourth set, we fix $\delta = 2$ across n. As expected, power improves as δ gets larger and as n increases. Power is better in setting C $(Q_i \neq X_i)$ than setting B $(Q_i = X_i)$, which reflects the larger departure of C from A, compared to that of B. Even in setting B, the reported power results are promising, with the power being practically 1 for $\delta = 2$ with $n = 250, 500$.

Table 2. Monte Carlo Power of Test of Continuity, Setting B.

δ	δ	$n\backslash s$	0.1	0.05	0.01
$n^{-1/4}\sqrt{10}/4$	0.25	100	0.1313	0.0661	0.0134
	0.1988	250	0.1205	0.0564	0.0097
	0.1672	500	0.1151	0.0574	0.0089
$n^{-1/4}\sqrt{10}/2$	0.5	100	0.1525	0.0726	0.013
	0.3976	250	0.1502	0.068	0.0098
	0.3344	500	0.1656	0.0787	0.0117
$n^{-1/4}\sqrt{10}$	1	100	0.3282	0.1918	0.0365
	0.7953	250	0.4684	0.3028	0.0623
	0.6687	500	0.637	0.4797	0.1685
Fixed	2	100	0.9471	0.8854	0.6293
	2	250	1	0.9997	0.9986
	2	500	1	1	1

Table 3. Monte Carlo Power of Test of Continuity, Setting C.

δ	δ	$n\backslash s$	0.1	0.05	0.01
$n^{-1/4}\sqrt{10}/4$	0.25	100	0.3756	0.2452	0.0635
	0.1988	250	0.4014	0.2535	0.069
	0.1672	500	0.4531	0.2783	0.089
$n^{-1/4}\sqrt{10}/2$	0.5	100	0.5779	0.4076	0.1365
	0.3976	250	0.7116	0.54	0.2212
	0.3344	500	0.8516	0.7071	0.3729
$n^{-1/4}\sqrt{10}$	1	100	0.9638	0.9194	0.709
	0.7953	250	0.9978	0.9939	0.9546
	0.6687	500	1	0.9998	0.9988
Fixed	2	100	1	1	0.9999
	2	250	1	1	1
	2	500	1	1	1

6. EMPIRICAL APPLICATION: GROWTH AND DEBT

Reinhart and Rogoff (2010) suggest that above some threshold, the higher debt-to-GDP ratio is related to a lower GDP growth rate, reporting 90% as their estimate for the threshold. There have been many studies that investigate the Reinhart–Rogoff hypothesis with the threshold regression models; see Hansen (2017) for references on earlier studies that utilize discontinuous threshold regression models. Hansen (2017) fitted a kink threshold model to a time series of US annual data and Hidalgo et al. (2019) applied their robust inference procedure that is valid for both kink and jump design to Sweden, UK, and Australia data as well as US data used in Hansen (2017). Hansen (2017) mentions that

> one could imagine testing the assumption of continuity within the threshold model class. This is a difficult problem, one to which we are unaware of a solution, and therefore is not pursued in this paper.

As we have developed testing procedures for continuity in this chapter, we follow up on Hansen's (2017) investigation and present complementary analysis to Hidalgo et al. (2019).

Hansen (2017) used long-span US annual data (1792–2009, $n = 218$) on real GDP growth rate in year t (y_t) and debt-to-GDP ratio of the previous year (q_t) and reported the following estimated equation with standard errors in parentheses:

$$\hat{y}_t = \underset{(0.69)}{3.78} + \underset{(0.09)}{0.28}\, y_{t-1} + \begin{cases} \underset{(0.026)}{0.033}(q_t - 43.8), & \text{if } q_t \le \underset{(12.1)}{43.8} \\ -\underset{(0.048)}{0.067}(q_t - 43.8), & \text{if } q_t > \underset{(12.1)}{43.8} \end{cases}$$

We carried out our test of continuity given in Section 4 with 10,000 bootstraps and obtained p-value of 0.029, hence reject the null of continuity at 5% nominal level. This result is in line with Hansen (2017) that reported p-value of 0.15 for the test of the presence of a kink threshold effect. We remark that Hidalgo et al. (2019) obtained p-value of 0.047 for the test of the presence of threshold effect using Hansen's (1996) test without imposing the kink model and rejected the null of no threshold effect at 5% nominal level.

The fitted jump model is given by:

$$\hat{y}_t = \begin{cases} \underset{(0.87)}{4.82} - \underset{(0.16)}{0.052}\, y_{t-1} - \underset{(0.049)}{0.114}\, q_t, & \text{if } q_t \le 17.2 \\ \underset{(0.74)}{2.78} + \underset{(0.082)}{0.49}\, y_{t-1} - \underset{(0.012)}{0.017}\, q_t, & \text{if } q_t > 17.2 \end{cases}$$

Lower regime contains 99 observations and upper regime contains 109 observations. Hidalgo et al. (2019) obtained grid bootstrap confidence intervals for τ_0 that are (10.8, 38.6) for 90% confidence level and (10.5, 39) for 95% confidence level. These confidence intervals do not contain the CLSE $\tilde{\tau} = 43.8$, which is not surprising as the null of continuity is rejected in our test.

Hidalgo et al. (2019) also conducts similar analysis with Sweden data for the period spanning 1881–2009 (n = 129). The p-value for Hansen's (1996) test of presence of threshold effect is reported to be 0.048. Applying our continuity tests based on 10,000 bootstraps yield *p*-value of 0.091. The estimated jump model is:

$$\hat{y}_t = \begin{cases} \underset{(2.17)}{1.12} - \underset{(0.24)}{0.2}\, y_{t-1} + \underset{(0.11)}{0.13} q_t, & \text{if } q_t \le 21.3 \\ \underset{(0.58)}{1.86} + \underset{(0.11)}{0.48}\, y_{t-1} - \underset{(0.0082)}{0.004} q_t, & \text{if } q_t > 21.3 \end{cases}$$

The number of observations of the lower regime is 61, and the upper regime has 68 observations.

The grid bootstrap confidence intervals for τ_0 obtained in Hidalgo et al. (2019) were (15.3, ∞) and (16.4, ∞) for 95% and 90% confidence levels. This is in line with our finding that the confidence interval for τ_0 tends to become much wider as the model becomes a kink model, as reflected by the cube-root convergence rate.

The coefficients of debt-to-GDP ratio were also not significant in the estimated kink model, which need to be read with caution in the light of the continuity test:

$$\hat{y}_t = \underset{(0.58)}{2.89} + \underset{(0.13)}{0.048}\, y_{t-1} + \begin{cases} \underset{(0.3)}{0.24}(q_t - 15.5), & \text{if } q_t \le \underset{(5.75)}{15.5} \\ -\underset{(0.014)}{0.0008}(q_t - 15.5), & \text{if } q_t > \underset{(5.75)}{15.5} \end{cases}$$

whereby the lower regime had 15 observations and the upper regime contained 114 observations. Note that CLSE $\tilde{\tau} = 15.5$ is contained in the confidence interval.

We conclude that there is substantial heterogeneity across countries in the relationship between the GDP growth and the debt-to-GDP ratio, not only in the values of model parameters but also in the type of suitable models.[2]

7. CONCLUSION

This chapter has developed the continuity test that concerns an interesting hypothesis involving both the regression coefficients and the threshold. The continuity test is complementary to the robust inference presented in Hidalgo et al. (2019). The robust inference concerns inference for each type of parameter separately.

There are several interesting future research topics. First, we have considered the continuity of mean regression function. However, the same issue of continuity also arises in the quantile regression with a threshold. As the continuity of quantile function is not guided by the economic theory, it would be useful to develop a data-driven method for detecting discontinuity of quantile function. Another direction could be to study the high-dimensional model with a threshold. This model has been considered in Lee et al. (2016, 2018). Finally, it would be interesting to find an estimator that matches with the minimax lower bound in Proposition 2.

APPENDIX

A. PROOFS OF MAIN THEOREMS

A.1. Proof of Proposition 1

Hidalgo et al. (2019) considers the case where the threshold is fixed over the sample size, namely, $\varphi = 0$. We generalize this result to the diminishing threshold, $0 < \varphi < 1/2$. Without loss of generality, we may assume that $\tau_0 = 0$. Let $\overline{\psi} := \psi - \psi_0$ for any parameter ψ and $\mathbb{I}_i(a;b) = \mathbb{I}\{a < Q_i < b\}$. Denote $v := \beta + \delta$.

We derive the convergence rate of the LSE, that is, we show that

$$\left(\sqrt{n}\left(\hat{\delta}_1 - \delta_{01}\right), \sqrt{n}\left(\hat{\delta}_3 - \delta_{03}\right), n^{\frac{1-2\varphi}{3}}\left(\hat{\tau} - \tau_0\right)\right) = O_p(1).$$

Note that we can write

$$\mathbb{S}_n(\theta) - \mathbb{S}_n(\theta_0) = \mathbb{A}_{n1}(\theta) + \mathbb{A}_{n2}(\theta) + \mathbb{A}_{n3}(\theta) + \mathbb{B}_{n1}(\theta) + \mathbb{B}_{n2}(\theta) + \mathbb{B}_{n3}(\theta),$$

where

$$\mathbb{A}_{n1}(\theta) = \overline{v}' \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \mathbb{I}_i(\tau)\overline{v}; \qquad \mathbb{A}_{n2}(\theta) = \overline{\beta}' \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \mathbb{I}_i(-\infty;0)\overline{\beta}$$

$$\mathbb{A}_{n3}(\theta) = \left(\overline{\beta} + \delta_0\right)' \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \mathbb{I}_i(0;\tau)\left(\overline{\beta} + \delta_0\right)$$

$$\mathbb{B}_{n1}(\theta) = \overline{v}' \frac{2}{n}\sum_{i=1}^{n} X_i U_i \mathbb{I}_i(\tau); \qquad \mathbb{B}_{n2}(\theta) = \overline{\beta} \frac{2}{n}\sum_{i=1}^{n} X_i U_i \mathbb{I}_i(-\infty;0)$$

$$\mathbb{B}_{n3}(\theta) = \left(\overline{\beta} + \delta_0\right)' \frac{2}{n}\sum_{i=1}^{n} X_i U_i \mathbb{I}_i(0;\tau),$$

and $\tau > 0$. The case where $\tau < 0$ can be handled similarly. We follow the approach taken in Hidalgo et al. (2019, Proposition 1), for which we need to verify that for any $\epsilon > 0$, there exist $C > 0$, $\eta > 0$ and n_0 such that for all $n > n_0$,

$$\Pr\left\{ \inf_{\frac{C}{n^{1/2}} < \|\overline{v}\|, \|\overline{\beta}\| < \eta; \frac{C}{n^{(1-2\varphi)/3}} < \|\overline{\tau}\| < \eta} \sum_{\ell=1}^{3}\left(\mathbb{A}_{n\ell}(\theta) + \mathbb{B}_{n\ell}(\theta)\right) \le 0 \right\} < \epsilon. \tag{19}$$

Note the change of the lower bound for $\overline{\tau}$ from $n^{1/3}$ to $n^{(1-2\varphi)/3}$. To prove (19), it suffices to show that for each $\ell = 1,2,3$,

$$\Pr\left\{ \inf_{\frac{C}{n^{1/2}} < \|\overline{v}\|, \|\overline{\beta}\| < \eta; \frac{C}{n^{(1-2\varphi)/3}} < \|\overline{\tau}\| < \eta} E\left(\mathbb{A}_{n\ell}(\theta)\right)/2 + \left(\mathbb{A}_{n\ell}(\theta) - E\left(\mathbb{A}_{n\ell}(\theta)\right)\right) \le 0 \right\} < \epsilon, \tag{20}$$

and

$$\Pr\left\{\inf_{\frac{C}{n^{1/2}}<\|\overline{v}\|,\|\overline{\beta}\|<\eta;\frac{C}{n^{(1-2\varphi)/3}}<\|\overline{\tau}\|<\eta} E\left(\mathbb{A}_{n\ell}(\theta)\right)/2+\mathbb{B}_{n\ell}(\theta)\leq 0\right\}<\epsilon. \tag{21}$$

Notice that the only difference from the Proof of Proposition 1 in Hidalgo et al. (2019) due to the assumption of $\delta_0 = d_0 \cdot n^{-\varphi}$ lies in the case $\ell = 3$. Therefore, it is sufficient to handle the contribution from $\mathbb{A}_{n3}(\theta)$ and $\mathbb{B}_{n3}(\theta)$. Since $E\left(X_i X_i{}'\mathbb{I}_i(0;\tau)\right)$ is positive definite, we may consider

$$\widetilde{\mathbb{A}}_{n3}(\theta)=(\beta_3-\beta_{30}+\delta_{30})^2\frac{1}{n}\sum_{i=1}^{n}Q_i^2\mathbb{I}_i(0;\tau);\quad \widetilde{\mathbb{B}}_{n3}(\theta)=(\beta_3-\beta_{30}+\delta_{30})\frac{2}{n}\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i(0;\tau).$$

Accordingly, we decompose the parameter space over which the infimum is taken as

$$\Xi_k=\left\{\theta:\frac{C}{n^{1/2}}<\|\overline{v}\|,\|\overline{\beta}\|<\eta,\frac{C2^{k-1}}{n^{(1-2\varphi)/3}}<\overline{\tau}<\frac{C2^{k}}{n^{(1-2\varphi)/3}}\right\};k=1,\ldots,\log_2\left(\frac{\eta}{C}n^{(1-2\varphi)/3}\right).$$

Recall that we have assumed that $\tau \geq 0$, as the case $\tau \leq 0$ follows similarly.

Also recall that we impose that $\delta_{30} = d_3 \cdot n^{-\varphi}$. Choose a positive real number C_1 such that $E\left(Q_i^2\mathbb{I}_i(0;\xi)\right)\geq C_1\xi^3$ and $|d_3| > C_1 > 0$. Then, we have

$$\begin{aligned}
&\Pr\left\{\inf_{\Xi_k} E\left(\widetilde{\mathbb{A}}_{n3}(\theta)\right)/2+\widetilde{\mathbb{B}}_{n3}(\theta)\leq 0\right\}\\
&\leq \Pr\left\{\inf_{\Xi_k}|d_3|n^{-\varphi}E\left(Q_i^2\mathbb{I}_i(0;\tau)\right)\leq\sup_{\Xi_k}\left\|\frac{4}{n}\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i(0;\tau)\right\|\right\}\\
&\leq \Pr\left\{\frac{C_1^2C^3}{32n^{(1-2\varphi)/2}}2^{3k}\leq\sup_{\Xi_k}\left\|\frac{1}{n^{1/2}}\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i(0;\tau)\right\|\right\}\\
&\leq (32C^{-3/2}C_1^{-2}C_2)2^{-3k/2},
\end{aligned}$$

by Lemma 2 and Markov's inequality where C_2 is a constant in Lemma 2. Letting C be sufficiently large, we obtain the inequality (21) from the summability of $2^{-3k/2}$. The remaining steps to obtain the convergence rate of $n^{(1-2\varphi)/3}$ are identical to Hidalgo et al. (2019, Proposition 1 and Theorem 1). ■

A.2. Proof of Proposition 2

The proof for the lower bound in Proposition 2 relies on Le Cam's method (Le Cam, 1973). Before proceeding to the proof, we collect some notations and basic properties of divergence measures. Let $\mathbb{P}, \mathbb{Q}$ be any probability measures on the measurable space $(\mathcal{X},\mathcal{A})$, where $\mathcal{A}$ is a σ-field on $\mathcal{X}$. Then the total variation distance between $\mathbb{P}$ and $\mathbb{Q}$ is defined as $d_{TV}(\mathbb{P},\mathbb{Q}) := \sup_{A\in\mathcal{A}} |\mathbb{P}(A)-\mathbb{Q}(A)|$ and Kullback–Leibler(KL) divergence from $\mathbb{P}$ to $\mathbb{Q}$ is $d_{KL}(\mathbb{P},\mathbb{Q}) = \int \log\frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P}$ if $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$, or $+\infty$, otherwise. It is known that for all probability measures $\mathbb{P}$ and $\mathbb{Q}$,

$$d_{TV}(\mathbb{P},\mathbb{Q}) \le \sqrt{\frac{1}{2} d_{KL}(\mathbb{P},\mathbb{Q})}, \tag{22}$$

which is called Pinsker's inequality. Finally, consider a regression model, $Y = g(X)+U$, where $U \sim \mathcal{N}(0,\sigma^2(X))$ given X. We write $\mathbb{P}_g$ for the joint distribution of (Y, X). Assume that $\sigma^2(X) \ge \underline{\sigma}^2 > 0$. Then $d_{KL}(\mathbb{P}_{g_0},\mathbb{P}_{g_1}) \le \frac{1}{2\underline{\sigma}^2} \| g_1 - g_0 \|^2_{\ell_2(\mathbb{P}_X)}$.

We state a version of Le Cam's method from Yu (1997). Let $\mathcal{P}$ be a class of probability measures. Let $X_1, X_2, \cdots, X_n$ be random variables sampled from $\mathbb{P}\in\mathcal{P}$ in *i.i.d.* manner and $\mathbb{P}^n$ denote the corresponding product measure. Define a function θ which maps a probability measure in $\mathcal{P}$ into the metric space Θ with a metric ρ. We write $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ for an estimator of θ. For any probability measure $\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}$, the ρ-minimax risk is lower bounded as follows:

$$\inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}\left(\rho(\theta(\mathbb{P}), \hat{\theta}(X_1,\cdots,X_n))\right) \ge \rho(\theta(\mathbb{P}_0),\theta(\mathbb{P}_1)) \frac{1-d_{TV}(\mathbb{P}_0^n,\mathbb{P}_1^n)}{2}, \tag{23}$$

where the inifimum is taken over all estimators $\hat{\theta}$.

Note that $d_{KL}(\mathbb{P}_0^n,\mathbb{P}_1^n) = n d_{KL}(\mathbb{P}_0,\mathbb{P}_1)$. Combining (23) with Pinsker's inequality, it is straightforward to see that the minimax risk is lower bounded as follows:

Lemma 1. *Let* $\{\mathbb{P}_{0,n}\}_{n\in\mathbb{N}}$ *and* $\{\mathbb{P}_{1,n}\}_{n\in\mathbb{N}}$ *be any two sequences of probability measures in* $\mathcal{P}$. *Let* $\{\mathbb{P}_{0,n}\}_{n\in\mathbb{N}}$ *and* $\{\mathbb{P}_{1,n}\}_{n\in\mathbb{N}}$ *satisfy*

$$d_{KL}(\mathbb{P}_{0,n},\mathbb{P}_{1,n}) \le \frac{1}{2n}$$

for all $n\in\mathbb{N}$. *Then,*

$$\inf_{\hat{\theta}} \sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}\left(\rho(\theta(\mathbb{P}), \hat{\theta}(X_1,\cdots,X_n))\right) \ge \frac{1}{4}\rho(\theta(\mathbb{P}_{0,n}),\theta(\mathbb{P}_{1,n}))$$

for all $n\in\mathbb{N}$, *where the infimum is taken over all estimators* $\hat{\theta}$.

We prove Proposition 2 with this lemma. Since we are considering *i.i.d.* sampling, we drop the subscript i of random variables. Let $\mathbb{P}_{(\alpha,\gamma)}$ denote the joint distribution of (Y, X) where $Y = X'(\tau)\alpha + U$, and $U \sim \mathcal{N}(0,\sigma^2(Q))$ given X. Let

$\xi = \inf \mathbb{T}$, $\beta = 0 \in \mathbb{R}^d$, $\delta = (-\kappa\xi, 0, \cdots, 0, \kappa)$ and $\alpha = (\beta', \delta')'$. First, we consider the case that $n \geq \frac{3\underline{\sigma}^2}{\overline{f}\kappa^2\eta^3}$. Let $\tau_0 = \xi$ and $\tau_{1,n} = \xi + \left(\frac{3\underline{\sigma}^2}{n\overline{f}\kappa^2}\right)^{1/3}$, then $\tau_0, \tau_{1,n} \in \mathbb{T}$ for all n. We can obtain the following inequality under this choice of parameter sequences,

$$
\begin{aligned}
d_{KL}\left(\mathbb{P}_{(\alpha,\tau_0)}, \mathbb{P}_{(\alpha,\tau_{1,n})}\right) &\leq \frac{1}{2\underline{\sigma}^2} \| \kappa(Q - \tau_0)\mathbb{I}\{\tau_0 < Q \leq \tau_{1,n}\} \|^2_{\ell_2(\mathbb{P}_X)} \\
&= \frac{\kappa^2}{2\underline{\sigma}^2} \int_{\tau_0}^{\tau_{1,n}} (q - \tau_0)^2 f(q) dq \leq \frac{\overline{f}\kappa^2}{6\underline{\sigma}^2} (\tau_{1,n} - \tau_0)^3 = \frac{1}{2n}
\end{aligned}.
$$

Applying Lemma 1 with $\{\mathbb{P}_{(\alpha,\tau_0)}\}_{n\in\mathbb{N}}$ and $\{\mathbb{P}_{(\alpha,\tau_{1,n})}\}_{n\in\mathbb{N}}$, we get the desired result.

Next, assume that $n < \frac{3\underline{\sigma}^2}{\overline{f}\kappa^2\eta^3}$. In this case, we let $\tau_1 = \sup \mathbb{T}$. Then,

$$
\begin{aligned}
d_{KL}(\mathbb{P}_{(\alpha,\tau_0)}, \mathbb{P}_{(\alpha,\tau_1)}) &\leq \frac{\kappa^2}{2\underline{\sigma}^2} \int_{\tau_0}^{\tau_1} (q - \tau_0)^2 f(q) dq \\
&\leq \frac{\overline{f}\kappa^2\eta^3}{6\underline{\sigma}^2} \leq \frac{1}{2n}
\end{aligned}.
$$

Therefore, the minimax risk is lower bounded by $\frac{\eta}{4}$ as desired. ■

A.3. Proof of Theorem 1

Theorem 1 is parallel to the Hidalgo et al. (2019, Theorem 1). Therefore, we briefly review the proof and emphasize the difference caused by the diminishing threshold assumption.

Observing the continuity of "argmin" function and the convergence rates in Proposition 1, we only need to consider the weak limit of

$$
\mathbb{G}_n(h, g) = n\left(\mathbb{S}_n\left(\alpha_0 + \frac{h}{n^{1/2}}, \frac{g}{n^{(1-2\varphi)/3}}\right) - \mathbb{S}_n(\alpha_0, 0)\right), \tag{24}
$$

where τ_0 is assumed to be 0. Note that

$$
\sup_{\|h\|, |g| \leq C} \left| \mathbb{G}_n(h, g) - \widetilde{\mathbb{G}}_n(h, g) \right| = o_p(1),
$$

where

$$\begin{aligned}\widetilde{\mathbb{G}}_n(h,g) &= \left\{h'\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'h - h'\frac{2}{n^{1/2}}\sum_{i=1}^{n}\mathbf{X}_iU_i\right\} \\ &\quad +\delta_{30}\left\{\delta_{30}\sum_{i=1}^{n}Q_i^2\mathbb{I}_i\left(0;\frac{g}{n^{(1-2\varphi)/3}}\right) - 2\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i\left(0;\frac{g}{n^{(1-2\varphi)/3}}\right)\right\} . \\ &=: \widetilde{\mathbb{G}}_n^1(h)+\widetilde{\mathbb{G}}_n^2(g)\end{aligned}$$

Comparing to Proposition 1 of Hidalgo et al. (2019), it suffices to examine $\widetilde{\mathbb{G}}_n^2(g)$. Note that the first term of $\widetilde{\mathbb{G}}_n^2(g)$ uniformly converges to $3^{-1}d_{30}^2 f(0)|g|^3$ due to the Lemma 2.

Next, we show that the second term, $-2\delta_{30}\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i\left(0;\frac{g}{n^{(1-2\varphi)/3}}\right)$, weakly converges to $2d_{30}\sqrt{\frac{\sigma^2(0)f(0)}{3}}W(g^3)$. Let $Z_{ni}=n^{(1-2\varphi)/2}Q_iU_i\mathbb{I}_i\left(0;\frac{g}{n^{(1-2\varphi)/3}}\right)$. Note that

$$\frac{1}{n}\sum_{i=1}^{n}Z_{ni}^2\xrightarrow{p}\frac{\sigma^2(0)f(0)}{3}|g|^3 .$$

Covariances are calculated to be

$$n^{1-2\varphi}E\left(Q_i^2U_i^2\mathbb{I}_i\left\{\frac{g_1}{n^{(1-2\varphi)/3}},\frac{g_2}{n^{(1-2\varphi)/3}}\right\}\right)=\frac{\sigma^2(0)f(0)}{3}\left(g_2^3-g_1^3\right)+o(1),$$

where $g_2>g_1$, other cases can be treated similarly. Therefore, the second term of $\widetilde{\mathbb{G}}_n^2(g)$ converges to $2d_{30}\sqrt{\frac{\sigma^2(0)f(0)}{3}}W(g^3)$ from the martingale CLT.

Similar analysis on the covariance shows the aysmptotic independence between $\widetilde{\mathbb{G}}_n^1(h)$ and $\widetilde{\mathbb{G}}_n^2(g)$ Remaining details are identical to Hidalgo et al. (2019). ■

A.4. Proof of Theorem 2

From the convergence rate in Proposition 1, we examine the weak limit of

$$\mathbb{G}_n(h,g)=n\left\{\mathbb{S}_n\left(\alpha_0+\frac{h}{n^{1/2}},\tau_0+\frac{g}{n^{(1-2\varphi)/3}}\right)-\mathbb{S}_n\left(\alpha_0,\tau_0\right)\right\},$$

for $0\le\varphi<1/2$. From the proof of Theorem 1,

$$\sup_{\|h\|,|g|\le C}\left|\mathbb{G}_n(h,g)-\widetilde{\mathbb{G}}_n^1(h)-\widetilde{\mathbb{G}}_n^2(g)\right|=o_p(1),$$

where

$$\widetilde{\mathbb{G}}_n^1(h) = \left\{h'\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'h - 2h'\frac{1}{n^{1/2}}\sum_{i=1}^{n}\mathbf{X}_iU_i\right\}$$

$$\widetilde{\mathbb{G}}_n^2(g) = \delta_{30}\left\{\delta_{30}\sum_{i=1}^{n}Q_i^2\mathbb{I}_i\left(0;\frac{g}{n^{1/3}}\right) - 2\sum_{i=1}^{n}Q_iU_i\mathbb{I}_i\left(0;\frac{g}{n^{(1-2\varphi)/3}}\right)\right\}$$

and $\tilde{\mathbb{G}}_n^j(\cdot), j = 1,2,$ are mutually independent. Therefore, for the unconstrained estimator $\hat{\alpha}$ and $\hat{\tau}$ we could write

$$n\left(\mathbb{S}_n(\hat{\alpha},\hat{\tau}) - \mathbb{S}_n(\alpha_0,\tau_0)\right) = \min_h \tilde{\mathbb{G}}_n^1(h) + \min_g \tilde{\mathbb{G}}_n^2(g) + o_p(1).$$

Similarly, for the constrained estimator $\tilde{\alpha}$ and $\tilde{\tau}$ we can write, see, for example, Chan and Tsay (1998) or Hansen (2017), that

$$n\left(\mathbb{S}_n(\tilde{\alpha},\tilde{\tau}) - \mathbb{S}_n(\alpha_0,\tau_0)\right) = \min_l \mathbb{H}_n(\ell) + o_p(1),$$

where

$$\mathbb{H}_n(\ell) = \ell'\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i'\ell - 2\ell'\frac{1}{n^{1/2}}\sum_{i=1}^{n}\bar{\mathbf{X}}_iU_i$$

and $\bar{\mathbf{X}}_i = \left(X_i', (Q_i - \tau_0)\mathbb{I}_i(\tau_0), -(\beta_{30} + \delta_{30})\mathbb{I}_i(\tau_0)\right)' = R\mathbf{X}_i$. Note that $\mathbb{H}_n(\ell) + \tilde{\mathbb{G}}_n^1(h)$ converges weakly as a function of h and ℓ since both $\ell'\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i'\ell$ and $h'\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'h$ converges uniformly in probability by uniform law of large numbers (ULLN) and $\ell'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{X}}_iU_i + h'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}_iU_i$ converges weakly by the linearity, the CLT and Cramer-Rao device. The weak convergence of $\tilde{\mathbb{G}}_n^2(g)$ to $2d_{30}\sqrt{\frac{\sigma^2(\tau_0)f(\tau_0)}{3}}W(g^3) + \frac{d_{30}^2}{3}f(\tau_0)|g|^3$ where the Gaussian process W is defined in Theorem 1 and its asymptotic independence from $\tilde{\mathbb{G}}_n^1(h)$ are given in the Proof of Theorem 1. By the same argument it is asymptotically independent of $\mathbb{H}_n(\ell)$. To sum up, let

$$\begin{aligned}\mathbb{K}(h,g,\ell) &= \ell' E\bar{\mathbf{X}}_i\bar{\mathbf{X}}_i\ell + h' E\mathbf{X}_i\mathbf{X}_ih - 2\left(\ell'R + h'\right)B \\ &+ \left(2d_{30}\sqrt{\frac{\sigma^2(\tau_0)f(\tau_0)}{3}}W(g^3) + \frac{d_{30}^2}{3}f(\tau_0)|g|^3\right),\end{aligned}$$

where B is a $\mathcal{N}(0,\Omega)$ and independent of the Gaussian process W, and let

$$\mathbb{K}_n(h,g,\ell)=\tilde{\mathbb{G}}_n^1(h)+\tilde{\mathbb{G}}_n^2(g)+\mathbb{H}_n(\ell).$$

Then, it follows from the preceding discussion that

$$\mathbb{K}_n(h,g,\ell)\Rightarrow\mathbb{K}(h,g,\ell).$$

Furthermore,

$$\begin{aligned}
&n\left(\mathbb{S}_n(\tilde{\alpha},\tilde{\tau})-\mathbb{S}_n(\hat{\alpha},\hat{\tau})\right)\\
&=\min_{\ell:g=0,h=0}\mathbb{K}_n(h,g,\ell)-\min_{h:g=0,\ell=0}\mathbb{K}_n(h,g,\ell)-\min_{g:h=0,\ell=0}\mathbb{K}_n(h,g,\ell)+o_p(1),\\
&\xrightarrow{d}\min_{\ell:g=0,h=0}\mathbb{K}(h,g,\ell)-\min_{h:g=0,\ell=0}\mathbb{K}(h,g,\ell)-\min_{g:h=0,\ell=0}\mathbb{K}(h,g,\ell)
\end{aligned}$$

due to the continuous mapping theorem as the (constrained) minimum is a continuous operator and the fact that $\tilde{\mathbb{G}}_n^1(h),\tilde{\mathbb{G}}_n^2(g)$, and $\mathbb{H}_n(\ell)$ are zero at the origin. Certainly this limit is $O_p(1)$ and does not degenerate since $\tilde{\mathbb{G}}_n^2(g)$ is asymptotically independent of the other terms. The convergence of $\hat{\mathbb{S}}_n$ is straightforward by standard algebra and the ULLN and CLT and thus details are omitted.

A.5. Proof of Theorem 4

Recalling the meaning of the superscript "*" in Section 4.3, we begin by observing the consistency and rate of convergence of $\left(\hat{\alpha}^*,\hat{\tau}^*\right)$.

Proposition 3. *Suppose that Assumptions 1*, 2, and C *hold. Let* $\delta_0=d_0\cdot n^{-\varphi}$. *If* $0\le\varphi<1/2$,

(a) $\hat{\alpha}^*-\tilde{\alpha}=O_{p*}\left(n^{-1/2}\right)$ and $\hat{\tau}*-\tau_0=\mathrm{O}_{p*}\left(n^{-(1-2\varphi)/3}\right)$,

(b) $\hat{\alpha}^*$ *and* $\hat{\tau}^*$ *are asymptotically independent and (in probability)*

$$n^{1/2}(\hat{\alpha}^*-\tilde{\alpha})\xrightarrow{d^*}\mathcal{N}\left(0,M^{-1}\Omega M^{-1}\right)$$

$$n^{(1-2\varphi)/3}(\hat{\tau}*-\tau_0)\xrightarrow{d*}\operatorname*{argmax}_{g\in\mathbb{R}}\left(2d_{30}\sqrt{\frac{\sigma^2(\tau_0)f(\tau_0)}{3}}W\left(g^3\right)+\frac{d_{30}^2}{3}f(\tau_0)|g|^3\right).$$

Proposition 3(a) is similar to Proposition 5(a) of Hidalgo et al. (2019). The only difference is that the centering term of the resampling scheme is $(\tilde{\alpha},\tilde{\tau})$

instead of $(\hat{\alpha},\hat{\tau})$. Following the proof of Hidalgo et al. (2019, Theorem 3(a)), we obtain Proposition 3(b). Theorem 4 is a direct consequence of Proposition 3 and the same argument as the Proof of Theorem 2. ■

B. AUXILIARY LEMMA

Refer to Hidalgo et al. (2019) for the proofs of the lemmas in this section. For $j = 1$ or 2, let

$$
\begin{aligned}
J_n(\tau,\tau') &= \frac{1}{n^{1/2}}\sum_{i=1}^{n} U_i X_i \mathbb{I}_i(\tau;\tau') \\
J_{1n}(\tau,\tau') &= \frac{1}{n^{1/2}}\sum_{i=1}^{n} U_i |Q_i-\tau|^j\, \mathbb{I}_i(\tau;\tau') \\
J_{2n}(\tau) &= \frac{1}{n^{1/2}}\sum_{i=1}^{n} \left\{ |Q_i-\tau_0|^j\, \mathbb{I}_i(\tau_0;\tau) - E|Q_i-\tau_0|^j\, \mathbb{I}_i(\tau_0;\tau) \right\}
\end{aligned}
$$

and for some sequence $\{Z_i\}_{i=1}^{n}$,

$$
J_{3n}(\tau) = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\left(Z_i \mathbb{I}_i(\tau_0;\tau) - EZ_i\mathbb{I}_i(\tau_0;\tau)\right).
$$

Lemma 2. *Suppose Assumptions 1 and 2 hold for the sequence* $\{X_i,U_i\}_{i=1}^{n}$. *In addition, for* $J_{3n}(\tau)$, *assume that* $\{Z_i,Q_i\}_{i=1}^{n}$ *be a sequence of strictly stationary, ergodic, and* ρ *-mixing with* $\sum_{m=1}^{\infty}\rho_m^{1/2}<\infty$, $E|Z_i|^4<\infty$ *and, for all* $\tau\in\mathbb{T}$, $E\ E\left(|Z_i|^4 | Q_i=\tau\right)<C<\infty$. *Then, there exists* $n_0<\infty$ *such that for all* τ' *in a neighborhood of* τ_0 *and for all* $n>n_0$ *and* $\epsilon\geq n_0^{-1}$,

(a) $E \sup\limits_{\tau'<\tau<\tau'+\epsilon} \left|J_n(\tau',\tau)\right| \leq C\epsilon^{1/2}$

(b) $E \sup\limits_{\tau'<\tau<\tau'+\epsilon} \left|J_{1n}(\tau',\tau)\right| \leq C\epsilon^{1/2}\left(\epsilon + |\tau_0-\tau'|\right)^j$

(c) $E \sup\limits_{\tau_0<\tau<\tau_0+\epsilon} \left|J_{2n}(\tau)\right| \leq C\epsilon^{j+1/2}$

(d) $E \sup\limits_{\tau_0<\tau<\tau_0+\epsilon} \left|J_{3n}(\tau)\right| \leq C\epsilon^{1/2}$,

where $j = 1$ *or* 2.

C. FIGURES FOR EMPIRICAL APPLICATION IN SECTION 7

Figs. A1–A6 are scatter plots of residuals from fitting AR(1) model on y_t, plotted against q_t, superimposed with the estimated jump and kink models for the USA and Sweden.

As is made clear by these figures, one cannot expect to spot presence of discontinuity visually by examining the scatter plots, let alone discern if the kink or jump models better fits the data. To illustrate this point, in Figs. A5 and A6 we present the same scatter plots based on simulated data that were generated from the estimated jump equations of the two countries, which used y_0, q_t from the data and $u_t \sim \mathcal{N}(0, s^2)$, with sample variance of the residuals s^2, to reconstruct y_t. They are both superimposed with the jump equation that is the true data generating process for the simulated data.

This lack of visual guidance is indeed why the testing procedures of presence of threshold effect of, for example, Andrews (1993), Hansen (1996),

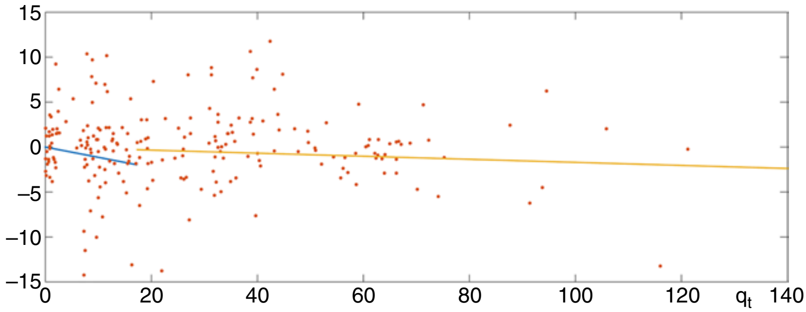


Fig. A1. Scatter Plot of AR(1) Residuals of y_t Against q_t and Estimated Jump Equation, USA.

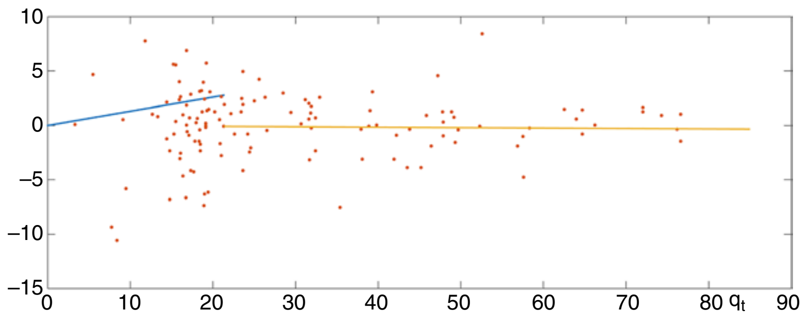


Fig. A2. Scatter Plot of AR(1) Residuals of y_t Against q_t and Estimated Jump Equation, Sweden.

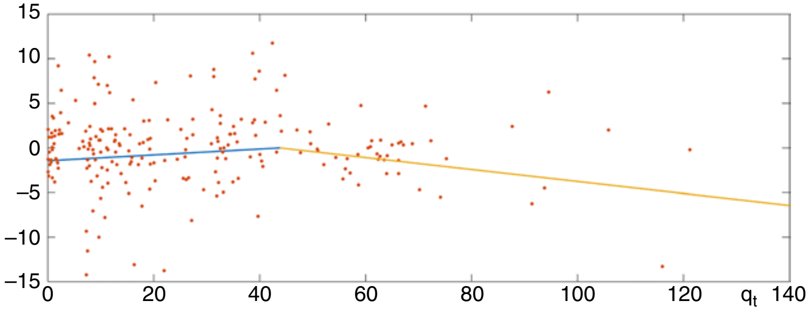


Fig. A3. Scatter Plot of AR(1) Residuals of y_t Against q_t and Estimated Kink Equation, USA.

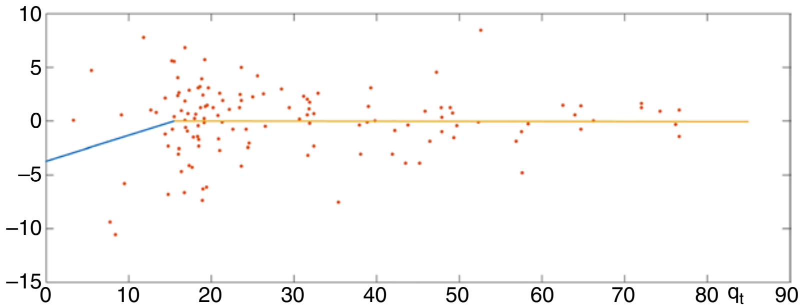


Fig. A4. Scatter Plot of AR(1) Residuals of y_t Against q_t and Estimated Kink Equation, Sweden.

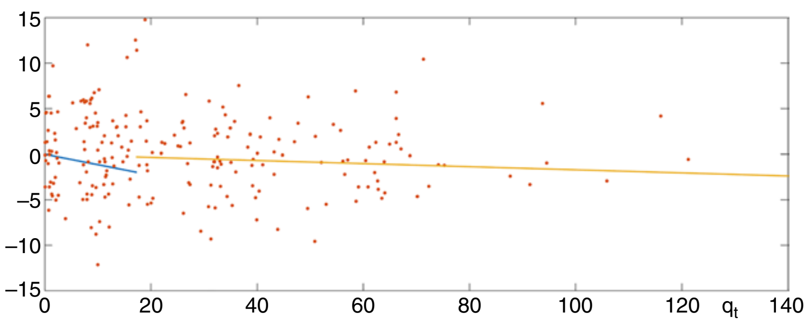


Fig. A5. Scatter Plot of AR(1) Residuals of y_t Against q_t From Reconstructed Data for the USA, and True Jump Equation.

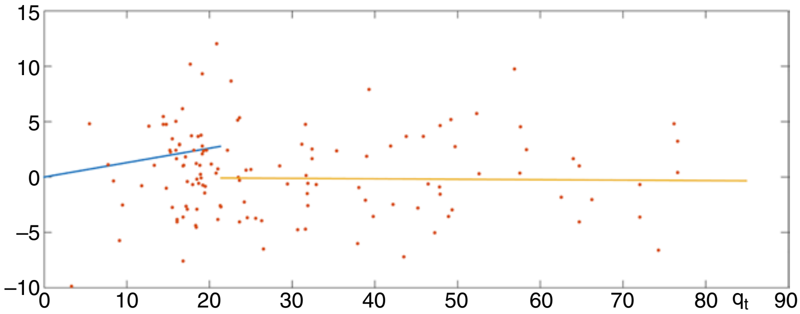


Fig. A6. Scatter Plot of AR(1) Residuals of y_t Against q_t From Reconstructed Data for Sweden, and True Jump Equation.

Lee et al. (2011), and our continuity testing procedure of this chapter are very much needed, and should be deployed in data analysis.

NOTES

1. $\delta = 2$ was the largest value of δ tried in Hansen (2000), although this chapter only looks at inference on τ in jump setups B and C.

2. Figs. A1–A6 present scatterplots of residuals from autoregression of y_t on y_{t-1} against q_t for the two countries, highlighting the importance of deploying the aforementioned tests in practice. Often neither the economic model nor data plots can tell us much about the true specification, and one should not expect to be able to spot the presence of discontinuity from visual inspection of data plots, let alone discern kink from jump. See Section C of Appendix for some further discussion.

ACKNOWLEDGMENTS

Seo gratefully acknowledges the support from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5A2A03046422) and from the Research Grant of the Center for Distributive Justice at the Institute of Economic Research, Seoul National University. Hidalgo acknowledges financial support from STICERD under the grant "Testing Economic Shape Restrictions."

REFERENCES

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*, 821–856.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, *21*, 520–533.

Chan, K. S., & Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, *85*, 413–426.

Chang, Y., Choi, Y., & Park, J. Y. (2017). A new approach to model regime switching. *Journal of Econometrics*, *196*(1), 127–143.

Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems-identified case. *The Annals of Statistics*, *3*, 49–83.

Giacomini, R., Dimitris, N. P., & White, H. (2013). A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, *29*, 567–589.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, *64*, 413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, *68*, 575–603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, *35*, 228–240.

Hansen, B. E., & Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics*, *110*(2), 293–318.

Hidalgo, J., Lee, J., & Seo, M. H. (2019). Robust inference for threshold regression models. *Journal of Econometrics*, *210*, 291–309.

Kim, Y. J., & Seo, M. H. (2017). Is there a jump in the transition? *Journal of Business & Economic Statistics*, *35*, 241–249.

Korostelev, A. (1987). On minimax estimation of a discontinuous signal. *Theory of Probability and its Applications*, *24*, 727–730.

Le Cam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, *1*, 38–53.

Lee, S., Liao, Y., Seo, M. H., & Shin, Y. (2018). Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association*, *113*(523), 1184–1194.

Lee, S., Liao, Y., Seo, M. H., & Shin, Y. (2021). Factor-driven two-regime regression. *The Annals of Statistics*, *49*(3), 1656–1678.

Lee, S., Seo, M. H., & Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association*, *106*, 220–231.

Lee, S., Seo, M. H., & Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(1), 193–210.

Park, J. Y., & Shintani, M. (2016). Testing for a unit root against transitional autoregressive models. *International Economic Review*, *57*(2), 635–664.

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review: Papers and Proceedings*, *100*, 573–578.

Seo, M. H. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *Journal of Econometrics*, *134*(1), 129–150.

Seo, M. H. (2008). Unit root test in a threshold autoregression: Asymptotic theory and residual-based block bootstrap. *Econometric Theory*, *24*(6), 1699–1716.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Oxford University Press.

Yu, B. (1997). Assouad, Fano, and Le Cam. In D. Pollard, E. Torgersen, & G. L. Yang (Eds.), *Research papers in probability and statistics: Festschrift in Honor of Lucien Le Cam* (pp. 423–435). Springer.

PART III

INFERENCE AND PREDICTION USING MODELS WITH TRENDING SERIES

This page intentionally left blank

CHAPTER 9

SEMIPARAMETRIC INDEPENDENCE TESTS BETWEEN TWO INFINITE-ORDER COINTEGRATED SERIES

Chafik Bouhaddioui[a], Jean-Marie Dufour[b] and Masaya Takano[b]

[a] *United Arab Emirates University, Al Ain, UAE*
[b] *McGill University, Montreal, QC, Cananda*

ABSTRACT

The authors propose a semiparametric approach for testing independence between two infinite-order cointegrated vector autoregressive series (IVAR(∞)). The procedures considered can be viewed as extensions of classical methods proposed by Haugh (1976, JASA*) and Hong (1996b,* Biometrika*) for testing independence between stationary univariate time series. The tests are based on the residuals of long autoregressions, hence allowing for computational simplicity, weak assumptions on the form of the underlying process, and a direct interpretation of the results in terms of innovations (or shocks). The test statistics are standardized versions of the sum of weighted squares of residual cross-correlation matrices. The weights depend on a kernel function and a truncation parameter. Multivariate portmanteau statistics can be viewed as a special case of our procedure based on the truncated uniform kernel. The asymptotic distributions of the test statistics under the null hypothesis are derived, and consistency is established against fixed alternatives of serial*

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 263–294

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A009

cross-correlation of unknown form. A simulation study is presented which indicates that the proposed tests have good size and power properties in finite samples.

Keywords: Infinite-order cointegrated vector autoregressive process; independence; causality; residual cross-correlation; consistency; asymptotic power

1. INTRODUCTION

Studying the dynamic relationship between two multivariate series is a fundamental objective of time-series analysis in statistics and econometrics. For example, in econometrics, this can help one to understand the associated economic mechanisms. In this context, a basic problem consists in testing independence (or the absence of serial cross-correlation) between two vector processes. The seminal paper on this problem is due to Haugh (1976), who proposed a general procedure for testing independence between two covariance-stationary autoregressive moving-average (ARMA) time series. His method is based on considering cross-correlations between residuals obtained after fitting univariate ARMA models on each series. Since the innovations of an ARMA model follow a white noise by assumption, this considerably simplifies the underlying distributional theory, and the corresponding tests are relatively simple to apply. Further, the corresponding statistics have a direct interpretation in terms of process innovations (or reduced-form shocks), a feature of interest in econometrics since innovations can often be interpreted as "shocks" to economic systems. Consequently, the possibility of focusing on "shock cross-correlations" should be useful in econometric research.

The work of Haugh (1976) has been extended by several authors; see Bouhaddioui and Dufour (2008), Bouhaddioui and Roy (2006a, 2006b), El Himdi and Roy (1997), Hallin and Saidi (2005, 2007), Hong (1996a), Pham et al. (2003), and Saidi (2007). Most of these studies focus on independence between two multivariate finite-order VAR or vector autoregressive moving-average (VARMA) models. El Himdi and Roy (1997) extended the procedure developed by Haugh (1976) in order to test non-correlation between two time series in the context of multivariate stationary and invertible VARMA models. This result was used by Hallin and Saidi (2005) to develop a test which takes into account a possible pattern in the signs of cross-correlations at different lags. In a non-parametric setup, Hallin et al. (1999) proposed a test for independence between two autoregressive time series which is based on autoregressive rank scores, while Hong (1998) proposed a test based on empirical distribution functions.

The stationarity condition is often unrealistic and constitutes a strong limitation. Even though stationarity may be achieved in many cases by differencing each series (so that distributional complications are avoided), this type of transformation can distort our ability to identify or accurately measure parameters and relations of interest. It is typically more interesting to be able to work with the original series without prefiltering (like differencing). This is especially important if we wish to study cointegrating relationships.

Engle and Granger (1987) introduced the concept of cointegration, which is used in many studies across several fields. In the case of a finite-order autoregressive cointegrated vector, Ahn and Reinsel (1990) developed an efficient estimation method for Gaussian processes. Yap and Reinsel (1995) proposed full- and reduced-rank Gaussian estimation procedures for cointegrated VARMA processes. For a good discussion of the related models, see Lütkepohl (2001). By exploiting the estimation methods proposed by Yap and Reinsel (1995), Pham et al. (2003) generalized the main result of El Himdi and Roy (1997) to the case of two cointegrated (or partially non-stationary) VARMA series. They proposed test statistics based on residual cross-correlation matrices $\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(12)}(j), |j| \leq M$ (where M does not depend on the sample size n) between the two residual series $\hat{\boldsymbol{a}}_t^{(1)}$ and $\hat{\boldsymbol{a}}_t^{(2)}$ resulting from fitting the *true* VARMA models to each of the original series $\boldsymbol{X}_t^{(1)}$ and $\boldsymbol{X}_t^{(2)}$. Under the hypothesis of non-correlation between the two series, they show that an arbitrary vector of residual cross-correlations asymptotically follows a multivariate normal distribution.

In practice, a finite-order VAR model can be a rough approximation to the true data-generating process of a multivariate time series. The "true" model may easily not be reducible to a parsimonious model with a small number of unknown parameters. From this perspective, a more flexible alternative approach assumes that the data are generated by an infinite-order autoregressive process. Such models lead one to consider a truncated (potentially long) autoregression as an approximation of the underlying process. In statistics and econometrics, one typically derives the properties of estimators and test criteria under the assumption of correct specification, even if model assumptions are clearly not fulfilled. For example, in VARMA estimation, it is well known that misspecification of the AR or MA orders can lead to inconsistent estimators. Further, the estimation of VARMA models is highly non-linear and raises difficult identification complications (in the sense of multiple observationally equivalent representations).

The autoregressive model fitting approach has been successfully applied by several authors: Akaike (1969), Berk (1974), and Parzen (1974) for spectral density estimation; Bhansali (1996), Lewis and Reinsel (1985), Lütkepohl (1985), and Parzen (1974) for prediction; Park (1990) and Saikkonen (1992) for inference in cointegrated systems; see also Lütkepohl (2005), Park et al. (2010), and Reinsel (1997). In previous work (Bouhaddioui & Roy, 2006b), we have generalized the work of El Himdi and Roy (1997) to the case of two stationary multivariate infinite-order autoregressive series VAR(∞). This result allows one to develop tests against serial cross-correlation at a particular lag or at a fixed number of lags j such as $|j| \leq M$, where M does not depend on the sample size n.

In the univariate stationary case, Hong (1996b) introduced an important extension of Haugh's procedure by proposing a class of spectral test statistics. His approach is semiparametric and valid for two infinite-order autoregressive series AR(∞). It is based on fitting an autoregressive model of order p to a series of n observations from each infinite-order autoregressive process. Following Berk (1974), the order p of the fitted autoregression is a function of the sample size.

This approach was also used by Duchesne (2005), Duchesne and Roy (2003), Hong (1999), and Shao (2009) for the case of two univariate long memory processes. In Bouhaddioui and Roy (2006a), it is extended to VAR (∞) models, hence protecting against misspecification of the underlying VARMA model. In contrast with Haugh's test, which is based on the residual cross-correlations at lag j such that $|j| \leq M$, the portmanteau test $\mathcal{Q}_n$ is consistent for a large class of serial cross-correlations alternatives of an arbitrary form between the two series.

The main objective of this chapter is to propose a semiparametric approach to test independence between two infinite-order cointegrated autoregressive [IVAR(∞)] models against alternatives where they would be correlated. These models were introduced by Saikkonen (1992) and involve much weaker conditions than those considered by Hallin and Saidi (2005), Pham et al. (2003), Saidi (2007), and Yap and Reinsel (1995); for further, discussion of this setup, see Lütkepohl and Saikkonen (1997), Saikkonen and Lütkepohl (1996), and Saikkonen and Luukkonen (1997). The problem of testing the absence of correlation between two IVAR(∞) was first considered in Bouhaddioui and Dufour (2008), where the asymptotic distribution of an arbitrary vector of residual cross-correlations and partial cross-correlations under the hypothesis of non-correlation of the two series is derived under the assumption that innovations are a strong white noise. However, the test statistics proposed in the latter paper only consider one lag at a time or a fixed number of lags j (e.g., $|j| \leq M$).

In this chapter, we propose a multivariate version of the weighted portmanteau statistic $\mathcal{Q}_n$ based on the sample cross-correlation matrices $\boldsymbol{R}_{\hat{a}}^{(12)}(j), |j| \leq n-1$, between the residuals $\hat{\boldsymbol{a}}_t^{(1)}$ and $\hat{\boldsymbol{a}}_t^{(2)}$. The latter are obtained by approximating two multivariate IVAR(∞) series with finite-order autoregressions whose order increases with the sample size at an appropriate rate. The test statistics continue to have a $\mathcal{N}(0,1)$ asymptotic distribution under the hypothesis of independence of the two series. The tests are consistent against serial cross-correlation of arbitrary form.

This chapter is organized as follows. Section 2 describes the statistical framework as well as some preliminary results. The new test statistics are introduced in Section 3. We show that their asymptotic distributions under the null hypothesis are $\mathcal{N}(0,1)$. In section 4, we establish the consistency of the tests. In Section 6, we present the results of a small Monte Carlo experiment on the level and power of the tests in finite samples, including the effect of the kernel. We conclude in Section 7. The proofs of all results are given in Appendix.

2. FRAMEWORK AND PRELIMINARY RESULTS

Following the notations of Saikkonen (1992), Saikkonen and Lütkepohl (1996), and Bouhaddioui and Dufour (2008), we consider a d-dimensional process $\boldsymbol{X} = \{X_t : t \in \mathbb{Z}\}$ partitioned into two subprocesses $\boldsymbol{X}_i = \{\mathbf{X}_{it} : t \in \mathbb{Z}\}, i = 1,2$, with d_1 and d_2 components respectively $(d_1 + d_2 = d)$. The data-generating process has the form:

$$\boldsymbol{X}_{1t} = \boldsymbol{C}_1 \boldsymbol{X}_{2t} + \varepsilon_{1t}, \tag{2.1}$$

$$\Delta \boldsymbol{X}_{2t} = \varepsilon_{2t}, \tag{2.2}$$

where $\boldsymbol{C}_1$ is a fixed $d_1 \times d_2$ matrix, Δ is the usual difference operator, and $\varepsilon_t = (\varepsilon_{1t}', \varepsilon_{2t}')'$ is a stationary process with zero mean and continuous spectral density matrix positive definite at frequency zero. $\boldsymbol{X}_{2t}$ is an integrated vector process of order one (with no cointegrating relationship), while $\boldsymbol{X}_{1t}$ and $\boldsymbol{X}_{2t}$ are cointegrated. By taking first differences in (2.1), we see that

$$\Delta \boldsymbol{X}_t = \begin{bmatrix} -\mathbb{I}_{d_1} & \boldsymbol{C}_1 \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \boldsymbol{X}_{t-1} + \boldsymbol{\nu}_t = \boldsymbol{J}\Theta' \boldsymbol{X}_{t-1} + \boldsymbol{\nu}_t \tag{2.3}$$

where $\mathbb{I}_d$ is the identity matrix of order d, $\boldsymbol{J}' = [-\mathbb{I}_{d_1} \vdots \boldsymbol{0}]$, $\Theta' = [\mathbb{I}_{d_1} \vdots -\boldsymbol{C}_1]$, $\boldsymbol{\nu}_t = (\boldsymbol{\nu}_{1t}', \boldsymbol{\nu}_{2t}')'$ is a non-singular transformation of $\boldsymbol{\varepsilon}_t$ defined by

$$\boldsymbol{\nu}_{1t} = \varepsilon_{1t} + \boldsymbol{C}_1 \varepsilon_{2t}, \quad \boldsymbol{\nu}_{2t} = \varepsilon_{2t}, \tag{2.4}$$

$$\boldsymbol{X}_t := \begin{bmatrix} \boldsymbol{X}_{1t} \\ \boldsymbol{X}_{2t} \end{bmatrix}, \quad \boldsymbol{\nu}_t := \begin{bmatrix} \boldsymbol{\nu}_{1t} \\ \boldsymbol{\nu}_{2t} \end{bmatrix}. \tag{2.5}$$

The notation $\boldsymbol{A} = [\boldsymbol{A}_1 \vdots \boldsymbol{A}_2]$ means that the matrix $\boldsymbol{A}$ is partitioned into a matrix $\boldsymbol{A}_1$ consisting of the first d_1 columns and a matrix $\boldsymbol{A}_2$ with d_2 columns.

We suppose that $\boldsymbol{\nu}_t$ (hence also ε_t) has an infinite-order autoregressive representation

$$\sum_{l=0}^{\infty} \boldsymbol{G}_l \nu_{t-l} = \boldsymbol{a}_t \tag{2.6}$$

where $\boldsymbol{G}_0 = \mathbb{I}_d$, $\boldsymbol{a}_t$ is a sequence of independent and identically distributed random vectors such that $\mathbb{E}(\boldsymbol{a}_t) = 0$ and $\mathbb{E}(\boldsymbol{a}_t \boldsymbol{a}_t') = \boldsymbol{\Sigma}_{\boldsymbol{a}}$ is positive definite, and the roots of the equation

$$\det\left\{ \boldsymbol{I}_d - \sum_{l=1}^{\infty} \boldsymbol{G}_l z^l \right\} = 0 \tag{2.7}$$

all lie outside the unit circle $|z| = 1$; $\det\{\boldsymbol{A}\}$ denotes the determinant of the square matrix $\boldsymbol{A}$. We also assume that the following summability condition holds:

$$\sum_{l=1}^{\infty} l^{\bar{\delta}} \| \boldsymbol{G}_l \| < \infty \quad \text{for some } \bar{\delta} \geq 1 \tag{2.8}$$

where $\|\cdot\|$ is the Euclidean matrix norm defined by $\| \boldsymbol{A} \|^2 = \operatorname{tr}(\boldsymbol{A}'\boldsymbol{A})$. This is a standard condition for weakly stationary processes, which ensures that the process is well defined. It also implies that the process $\boldsymbol{\nu}_t$ and, consequently $\boldsymbol{X}_t$, can be approximated by an autoregression of finite order $p_n = p(n)$ where n is the sample size and p_n can grow with n. More explicitly, we assume that p_n satisfies the following condition.

Assumption 2.1. *There is a sequence of positive integers* p_n *such that*

$$n^{-1/3} p_n \to 0 \quad \text{and} \quad \sqrt{p_n} \sum_{l=p_n+1}^{\infty} \| \boldsymbol{G}_l \| \to 0 \quad \text{as } n \to \infty. \tag{2.9}$$

The condition $p_n = o(n^{1/3})$ for the rate of increase of p_n ensures that enough sample information is asymptotically available for estimators to have standard limiting distributions. The condition $\sqrt{p_n} \sum_{j=p_n+1}^{\infty} \| \boldsymbol{G}_j \| \to 0$ imposes a lower bound on the growth rate of p_n, which ensures that the approximation error of the true underlying model by a finite-order autoregression gets small when the sample size increases. A more detailed discussion of these conditions is available in Burnham and Anderson (2002) and Lütkepohl (2005).

Using the equations (2.3)–(2.6) and rearranging terms, we obtain the autoregressive *error correction model* (ECM) representation

$$\Delta \boldsymbol{X}_t = \boldsymbol{\Psi} \Theta' \boldsymbol{X}_{t-1} + \sum_{l=1}^{p_n} \boldsymbol{\Pi}_l \Delta \boldsymbol{X}_{t-l} + \boldsymbol{e}_t(n), \quad t = p_n + 1, p_n + 2, \ldots, \tag{2.10}$$

$$\boldsymbol{e}_t(n) = \boldsymbol{a}_t - \sum_{l=p_n+1}^{\infty} \boldsymbol{G}_l \boldsymbol{\nu}_{t-l}, \quad \boldsymbol{\Psi} = -\sum_{l=0}^{p_n} \boldsymbol{G}_l \boldsymbol{J}, \tag{2.11}$$

where $\boldsymbol{\Psi}$ is a $d \times d_1$ a full-column rank matrix (at least for p_n large enough). Details for this derivation can be found in Saikkonen and Lütkepohl (1994, 1997). Note the coefficient matrices $\boldsymbol{\Pi}_l (l = 1, \ldots, p_n)$ are functions of Θ and $\boldsymbol{G}_l$ $(l = 1, 2, \ldots)$, and they depend on p_n. Furthermore, the sequence $\boldsymbol{\Pi}_l (l = 1, \ldots, p_n)$ is absolutely summable as $p_n \to \infty$.

The autoregressive ECM in (2.10) can be rewritten in a pure VAR form

$$\boldsymbol{X}_t = \sum_{l=1}^{p_n+1} \boldsymbol{\Phi}_l \boldsymbol{X}_{t-l} + \boldsymbol{e}_t(n) \tag{2.12}$$

where $\boldsymbol{\Phi}_1 = \mathbb{I}_d + \boldsymbol{\Psi}\boldsymbol{\Theta}' + \boldsymbol{\Pi}_1$, $\boldsymbol{\Phi}_l = \boldsymbol{\Pi}_l - \boldsymbol{\Pi}_{l-1}$, $l = 2, \ldots, p_n$ and $\boldsymbol{\Phi}_{p_n+1} = -\boldsymbol{\Pi}_{p_n}$. Although the $\boldsymbol{\Pi}_l$ depend on p_n, the same is not true for the $\boldsymbol{\Phi}_l$ except for $\boldsymbol{\Phi}_{p_n+1}$.

Saikkonen and Lütkepohl (1996) derived the asymptotic properties of the multivariate least square (LS) estimators of the VAR coefficients under a standard assumption. Let

$$\boldsymbol{\Phi}(p_n) = [\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_{p_n}] \tag{2.13}$$

be the matrix of the first p_n autoregressive parameter matrices in the representation (2.12), and denote by $\hat{\boldsymbol{\Phi}}(p_n) = [\hat{\boldsymbol{\Phi}}_1, \ldots, \hat{\boldsymbol{\Phi}}_{p_n}]$ the corresponding LS estimator.

The following proposition gives a direct result on the asymptotic properties of the estimator $\hat{\boldsymbol{\Phi}}(p_n)$. It can be proved using the techniques very similar to those used by Saikkonen (1992, part (i) of Theorem 3.2), see also (Saikkonen & Lütkepohl, 1996, Theorem 2).

Proposition 2.1. Asymptotic properties of the autoregressive parameter estimators. *Let* $\{\boldsymbol{X}_t\}$ *be a process which satisfies* (2.3)–(2.6) *with*

$$\mathbb{E}\,|\,a_{it}a_{jt}a_{kt}a_{lt}\,| < \gamma_4 < \infty, \quad 1 \le i,j,k,l \le d. \tag{2.14}$$

where $\boldsymbol{a}_t := (a_{1t},\ldots,a_{dt})'$. *If Assumption* 2.1 *holds, then*

$$\left\|\hat{\boldsymbol{\Phi}}(p_n) - \boldsymbol{\Phi}(p_n)\right\| = O_p(p_n^{1/2} / n^{1/2}). \tag{2.15}$$

This proposition is formulated for the first p_n coefficient matrices, whereas the fitted model is a VAR (p_n+1) where p_n goes to infinity with the sample size n. Dropping the last lag in deriving the consistency of the estimators will not affect the asymptotic distribution of the test statistic (see Lütkepohl, 2005). Details on the estimates of the $\boldsymbol{\Phi}_l$ matrices are given in Saikkonen and Lütkepohl (1996). This result can be viewed as a generalization of Theorem 1 in Lewis and Reinsel (1985) to IVAR processes.

Let us now consider two processes $\boldsymbol{X}^{(h)} = \left\{\boldsymbol{X}_t^{(h)} : t \in \mathbb{Z}\right\}, h = 1,2$, with m_1 and m_2 components, respectively, each of which satisfies an IVAR(∞) model of the form (2.3)–(2.6) with $m_h = d_1^{(h)} + d_2^{(h)}, h = 1,2$, where $d_1^{(h)}$ and $d_2^{(h)}$ replace d_1 and d_2 for $\boldsymbol{X}^{(h)}$. The coefficients of the two processes may differ. We wish to decide whether $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent against an alternative where they are correlated at some lag. Following Pham et al. (2003), the independence between $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ can be tested by testing non-correlation between the corresponding innovation processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$. This leads one to consider the hypothesis:

$$\mathcal{H}_0 : \boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j) = \boldsymbol{0}\,, \text{for all } j \in \mathbb{Z}, \tag{2.16}$$

where

$$\boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j) = \left[\boldsymbol{D}(\boldsymbol{\Sigma}_1)\right]^{-1/2} \boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j)[\boldsymbol{D}(\boldsymbol{\Sigma}_2)]^{-1/2}, \quad \boldsymbol{\Gamma}_{\boldsymbol{a}}^{(hi)}(j) = \mathbf{E}\left[\boldsymbol{a}_t^{(h)}\left(\boldsymbol{a}_{t-j}^{(i)}\right)'\right], j \in \mathbb{Z}, \tag{2.17}$$

$$\boldsymbol{\Sigma}_h = \boldsymbol{\Gamma}_{\boldsymbol{a}}^{(hh)}(0), \boldsymbol{D}(\boldsymbol{\Sigma}_h) = \text{diag}\{\boldsymbol{\Sigma}_h\}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}, \quad h,i = 1,2. \tag{2.18}$$

$\boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j)$ represents the cross-correlation matrix at lag j between the two innovation processes. On setting

$$\boldsymbol{b}_t := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{a}_t = \begin{bmatrix} \boldsymbol{\Sigma}_1^{-1/2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2^{-1/2} \end{bmatrix}\begin{bmatrix} \boldsymbol{a}_{1t} \\ \boldsymbol{a}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{a}_{1t} \\ \boldsymbol{\Sigma}_2^{-1/2}\boldsymbol{a}_{2t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{1t} \\ \boldsymbol{b}_{2t} \end{bmatrix}, \tag{2.19}$$

we see that $\boldsymbol{\rho}_{\boldsymbol{a}}^{(12)}(j)=\boldsymbol{\Gamma}_{\boldsymbol{b}}^{(12)}(j)=\boldsymbol{\rho}_{\boldsymbol{b}}^{(12)}(j)$, for all $j\in\mathbb{Z}$, so that $\mathcal{H}_0$ is equivalent to

$$\boldsymbol{\rho}_{\boldsymbol{b}}^{(12)}(j)=0, \text{ for all } j\in\mathbb{Z}. \tag{2.20}$$

This equivalence plays a central role for proving the required distributional results stated below.

3. TEST STATISTICS AND ASYMPTOTIC NULL DISTRIBUTIONS

Based on a realization $\boldsymbol{X}_1^{(h)},\ldots,\boldsymbol{X}_n^{(h)}$ of length n, for $h=1,2$, a finite-order autoregressive model VAR($p_n^{(h)}+1$) is fitted to each one of these two series. The order $p_n^{(h)}$ depends on the sample size n. The resulting residuals are given by

$$\hat{\boldsymbol{a}}_t^{(h)}=\begin{cases}\boldsymbol{X}_t^{(h)}-\sum_{l=1}^{p_n^{(h)}+1}\hat{\boldsymbol{\Phi}}_l^{(h)}(n)\boldsymbol{X}_{t-l}^{(h)} & \text{if } t=p_n^{(h)}+2,\ldots,n,\\ \mathbf{0} & \text{if } t\le p_n^{(h)}+1,\end{cases} \tag{3.1}$$

where the matrices $\hat{\boldsymbol{\Phi}}_l^{(h)}(n)$ are the OLS estimators of $\boldsymbol{\Phi}_l^{(h)}(n)$, and $h=1,2$. We can also use the conditional maximum likelihood estimator of the error correction form of the model as discussed by Ahn and Reinsel (1990) and Reinsel (1993), or some other estimator with the same rate of convergence. We now consider the residual sample (cross-)covariance matrices

$$\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hi)}(j)=\begin{cases} n^{-1}\sum_{t=j+1}^{n}\hat{\boldsymbol{a}}_t^{(h)}\left(\hat{\boldsymbol{a}}_{t-j}^{(i)}\right)' & \text{if } 0\le j\le n-1\\ n^{-1}\sum_{t=-j+1}^{n}\hat{\boldsymbol{a}}_{t+j}^{(h)}\left(\hat{\boldsymbol{a}}_t^{(i)}\right)' & \text{if } -n+1\le j\le 0\end{cases} \tag{3.2}$$

where $h,\ i=1,2$, and the corresponding cross-correlation matrices

$$\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(hi)}(j)=\left[\boldsymbol{D}\left(\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hh)}(0)\right)\right]^{-1/2}\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hi)}(j)\left[\boldsymbol{D}\left(\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(ii)}(0)\right)\right]^{-1/2} \tag{3.3}$$

where $\boldsymbol{D}\left(C_{\hat{\boldsymbol{a}}}^{(hh)}(0)\right)=\operatorname{diag}\left\{C_{\hat{\boldsymbol{a}}}^{(hh)}(0)\right\}$. The orthogonality tests we consider are based on $C_{\hat{\boldsymbol{a}}}^{(12)}(j)$ and $\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)$. In the sequel, we suppose that $\boldsymbol{X}^{(h)}$ satisfies (2.3) for $h=1,2$. We wish to test the null hypothesis $\mathcal{H}_0$ using the cross-correlation matrices $\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(hi)}(j),\ j\in\mathbb{Z}$.

In the univariate case, Hong (1996b) proposed a portmanteau-type statistic based on the sum of the weighted squared cross-correlations $r_{\hat{a}}^{(12)}(j)$ at all possible lags between the residual series:

$$\mathcal{Q}_n = \frac{n\sum_{j=1-n}^{n-1} k^2(j/M) r_{\hat{a}}^{(12)}(j)^2 - S_n(k)}{\{2D_n(k)\}^{1/2}} \tag{3.4}$$

where $k(\cdot)$ is an arbitrary kernel function (see, e.g., Table 2) and M is a smoothing parameter, while $S_n(k)$ and $D_n(k)$ are normalization coefficients which depend on the kernel $k(\cdot)$:

$$S_n(k) = \sum_{j=1-n}^{n-1}\left(1-\frac{|j|}{n}\right)k^2(j/M),$$

$$D_n(k) = \sum_{j=2-n}^{n-2}\left(1-\frac{|j|}{n}\right)\left(1-\frac{|j|+1}{n}\right)k^4(j/M). \tag{3.5}$$

They correspond to the asymptotic mean and variance of the weighted sum. In multivariate time series, the squared cross-correlation $r_{\hat{a}}^{(12)}(j)^2$ in (3.4) is replaced by a quadratic form in the vector $\boldsymbol{r}_{\hat{\boldsymbol{a}}}^{(12)}(j) = \operatorname{vec}\left[\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)\right]$. For $\mathcal{H}_0$, the test statistic is based on the following sum of weighted quadratic forms at all possible lags:

$$\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) = \sum_{j=1-n}^{n-1} k^2(j/M) Q_{\hat{\boldsymbol{a}}}^{(12)}(j), \tag{3.6}$$

$$Q_{\hat{\boldsymbol{a}}}^{(12)}(j) := n\boldsymbol{r}_{\hat{\boldsymbol{a}}}^{(12)}(j)'\left[\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(22)}(0)^{-1} \otimes \boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(11)}(0)^{-1}\right]\boldsymbol{r}_{\hat{\boldsymbol{a}}}^{(12)}(j), \quad \hat{\boldsymbol{\Sigma}} := \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{\Sigma}}_2 \end{bmatrix}, \tag{3.7}$$

where $\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(hh)}(0)$ is a consistent estimator of the correlation matrix $\boldsymbol{\rho}_{\boldsymbol{a}}^{(h)}$ of the process $\boldsymbol{a}^{(h)}$, and $k(\cdot)$ is a suitable kernel function. The parameter M is a truncation point when the kernel has compact support, or a smoothing parameter when the kernel support is unbounded. We suppose that M is function of n ($M = M_n$) such that $M_n \to \infty$ and $M_n / n \to 0$ as $n \to \infty$. The most commonly used kernels typically give more weight to lower lags and less weight to higher ones. An exception is the truncated uniform kernel $k_T(z) = \mathbb{I}[|z| \leq 1]$, where $\mathbb{I}(A)$ represents the indicator function of the set A, which gives the same weight to all lags. The asymptotic distribution of $Q_{\hat{\boldsymbol{a}}}(j)$ is given in Bouhaddioui and Dufour (2008). In the sequel, we suppose that the kernel function k and the order $p_n^{(h)}$, respectively, satisfy the following assumptions.

Assumption 3.1. *The kernel function* $k:\mathbb{R}\to[-1,1]$ *is a symmetric function, continuous at zero, with at most a finite number of discontinuity points, such that* $k(0)=1$ *and* $\int_{-\infty}^{+\infty} k^2(z)dz<\infty$.

Assumption 3.2. *The orders* $p_n^{(h)}$, $h=1,2$, *satisfy the following conditions*:

$$(i) \quad p_n^{(h)} = o\left(n^{1/2} / M^{1/4}\right), \quad (ii) \quad n\sum_{j=p_n^{(h)}+1}^{\infty} \|\boldsymbol{\Phi}_j^{(h)}\|^2 = o\left(n^{1/2} / M^{1/4}\right). \tag{3.8}$$

Note that the two conditions (i) and (ii) imply that the order $p_n^{(h)}$ satisfies Assumption 2.1. The property $k(0)=1$ implies that the weights assigned to the lower lags are close to unity. The square integrability of $k(\cdot)$ implies that $k(z)\to 0$ as $|z|\to\infty$, so that less weight is given to $\boldsymbol{R}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ as j increases. Note that all the kernels used in spectral analysis satisfy Assumption 3.1 (see, Priestley, 1981, Section 6.2.3). For hypothesis $\mathcal{H}_0$, the test statistic is a standardized version of $\mathcal{T}(\hat{\mathbf{a}},\hat{\boldsymbol{\Sigma}})$:

$$\mathcal{Q}_n = \frac{\mathcal{T}(\hat{\mathbf{a}},\hat{\boldsymbol{\Sigma}}) - m_1 m_2 S_n(k)}{\sqrt{2m_1 m_2 D_n(k)}}, \tag{3.9}$$

where the smoothing parameter $M_n\to\infty$ and $M_n/n\to 0$ when $n\to\infty$.

This test statistic can be viewed as a normalized version of the $\mathcal{L}_2$-norm of a kernel-based estimator of the cross-coherency function between the two innovation series. $S_n(k)$ and $D_n(k)$ represent the asymptotic mean and variance of $\mathcal{T}(\hat{\text{a}},\hat{\boldsymbol{\Sigma}})$ under $\mathcal{H}_0$. If $k(\cdot)$ is the truncated uniform kernel, apart from the standardization factors $S_n(k)$ and $D_n(k)$, $\mathcal{Q}_n$ corresponds to the multivariate version of Haugh's statistic used by Pham et al. (2003) for the finite-order cointegrated case, and by Bouhaddioui and Dufour (2008) for the infinite-order case, namely

$$P_M = \sum_{j=-M}^{M} Q_{\hat{\mathbf{a}}}(j). \tag{3.10}$$

In this case, M is a fixed integer that does not depend on the sample size n. The properties of P_M in the stationary VAR(∞) context and cointegrated IVAR(∞) are studied, respectively, in Bouhaddioui and Roy (2006b) and Bouhaddioui and Dufour (2008). As it will be seen below, many kernels k yield tests that are more powerful than P_M.

In the case of testing independence, under some conditions on the smoothing parameter M and if the kernel k verifies Assumption 3.1, one sees easily that

$$M^{-1}S_n(k)\to S(k)\,,\, M^{-1}D_n(k)\to D(k), \tag{3.11}$$

where

$$S(k)=\int_{-\infty}^{+\infty} k^2(z)dz,\, D(k)=\int_{-\infty}^{+\infty} k^4(z)dz. \tag{3.12}$$

An alternative statistic is obtained by replacing $S_n(k)$ and $D_n(k)$ by their asymptotic approximations $MS(k)$ an $MD(k)$ respectively and is defined by

$$\mathcal{Q}_n^* = \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - Mm_1m_2S(k)}{\sqrt{2Mm_1m_2D(k)}}. \tag{3.13}$$

Both $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ have the same asymptotic null distribution and power properties.

The statistic $\mathcal{Q}_n$ can also be expressed in terms of the autocovariances $\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hh)}(0)$ and the cross-covariances $\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ of the same residual series. Invoking Lemma 4.1 of El Himdi and Roy (1997), the quadratic form $\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}})$ can be written as follows in terms of the residual covariances:

$$\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) = n\sum_{j=1-n}^{n-1} k^2(j/M)\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)'\left[\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(22)}(0)^{-1} \otimes \boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(11)}(0)^{-1}\right]\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) \tag{3.14}$$

with $\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) = \text{vec}[\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j)]$. Let us now consider the "pseudo-statistic"

$$\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) = n\sum_{j=1-n}^{n-1} k^2(j/M)\boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)'\left(\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}\right)\boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j) \tag{3.15}$$

where $\boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)$ is defined as $\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)$ with the residuals $\hat{\boldsymbol{a}}_t^{(1)}$ and $\hat{\boldsymbol{a}}_t^{(2)}$ replaced by the unobservable innovation series $\boldsymbol{a}_t^{(1)}$ and $\boldsymbol{a}_t^{(2)}, t = 1, \ldots, n$, and

$$\mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) = n\sum_{j=1-n}^{n-1} k^2(j/M)\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)'\left(\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}\right)\boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j). \tag{3.16}$$

Thus, with $\hat{\boldsymbol{\Sigma}}_h = \boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hh)}(0)$, $h = 1, 2$, we can write the statistic $\mathcal{Q}_n$ as

$$\begin{aligned} \mathcal{Q}_n &= \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\Sigma}) - m_1m_2S_n(k)}{\sqrt{2m_1m_2D_n(k)}} \\ &= \frac{\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) - m_1m_2S_n(k)}{\sqrt{2m_1m_2D_n(k)}} + \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) - \mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma})}{\sqrt{2m_1m_2D_n(k)}} + \frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma})}{\sqrt{2m_1m_2D_n(k)}}. \end{aligned} \tag{3.17}$$

Since the quantity $\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma})$ depends only on the stationary process $\boldsymbol{a}$, the result of Lemma 3.1 in Bouhaddioui and Roy (2006a) is still valid. We conclude that

$$\frac{\mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) - m_1m_2S_n(k)}{\sqrt{2m_1m_2D_n(k)}} \xrightarrow{L} \mathcal{N}(0,1). \tag{3.18}$$

The asymptotic distribution of $\mathcal{Q}_n$ follows from the next two propositions.

Proposition 3.1. APPROXIMATION OF THE PSEUDO-STATISTIC. *Suppose* $X^{(1)} = \{X_t^{(1)} : t \in \mathbb{Z}\}$ *and* $X^{(2)} = \{X_t^{(2)} : t \in \mathbb{Z}\}$ *satisfy the IVAR*(∞) *models* (2.3)–(2.6) *along with Assumption* 3.1 *and the bounded moment condition*

$$\mathbb{E} \mid a_{it}^{(h)} a_{jt}^{(h)} a_{kt}^{(h)} a_{lt}^{(h)} \mid < \gamma_4 < \infty, 1 \le i, j, k, l \le m_h. \tag{3.19}$$

Let $M = M_n$, *with* $M_n \to \infty$ *and* $M_n / n \to 0$ *as* $n \to \infty$, *and suppose that* $p_n^{(h)}$, $h = 1,2$, *satisfy Assumption* 3.2. *If the processes* $\boldsymbol{a}^{(1)}$ *and* $\boldsymbol{a}^{(2)}$ *are independent, then*

$$\mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) - \mathcal{T}(\boldsymbol{a}, \boldsymbol{\Sigma}) = o_p(M^{1/2}). \tag{3.20}$$

Proposition 3.2. ASYMPTOTIC EQUIVALENCE OF THE TEST STATISTIC *Under the assumptions of Proposition* 3.1, *we have*

$$\frac{\mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma})}{\sqrt{2 m_1 m_2 D_n(k)}} \xrightarrow{p} 0. \tag{3.21}$$

Our main result is a simple consequence of Propositions 3.1–3.2, as follows.

Theorem 3.3. NULL ASYMPTOTIC DISTRIBUTION *Under the assumptions of Proposition 3.1, the statistic* $\mathcal{Q}_n$ *defined by (3.9) has an asymptotic* $\mathcal{N}(0,1)$ *distribution, that is,* $\mathcal{Q}_n \xrightarrow{L} \mathcal{N}(0,1)$.

4. CONSISTENCY OF THE GENERALIZED TESTS

We now investigate the asymptotic power of the test $\mathcal{Q}_n$ under fixed alternatives. We consider a fixed alternative $\mathcal{H}_1$ of serial cross-correlation between the two innovation processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ with the following assumption.

Assumption 4.1. *The two innovation processes*

$$\boldsymbol{a}_t^{(1)} = \left(a_{1,t}^{(1)}, \ldots, a_{m_1,t}^{(1)}\right)' \text{ and } \boldsymbol{a}_t^{(2)} = \left(a_{1,t}^{(2)}, \ldots, a_{m_2,t}^{(2)}\right)', \quad t \in \mathbb{Z}, \tag{4.1}$$

are jointly fourth-order stationary, and their cross-correlation structure is such that $\boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j) \neq \boldsymbol{0}$ *for at least one value of j, with*

$$\sum_{j=-\infty}^{+\infty} \| \boldsymbol{\Gamma}_{\boldsymbol{a}}^{(12)}(j) \|^2 < \infty, \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |\kappa_{\text{uvuv}}(0,i,j,l)| < \infty, \tag{4.2}$$

where $\kappa_{\text{uvuv}}(0,i,j,l)$ *is the fourth cumulant of the joint distribution of* $a_{u,t}^{(1)}, a_{v,t+i}^{(2)}, a_{u,t+j}^{(1)}, a_{v,t+l}^{(2)}$.

The following theorem gives conditions for the consistency of $\mathcal{Q}_n$ under a fixed alternative.

Theorem 4.1. GLOBAL POWER. *Let $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ be two multivariate processes which follow the IVAR(∞) models* (2.3)–(2.6), *and suppose that their innovation processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ satisfy Assumption* 4.1. *If the kernel $k(\cdot)$ satisfies Assumption* 3.1 *and if $p_n^{(h)}$, $h = 1, 2$, satisfy*

$$p_n^{(h)^2} = o\left(\frac{n}{M}\right), \qquad \sum_{j=p_n^{(h)}+1}^{\infty} \left\|\boldsymbol{\Phi}_j^{(h)}\right\|^2 = o\left(M^{-1}\right), \tag{4.3}$$

then, for any sequence of constants $C(n,M)$ such that $C(n,M) = o(n/M^{1/2})$,

$$P[\mathcal{Q}_n > C(n,M)] \to 1. \tag{4.4}$$

This theorem entails that the test based on $\mathcal{Q}_n$ is consistent against the general class of dependence alternatives described by Assumption 4.1. The slower M grows, the faster $\mathcal{Q}_n$ goes to infinity. To investigate the relative efficiency of $\mathcal{Q}_n$, one can use the Bahadur's asymptotic slope criterion defined in Bahadur (1960) (see also Bouhaddioui & Roy, 2006a; Hong, 1996a, 1996b). As in Bouhaddioui and Roy (2006a), we can show that the relative efficiency of the kernel $k_2(\cdot)$ with respect to $k_1(\cdot)$ when $M = n^{\nu}$ is given by

$$\mathrm{ARE}_B(k_2,k_1) = \left\{\frac{D(k_1)}{D(k_2)}\right\}^{1/(2-\nu)}. \tag{4.5}$$

We can then proceed like Bouhaddioui (2002) and Hong (1996a, 1996b) to derive the kernel that maximizes the asymptotic slope over appropriate classes of kernel functions. For example, consider the following class of kernels:

$$\kappa(\tau) = \{k(\cdot): \text{Assumption 3.1 is satisfied, } k^{(2)} = \tau^2/2,\ K(\lambda) \geq 0 \text{ for } \lambda \in (-\infty,+\infty)\} \tag{4.6}$$

where

$$k^{(2)} = \lim_{z\to 0}[1-k(z)]/z^2 \text{ and } K(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} k(z)e^{-iz\lambda}dz. \tag{4.7}$$

This class contains the Daniell, Parzen, and quadratic-spectral kernels (among others). Using Theorem 1 of Ghosh and Huang (1991), we can see that the Daniell kernel (see Table 2) maximizes the asymptotic slope of $\mathcal{Q}_n$ over $\kappa(\tau)$; for a similar argument, see Bouhaddioui (2002). As mentioned in Bouhaddioui and Roy (2006a), a test with a greater asymptotic slope may be expected to have a greater power for a fixed alternative than one with a smaller asymptotic slope. However, there is no clear analytical relationship between the slope of a test and its power function. For a specific alternative, we cannot conclude that a test with greater asymptotic slope should be automatically preferred to one with a smaller asymptotic slope without further analysis of the finite-sample properties of the two test statistics.

5. LOCAL POWER ANALYSIS

In this section, we study the power of the test proposed above against a class of local alternatives of the form

$$\mathcal{H}_{a}(\boldsymbol{\Lambda}_{b}^{(12)}) : \boldsymbol{\Gamma}_{b}^{(12)}(j) = \frac{M^{1/4}}{n^{1/2}} \boldsymbol{\Lambda}_{b}^{(12)}(j), \text{ for all } j \in \mathbb{Z},$$

where $\boldsymbol{\Lambda}_{b}^{(12)} = \{\boldsymbol{\Lambda}_{b}^{(12)}(j)\}_{j \in \mathbb{Z}}$ is a sequence of $m_1 \times m_2$ cross-correlation matrices such that only finite elements of $\boldsymbol{\Lambda}_{b}^{(12)}$ are non-zero elements. Let

$$\lambda_{b}^{(12)}(j) = \text{vec}[\boldsymbol{\Lambda}_{b}^{(12)}(j)], \tag{5.1}$$

$$\beta(\boldsymbol{\Lambda}_{b}^{(12)}) = \sum_{j=-\infty}^{\infty} \lambda_{b}^{(12)}(j)' \lambda_{b}^{(12)}(j). \tag{5.2}$$

The following theorem establishes the asymptotic distribution of $\mathcal{Q}_n$ under the local alternative $\mathcal{H}_{a}(\boldsymbol{\Lambda}_{b}^{(12)})$.

Theorem 5.1. LOCAL POWER. *Let $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ be two multivariate processes which follow the IVAR*(∞) *models* (2.3)–(2.6), *and suppose that their innovation processes $\boldsymbol{a}^{(1)}$ and $\boldsymbol{a}^{(2)}$ satisfy Assumption 4.1. If the kernel $k(\cdot)$ satisfies Assumption 3.1 and if $p_n^{(h)}$, $h = 1, 2$, satisfy*

$$p_n^{(h)^2} = o(n / M), \qquad \sum_{j=p_n^{(h)}+1}^{\infty} \| \boldsymbol{\Phi}_j^{(h)} \|^2 = o\left(M^{-1}\right), \tag{5.3}$$

then, under $\mathcal{H}_{a}(\boldsymbol{\Lambda}_{b}^{(12)})$,

$$\mathcal{Q}_n \xrightarrow{L} \mathcal{N}\left[\beta\left(\boldsymbol{\Lambda}_{b}^{(12)}\right) / \sqrt{2 m_1 m_2 D(k)}, 1\right]. \tag{5.4}$$

where $\beta\left(\boldsymbol{\Lambda}_{b}^{(12)}\right)$ is defined in (5.2).

Theorem 5.1 shows that the test $\mathcal{Q}_n$ has non-trivial power against a class of local alternatives converging to $\mathcal{H}_0$ at the rate of $\frac{M^{1/4}}{n^{1/2}}$. The power depends on the kernel function k through $D(k)$. Similarly, we note that increasing slowly the parameter M, the divergence of the test statistic to infinity is faster and, consequently, the test is more powerful.

6. SIMULATION STUDY

In the previous sections, we have studied the asymptotic distribution of the test statistics. Here we investigate the finite-sample properties of the proposed test statistics, in particular their exact level and power. To do this, we performed a small Monte Carlo study. In addition to the test statistics discussed in the preceding sections, we also consider the non-stationary multivariate version of the Haugh statistic (previously studied by Pham et al., 2003):

$$P_M^* = \sum_{j=-M}^{M} \frac{n}{n-|j|} Q_{\hat{a}}(j) \tag{6.1}$$

where $Q_{\hat{a}}^{(12)}(j)$ is given by (3.7). P_M^* is a slightly modified version of P_M defined by (3.10).

6.1. Description of the Experiment

In the simulation experiment, we considered bivariate series $\{\boldsymbol{X}_t^{(1)}\}$ and $\{\boldsymbol{X}_t^{(2)}\}$ generated from (joint) four-dimensional **VAR**(2), **VARMA**(1,1), and **VAR**$_\delta$(1) models (see Table 1). In the first two models, the two subprocesses $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ are independent bivariate **VAR**(2) or **VARMA**(1,1) and served for the level study and the corresponding submodels are partially non-stationary and invertible.

Table 1. Time-series Models Used in the Simulation Study.

Models	Equations
VAR(2)	$\begin{bmatrix} \boldsymbol{X}_t^{(1)} \\ \boldsymbol{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi_1^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \Phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1}^{(1)} \\ \boldsymbol{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \Phi_2^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \Phi_2^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-2}^{(1)} \\ \boldsymbol{X}_{t-2}^{(2)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_t^{(1)} \\ \boldsymbol{a}_t^{(2)} \end{bmatrix}$
VARMA(1,1)	$\begin{bmatrix} \boldsymbol{X}_t^{(1)} \\ \boldsymbol{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi_1^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \Phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1}^{(1)} \\ \boldsymbol{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \Psi^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \Psi^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{t-1}^{(1)} \\ \boldsymbol{a}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_t^{(1)} \\ \boldsymbol{a}_t^{(2)} \end{bmatrix}$
VAR$_\delta$(1)	$\begin{bmatrix} \boldsymbol{X}_t^{(1)} \\ \boldsymbol{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \Phi_1^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \Phi_1^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{t-1}^{(1)} \\ \boldsymbol{X}_{t-1}^{(2)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a}_t^{(1)} \\ \boldsymbol{a}_t^{(2)} \end{bmatrix}$

Noise covariance matrices

$$\boldsymbol{\Sigma}_{\boldsymbol{a}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\boldsymbol{a}}^{(2)} \end{bmatrix} \qquad \boldsymbol{\Sigma}_{\boldsymbol{a},\delta} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{a}}^{(1)} & \boldsymbol{\Sigma}_{\boldsymbol{a},\delta}^{(12)} \\ \boldsymbol{\Sigma}_{\boldsymbol{a},\delta}^{(21)} & \boldsymbol{\Sigma}_{\boldsymbol{a}}^{(2)} \end{bmatrix}$$

Parameters values

$$\Phi_1^{(1)} = \begin{bmatrix} 0.4 & 0.0 \\ -1.0 & 1.0 \end{bmatrix} \quad \Phi_1^{(2)} = \begin{bmatrix} 1.0 & 0.0 \\ -0.8 & 0.5 \end{bmatrix} \quad \Phi_2^{(1)} = \begin{bmatrix} 0.6 & -0.5 \\ 0.3 & 0.4 \end{bmatrix} \quad \Phi_2^{(2)} = \begin{bmatrix} -0.5 & -0.8 \\ -0.4 & 0.2 \end{bmatrix}$$

$$\Psi^{(1)} = \begin{bmatrix} -0.2 & 0.3 \\ -0.6 & 1.1 \end{bmatrix} \quad \Psi^{(2)} = \begin{bmatrix} 0.8 & 0.3 \\ 0.1 & 0.6 \end{bmatrix} \quad \Sigma_a^{(1)} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \quad \Sigma_a^{(2)} = \begin{bmatrix} 1.0 & 0.75 \\ 0.75 & 1.0 \end{bmatrix}$$

$$\Sigma_{a,\delta}^{(12)} = \begin{bmatrix} 0.1\delta & 0 \\ 0 & 0.05\delta \end{bmatrix}$$

Table 2. Kernels Used With the Test Statistics Q_n and Q_n^*.

Truncated Uniform (TR):	$k(z)=\begin{cases} 1, & \lvert z\rvert\leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Bartlett (BAR):	$k(z)=\begin{cases} 1-\lvert z\rvert, & \lvert z\rvert\leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Daniell (DAN):	$k(z)=\dfrac{\sin(\pi z)}{\pi z}, z\in\mathbb{R}$.
Parzen (PAR):	$k(z)=\begin{cases} 1-6z^2+6\lvert z\rvert^3, & \text{if } \lvert z\rvert\leq 0.5, \\ 2(1-\lvert z\rvert)^3, & \text{if } 0.5\leq\lvert z\rvert\leq 1, \\ 0, & \text{otherwise.} \end{cases}$
Bartlett-Priestley (BP):	$k(z)=\dfrac{3}{(\pi z)^2}\left\{\dfrac{\sin(\pi z)}{\pi z}-\cos(\pi z)\right\}, z\in\mathbb{R}$.

The third one, in which there is instantaneous correlation between the two innovation series, was used for the power study. The correlation depends on a parameter δ and the values $\delta = 1.0, 1.5$, and 2 were chosen. For each model, two series lengths ($n = 100, 200$) were considered. With the statistics Q_n and Q_n^* defined by (3.9) and (3.13), we used the four kernels described in Table 2. For each kernel, the following three truncation values M were employed: $M = [\ln(n)], [3n^{0.2}]$, and $[3n^{0.3}]$ ($[a]$ denotes the integer part of a). These rates are discussed in Hong (1996a, p. 849). They lead to $M = 5, 8, 12$, respectively, for the series length $n = 100$, and to $M = 5,\ 9, 15$ for $n = 200$. The same truncation values were used for P_M^*.

In the level study, 5,000 independent realizations were generated from both models **VAR**(2) and **VARMA**(1,1) for each series length n. Computations were made in the following way.

1. First, pseudo-random variables from the $\mathcal{N}(0,1)$ distribution were obtained with the pseudo-random normal generator of the S-plus package and were transformed into independent $\mathcal{N}[\mathbf{0}, \mathbf{\Sigma}_{\mathbf{a}}]$. pseudo-random vectors using the Cholesky decomposition. Second, the $\boldsymbol{X}_t$ values were obtained by directly solving the model difference equation.
2. For the **VAR**(2) model, the LS estimates of the coefficients of the true models were obtained using the procedure described in Reinsel (1993). The autoregressive order was obtained by minimizing the AIC criterion for $p \leq P$, where P is set to $n^{1/3}$. We chose Akaike information criterion (AIC) criterion which seems behave better than other criteria such HQ or SC specifically in LR cointegration tests (see Lütkepohl & Saikkonen, 1999). For the **VARMA**(1,1), each subseries was approximated by a possible high-order **VAR** model. From Pham et al. (2003), the value of the **VAR** order was obtained by minimizing the Hannan–Quinn criterion using conditional LS estimation. The residual cross-correlation matrix $\boldsymbol{R}_{\hat{a}}^{(12)}(j)$'s as defined by (3.3) is then computed.
3. For each realization, the test statistics Q_n and Q_n^* were compared for each of the four kernels and the three values of M. The same values of M were used for

the statistic P_M^*. The values of the statistics $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ were compared with the $\mathcal{N}(0,1)$ critical values and those of P_M^* to the $\chi^2_{4(2M+1)}$ critical values.

4. Finally, for each model, each series length and nominal level, the empirical frequencies of rejection of the null hypothesis of non-correlation were obtained from the 5,000 realizations. The results in percentage are reported in Table 3. The standard error of the empirical level is 0.14% for the nominal level 1%, 0.31% for 5%, and 0.42% for 10%.

Table 3. Empirical Level (%) of the $\mathcal{Q}_n$, $\mathcal{Q}_n^*$, and P_M^* Tests, With Different Kernels and Truncation Values.

Gaussian **VAR**(2) and **VARMA**(1, 1) Models and the Number of Realizations: 5,000

				$\mathcal{Q}_n$					$\mathcal{Q}_n^*$					
	n	M	α%	DAN	PAR	BAR	BP	TR	DAN	PAR	BAR	BP	TR	P_M^*
VAR(2)	100	5	1	0.7	0.6	0.8	0.7	0.6	1.2	0.9	0.7	0.7	1.3	0.7
			5	5.8	3.9	5.7	5.2	4.4	5.9	4.3	5.8	6.1	3.7	4.2
			10	9.6	8.0	9.5	10.6	8.3	10.3	8.8	9.4	10.7	9.0	8.8
		8	1	1.3	0.6	0.9	1.2	0.7	1.4	1.2	1.0	1.5	0.6	0.8
			5	5.6	4.1	5.9	5.6	4.0	5.4	4.0	5.2	4.8	4.0	4.3
			10	10.7	9.2	10.8	10.7	7.4	10.6	9.6	11.0	10.4	8.2	8.4
		12	1	0.8	0.7	0.8	1.2	0.6	1.3	0.8	1.4	1.5	0.7	0.8
			5	5.4	4.8	5.3	5.4	4.2	5.6	4.5	4.9	5.7	4.2	4.5
			10	10.4	8.7	11.2	10.8	7.8	10.8	10.4	11.2	10.5	8.1	8.4
	200	5	1	0.8	1.2	0.8	1.2	0.8	0.7	0.8	1.2	1.3	0.7	0.9
			5	5.7	5.2	5.8	5.5	4.1	5.5	4.2	5.9	5.7	4.4	4.2
			10	9.1	9.2	10.4	10.6	8.3	8.4	10.2	10.6	10.2	8.7	8.9
		9	1	1.2	1.1	0.9	0.8	0.7	1.4	0.9	0.8	1.2	0.7	0.7
			5	6.1	4.3	5.5	5.7	4.4	6.3	4.6	5.5	5.9	4.5	4.1
			10	10.9	9.5	10.5	11.0	7.6	11.2	9.3	10.6	10.7	8.6	9.2
		15	1	1.4	0.8	1.2	1.4	1.2	0.9	1.2	1.4	0.8	0.6	0.6
			5	6.0	4.5	6.2	5.4	4.1	5.8	4.7	5.8	5.6	4.3	4.5
			10	10.6	10.3	11.2	10.6	7.9	11.0	10.5	10.8	10.4	8.2	8.9
VARMA(1,1)	100	5	1	1.3	1.1	0.7	0.8	0.7	1.2	0.7	1.4	1.2	0.6	0.8
			5	5.7	4.7	6.2	4.5	4.3	5.8	4.4	5.8	4.6	3.9	4.3
			10	9.6	8.6	9.3	10.4	8.3	9.6	9.0	9.5	10.8	8.2	8.4
		8	1	1.4	0.7	0.8	1.2	0.7	1.3	0.8	1.2	0.9	0.8	1.3
			5	5.6	4.4	5.9	5.6	3.9	5.4	4.1	5.5	5.5	4.3	5.6
			10	10.6	8.5	11.3	10.6	7.3	9.4	9.0	11.0	10.7	8.0	9.4
		12	1	0.9	1.2	0.7	0.8	0.6	1.1	0.9	0.9	1.3	0.7	1.4
			5	5.4	5.1	6.0	5.6	4.2	5.6	5.4	5.8	5.6	4.1	4.5
			10	9.4	8.8	10.4	10.2	7.9	9.1	8.2	9.1	10.6	7.5	8.3
	200	5	1	0.8	1.3	0.7	0.9	0.7	1.2	0.8	1.2	1.2	0.7	1.3
			5	5.6	4.7	5.4	5.9	4.0	6.2	4.8	5.7	6.3	4.6	5.9
			10	9.0	9.3	10.6	11.0	8.9	10.5	9.2	10.5	9.6	8.2	8.9
		9	1	1.3	0.7	1.2	1.1	0.8	0.9	0.8	1.3	0.8	0.8	0.9
			5	6.1	5.2	4.2	6.1	4.3	5.7	5.1	5.5	6.3	4.3	5.6
			10	9.4	10.5	11.0	10.7	8.4	10.7	9.5	10.8	10.3	8.7	8.9
		15	1	1.4	1.1	0.8	0.9	0.7	1.3	0.9	0.9	0.8	0.7	0.8
			5	6.2	4.6	5.2	6.0	4.3	5.3	5.1	5.3	6.0	4.6	5.5
			10	10.3	10.5	10.8	10.6	7.9	10.7	10.2	11.2	10.7	8.4	9.1

Computations for the power analysis were made in a similar way using the $\mathbf{VAR}_\delta(1)$ model with different values of δ.

6.2. Level

6.2.1. Gaussian Innovations

Results from the level study are presented in Table 3. We make the following observations. The asymptotic $\mathcal{N}(0,1)$ distribution provides a good approximation of the exact distributions of $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ at all nominal levels considered, kernels and truncation values. Almost all empirical levels are within three standard errors of the corresponding nominal levels and the majority are within two standard errors. The statistic $\mathcal{Q}_n^*$ is slightly better approximated than $\mathcal{Q}_n$ since most of its empirical levels are within two standard errors of the nominal level.

These results are similar to those obtained for orthogonality tests between stationary series (see Bouhaddioui & Roy, 2006a). At the 1% and 10% nominal levels, both statistics have a small tendency to under or over-reject. There is no significant difference between the kernels. The best approximations are obtained with the Bartlett and BP kernels, while the performance of the Parzen kernel is inferior. With the Bartlett kernel, the empirical size is always within two standard errors of the nominal size. For the truncated uniform kernel, the size of $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ are very close to the size of P_M^*, which is normal since $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ are linear transformations of P_M and P_M^* is a finite-sample version of P_M. For the models considered, the values of the truncation parameter M has no significant effect on the size of the tests. Finally, when the series length n goes from 100 to 200, the approximation improves very slightly.

6.2.2. Non-Gaussian Innovations

We now examine simulation results where innovations follow a multivariate contaminated normal distribution. We consider the distribution

$$p\mathcal{N}_m[\boldsymbol{0},\boldsymbol{\Gamma}]+(1-p)\mathcal{N}_m[\boldsymbol{0},\boldsymbol{\Lambda}]$$

to denote the m-dimensional contaminated normal distribution in which the $\mathcal{N}_m(0,\boldsymbol{\Gamma})$ distribution is contaminated with probability $1-p$, by the $\mathcal{N}_m[0,\boldsymbol{\Lambda}]$ distribution. We can verify that the fourth-order cumulants of this distribution depend on p, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Lambda}$. Thus, we consider in this part of the simulation two innovations series $\boldsymbol{a}_t^{(1)}$ and $\boldsymbol{a}_t^{(2)}$ generated independently according to the following two distributions:

$$p_1\mathcal{N}_{m_1}[\boldsymbol{0},\mathbb{I}_{m_1}]+(1-p_1)\mathcal{N}_{m_1}[\boldsymbol{0},\Omega_{\boldsymbol{a}}^{(1)}],\ p_2\mathcal{N}_{m_2}[\boldsymbol{0},\mathbb{I}_{m_2}]+(1-p_2)\mathcal{N}_{m_2}[\boldsymbol{0},\Omega_{\boldsymbol{a}}^{(2)}],$$
$$\Omega_{\boldsymbol{a}}^{(1)}=\begin{bmatrix}25 & 5\\ 5 & 4\end{bmatrix} \text{ and } \Omega_{\boldsymbol{a}}^{(2)}=\begin{bmatrix}25 & 7.5\\ 7.5 & 4\end{bmatrix}. \tag{6.2}$$

Simulations were made for different values of the pair (p_1, p_2) and for two models of Table 1, where $\Sigma_a^{(1)}$ and $\Sigma_a^{(2)}$ are now the covariance matrices of the two contaminated normal distributions in (6.2). The results in Table 4 are obtained by using $(p_1, p_2) = (0.7, 0.9)$; the results for the other values of (p_1, p_2) are similar. From Table 4, we see that the non-normality of the innovations does not significantly affect the behavior of the test statistic Q_n with the associate kernel function and truncation parameter for the two sizes $n = 100$ and $n = 200$.

Table 4. Empirical Level (%) of the Test Q_n, Q_n^*, and P_M^* With Different Kernels and Truncation Values.

VAR(2) and VARMA(1, 1) Models With Non-Gaussian Innovations and the Number of Realizations: 5,000

				Q_n					Q_n^*					
	n	M	$\alpha\%$	DAN	PAR	BAR	BP	TR	DAN	PAR	BAR	BP	TR	P_M^*
VAR(2)	100	5	1	1.3	0.7	1.2	1.3	0.6	0.8	1.3	0.9	0.8	1.4	1.3
			5	5.4	4.6	5.8	5.3	4.1	5.5	4.4	5.9	5.8	4.0	4.2
			10	9.8	8.4	10.5	10.7	8.2	10.5	9.0	9.1	9.3	8.5	8.9
		8	1	0.7	1.2	0.8	1.3	0.7	1.2	0.8	0.8	1.3	0.7	0.8
			5	6.0	5.4	4.6	5.8	3.8	5.7	4.2	5.6	4.4	4.0	4.2
			10	11.0	9.4	10.6	9.5	8.2	10.8	9.4	10.8	10.6	8.4	8.8
		12	1	1.2	0.9	0.7	1.3	0.7	1.4	1.2	0.8	1.3	0.6	0.8
			5	5.8	5.6	5.2	5.6	4.0	4.6	4.8	5.3	5.4	3.8	4.2
			10	11.3	10.9	11.0	10.6	8.4	10.6	9.8	10.8	9.5	8.3	8.8
	200	5	1	1.2	0.9	0.8	1.3	0.7	0.8	1.3	1.1	0.8	0.8	1.2
			5	6.0	5.8	5.4	5.6	3.9	6.1	5.9	5.5	5.3	4.0	4.4
			10	10.6	9.0	10.2	10.4	8.4	9.4	10.8	11.0	10.6	8.4	9.2
		9	1	0.7	0.9	0.7	0.8	0.8	1.3	0.7	0.7	1.1	0.8	0.8
			5	5.8	5.6	5.2	4.7	4.2	6.0	4.8	5.8	5.8	4.2	4.6
			10	11.2	9.3	9.6	10.6	8.8	11.4	9.7	10.3	10.9	8.6	9.4
		15	1	1.3	1.1	0.8	0.7	0.7	1.1	1.3	0.9	0.8	0.6	0.7
			5	5.6	5.8	6.0	5.6	4.2	5.6	4.4	6.0	6.2	4.1	4.6
			10	11.2	10.6	10.2	10.8	8.6	11.0	10.8	10.3	10.2	8.6	9.0
VARMA(1,1)	100	5	1	0.8	1.2	1.3	0.7	0.6	1.1	0.8	1.2	1.2	0.6	0.7
			5	5.9	6.1	5.6	4.4	4.0	5.7	5.9	4.8	4.8	4.0	4.4
			10	10.6	9.2	9.6	11.0	8.5	10.9	10.4	9.2	11.0	8.0	9.0
		8	1	1.4	1.2	1.2	0.8	0.7	1.2	1.4	1.3	0.8	0.7	1.4
			5	6.0	4.2	5.6	5.8	3.8	6.2	4.0	6.1	6.3	4.2	6.0
			10	11.6	9.6	10.4	10.8	8.0	11.2	9.4	11.2	10.6	8.0	9.6
		12	1	0.8	1.3	0.8	0.9	0.7	1.2	1.1	0.9	1.1	0.8	1.3
			5	5.8	5.3	5.8	6.0	4.4	6.0	5.2	5.4	5.8	4.0	5.8
			10	10.8	9.2	11.4	10.6	8.1	11.2	9.4	9.3	11.0	8.4	8.8
	200	5	1	1.1	1.2	0.9	1.3	0.7	1.2	1.3	1.1	0.8	0.8	1.2
			5	6.1	5.4	4.8	6.1	4.2	5.9	4.7	5.4	6.0	4.4	5.8
			10	10.6	10.3	11.3	11.5	8.4	11.3	10.4	11.0	10.8	8.4	9.2
		9	1	1.3	1.2	0.9	1.2	0.8	1.2	1.3	1.1	0.9	0.7	1.3
			5	5.9	5.9	4.6	5.4	4.1	5.7	6.1	5.2	5.8	4.4	5.8
			10	11.4	10.8	10.6	10.6	8.8	11.2	10.8	10.4	9.8	8.6	9.3
		15	1	0.9	1.3	0.8	1.2	0.8	1.3	1.2	1.3	1.1	0.7	1.3
			5	5.4	5.8	6.2	5.6	4.0	5.5	5.6	5.8	5.4	4.2	5.8
			10	11.0	10.8	9.8	10.2	8.2	10.6	10.6	10.2	10.4	8.6	9.3

Table 5. Power of the Tests $\mathcal{Q}_n$, $\mathcal{Q}_n^*$, and P_M^*. Based on Asymptotic Critical Values With Different Kernels and Different Truncation Values.

VAR$_\delta$(1) Model With $\delta = 2$

			$\mathcal{Q}_n^*$					
n	M	α%	DAN	PAR	BAR	BP	TR	P_M^*
		1	57.3	53.5	54.6	52.6	35.3	24.6
	5	5	63.2	60.1	56.4	58.6	36.8	26.8
		10	72.6	70.8	62.5	64.3	38.2	27.5
		1	49.6	46.1	51.4	48.0	27.5	22.6
100	8	5	58.4	53.2	55.8	51.6	31.2	23.8
		10	63.7	60.8	62.6	61.7	34.6	25.8
		1	43.6	38.5	41.8	42.6	23.3	18.9
	12	5	50.2	44.7	40.3	43.0	26.4	21.2
		10	56.8	50.6	48.8	46.5	28.8	23.7
		1	78.4	74.5	74.8	76.2	54.8	50.6
	5	5	85.6	82.6	81.6	85.8	56.4	54.1
		10	93.4	89.5	87.5	90.2	60.4	56.8
		1	69.5	65.2	63.0	66.8	42.4	40.7
200	9	5	75.6	76.6	72.4	78.2	46.2	44.6
		10	80.8	78.5	77.6	82.8	50.4	46.4
		1	56.8	52.4	54.8	56.1	36.8	32.8
	15	5	60.1	57.4	53.2	60.1	40.2	35.0
		10	64.8	54.4	54.2	62.6	44.8	40.4

6.3. Power

Results on power are presented in Table 5. In **VAR**$_\delta$(1), the cross-correlation at lag 0 between the two innovation series increases with δ and, as expected, the powers of the three tests increase with δ. Since the relative behaviors of the various tests are similar for the three values of δ considered ($\delta = 1, 1.5, 2$), we only present the results for $\delta = 2$. Similarly, we only present results for $\mathcal{Q}_n^*$, since $\mathcal{Q}_n$ and $\mathcal{Q}_n^*$ have exhibit similar behaviors with respect to kernels and truncation values.

From Table 5, we draw the following observations. First, power decreases as M increases. Indeed, the model considered here is characterized by the lag 0 serial correlation. In such a situation, we expect that the tests assigning more weight to small lags will be more powerful than those assigning weights to a large number of lags. For the three significant levels and the three truncation values, the Daniell kernel provides the most powerful test, while the Parzen, Bartlett, and BP kernels yield similar powers for $\mathcal{Q}_n^*$. However, the power of $\mathcal{Q}_n^*$ with the truncated uniform kernel is much smaller and is comparable to the power of P_M^*. For the chosen model, the new tests $\mathcal{Q}_n$ or $\mathcal{Q}_n^*$ with kernels other than the truncated uniform are preferable to the non-stationary multivariate version of Haugh's test P_M^*. Finally, the powers of all tests increase when the sample size varies from 100 to 200.

7. CONCLUSION

In this chapter, we have proposed a semiparametric approach to test the non-correlation (or independence in the Gaussian case) between infinite-order cointegrated series IVAR(∞). The approach is semiparametric in the sense that if the two series are VARMA, we do not need to separately estimate the *true* model for each of the series. Instead, we fit a vector autoregression to each series, and the test statistics are based on residual cross-correlations at all possible lags. The weights assigned to the lags are defined by a kernel function and a smoothing parameter. Under the hypothesis of independence or non-causality of the two series, the asymptotic normality of the tests statistics are established. The finite-sample properties of the test were investigated by a Monte Carlo experiment which shows that the level is reasonably well controlled for both series lengths 100 and 200. Furthermore, with the model considered, the four kernels DAN, PAR, BAR, and BP lead to similar powers and are more powerful than the truncated uniform kernel which corresponds to the multivariate version of the portmanteau test proposed by Bouhaddioui and Dufour (2008).

ACKNOWLEDGMENTS

The authors thank Yoosoon Chang, Kilani Ghoudi, Ian McLeod, Abderrahim Taamouti, and an anonymous referee for several useful comments. Earlier versions of this chapter were presented at the Statistical Society of Canada annual meeting and the University of Western Ontario (Department of Statistical and Actuarial Sciences). This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

REFERENCES

Ahn, S. K., & Reinsel, G. C. (1990). Estimation of partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, *85*, 813–823.

Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting. *Annals of the Institute of Statistical Mathematics*, *21*, 407–419.

Bahadur, R. R. (1960). Stochastic comparison of tests. *Annals of Mathematical Statistics*, *31*, 276–295.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Annals of Statistics*, *2*, 489–502.

Bhansali, R. J. (1996). Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics*, *48*, 577–602.

Bouhaddioui, C. (2002). *Tests d'indépendance de deux séries multivariées autorégressives d'ordre infini.* Ph.D. thesis. Département de Mathématiques et de Statistique, Université de Montréal.

Bouhaddioui, C., & Dufour, J.-M. (2008). Tests for non-correlation of two infinite-order cointegrated vector autoregressive series. *Journal of Applied Probability and Statistics*, *3*(1), 78–94.

Bouhaddioui, C., & Roy, R. (2006a). A generalized portmanteau test for independence of two infinite order vector autoregressive series. *Journal of Time Series Analysis*, *27*(4), 505–544.

Bouhaddioui, C., & Roy, R. (2006b). On the distribution of residual cross-correlations between two uncorrelated infinite order vector autoregressive series. *Statistics and Probability Letters*, *76*(1), 58–68.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag.

Duchesne, P. (2005). Testing for serial correlation of unknown form in cointegrated time series models? *Annals of the Institute of Statistical Mathematics*, *57*, 575–595.

Duchesne, P., & Roy, R. (2003). Robust tests for independence of two time series. *Statistica Sinica*, *13*, 827–852.

El Himdi, K., & Roy, R. (1997). Tests for noncorrelation of two multivariate ARMA time series. *Canadian Journal of Statistics*, *25*, 233–256.

Engle, R., & Granger, C. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251–276.

Ghosh, B. K., & Huang, W. M. (1991). The power and optimal kernel of the Bickel–Rosenblatt test for goodness of fit. *The Annals of Statistics*, *17*, 999–1009.

Hallin, M., & Saidi, A. (2005). Testing independence and causality between multivariate ARMA times series. *Journal of Time Series Analysis*, *26*, 83–106.

Hallin, M., & Saidi, A. (2007). Optimal tests for non-correlation between multivariate time series. *Journal of the American Statistical Association*, *102*(479), 938–951.

Hallin, M., Jurečkovà, J., Picek, J., & Zahaf, T. (1999). Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores. *Journal of Statistical Planning and Inference*, *75*, 319–330.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: A univariate residual cross-correlation approach. *Journal of the American Statistical Association*, *71*, 378–385.

Hong, Y. (1996a). Consistent testing for serial correlation of unknown form. *Econometrica*, *64*, 837–864.

Hong, Y. (1996b). Testing for independence between two covariance stationary time series. *Biometrika*, *83*(3), 615–625.

Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *Journal of the Royal Statistical Society, Series B*, *60*(2), 271–496.

Hong, Y. (1999). Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach. *Journal of the American Statistical Association*, *94*(448), 1201–1220.

Lewis, R., & Reinsel, G. C. (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis*, *16*, 393–411.

Lütkepohl, H. (1985). The joint asymptotic distribution of multistep prediction errors of estimated vector autoregressions. *Economics Letters*, *17*, 103–106.

Lütkepohl, H. (2001). Vector autoregressions. In B. Baltagi (Ed.), *'Companion to theoretical econometrics', Blackwell companions to contemporary economics* (chapter 32, pp. 678–699). Basil Blackwell.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag.

Lütkepohl, H., & Saikkonen, P. (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics*, *81*, 127–157.

Lütkepohl, H., & Saikkonen, P. (1999). Order selection in testing for the cointegrating rank of a VAR process. In *Cointegration, causality, and forecasting. A Festschrift in Honour of Clive W.J. Granger, Engle, R.F., White, H*. (pp. 168–199). Oxford University Press.

Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition. *Advances in Econometrics*, *8*, 107–133.

Park, J. Y., Shin, K., & Wang, Y.-J. (2010). A semiparametric cointegrating regression: Investigating the effects of age distributions on consumption and saving. *Journal of Econometrics*, *157*, 165–178.

Parzen, E. (1974). Some recent advances in time series modeling. *IEEE Transactions on Automatic Control*, *19*, 723–730.

Pham, D. T., Roy, R., Cédras, L. (2003). Tests for non-correlation of two cointegrated ARMA time series. *Journal of Time Series Analysis*, *24*, 553–577.

Priestley, M. B. (1981). *Spectral analysis and time series*, Vol. *1*. Academic Press.

Reinsel, G. C. (1993). *Elements of multivariate time series analysis*. Springer-Verlag.

Reinsel, G. C. (1997). *Elements of multivariate time series analysis* (2nd ed.), Springer-Verlag.

Saidi, A. (2007). Consistent testing for non-correlation of two cointegrated ARMA time series. *Canadian Journal of Statistics*, *35*(1), 169–188.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory*, *8*, 1–27.

Saikkonen, P., & Lütkepohl, H. (1994). Infinite order cointegrated autoregressive processes: Estimation and inference. Technical Report 5. Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.

Saikkonen, P., & Lütkepohl, H. (1996). Infinite-order cointegrated vector autoregressive processes: Estimation and inference. *Econometric Theory*, *12*, 814–844.

Saikkonen, P., & Luukkonen, R. (1997). Testing cointegration in infinite order vector autoregressive proceses. *Journal of Econometrics*, *81*, 93–126.

Shao, X. (2009). A generalized portmanteau test for independence between two stationary time series. *Econometric Theory*, *25*, 195–210.

Yap, S. F., & Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *Journal of the American Statistical Analysis*, *90*, 253–267.

APPENDIX: PROOFS

The following notations are adopted. The Euclidean scalar product of $\mathbf{x}_t$ and $\mathbf{x}_s$ is defined by $\langle \mathbf{x}_t, \mathbf{x}_s \rangle = \mathbf{x}_t' \mathbf{x}_s$ and the Euclidean norm of $\mathbf{x}_t$ by $\|\mathbf{x}_t\| = \sqrt{\langle \mathbf{x}_t, \mathbf{x}_t \rangle}$. The scalar Δ denotes a generic positive bounded constant which may differ from place to place.

Proof of Proposition 3.1. First, let

$$\boldsymbol{\Xi} := [\boldsymbol{\Xi}_1 \vdots \cdots \vdots \boldsymbol{\Xi}_p \vdots \boldsymbol{\Xi}_{p+1,1}] = [\boldsymbol{\Psi} \vdots \boldsymbol{\Pi}_1 \vdots \cdots \vdots \boldsymbol{\Pi}_p] \boldsymbol{D}_p := \boldsymbol{\Pi} \boldsymbol{D}_p \tag{A.1}$$

where $\boldsymbol{D}_p$ is a suitable non-singular transformation matrix containing the unknown matrix $\boldsymbol{C}_1$. The ECM representation (2.10) can be written as

$$\Delta \boldsymbol{X}_t = \boldsymbol{\Psi}_0 \boldsymbol{X}_{2,t-1} + \sum_{l=1}^{p} \boldsymbol{\Xi}_l \varepsilon_{t-j} + \boldsymbol{\Xi}_{p+1,1} \varepsilon_{1,t-p-1} + \boldsymbol{e}_t(n). \tag{A.2}$$

The matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Psi}_0$ are defined in Saikkonen (1992, equation A.2). Set

$$\boldsymbol{\Lambda} := [\boldsymbol{\Xi} \vdots \boldsymbol{\Psi}_0], \quad \boldsymbol{W}_t := \boldsymbol{W}_t(p) := [\boldsymbol{\Upsilon}_t', \boldsymbol{X}_{2,t-1}'], \tag{A.3}$$

$$\boldsymbol{\Upsilon}_t' := \boldsymbol{\Upsilon}_t(p)' := [\varepsilon_{t-1}', \ldots, \varepsilon_{t-p}', \varepsilon_{1,t-p-1}']. \tag{A.4}$$

Consider the linear transformation

$$\boldsymbol{b}_t := \boldsymbol{\Sigma}^{-1/2} \boldsymbol{a}_t, \quad \hat{\boldsymbol{b}}_t = \boldsymbol{\Sigma}^{-1/2} \hat{\boldsymbol{a}}_t, \tag{A.5}$$

where $\boldsymbol{\Sigma}$ is defined in (2.18). Since $C_{\hat{\boldsymbol{b}}}^{(12)}(j) = \boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(12)}(j) \boldsymbol{\Sigma}_2^{-1/2}$. Using the property $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}(B)$, we have:

$$\begin{aligned} \mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) &= n \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' \left(\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1} \right) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) \\ &= n \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j)' \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) = \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)}. \end{aligned} \tag{A.6}$$

Thus, to prove the result, it is sufficient to show that

$$\mathcal{T}_{\boldsymbol{b}}^{(12)} - \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)} = o_p(M^{1/2}). \tag{A.7}$$

The result follows by decomposing the latter difference in two parts:

$$\mathcal{T}_{\boldsymbol{b}}^{(12)} - \mathcal{T}_{\hat{\boldsymbol{b}}}^{(12)} = n\sum_{j=1-n}^{n-1} k^2(j/M)\Big(\| \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 + 2\langle \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j), \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\rangle\Big)$$
$$= \left[T_n^{(1)} + T_{n-}^{(1)}\right] + T_n^{(2)} \tag{A.8}$$

where

$$T_n^{(1)} := n\sum_{j=0}^{n-1} k^2(j/M)\| \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2, \tag{A.9}$$

$$T_{n-}^{(1)} := n\sum_{j=1-n}^{-1} k^2(j/M)\| \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2, \tag{A.10}$$

$$T_n^{(2)} := n\sum_{j=1-n}^{n-1} k^2(j/M)\left\langle \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j), \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\rangle, \tag{A.11}$$

and then showing that each part is $o_p(M^{1/2})$. Consider the positive lags $j \geq 0$, since for negative lags, the proof is similar by symmetry.

Define $\hat{\delta}_t = \boldsymbol{b}_t^{(1)} - \hat{\boldsymbol{b}}_t^{(1)}$ and $\hat{\eta}_t = \boldsymbol{b}_t^{(2)} - \hat{\boldsymbol{b}}_t^{(2)}$. From (3.2), we have

$$\begin{aligned} T_n^{(1)} &= n\sum_{j=0}^{n-1} k^2(j/M)\| \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\|^2 \\ &= n\sum_{j=0}^{n-1} k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\left(\boldsymbol{b}_t^{(1)}\boldsymbol{b}_{t-j}^{(2)'} - \hat{\boldsymbol{b}}_t^{(1)}\hat{\boldsymbol{b}}_{t-j}^{(2)'}\right)\right\|^2, \end{aligned} \tag{A.12}$$

and using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} T_n^{(1)} &= n\sum_{j=0}^{n-1} k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\left(\boldsymbol{b}_t^{(1)}\hat{\eta}_{t-j}' + \hat{\delta}_t\boldsymbol{b}_{t-j}^{(2)'} - \hat{\delta}_t\hat{\eta}_{t-j}'\right)\right\|^2 \\ &\leq 4n(T_{1n} + T_{2n} + T_{3n}) \end{aligned} \tag{A.13}$$

with $T_{1n} = \sum_{j=0}^{n-1} k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{b}_t^{(1)}\hat{\eta}_{t-j}'\right\|^2$, $T_{2n} = \sum_{j=0}^{n-1} k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\hat{\delta}_t\boldsymbol{b}_{t-j}^{(2)'}\right\|^2$, and $T_{3n} = \sum_{j=0}^{n-1} k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\hat{\delta}_t\hat{\eta}_{t-j}'\right\|^2$. It suffices to show that the terms $T_{jn}, j = 1, 2, 3$, are $o_p(M^{1/2}/n)$. We can then write:

$$\begin{aligned} \hat{\delta}_t &= \left(\hat{\boldsymbol{b}}_t^{(1)} - \boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{e}_t^{(1)}\right) + \left(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{e}_t^{(1)} - \boldsymbol{b}_t^{(1)}\right) = \boldsymbol{\Sigma}_1^{-1/2}\left\{\left[\hat{\boldsymbol{a}}_t^{(1)} - \boldsymbol{e}_t^{(1)}\right] + \left[\boldsymbol{e}_t^{(1)} - \boldsymbol{a}_t^{(1)}\right]\right\} \\ &= \boldsymbol{\Sigma}_1^{-1/2}\left\{\left(\hat{\boldsymbol{\Lambda}}^{(1)} - \boldsymbol{\Lambda}^{(1)}\right)\boldsymbol{W}_t^{(1)} + \xi_t(p_1)\right\} \end{aligned} \tag{A.14}$$

where $\boldsymbol{e}_t^{(1)} := \boldsymbol{e}_t^{(1)}(n)$, $\boldsymbol{\Lambda}^{(h)}$ and $\boldsymbol{W}_t^{(h)}, h = 1, 2$ are defined as in (A.2) for each process, $\hat{\boldsymbol{\Lambda}}$ is the LS estimator of $\boldsymbol{\Lambda}$ and $\xi_t(p_1) = \sum_{l=p_1+1}^{\infty}\boldsymbol{\Phi}_l\boldsymbol{X}_{t-l}^{(1)}$ represents the bias of the

VAR(p_1) approximation of $\{\boldsymbol{X}_t^{(1)}\}$. The second equality is from Saikkonen and Lütkepohl (1996, p. 832). Also, using the result of Proposition 2.1, we deduce that

$$\|\hat{\boldsymbol{\Lambda}}^{(1)}-\boldsymbol{\Lambda}^{(1)}\|^2=O_p\left(\frac{p_1}{n}\right). \tag{A.15}$$

By equation (3.15) in Bouhaddioui and Roy (2006b), we have $\mathbb{E}\left(\left\|\xi_t\left(p_n^{(h)}\right)\right\|^2\right)=O\left(\sum_{l=p_n^{(h)}+1}^{\infty}\left\|\boldsymbol{\Phi}_l^{(h)}\right\|\right)^2$, $h=1,2$. Based on the result (3.17) in Bouhaddioui and Roy (2006b) and equation (2.15), we obtain:

$$T_{1n}=\sum_{j=0}^{n-1}k^2(j/M)\left\|\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{b}_t^{(1)}\hat{\eta}_{t-j}'\right\|^2=O_p\left(\frac{p_n^{(2)}M}{n^2}\right)\left\{\frac{1}{M}\sum_{j=0}^{n-1}k^2(j/M)\right\}. \tag{A.16}$$

Since $p_n^{(2)}=o(n/M^{1/2})$, we have $T_{1n}=o_p(M^{1/2}/n)$. By symmetry, we can prove that $T_{2n}=o_p\left(\frac{M^{1/2}}{n}\right)$. For the third term T_{3n}, using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
T_{3n} &= \sum_{j=0}^{n-1}k^2(j/M)\left\|n^{-1}\sum_{t=j+1}^{n}\hat{\eta}_t\hat{\delta}_{t-j}'\right\|^2\\
&\le \left\|\boldsymbol{\Lambda}^{(1)}-\hat{\boldsymbol{\Lambda}}^{(1)}\right\|^2\left\|\boldsymbol{\Lambda}^{(2)}-\hat{\boldsymbol{\Lambda}}^{(2)}\right\|^2\sum_{j=0}^{n-1}k^2(j/M)\left\|n^{-1}\sum_{t=j+1}^{n}\boldsymbol{W}_t^{(1)}(p_1)\boldsymbol{W}_{t-j}^{(2)}(p_2)'\right\|^2\\
&+ \left\|\boldsymbol{\Lambda}^{(1)}-\hat{\boldsymbol{\Lambda}}^{(1)}\right\|^2\sum_{j=0}^{n-1}k^2(j/M)\left\|n^{-1}\sum_{t=j+1}^{n}\boldsymbol{W}_t^{(1)}(p_1)\xi_{t-j}(p_2)'\right\|^2\\
&+ \left\|\boldsymbol{\Lambda}^{(2)}-\hat{\boldsymbol{\Lambda}}\right\|^2\sum_{j=0}^{n-1}k^2(j/M)\left\|n^{-1}\sum_{t=j+1}^{n}\xi_t(p_1)\boldsymbol{W}_{t-j}^{(2)}(p_2)'\right\|^2\\
&+ \sum_{j=0}^{n-1}k^2(j/M)\left\|n^{-1}\sum_{t=j+1}^{n}\xi_t(p_1)\xi_{t-j}(p_2)'\right\|^2.
\end{aligned} \tag{A.17}$$

Using the equations (3.19)–(3.22) in Bouhaddioui and Roy (2006a), the assumptions $p_n^{(h)}=o(n^{1/2}/M^{1/4})$, $n\sum_{l=p_n^{(h)}+1}^{\infty}\left\|\boldsymbol{\Phi}_l^{(h)}\right\|^2=o(n^{1/2}/M^{1/4})$ and the result (2.15), we conclude that $T_{3n}=o_p(M^{1/2}/n)$. Therefore, we obtain

$$T_n^{(1)}=n\sum_{j=0}^{n-1}k^2(j/M)\left\|\boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j)-\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|^2=o_p\left(M^{1/2}\right). \tag{A.18}$$

Finally, using Cauchy–Schwarz inequality once more, we have

$$\left|T_n^{(2)}\right| \leq 2n\sum_{j=1-n}^{n-1} k^2(j/M)\left|\left\langle \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j), \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j)-\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\rangle\right| \leq 2n\sum_{l=4}^{6} T_{\ln}, \quad \text{(A.19)}$$

with

$$T_{4n}=\sum_{j=0}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|\left\|\frac{1}{n}\sum_{t=j+1}^{n}\hat{\delta}_t\left(\boldsymbol{b}_{t-j}^{(2)}\right)'\right\|, \quad \text{(A.20)}$$

$$T_{5n}=\sum_{j=0}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|\left\|\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{b}_t^{(1)}\hat{\eta}_{t-j}'\right\|, \quad \text{(A.21)}$$

$$T_{6n}=\sum_{j=0}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|\left\|\frac{1}{n}\sum_{t=j+1}^{n}\hat{\delta}_t\hat{\eta}_{t-j}'\right\|. \quad \text{(A.22)}$$

Thus, it is sufficient to show that the terms T_{jn}, $j = 4, 5, 6$, are $o_p(M^{1/2}/n)$. By conditioning on $(\boldsymbol{b}_s^{(2)})_{s=-\infty}^{n}$ and using Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E}\left[T_{4n}\,|\left(\boldsymbol{b}_s^{(2)}\right)_{s=-\infty}^{n}\right] &\leq \sum_{j=1-n}^{n-1} k^2(j/M) \\
&\times \left[\mathbb{E}\left(\left\{\left(\frac{1}{n}\sum_{\tau=1}^{n}\left\|\boldsymbol{b}_\tau^{(1)}\boldsymbol{b}_{\tau-j}^{(2)^{\mathrm{T}}}\right\|\right)\left(\frac{1}{n}\sum_{t=j+1}^{n}\left\|\hat{\delta}_t\boldsymbol{b}_{t-j}^{(2)^{\mathrm{T}}}\right\|\right)\right\}^2 |\left(\boldsymbol{b}_s^{(2)}\right)_{s=-\infty}^{n}\right)\right]^{1/2} \\
&\leq \frac{M\Delta}{n^2}\left\{\frac{1}{M}\sum_{1-n}^{n-1}k^2(j/M)\right\}\left(\frac{1}{n}\sum_{\tau=1}^{n}\|\boldsymbol{b}_\tau^{(2)}\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left\|\hat{\delta}_t\right\|^2\right)^{1/2} \\
&= O_p\left(\frac{M\left(p_n^{(2)}\right)^{1/2}}{n^{5/2}}\right)=o_p\left(\frac{M^{1/2}}{n^{3/2}}\right).
\end{aligned} \quad \text{(A.23)}$$

The first equality is obtained by using the conditions on $p_n^{(2)}$, $\boldsymbol{\Phi}^{(2)}$, and the assumption of independence of the two innovation series. Then, $T_{4n}=o_p(M^{1/2}/n)$. By symmetry, we have also $T_{5n}=o_p(M^{1/2}/n)$. Finally, from Markov inequality, we have

$$\sum_{j=1}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|^2=O_p(M/n) \quad \text{(A.24)}$$

hence, using the Cauchy–Schwarz inequality and the result for T_{3n}, we obtain that $T_{6n}=o_p(M/n)$. Thus, $T_n^{(2)}=o_p(M^{1/2})$ and the Proof of Proposition 3.1 is completed.

Proof of Proposition 3.2. Since $D_n(k)=MD(k)\{1+o(1)\}$, it is sufficient to show that

$$\mathcal{T}(\hat{\boldsymbol{a}},\hat{\boldsymbol{\Sigma}})-\mathcal{T}(\hat{\boldsymbol{a}},\boldsymbol{\Sigma})=O_p(M/n^{1/2}). \quad \text{(A.25)}$$

Using the fact that $\boldsymbol{C}_{\hat{\boldsymbol{a}}}^{(hh)}(0) - \boldsymbol{\Sigma}_n^h = O_p(n^{-1/2})$ for $h = 1, 2$ (see Lütkepohl & Saikkonen, 1997, p. 133), it follows that

$$\left[\mathrm{C}_{\hat{\mathrm{a}}}^{(22)}(0)^{-1} \otimes \mathrm{C}_{\hat{\mathrm{a}}}^{(11)}(0)^{-1}\right] - \left[\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_1^{-1}\right] = \mathrm{O}_p(n^{-1/2}). \tag{A.26}$$

Thus,

$$\begin{aligned} \mathcal{T}(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}) - \mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) &= n \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' O_p(n^{-1/2}) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) \\ &= O_p(n^{1/2}) \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j). \end{aligned} \tag{A.27}$$

To complete the proof, it remains to prove that

$$\mathcal{B}(n) = \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) = O_p(M/n). \tag{A.28}$$

First, let us decompose $\mathcal{B}(n)$ in two parts

$$\begin{aligned} \mathcal{B}(n) &= \sum_{j=1-n}^{n-1} k^2(j/M) \left\{ \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)' \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j) \right\} \\ &\quad + \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)' \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j) \\ &= \mathcal{B}_1 + \mathcal{B}_2. \end{aligned} \tag{A.29}$$

By an argument similar to the one used to prove (A.7) in Proposition 3.1, we have:

$$\mathcal{B}_1(n) = \sum_{j=1-n}^{n-1} k^2(j/M) \left\{ \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j)' \boldsymbol{c}_{\hat{\boldsymbol{a}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)' \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j) \right\} = o_p(M^{1/2}/n), \tag{A.30}$$

and, by Markov inequality, it follows that

$$\mathcal{B}_2(n) = \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j)' \boldsymbol{c}_{\boldsymbol{a}}^{(12)}(j) = O_p(M/n). \tag{A.31}$$

Combining the results for $\mathcal{B}_1(n)$ and $\mathcal{B}_2(n)$, we obtain that

$$\mathcal{T}\left(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\Sigma}}\right) - \mathcal{T}(\hat{\boldsymbol{a}}, \boldsymbol{\Sigma}) = O_p\left(n^{1/2}\right) O_p(M/n) = O_p(M/n^{1/2}), \tag{A.32}$$

and the *Proof of Proposition 3.2* is completed.

Proof of Theorem 4.1. First, we note that the statistic $\mathcal{Q}_n$ is a normalized version of $\mathcal{T}(\hat{\boldsymbol{a}},\hat{\boldsymbol{\Sigma}})$ which can be viewed as the $\mathcal{L}_2$-norm of a kernel-based estimator of the cross-coherency function between the two innovations processes. Thus, the statistic $\mathcal{Q}_n$ can be expressed as

$$\mathcal{Q}_n = \frac{n\left\|s_{\hat{\boldsymbol{a}}}^{(12)}\right\|_2^2 - m_1 m_2 S_n(k)}{\sqrt{2 m_1 m_2 D_n(k)}} \tag{A.33}$$

where $s_{\hat{\boldsymbol{a}}}^{(12)}$ is the estimator of the cross-coherency function between the two innovations processes given by

$$\left\|s_{\boldsymbol{a}}^{(12)}\right\|_2^2 = \sum_{j=-\infty}^{\infty} \gamma_{\boldsymbol{a}}^{(12)}(j)'\left(\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1\right)^{-1} \gamma_{\boldsymbol{a}}^{(12)}(j) \tag{A.34}$$

where $\gamma_{\boldsymbol{a}}^{(12)}(j) := \mathrm{vec}[\Gamma_{\boldsymbol{a}}^{(12)}(j)]$. For details, see Section 4 in Bouhaddioui and Roy (2006a). By definition of $\mathcal{Q}_n$, we can write

$$\begin{aligned}\left(\frac{M^{1/2}}{n}\right)\mathcal{Q}_n &= \frac{M^{1/2}\left\|s_{\hat{\boldsymbol{a}}}^{(12)}\right\|_2^2 - \dfrac{M^{1/2}}{n} m_1 m_2 S_n(k)}{\left\{2 m_1 m_2 D(k)\right\}^{1/2}} \\ &= \frac{\left\|s_{\hat{\boldsymbol{a}}}^{(12)}\right\|_2^2}{\left\{2 m_1 m_2 M^{-1} D_n(k)\right\}^{1/2}} - \frac{n^{-1} S_n(k)}{\left\{2 M^{-1} D_n(k)\right\}^{1/2}}\left(m_1 m_2\right)^{1/2}.\end{aligned} \tag{A.35}$$

From (3.11), the last term of the previous equation goes to zero when $M/n \to 0$ as $n \to \infty$. Using the linear transformation $\boldsymbol{b}_t = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{a}_t$, as in Proposition 3.1, we have $\left\|s_{\boldsymbol{a}}^{(12)}\right\| = \left\|s_{\boldsymbol{b}}^{(12)}\right\|$. Also, since the processes $\boldsymbol{b}^{(1)}$ and $\boldsymbol{b}^{(2)}$ are stationary and by Lemma A.7 in Bouhaddioui and Roy (2006a), we have that

$$\left\|\tilde{s}_{\boldsymbol{b}}^{(12)}\right\|^2 - \left\|s_{\boldsymbol{b}}^{(12)}\right\|^2 \xrightarrow{p} 0 \tag{A.36}$$

where $\left\|\tilde{s}_{\boldsymbol{b}}^{(12)}\right\|$ is defined as $\left\|s_{\hat{\boldsymbol{b}}}^{(12)}\right\|$, the residual series $\left\{\hat{\boldsymbol{b}}_t^{(1)}, \hat{\boldsymbol{b}}_t^{(2)}\right\}_{t=1}^n$ being replaced by the innovation series $\left\{\boldsymbol{b}_t^{(1)}, \boldsymbol{b}_t^{(2)}\right\}_{t=1}^n$. Thus, to prove the consistency result (4.4), it is sufficient to verify that $\left\|s_{\hat{\boldsymbol{b}}}^{(12)}\right\|_2^2 - \left\|\tilde{s}_{\boldsymbol{b}}^{(12)}\right\|_2^2 \xrightarrow{p} 0$, which follows from the following lemma.

Lemma A.1. *Under the assumptions of Theorem 4.1, we have*

$$\left\|\tilde{s}_{\hat{\boldsymbol{b}}}^{(12)}\right\|_2^2 - \left\|\tilde{s}_{\boldsymbol{b}}^{(12)}\right\|_2^2 \xrightarrow{p} 0 \tag{A.37}$$

Proof of Lemma A.1. By definition of $s_{\hat{\boldsymbol{b}}}^{(12)}$ and $\tilde{s}_{\boldsymbol{b}}^{(12)}$, and by similar calculations to those for the proof in Proposition 3.1, we obtain

$$\begin{aligned} \left\|s_{\hat{\boldsymbol{b}}}^{(12)}\right\|_2^2 - \left\|\tilde{s}_{\boldsymbol{b}}^{(12)}\right\|_2^2 &= \sum_{j=1-n}^{n-1} k^2(j/M)\left(\left\|\boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j)\right\|^2 - \left\|\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|^2\right) \\ &= \sum_{j=1-n}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|^2 \\ &+ 2\sum_{j=1-n}^{n-1} k^2(j/M)\left\langle \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j), \boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\rangle \end{aligned}. \quad \text{(A.38)}$$

It is sufficient to prove that the first term goes to zero in probability, because the second term can be bounded by a product of the first term and a finite quantity, using the Cauchy–Schwarz inequality. With the notations of Proposition 3.1, we can write

$$\sum_{j=1-n}^{n-1} k^2(j/M)\left\|\boldsymbol{c}_{\hat{\boldsymbol{b}}}^{(12)}(j) - \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)\right\|^2 \le 4\sum_{l=1}^{3} T_{ln}, \quad \text{(A.39)}$$

where T_{ln}, $l = 1, 2, 3$, are defined in Proposition 3.1. We first prove that $T_{1n} \to 0$ in probability. By the Cauchy–Schwarz inequality, we obtain

$$T_{1n} \le M\left\{\frac{1}{M}\sum_{j=0}^{n-1} k^2(j/M)\right\}\left\{\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{b}_t^{(1)}\right\|^2\right\}\left\{\frac{1}{n}\sum_{t=1}^{n}\left\|\hat{\eta}_t\right\|^2\right\}. \quad \text{(A.40)}$$

By definition of $\hat{\eta}_t$, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\left\|\hat{\eta}_t\right\|^2 \le \frac{1}{n}\sum_{t=1}^{n}\left\{\left\|\left(\boldsymbol{\Lambda}^{(2)} - \hat{\boldsymbol{\Lambda}}^{(2)}\right)\boldsymbol{W}_t^{(2)}\right\|^2 + \left\|\xi_t\left(p_n^{(2)}\right)\right\|^2\right\}. \quad \text{(A.41)}$$

Since $\|\boldsymbol{\Gamma}_{\boldsymbol{a}}^{(11)}(l)\|$ is uniformly bounded by a positive constant Δ, and the parameters $\{\boldsymbol{\Phi}_l\}$ are a linear function of the original parameters $\{\boldsymbol{G}_l\}$, then the bias approximation can be bounded by

$$\mathbb{E}\left\|\xi_t\left(p_n^{(2)}\right)\right\|^2 \le \Delta\left(\sum_{l=p_n^{(2)}+1}^{\infty}\left\|\boldsymbol{\Phi}_l^{(2)}\right\|^2\right) = o(n^{-1}). \quad \text{(A.42)}$$

See also the result (A.12) in Saikkonen (1992). Under the assumptions on the process $\boldsymbol{b}_t^{(1)}$, on $p_n^{(2)}$ and on the parameters $\left\{\boldsymbol{\Phi}_l^{(2)}\right\}$, we have

$$T_{1n} = O_p\left(\frac{M\left(p_n^{(2)}\right)^2}{n}\right) + O_p\left(M\left(\sum_{l=p_n^{(2)}+1}^{\infty}\left\|\boldsymbol{\Phi}_l^{(2)}\right\|\right)^2\right) = o_p(1). \quad \text{(A.43)}$$

By symmetry, we can verify that $T_{2n} = o_p(1)$. For T_{3n}, we can write

$$\begin{aligned} T_{3n} &= \sum_{j=0}^{n-1} k^2(j/M) \left\| \frac{1}{n} \sum_{t=j+1}^{n} \hat{\delta}_t \hat{\eta}'_{t-j} \right\|^2 \\ &\leq M \left\{ \frac{1}{M} \sum_{j=0}^{n-1} k^2(j/M) \right\} \left\{ \frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\delta}_t \right\|^2 \right\} \left\{ \frac{1}{n} \sum_{t=1}^{n} \left\| \hat{\eta}_t \right\|^2 \right\}. \end{aligned} \tag{A.44}$$

By symmetry, we can prove that $\frac{1}{n}\sum_{t=1}^{n} \left\| \hat{\delta}_t \right\|^2 = O_p\left(\left(p_n^{(1)} \right)^2 / n \right) + O_p(1) \left(\sum_{l=p_n^{(1)}+1}^{\infty} \left\| \boldsymbol{\Phi}_l^{(1)} \right\| \right)^2$, and using the same assumptions as those for T_{1n}, we obtain that $T_{3n} = o_p(1)$. Finally, we conclude that

$$\left\| \tilde{s}_{\hat{\boldsymbol{b}}}^{(12)} \right\|^2 - \left\| \tilde{s}_{\boldsymbol{b}}^{(12)} \right\|^2 = o_p(1). \tag{A.45}$$

This completes the Proof of Lemma A.1 and then Theorem 4.1.

Proof of Theorem 5.1. By the Proof of Theorem 4.1,

$$\mathcal{Q}_n = \frac{n \left\| s_{\mathrm{b}}^{(12)} \right\|_2^2 - m_1 m_2 S_n(k)}{\sqrt{2 m_1 m_2 D_n(k)}} + o_p(1). \tag{A.46}$$

where

$$\begin{aligned} n \left\| s_{\boldsymbol{b}}^{(12)} \right\|_2^2 &= n \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)' \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) \\ &= n \sum_{j=1-n}^{n-1} k^2(j/M) \left(\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \gamma_{\boldsymbol{b}}^{(12)}(j) \right)' \left(\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \gamma_{\boldsymbol{b}}^{(12)}(j) \right) \\ &\quad + 2n \sum_{j=1-n}^{n-1} k^2(j/M) \boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j)' \gamma_{\boldsymbol{b}}^{(12)}(j) - n \sum_{j=1-n}^{n-1} k^2(j/M) \gamma_{\boldsymbol{b}}^{(12)}(j)' \gamma_{\boldsymbol{b}}^{(12)}(j) \\ &= n \sum_{j=1-n}^{n-1} k^2(j/M) \left(\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \gamma_{\boldsymbol{b}}^{(12)}(j) \right)' \left(\boldsymbol{c}_{\boldsymbol{b}}^{(12)}(j) - \gamma_{\boldsymbol{b}}^{(12)}(j) \right) \\ &\quad + M^{1/2} \sum_{j=1-n}^{n-1} k^2(j/M) \lambda_{\boldsymbol{b}}^{(12)}(j)' \lambda_{\boldsymbol{b}}^{(12)}(j) + o_p(M^{1/4}). \end{aligned}$$

Since there exists $j^* \in \mathbb{Z}$ such that $\boldsymbol{\Lambda}_{\boldsymbol{b}}^{(12)}(j) = \boldsymbol{0},: \forall j : (|j| > j^*)$, we have

$$\begin{aligned} \sum_{j=-\infty}^{\infty} k^2(j/M) \lambda_{\boldsymbol{b}}^{(12)}(j)' \lambda_{\boldsymbol{b}}^{(12)}(j) &= \sum_{j=-j^*}^{j^*} k^2(j/M) \lambda_{\boldsymbol{b}}^{(12)}(j)' \lambda_{\boldsymbol{b}}^{(12)}(j) \\ &\to \sum_{j=-\infty}^{\infty} \lambda_{\boldsymbol{b}}^{(12)}(j)' \lambda_{\boldsymbol{b}}^{(12)}(j) := \beta\left(\boldsymbol{\Lambda}_{\boldsymbol{b}}^{(12)} \right). \end{aligned}$$

Thus,

$$\frac{n\left\|s_{\boldsymbol{b}}^{(12)}\right\|-m_1m_2S_n(k)}{\sqrt{2m_1m_2D_n(k)}}\xrightarrow{L}Z+\frac{\beta\left(\boldsymbol{\Lambda}_{\boldsymbol{b}}^{(12)}\right)}{\sqrt{2m_1m_2D(k)}}$$

where $Z \sim \mathcal{N}(0,1)$.

CHAPTER 10

INFERENCE IN CONDITIONAL VECTOR ERROR CORRECTION MODELS WITH A SMALL SIGNAL-TO-NOISE RATIO*

Nikolay Gospodinov[a], Alex Maynard[b] and Elena Pesavento[c]

[a]*Federal Reserve Bank of Atlanta, Atlanta, GA, USA*
[b]*University of Guelph, Guelph, ON, Canada*
[c]*Emory University, Atlanta, GA, USA*

ABSTRACT

It is widely documented that while contemporaneous spot and forward financial prices trace each other extremely closely, their difference is often highly persistent and the conventional cointegration tests may suggest lack of cointegration. This chapter studies the possibility of having cointegrated errors that are characterized simultaneously by high persistence (near-unit root behavior) and

*This chapter is prepared for the conference and festschrift in honor of Joon Y. Park whose work on nonstationarity in economic and financial data had an enormous influence on our research and analytical approach. The authors would like to thank the volume Co-Editor (Zack Miller) and an anonymous referee for insightful suggestions. We are also grateful to Gregory Cox, Christian Gourieroux, Joann Jasiak, and participants at the 2018 IAAE Meetings and the 26th Annual Meeting of the Midwest Econometrics Group for useful comments and discussions on an earlier version of this chapter. The authors gratefully acknowledge funding under SSHRC Grant 430417. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 295–318

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A010

very small (near zero) variance. The proposed dual parameterization induces the cointegration error process to be stochastically bounded which prevents the variables in the cointegrating system from drifting apart over a reasonably long horizon. More specifically, this chapter develops the appropriate asymptotic theory (rate of convergence and asymptotic distribution) for the estimators in unconditional and conditional vector error correction models (VECM) when the error correction term is parameterized as a dampened near-unit root process (local-to-unity process with local-to-zero variance). The important differences in the limiting behavior of the estimators and their implications for empirical analysis are discussed. Simulation results and an empirical analysis of the forward premium regressions are also provided.

Keywords: Cointegration; vector error correction models; local-to-unity asymptotics; local-to-zero variance; spot and forward exchange rates; forward premium regression

JEL classifications: C12; C15; C22; F31

1. INTRODUCTION AND MOTIVATION

This chapter considers some important inference issues that arise in the analysis of nearly cointegrated processes in the presence of highly persistent cointegrating errors whose variability is only a small fraction of the variance of the original variables. Equivalently, in the vector error correction (VEC) representation of the cointegrated system, the error correction term is near-integrated with low signal-to-noise ratio. Typical examples of this setup include models that study the unbiasedness of forward and futures prices (exchange rates, interest rates and commodity prices) for the expected future spot values. For instance, spot and one-month forward exchange rates trace each other very closely and are virtually indistinguishable from each other as illustrated in the left panel of Fig. 1 for the British pound (BP), German mark (DM), Swiss franc (SF) and Canadian dollar (CD) – all against the US dollar. And yet, the spot–forward spread (the difference between the two series) is characterized by high persistence which becomes visible when plotted in isolation in the right panel of Fig. 1. In fact, a formal unit root test on the spot–forward spread often cannot reject the null hypothesis of a lack of cointegration. The heuristic reason for this is that the spot–forward spread has a tiny variance compared to the variability of the individual variables and this prevents its near random walk component from forcing spot and forward rates to drift apart in the long run. Similar arguments apply to the time series behavior of cash and futures prices of other asset classes, such as commodities or bond yields.

To accommodate this empirical regularity without compromising the integrity of the cointegrating system, we model this component as a dampened (stochastically bounded) near-unit root process. More generally, this parameterization proves to be a useful device in reconciling the internal consistency of the statistical behavior with the widely held belief that many economic and financial time series

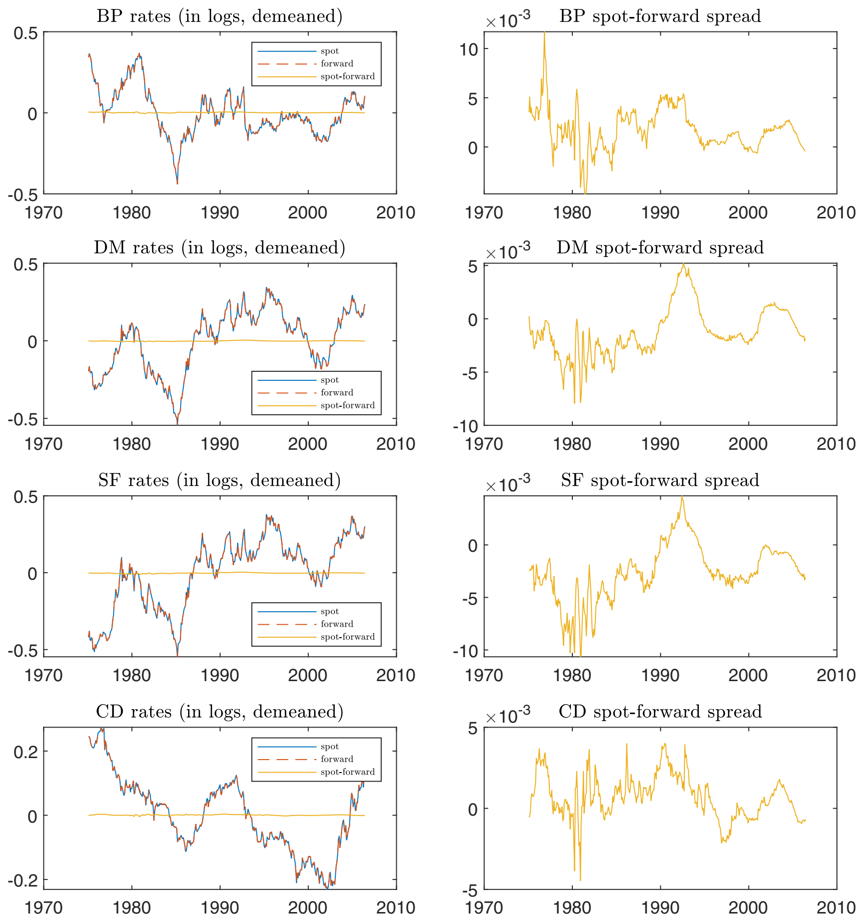


Fig. 1. The Left Charts Plot (on the Same Scale) the Spot Rate, One-month Forward Rate and Their Difference (Spot-forward Spread) for Four (BP, DM, SF and CD) Currencies against the US Dollar. The Right Charts Zoom on the Dynamics of the Spot-forward Spread for These Four Currencies.

are driven by a slowly moving, low-frequency persistent component (Bansal & Yaron, 2004; Gourieroux & Jasiak, 2020[1]; Phillips & Lee, 2013; among others).[2] Consider, for instance, the unobserved component (local level) model

$$x_t = \mu_t + u_t,$$
$$\mu_t = \rho\mu_{t-1} + \tau\xi_t,$$

where μ_t is a highly persistent and possibly unit root component (with ρ near 1), u_t and ξ_t are mutually uncorrelated white noise disturbances with variance σ^2, and τ is the signal-to-noise ratio. The observed variable x_t could be consumption growth, dividend growth or equity returns and the long-run risks associated with μ_t, which is largely interpreted as capturing the common variation in real activity (Bansal & Yaron, 2004), play a crucial role in explaining the equity premium puzzle. While the above representation is theoretically appealing, the empirical evidence on the existence of such a long-run component in stock returns is rather weak. There are two main questions that drive a wedge between the theoretical setup and the empirical justification of these low-frequency components. First, how come we do not detect this persistence in the data? And second, how can we reconcile the statistical behavior of the model and the data as the sample size increases? After all, the persistent component μ_t has to dominate the dynamics of the observed series as the number of time series observations grows.

To reconcile this tension, it is convenient to adopt a dual localization and model μ_t as a dampened near-unit root process (see Gospodinov, 2009). More specifically, let $\mu_t = \rho_T \mu_{t-1} + \tau_T \xi_t$ denote the low-frequency component. The dual localization involves (a) a local-to-unity parameterization $\rho_T = 1 + c / T$ for some fixed constant $c \leq 0$, and (b) a local-to-zero parameterization for the signal-to-noise ratio $\tau_T = \lambda / \sqrt{T}$ for some fixed constant $\lambda > 0$. This dual localization proves to be instrumental in producing a process that is stochastically bounded and hence consistent with both statistical and economic theory. Unlike regular near-unit root processes that are of order $O_p(T^{1/2})$, the local-to-zero variance localization dampens the stochastic trend behavior of x_t and keeps it stochastically bounded $\left(O_p(1)\right)$. The dual localization removes the economically unappealing possibility that the low-frequency component can wander off and induce non-stationarity in asset returns or consumption growth. The persistent and noise components of the model now have comparable orders of magnitude as both μ_t and the rest of the variables are stochastically bounded. While this statistical device renders the model congruent, the observed stock returns at the monthly or quarterly frequency are still overwhelmed by noise and the empirical detection of this small low-frequency component remains elusive.

With this background in mind, this chapter derives the theoretical implications of the simultaneous presence of high persistence, low variability and endogeneity of the cointegrating errors for the concept of cointegration, the properties of cointegrating regressions, estimation and testing in VECM, etc. More specifically, we develop the appropriate theory (rate of convergence and asymptotic distributions) for the estimators in unconditional and conditional VECM (Boswijk, 1994; Johansen, 1992; Phillips, 1991) when the error correction term is parameterized as a dampened near-unit root process (local-to-unity process with local-to-zero variance). In doing this, we combine the literatures on near cointegration (Elliott et al., 2005; Jansson & Haldrup, 2002; Pesavento, 2004; Zivot, 2000) and near-zero variance regressors (Deng, 2014; Gospodinov, 2009; Gourieroux & Jasiak, 2020; Moon et al., 2004; Torous & Valkanov, 2000). This double local parameterization of the persistence and variance of the cointegration errors provides a powerful tool for deriving limiting results by capturing the salient features of the data in the empirical examples. One important result that emerges from our analysis is that the estimator in the conventional VECM is characterized by a large bias, a reduced rate of convergence and a highly dispersed asymptotic distribution, while its conditional counterpart enjoys a substantially improved asymptotic behavior. This chapter provides a detailed investigation of the numerical properties of the estimators in unconditional and conditional VECM and the empirical size and power of tests for cointegration based on the corresponding test statistics. The practical importance of the analytical results is demonstrated in the context of exchange rate models.

The rest of the chapter is organized as follows. Section 2 introduces the analytical setup, modeling assumptions and appropriate limits. It also presents the main representations that characterize the asymptotic behavior of the estimators and their corresponding t-tests. Section 3 contains simulation results while Section 4 reports the results from the empirical application for spot and forward exchange rates. Section 5 concludes.

2. MODEL AND MAIN RESULTS

2.1. Assumptions and Parameterization

First, we discuss the model setup, assumptions and the proposed dual local parameterization. Suppose that $(x_t', y_t)'$ is a $(k+1)\times 1$ vector generated by the triangular system[3]

$$\begin{aligned} x_t &= \psi_x + \phi_x t + u_{x,t} \\ y_t &= \psi_y + \phi_y t + \gamma' x_t + u_{y,t} \end{aligned} \tag{1}$$

and

$$\begin{pmatrix} (1-L)u_{x,t} \\ (1-\rho_T L)u_{y,t} \end{pmatrix} = \begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} \end{pmatrix}$$

with

$$A(L)v_t = \varepsilon_t$$

for $t = 1,\ldots,T$. We make the following assumptions.

Assumption A. *Assume that* $A(L) = I_{k+1} - \sum_{i=1}^{p} A_i L^i$ *is a matrix polynomial of a finite (known) order p in the lag operator L, with roots that lie outside the unit circle.*

Assumption B. *Assume that* $\max_{-p\le t\le 0} \left\| \left(u_{x,t}', u_{y,t}\right)' \right\| = O_p(1)$, *where* $\|\cdot\|$ *is the Euclidean norm.*

Assumption C. *Assume that* $\varepsilon_t = (\varepsilon_{x,t}', \varepsilon_{y,t})'$ *is a homoskedastic martingale difference sequence with a variance matrix* Σ, $0 < \|\Sigma\| < \infty$, *and finite fourth moments*, $\max_i E(\varepsilon_{it}^4) < \infty$.

Our model resembles the standard triangular model for cointegration (Phillips, 1991; see also Engle & Granger, 1987; Park, 1992; Park & Phillips, 1989; Stock & Watson, 1993; among many others) but we allow the error in the cointegration regression to be persistent, yet still bounded. For clarity of exposition, the analytical results below are presented for the case of no deterministic terms in model (1); that is, $\psi_x = 0, \psi_y = 0, \phi_x = 0$ and $\phi_y = 0$. The generalization to deterministic terms is straightforward to obtain at the expense of additional notation (see Section 2.2 for further discussion).

Assumptions A–C are the same as in Elliott et al. (2005). Assumption A implies stationarity while Assumption B states that the initial values are asymptotically negligible. Assumption C ensures that ε_t satisfies the Functional Central Limit Theorem (FCLT) so that

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{[\mathrm{Tr}]}\varepsilon_s \Rightarrow \Sigma^{1/2}W(r),$$

where $W(r)$ is a standard vector Brownian motion, $\Rightarrow$ signifies weak convergence and $[\cdot]$ denotes the greatest lesser integer function. Furthermore, Assumptions A–C imply that $v_t = A(L)^{-1}\varepsilon_t$ has the following limit

$$\frac{1}{\sqrt{T}}\sum_{s=1}^{[\mathrm{Tr}]}v_s \Rightarrow \Omega^{1/2}W(r),$$

where $\Omega = A(1)^{-1}\Sigma A(1)^{-1'}$ is the spectral density at frequency zero of v_t scaled by 2π, $\Omega^{1/2} = \begin{pmatrix} \Omega_{11}^{1/2} & 0 \\ \omega_{21}\Omega_{11}^{-1/2} & \omega_{2.1}^{1/2} \end{pmatrix}$, $\omega_{2.1}^{1/2} = \omega_{22} - \omega_{21}\Omega_{11}^{-1}\omega_{12}$ and $W' = (W_1' \; W_2)$ is a vector of independent standard Brownian motions, partitioned conformably to $v_{x,t}$ and $v_{y,t}$.

Next, define the scalar $\theta^2 = \delta'\delta$, where $\delta = \Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ denotes a vector containing the bivariate correlations at frequency zero of each element of $v_{x,t}$ with $v_{y,t}$. The scalar θ^2 represents the contribution of the right-hand variables in the second equation of (1) and it takes a value of zero when $v_{x,t}$ are not correlated in the long run with the errors from the cointegration regression.

Assumption D. *Assume that* $0 \leq \theta^2 < 1$ *and* Ω_{11} *is non-singular*.

Assumption D restricts the squared long-run correlation θ^2 to be strictly less than one for technical reasons (see also Hansen, 1995). Also, the assumption that Ω_{11} is non-singular implies that elements of x_t are not individually cointegrated. Our next assumption follows Gospodinov (2009) and reparameterizes ρ_T and τ_T as local-to-unity and local-to-zero sequences to account for the possibility of highly persistent errors and low signal-to-noise ratio.

Assumption E. *Assume that* $\rho_T = 1 + c/T$ *for some fixed constant* $c \leq 0$, *and* $\tau_T = \lambda/\sqrt{T}$ *for some fixed constant* $\lambda > 0$.

The normalization factors T and $T^{1/2}$ for the local-to-unity and local-to-zero parameterizations are chosen to match the asymptotics of the estimators of

ρ_T and τ_T. The local-to-zero parameterization has been used in a predictive regression framework by Deng (2014), Gourieroux and Jasiak (2020), Moon et al. (2004), and Torous and Valkanov (2000). In a different context, Ng and Perron (1997) adopt a similar parameterization to study the effect of low signal-to-noise ratio of the regressor on the sampling properties of cointegrating vector estimators.

Thus, under our assumptions, both y_t and x_t in (1) have a unit root but are cointegrated, and the cointegration error $y_t - \gamma' x_t$ is persistent – potentially persistent enough that we would not detect cointegration with standard tests in small samples. Yet, we have that asymptotically τ_T is approaching zero so that the cointegration error remains stochastically bounded even when its persistence parameter ρ_T, that drives its dynamics, is near or at unity.

As pointed out in the introduction, the dual localization is key for ensuring that the cointegration error $u_{y,t} = y_t - \gamma' x_t$ is stochastically bounded and hence consistent with both statistical and economic theory. Unlike regular near-unit root processes that are of order $O_p(T^{1/2})$, the local-to-zero variance localization dampens the stochastic trend behavior of $u_{y,t}$ and keeps it stochastically bounded ($O_p(1)$). More specifically, $u_{y,t}$ converges weakly to an Ornstein–Uhlenbeck process without any normalization that depends on the sample size:[4]

$$\begin{aligned} u_{y,t} &= \lambda T^{-1/2} \sum_{i=1}^{t} (1 + c/T)^{t-i} v_{y,i} \\ &\Rightarrow \omega_{2.1}^{1/2} \lambda J_{12c}(r), \end{aligned}$$

where $J_{12c}(r) = W_{12}(r) + c\int_0^r e^{(r-s)c} W_{12}(s)ds$ with $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}} \tilde{W}_1(r) + W_2(r)$, where $\tilde{W}_1(r)$ is an univariate standard Brownian motion independent of $W_2(r)$. The dual localization removes the unappealing possibility for some economic series (e.g., spot and forward prices) that the errors $u_{y,t}$ can wander off and preserves the cointegration between y_t and x_t.

2.2. Limiting Distributions

We first consider the standard OLS estimator for the cointegration vector γ obtained from the regression of y_t on x_t. Theorem 1 presents the limiting distribution of the estimator $\hat{\gamma}$ and its standard error (SE).[5]

Theorem 1. *Under Assumptions A–E, and as $T \to \infty$,*

$$\sqrt{T}\left(\hat{\gamma} - \gamma_0\right) \Rightarrow \Omega_{11}^{-1/2\prime} \omega_{2.1}^{1/2} \lambda \left(\int W_1 W_1'\right)^{-1} \left(\int W_1 J_{12c}\right), \tag{2}$$

$$T \cdot \mathrm{SE}(\hat{\gamma}) \Rightarrow \left[\lambda^2 \Omega_{11}^{-1/2} \left(\int W_1 W_1'\right)^{-1} \Omega_{11}^{-1/2\prime} \omega_{2.1} \left(\int \tilde{J}_{12c}^2\right)\right]^{1/2}, \tag{3}$$

where $\tilde{J}_{12c} = J_{12c} - \left(\int W_1 J_{12c}\right)' \left(\int W_1 W_1'\right)^{-1} W_1, J_{12c}(r) = W_{12}(r) + c\int_0^r e^{(r-s)c} W_{12}(s)ds$, $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}} \tilde{W}_1(r) + W_2(r)$, *and* $\tilde{W}_1(r)$ *and* $W_2(r)$ *are independent univariate standard Brownian motions.*

Proof. See Appendix.

Interestingly, the estimator $\hat{\gamma}$ has an asymptotic distribution that resides in between the usual spurious and cointegration regressions. Unlike spurious regressions, $\hat{\gamma}$ is consistent but, in contrast to the usual cointegration regressions, it is not super-consistent as it has a slower $\left(\sqrt{T}\right)$ rate of convergence. Additionally, while the estimator is consistent, the conventional t-statistic of $H_0 : \gamma = \gamma_0$ diverges at rate $T^{1/2}$ as in a spurious regression. This can be easily seen from the results in Theorem 1; that is,

$$t_{\gamma=\gamma_0} = \frac{(\hat{\gamma}-\gamma_0)}{\mathrm{SE}(\hat{\gamma})} = \frac{\sqrt{T}\,(\hat{\gamma}-\gamma_0)}{T{\cdot}\mathrm{SE}(\hat{\gamma})}\sqrt{T} \to \pm\infty$$

as $T \to \infty$. Note also that an efficient estimator of γ can be obtained using a control variable approach (Phillips, 1991).

In what follows, we assume that γ is known, which is the case in our empirical application. We briefly discuss the setup when γ is estimated after we present our main results in Theorems 2 and 3. Consider the VEC representation of the model given by

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = (\rho-1)\begin{pmatrix} 0_k & 0 \\ -\gamma' & 1 \end{pmatrix}\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} I_k & 0 \\ \gamma' & 1 \end{pmatrix}\begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} \end{pmatrix}. \tag{4}$$

Premultiplying by $A(L)$ and using $A(L) = I_{k+1} - \sum_{i=1}^{p} A_i L^i = I_{k+1} - \left(A_1 + A_2 L + \ldots + A_p L^{p-1}\right)L = I_{k+1} - A^*(L)L$ and $A(L) = A(1) + (1-L)\bar{A}(L)$, where $\bar{A}(L)$ is another $(p-1)$-order lag polynomial, we obtain

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = (\rho-1)A(1)\begin{pmatrix} 0_k \\ u_{y,t-1} \end{pmatrix} + \tilde{A}(L)\begin{pmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + A(L)\begin{pmatrix} v_{x,t} \\ \tau_T v_{y,t} + \gamma' v_{x,t} \end{pmatrix},$$

where $\tilde{A}(L) = A^*(L) + \bar{A}(L)(\rho-1)\begin{pmatrix} 0_k & 0 \\ -\gamma' & 1 \end{pmatrix}$.

Here, we restrict our analysis to a single-equation ECM by imposing that the first to the second-to-last elements of the last column of $A(1)$ are equal to zero (see our discussion after Theorem 2); that is, $A(1) = \begin{pmatrix} A_{11}(1) & 0_k \\ A_{21}(1) & a_{22}(1) \end{pmatrix}$. In this case, the conditional ECM for Δy_t is given by

$$\begin{aligned} \Delta y_t &= (\rho-1)a_{22}(1)u_{y,t-1} + \tilde{A}_{21}(L)\Delta x_{t-1} + \tilde{a}_{22}(L)\Delta y_{t-1} \\ &\quad + (A_{21}(L) + a_{22}(L)\gamma')v_{x,t} + a_{22}(L)\tau_T v_{y,t} \end{aligned}$$

or, using that $\Delta x_t = v_{x,t}$ and $\tau_T a_{22}(L)v_{y,t} = \tau_T \varepsilon_{y,t} - A_{21}(L)\tau_T \Delta x_t$,

$$\Delta y_t = \beta u_{y,t-1} + \gamma'\Delta x_t + \pi_1^*(L)'\Delta x_{t-1} + \pi_2^*(L)\Delta y_{t-1} + \tau_T \varepsilon_{y,t},$$

where $\beta = (\rho - 1)a_{22}(1)$ and $\pi_1^*(L)$ and $\pi_2^*(L)$ are lag polynomials of order $p - 1$ that are functions of $A(L)$ and $\tilde{A}(L)$. Furthermore, define $\eta_t = \Sigma^{-1/2}\varepsilon_t$ such that $\varepsilon_{x,t} = \Sigma_{11}^{1/2}\eta_{x,t}$ and $\varepsilon_{y,t} = \sigma_{21}\Sigma_{11}^{-1/2}\eta_{xt} + \sigma_{2.1}^{1/2}\eta_{yt}$. Then, substituting $\eta_{x,t} = \Sigma_{11}^{-1/2}\varepsilon_{x,t}$ and noting that the terms in $\frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}\left(\Delta x_t - \varepsilon_{x,t}\right)$ can be expressed in terms of the lags of Δx_t and Δy_t, whose coefficients can be absorbed into $\pi_1^*(L)$ and $\pi_2^*(L)$, the conditional ECM for Δy_t takes the form

$$\Delta y_t = \beta u_{y,t-1} + \varphi'\Delta x_t + \pi_1(L)'\Delta x_{t-1} + \pi_2(L)\Delta y_{t-1} + e_t, \tag{5}$$

where $\varphi' = \gamma' + \tau_T\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_t = \tau_T\sigma_{2.1}^{1/2}\eta_{yt}$.

Notice that given our assumptions, the stationary-dependent variable in (5) is explained by a stationary regressor Δx_t, whose influence is not dominated by $u_{y,t-1}$, even when $u_{y,t-1}$ is persistent. Next, let $\tilde{\beta}$ be the OLS estimator in the conditional VECM (5), $\tilde{t}_{\beta=\beta_0}$ be the t-test of $H_0 : \beta = \beta_0$ and $\tilde{t}_{\beta=0}$ denote the t-test of the null hypothesis $H_0 : \beta = 0$ (or $\rho = 1$).

Theorem 2. *Suppose that Assumptions A–E hold. In addition, assume that* $A_{12}(1) = 0_k$. *Then, as* $T \to \infty$,

$$T\left(\tilde{\beta} - \beta_0\right) \Rightarrow \left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}dW_2\right), \tag{6}$$

$$\tilde{t}_{\beta=\beta_0} \Rightarrow \left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}dW_2\right), \tag{7}$$

$$\tilde{t}_{\beta=0} \Rightarrow \left(\int W_{12}^2\right)^{-1/2}\left(\int W_{12}dW_2\right), \tag{8}$$

where $J_{12c}(r)$ and $W_{12}(r)$ *are defined in Theorem 1.*

Proof. See Appendix.

Theorem 2 shows that the estimator in the conditional error correction equation converges at rate T – that is, it is super-consistent – and has a non-standard distribution which is an implicit function of the long-run correlation θ^2 and c. The asymptotic behavior of the estimator $\tilde{\beta}$ bears some similarities to the limiting representations derived in other contexts; see, for example, Hansen (1995), Pesavento (2004) and Zivot (2000). As expected, it depends on the long-run correlation θ^2 and c. By controlling for Δx_t, we remove the noisy component of the error term and both $u_{y,t-1}$ and e_t have shrinking innovations variances, even though $u_{y,t-1}$ is allowed to be persistent (local to unity). Importantly, the limiting distributions of the estimator and the t-statistic do not depend on the signal-to-noise ratio through the localizing constant λ.

The assumption $A_{12}(1) = 0_k$ warrants some remarks. It is imposed in this model (see also Zivot, 2000) to ensure that $H_0 : \beta = 0$ can be interpreted as a test for cointegration in the conditional ECM (5) by ignoring the information contained in the marginal model for Δx_t. While this assumption simplifies the asymptotic representations for $T\left(\tilde{\beta}-\beta_0\right)$ and $\tilde{t}_{\beta=\beta_0}$ in Theorem 2, it should be stressed that the condition $A_{12}(1)=0_k$ is not required for establishing the limiting behavior of the estimator and the t-test of $H_0:\beta=\beta_0$ and the limiting expressions in (6) and (7) can be readily modified by relaxing this assumption. Of course, this assumption is automatically satisfied in the case of no seral correlation; that is, $A(L)=I_{k+1}$. For further discussion on the trade-off between the weak exogeneity assumption in single-equation ECM and the system-based approach to testing for cointegration, see Elliott et al. (2005, p. 36). These remarks also apply to the results in Theorem 3.

It is often the case that the VECM is defined (for predictive purposes, for instance) as

$$\Delta y_t=\beta u_{y,t-1}+\pi_1(L)'\Delta x_{t-1}+\pi_2(L)\Delta y_{t-1}+\xi_t, \tag{9}$$

where $\xi_t=\varphi'\Delta x_t+e_t$.[6] We refer to model (9) as the unconditional VECM. Let $\hat{\beta}$ denote the OLS estimator in the unconditional VECM, $\hat{t}_{\beta=\beta_0}$ be the t-test of $H_0:\beta=\beta_0$ based on the estimator $\hat{\beta}$ and $\hat{t}_{\beta=0}$ be the t-test of the null hypothesis $H_0:\beta=0$ (or $\rho=1$). We then have the following result.

Theorem 3. *Suppose that Assumptions A–E hold. In addition, assume that $A_{12}(1)=0_k$ and $\gamma\neq 0$. Then, as $T\to\infty$,*

$$\sqrt{T}\left(\hat{\beta}-\beta_0\right)\Rightarrow\lambda^{-1}\omega_{2.1}^{-1/2}(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}d\tilde{W}_1+\Lambda^*\right), \tag{10}$$

$$\hat{t}_{\beta=\beta_0}\Rightarrow\frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,xx}\gamma)^{1/2}}\left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}d\tilde{W}_1+\Lambda^*\right), \tag{11}$$

$$\hat{t}_{\beta=0}\Rightarrow\frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,xx}\gamma)^{1/2}}\left(\int W_{12c}^2\right)^{-1/2}\left(\int W_{12c}d\tilde{W}_1+\Lambda^*\right), \tag{12}$$

where $\Lambda^*=\omega_{21}^{-1/2}\lambda^{-1}(\gamma'\Omega_{11}\gamma)^{-1/2}\gamma'\Lambda_{y,x}$, $\Lambda_{y,x}=\sum_{h=1}^{\infty}\Gamma'_{h,yx}$,

$$\Gamma_h=\begin{pmatrix} E\left(v_{x,t-h}v'_{x,t}\right) & E\left(v_{x,t-h}v_{y,t}\right) \\ E\left(v_{y,t-h}v'_{x,t}\right) & E\left(v_{y,t-h}v_{y,t}\right)\end{pmatrix}=\begin{pmatrix}\Gamma_{h,xx} & \Gamma_{h,xy}\\ \Gamma_{h,yx} & \Gamma_{h,yy}\end{pmatrix},$$

$\tilde{W}_1$ is an univariate Brownian motion independent of W_2, and $J_{12c}(r)$ and $W_{12}(r)$ are defined in Theorem 1.

Proof. See Appendix.

Unlike the estimator $\tilde{\beta}$ in the conditional VECM, the estimator $\hat{\beta}$ has a slower (root-T) rate of convergence and its limiting distribution depends inversely on λ so that low values of λ make the estimator highly volatile. The asymptotic distributions of the estimator and its t-statistic are still non-standard. While they are also functionals of Brownian motions as for the estimator in the conditional VECM, there is a sharp contrast in the limiting behavior of these two estimators and their corresponding t-tests. In particular, because we are not conditioning on Δx_t, the errors in the unconditional error correction equation will be serially correlated and there will be extra parameters for the short- and long-run variances that will enter the asymptotic distribution. When there is no serial correlation, that is, $A(L) = I_{k+1}$, the limit distribution of the t-statistics in the unconditional VECM are

$$\hat{t}_{\beta=\beta_0} \Rightarrow \left(\int J_{12c}^2\right)^{-1/2} \left(\int J_{12c} d\tilde{W}_1\right) \tag{13}$$

and

$$\hat{t}_{\beta=0} \Rightarrow \left(\int W_{12}^2\right)^{-1/2} \left(\int W_{12} d\tilde{W}_1\right). \tag{14}$$

For example, when $\theta^2 = 0$ (and $W_{12}(r) = W_2(r)$), the asymptotic distribution of the t-statistic for $H_0 : \beta = 0$ in the unconditional VECM reduces to the standard normal distribution while the limit of the t-test in the conditional VECM is characterized by the Dickey–Fuller distribution.

Critical values at the 5% significance level for the limiting distributions of the t-tests $\tilde{t}_{\beta=0}$ and $\hat{t}_{\beta=0}$ when there is no serial correlation ($p = 0$) are presented in Table 1. This is the setup of our empirical example and simulation experiment where these critical values are directly applicable. They are tabulated for different values of the scalar θ^2 that characterizes the degree of endogeneity in the model and determines implicit weights assigned to the standard normal and Dickey–Fuller distribution. The critical values for the case with no deterministic terms are obtained from the asymptotic representations in Theorems 2 and 3. For comparison purposes, we only use the case when there is no serial correlation and no nuisance parameters that need to be estimated, which is also the relevant

Table 1. Asymptotic Critical Values for $\tilde{t}_{\beta=0}$ (Conditional VECM) and $\hat{t}_{\beta=0}$ (Unconditional VECM) at 5% Significance Level.

θ^2	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$	$\tilde{t}_{\beta=0}$	$\hat{t}_{\beta=0}$
	No Determ. Terms		Constant, No Trend		Constant and Trend	
0	−1.941	−1.645	−2.863	−1.645	−3.413	−1.645
0.2	−1.939	−1.819	−2.775	−2.278	−3.274	−2.544
0.3	−1.927	−1.857	−2.721	−2.403	−3.192	−2.725
0.5	−1.900	−1.902	−2.584	−2.584	−2.995	−2.995
0.7	−1.853	−1.921	−2.398	−2.721	−2.730	−3.192
0.8	−1.822	−1.932	−2.274	−2.778	−2.548	−3.280
0.9	−1.173	−1.938	−2.098	−2.826	−2.301	−3.351

Notes: Critical values are computed by simulating the asymptotic distributions with 200,000 replications, $T = 30{,}000$ and $p = 0$. For the case of no deterministic terms ("No Determ. Terms"), the critical values are obtained from the limiting distributions in Theorems 2 and 3. For the other two cases, the standard Brownian motion in the limiting distributions is replaced by its demeaned and detrended analogs.

case in our empirical application. For the cases with deterministic terms in (5) and (9) ("constant, no trend" and "constant and trend"), the standard Brownian motion is replaced by its demeaned and detrended analogs. The critical values are obtained by simulation using 200,000 replications and $T = 30,000$.

Finally, while the results in Theorems 2 and 3 assume that the cointegration vector γ is known, the corresponding asymptotic representations can also be characterized when γ is estimated by OLS and $\hat{u}_{y,t-1}$ is used in the conditional and unconditional regressions. Despite the more complex form of these asymptotic distributions, the limiting behavior of the main quantities of interest is qualitatively unchanged: $\hat{u}_{y,t}$ is still stochastically bounded, $\tilde{\beta}$ is super-consistent and has an asymptotic distribution that does not depends on λ, and $\hat{\beta}$ continues to converge at a slower (root-T) rate with a limiting representation that depends on the signal-to-noise ratio.[7]

3. SIMULATION RESULTS

To gain further understanding of the combined effect of low signal-to-noise ratio and persistent cointegration errors, and to quantify the cost of using the unconditional VECM in this context, we simulate data from a bivariate version of (1) with no serial correlation:

$$\begin{aligned} x_t &= u_{x,t} \\ y_t &= \gamma x_t + u_{y,t} \end{aligned} \tag{15}$$

and

$$\begin{pmatrix} (1-L)u_{x,t} \\ (1-\rho_T L)u_{y,t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{x,t} \\ \tau_T \varepsilon_{y,t} \end{pmatrix}$$

with $\rho_T = 1 + c/T$ and $\tau_T = \lambda/\sqrt{T}$. We set $\gamma = 1$. Several values of λ, c, θ^2 and T are considered. In each case, we estimate the conditional and unconditional VECM and compute the bias and standard error of $\tilde{\beta}$ and $\hat{\beta}$ together with the rejection rates for $\tilde{t}_{\beta=0}$ and $\hat{t}_{\beta=0}$ using 100,000 Monte Carlo replications. We first choose a value for $\lambda = 0.05$ to match the implicit values for λ in the empirical application. The parameter θ^2 is set to 0, 0.3 or 0.7. In the bivariate model, θ^2 is the square of the correlation between $\varepsilon_{x,t}$ and $\varepsilon_{y,t}$, so that we can think of these values for θ^2 as corresponding to low, medium and high degrees of endogeneity. We set c equal to 0, -5 and -10: values that correspond to ρ between 1 and 0.95 depending on the sample size. All reported tests have a nominal level of 5%.

When $c = 0$ $(\rho = 1)$ and $\theta^2 = 0$, both the conditional and unconditional VEC estimators have negligible bias and correct size, although the standard error of the unconditional estimator can be large. As the degree of endogeneity increases, the size of both tests remains close to the nominal 5% level but the estimator from the unconditional regression, $\hat{\beta}$, has a large negative bias that increases as θ^2 increases.

Table 2. Simulation Results for $\lambda = 0.05$, and Various Values of c (ρ), θ^2 and T.

	$\theta^2 = 0$			$\theta^2 = 0.3$			$\theta^2 = 0.7$		
	Bias	s.d.	*t*-test	Bias	s.d.	*t*-test	Bias	s.d.	*t*-test
$c = 0$									
$T = 200$ ($\rho = 1$)									
CVECM	−0.007	0.008	0.052	−0.005	0.007	0.053	−0.002	0.004	0.050
UVECM	0.003	2.276	0.052	−1.035	2.279	0.051	−1.583	2.281	0.051
$T = 400$ ($\rho = 1$)									
CVECM	−0.004	0.004	0.052	−0.003	0.004	0.053	0.001	0.002	0.051
UVECM	0.000	1.771	0.052	−0.806	1.774	0.051	−1.240	1.774	0.050
$c = -5$									
$T = 200$ ($\rho = 0.97$)									
CVECM	−0.009	0.018	0.368	−0.006	0.015	0.566	−0.003	0.010	0.880
UVECM	−0.05	5.002	0.053	−1.396	5.011	0.044	−2.136	5.015	0.041
$T = 400$ ($\rho = 0.99$)									
CVECM	−0.004	0.009	0.364	−0.003	0.007	0.565	−0.001	0.005	0.880
UVECM	0.014	3.554	0.052	−0.981	3.557	0.043	−1.508	3.561	0.040
$c = -10$									
$T = 200$ ($\rho = 0.95$)									
CVECM	−0.009	0.024	0.793	−0.006	0.020	0.918	−0.003	0.013	0.994
UVECM	−0.011	6.671	0.052	−1.422	6.681	0.041	−2.174	6.684	0.039
$T = 400$ ($\rho = 0.97$)									
CVECM	−0.005	0.012	0.792	−0.003	0.010	0.918	−0.001	0.007	0.994
UVECM	0.020	4.742	0.053	−1.009	4.746	0.041	−1.556	4.751	0.037

Notes: This table presents the average bias and standard deviations of the β estimates as well as the rejection probabilities of the t-test for $H_0 : \beta = 0$. The results are based on 100,000 Monte Carlo replications.

For negative values of c ($\rho < 1$), we move away from the null hypothesis. In this case, we can see that both the bias and the standard error of $\hat{\beta}$ increase. Importantly, the probability of rejecting the null hypothesis in the conditional VECM correctly increases as ρ moves away from unity and as the degree of endogeneity increases. By contrast, the rejection probability of the t-test in the unconditional VECM remains around the 5% nominal level indicating a lack of power. As Theorem 3 shows, these asymptotic results are driven by the low signal-to-noise ratio: the asymptotic distribution of the unconditional VECM estimator $\hat{\beta}$ and, in particular, its variance[8] depend on λ, while the asymptotic distribution of the conditional VECM estimator $\tilde{\beta}$ is invariant to the value of λ. This suggests that the bias of the unconditional VECM estimator and the power of its test for significance will not improve even with large samples. This is in line with the theoretical results in Theorem 3. For example, the only difference that we observe between sample sizes of 200 and 400 is that, as T gets larger, the unconditional VECM estimates tend to have slightly smaller standard errors, although they are substantially larger than the standard errors for its conditional model counterpart.

As Theorem 3 highlights, the performance of the estimators from the unconditional VEC regression is inversely related to the localizing constant λ. A small

Table 3. Simulation Results for $\lambda = 10$, and Various Values of c (ρ), θ^2 and T.

	$\theta^2 = 0$			$\theta^2 = 0.3$			$\theta^2 = 0.7$		
	Bias	s.d.	t-test	Bias	s.d.	t-test	Bias	s.d.	t-test
$c = 0$									
$T = 200$ ($\rho = 1$)									
CVECM	−0.007	0.008	0.052	−0.005	0.007	0.053	−0.002	0.004	0.050
UVECM	−0.007	0.014	0.079	−0.012	0.017	0.061	−0.015	0.019	0.054
$T = 400$ ($\rho = 1$)									
CVECM	−0.004	0.004	0.052	−0.003	0.004	0.053	−0.001	0.002	0.051
UVECM	−0.004	0.010	0.073	−0.008	0.012	0.059	−0.010	0.013	0.052
$c = -5$									
$T = 200$ ($\rho = 0.97$)									
CVECM	−0.009	0.018	0.368	−0.006	0.015	0.566	−0.003	0.010	0.880
UVECM	−0.009	0.031	0.248	−0.016	0.038	0.139	−0.020	0.041	0.109
$T = 400$ ($\rho = 0.99$)									
CVECM	−0.004	0.009	0.364	−0.003	0.007	0.565	−0.001	0.005	0.880
UVECM	−0.004	0.020	0.182	−0.009	0.024	0.110	−0.012	0.026	0.088
$c = -10$									
$T = 200$ ($\rho = 0.95$)									
CVECM	−0.009	0.024	0.793	−0.006	0.020	0.918	−0.003	0.013	0.994
UVECM	−0.009	0.041	0.390	−0.016	0.050	0.219	−0.020	0.055	0.169
$T = 400$ ($\rho = 0.97$)									
CVECM	−0.005	0.012	0.792	−0.003	0.010	0.918	−0.001	0.007	0.994
UVECM	−0.004	0.027	0.269	−0.010	0.032	0.158	−0.012	0.034	0.125

Notes: This table presents the average bias and standard deviations of the β estimates as well as the rejection probabilities of the t-test for $H_0 : \beta = 0$. The results are based on 100,000 Monte Carlo replications.

value for λ makes the signal-to-noise negligible and induces high volatility in $\hat{\beta}$. To provide further evidence of the effect of λ on the behavior of the estimators and tests of significance, Table 3 presents results for a bigger value of λ ($\lambda = 10$) that renders the signal-to-noise ratio relatively large.

Table 3 reveals that when λ is larger, the estimator based on the unconditional regression performs better with the bias and standard error of $\hat{\beta}$ being much smaller than those for a small signal-to-noise ratio ($\lambda = 0.05$). However, the conditional VECM estimator continues to be more efficient which results in non-trivial power for its t-test. In contrast, while the power of the t-test in the unconditional VECM is improved compared to the case in Table 2, it is still dominated by the conditional VECM test.

4. EMPIRICAL APPLICATION: FORWARD PREMIUM MODEL

The main motivation for the theoretical analysis developed above has been some puzzling results in the forward premium regression models that link spot currency future returns to current forward premium, defined as the difference between the forward and spot exchange rates. The main properties of the data for the BP, DM, SF and CD – all against the US dollar – are visualized in Fig. 1 in the Introduction. The left charts illustrate the extremely small signal-to-noise ratio of the forward premium regression while the right panel of charts presents the near-unit root dynamics of the spot–forward spread. These data features strongly suggest that the conventional forward premium regression attempts to explain a noisy but stationary-dependent variable with a small, but persistent regressor. Consequently, the estimator is likely to exhibit non-standard finite sample and asymptotic behavior.

The two regression specifications that we consider are based on the unconditional and conditional VECM, for a cointegrating vector $\gamma = (1, -1)'$, and take the following form

$$\Delta s_t = \alpha_U + \beta_U (s_{t-1} - f_{t-1}) + \xi_t, \tag{16}$$

$$\Delta s_t = \alpha_C + \beta_C (s_{t-1} - f_{t-1}) + \varphi_C \Delta f_t + u_t, \tag{17}$$

where s_t denotes the log spot exchange rate and f_t is its corresponding one-month log forward rate. The parameters are indexed by U and C to signify their association with the unconditional and conditional VECM, respectively. Equation (16) has been used extensively in the forward premium literature following Fama (1984),[9] but note that to be consistent with our setup and notation in the methodological section, the error correction term is defined as $(s_{t-1} - f_{t-1})$ and not as $(f_{t-1} - s_{t-1})$ as in the forward premium literature. We should also stress that our setup and hypothesis of interest differ from the tests for forward rate unbiasedness in the forward premium puzzle (for the analysis of the forward premium puzzle, see Gospodinov, 2009; Maynard & Phillips, 2001; among others). Specifically, we test $\beta = \rho - 1 = 0$ whereas the unbiasedness hypothesis is a test of $\beta_U = 1$.

The data consist of monthly observations for the four exchange rates (BP, DM, SF and CD), mentioned above, for the period January 1975–May 2006 and the Japanese yen (JY) for the period August 1978–May 2006. The monthly spot rates are constructed by taking the observation on the last business day of each month (daily mid-market observation from Datastream). One-month forward rates are constructed from end-of-the month Eurocurrency rates for the USA, UK, Germany, Japan, Canada and Switzerland obtained from Datastream, using the covered interest parity.

Table 4 presents the regression estimates and their associated Newey-West standard errors (with 12 lags) from the model specifications (16) and (17), along with the R^2 from these regressions. But the table starts by reporting some salient features of the data that justify our dual parameterization and the system approach to estimation and inference. The ratio of volatilities between the regressor $(s_{t-1} - f_{t-1})$ and the dependent variable Δs_t in the unconditional VECM is very low (close to zero) with implied values of λ ranging between 0.026 and 0.052. At the same time, the spot–forward spread $(s_{t-1} - f_{t-1})$ is a highly persistent process with AR(1) coefficients near one. The unit root test (augmented Dickey–Fuller test for a model with a drift and 12 lags) cannot reject the null of a unit root at

5% significance level. This leads to the counter-intuitive conclusion that spot and forward rates drift apart in the long run despite being visually indistinguishable from each other in Fig. 1. Overall, the combination of these two data characteristics (low signal-to-noise ratio and high persistence) highlights the importance of using limiting distributions that are explicit functions of these parameters.

As our theory and simulation results suggest, the estimate of β_U appears to be highly volatile (large standard error) and thus has a substantial probability of being far away from its implied value under the null. Given the very low signal-to-noise ratio of this regression model, the explanatory power of $(s_{t-1} - f_{t-1})$ is very low which is reflected in values of R^2 ranging between 0.005 and 0.045 for the different currencies. As Theorem 3 highlights, the precision of these estimates is inversely related to the localizing constant λ whose proximity to zero makes the signal-to-noise ratio negligibly small and induces high volatility in the OLS estimate of β_U. Since the degree of endogeneity in this model (measured by the long-run correlation θ^2) is somewhat low, the large downward biases reported in the simulations (for large θ^2) are not expected to be an issue here. Instead, the sampling behavior of the unconditional VECM estimator of β_U is driven by the negligible signal-to-noise ratio and the incompatibility between the dependent and independent variables in terms of both their scale and persistence.

By contrast, the estimates of β_C in model specification (17) are in line with those predicted by theory (one minus the implied value of the persistence parameter ρ) with significantly reduced variability. The meaningful improvement in the sampling properties of the estimator in (17) arises from "balancing" the stationary, but noisy, dependent variable with the inclusion of an additional regressor, Δf_{t+1}, with similar scale and persistence.[10] This recalibrates the signal and noise components to put them on a more equal footing.

Table 4. Estimation Results for Unconditional and Conditional VECM (16) and (17).

	BP	DM	SF	CD	JY
Dual param.					
λ	0.0402	0.0444	0.0519	0.0261	0.0407
ρ	0.9188 (0.0239)	0.9585 (0.0221)	0.9635 (0.0161)	0.8869 (0.0413)	0.8579 (0.0297)
ADF p-value	0.0740	0.3855	0.4614	0.0619	0.0750
Model (16)					
β_U	1.7395 (0.9785)	0.9848 (0.7968)	1.4326 (0.6829)	1.1368 (0.5197)	3.3247 (0.7189)
R^2	0.0141	0.0050	0.0119	0.0100	0.0450
Model (17)					
β_C	−0.0811 (0.0271)	−0.0433 (0.0211)	−0.0358 (0.0166)	−0.1098 (0.0382)	−0.1410 (0.0292)
φ_C	0.9999 (0.0025)	1.0018 (0.0019)	0.9995 (0.0020)	0.9974 (0.0026)	0.9997 (0.0023)
R^2	0.9993	0.9996	0.9996	0.9984	0.9989

Notes: OLS estimates with Newey–West standard errors with 12 lags are in parentheses. The ADF test for the null of a unit root is based on a model with a drift and 12 lags. This table reports its p-value. R^2 denotes the goodness-of-fit R^2 statistic.

The OLS estimates of β_C vary between 0.036 (Swiss franc) and 0.141 (JY) that map well within the spectrum of plausible values for the implied persistence of the forward premium. Using the critical values in Table 1 for values of θ^2 close to zero, these results lend support to the alternative hypothesis that the estimate of β_C is statistically significant (and smaller than 0); that is, the persistent parameter of the forward premium is close to but strictly less than unity.

The large values of R^2 for this model provide additional evidence that all relevant information about Δs_t is reflected in the regressors $(s_{t-1} - f_{t-1})$ and Δf_t. It also reflects the good fit of our theoretical model to the exchange rate data. The high R^2 results from the strong correlation between Δs_t and Δf_t. This, in turn, is an implication of our modeling framework. Specifically, it is implied by the near-zero variance for $s_t - f_t$. The small variance of $s_t - f_t = s_{t-1} - f_{t-1} + (\Delta s_t - \Delta f_t)$ requires that Δs_t and Δf_t are typically close in value (i.e., highly correlated).

In summary, the magnitudes of the estimates from model (16) and their tests for significance should be interpreted with caution given their highly volatile behavior arising from the low signal-to-noise ratio of the regressor. On the other hand, model (17) is statistically balanced and its estimates and statistics are characterized by more appealing sampling properties.

5. CONCLUDING REMARKS

In this chapter, we proposed and studied a model of a non-stationary levels relationship in which the residual follows a local-unity process with a shrinking innovation variance. This setup captures empirical applications, such as spot and forward exchange rate and commodity prices, where the levels relationship appears tight despite a persistent, yet small residual. The asymptotic theory that we develop in the chapter offers some interesting insights. The limiting behavior of the levels regression lies in between the cointegrating and spurious regression cases. The estimated coefficients remain consistent, but not super-consistent, and their corresponding t-tests diverge with the sample size.

We also analyzed the VEC specifications of this model. Unfortunately, the unconditional VECM is characterized by an imbalance between a small but persistent error correction term and a large stationary component in its error term. This imbalance is reflected in a low signal-to-noise ratio, resulting in highly variable coefficient estimates. Conversely, the conditional VEC specification addresses this imbalance by explicitly controlling for the high variance component of the residual in the unconditional VECM. This is manifested in a higher signal-to-noise ratio and a super-consistent error correction coefficient estimate. The asymptotic distribution is non-standard, but the t-test depends on only a single endogeneity term, which can be consistently estimated and used to adjust the critical value.

Our simulations confirmed the superiority of the conditional VEC specification. While the unconditional and conditional VECM perform similarly in the standard cointegration setting, the relative performance of the unconditional VECM significantly deteriorates when error variance of the levels residual is small and persistent. By contrast, the conditional VECM continues to exhibit excellent size and power properties.

We illustrated the practical relevance of our theoretical results in the context of spot–forward exchange rate regressions. Our analytical framework rationalizes the otherwise conflicting observations that (i) the spot and forward rates move closely together and (ii) their difference, the forward premium, is highly persistent. Common spot return forward premium regressions correspond to the unconditional VECM. As predicted by the theory, this regression is imbalanced both in terms of its persistence and in terms of the magnitude of its innovation variance. Not surprisingly, the resulting estimates are imprecise with large standard errors. By contrast, the conditional VECM produces more precise estimates with tighter standard estimates. Using the conditional VECM, we can reject the hypothesis of an exact unit root in the spot–forward spread, providing additional support for the tight levels relationship observed between spot and forward exchange rates.

NOTES

1. Gourieroux and Jasiak (2020) employ a model similar to ours in order to explain the long-run predictability puzzle, whereas we focus primarily on explaining puzzles involving spot and forward rates that move very closely together, and yet are barely cointegrated according to traditional cointegration tests.
2. Müller and Watson (2008, 2018) provide a comprehensive analytical framework for studying low-frequency movements and co-movements in economic and financial time series. In this chapter, we consider the possibility that the low-frequency component is not readily detectable, as the explosive behavior arising from its high persistence is offset by its asymptotically vanishing variability.
3. For notational convenience, we suppress the dependence of u_t, x_t and y_t on T.
4. The proof of this result is provided in the Appendix and follows closely Pesavento (2004).
5. For ease of exposition, we follow the usual convention and suppress the (r) from the Brownian motion terms. All integrals are intended to be between 0 and 1, unless stated otherwise.
6. Strictly speaking, the parameters in model (9) should be denoted differently than those in the conditional ECM (5) since the unconditional model (9) is misspecified as it omits the term $\varphi\Delta x_t$. For notational simplicity, we do not index the parameters in the conditional and unconditional specifications in the theoretical part but we do so in the empirical application.
7. A sketch of this result is present in the Proof of Theorem 1. The limiting results for the t-tests when γ is estimated follow directly.
8. See the Appendix for details on the asymptotic variance of the estimators.
9. The vast majority of this literature does not include lags of Δs_t and Δf_t and we follow this tradition. We have also tried estimating the model with lagged differences included but found the lags to be insignificant (at 5% significance level) for BP, DM, SF and CD. Even in the JY regression where they are only borderline significant, the coefficients on the lagged Δs_t and Δf_t offset each other (with opposite signs and similar magnitude so that the sum of the coefficients is near zero). For this reason, we decided to maintain the specification of the forward premium regression that is commonly used in practice (with no lags of Δs_t and Δf_t).
10. The traditional definition of an unbalanced regression is that the regressand and the regressor are of different orders of integration. In our context, there can also be imbalance between the innovation variances, when one is fixed and the other shrinking, or between the

overall scale or magnitude of the regressor and regressand, which depends jointly on both the integration order and innovation variance. Given these multiple notions of balance, we avoid the generic use of the term "balanced/unbalanced" regression and instead clarify the notion of "balance/unbalance" intended at each point in which we use the term.

11. To simplify the notation, we will assume that the summations are over all the available data which will depend on the lag length.

REFERENCES

Bansal, R., & Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, *59*, 1447–1950.

Boswijk, H. P. (1994). Testing for an unstable root in conditional and structural error correction models. *Journal of Econometrics*, *63*, 37–60.

Deng, A. (2014). Understanding spurious regression in financial economics. *Journal of Financial Econometrics*, *12*, 122–150.

Elliott, G., Jansson, M., & Pesavento, E. (2005). Optimal power for testing potential cointegrating vectors with known parameters for nonstationarity. *Journal of Business and Economic Statistics*, *23*, 34–48.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error-correction: Representation, estimation and testing. *Econometrica*, *55*, 251–276.

Fama, E. (1984). Forward and spot exchange rates. *Journal of Monetary Economics*, *14*, 319–338.

Hansen, B. E. (1995). Rethinking the univariate approach to unit root tests: How to use covariates to increase power. *Econometric Theory*, *11*, 1148–1171.

Gospodinov, N. (2009). A new look at the forward premium puzzle. *Journal of Financial Econometrics*, *7*, 312–338.

Gourieroux, C., & Jasiak, J. (2020). *Inference for noisy long run component process*. York University. Unpublished manuscript.

Jansson, M., & Haldrup, N. (2002). Regression theory for nearly cointegrated time series. *Econometric Theory*, *18*, 1309–1335.

Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics*, *52*, 389–402.

Maynard, A., & Phillips, P. C. B. (2001). Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics*, *16*, 671–708.

Moon, R., Rubia, A., & Valkanov, R. (2004). *Long-horizon regressions when the predictor is slowly varying*. University of Southern California. Unpublished manuscript.

Müller, U. K., & Watson, M. W. (2008). Testing models of low-frequency variability. *Econometrica*, *76*, 979–1016.

Müller, U. K., & Watson, M. W. (2018). Long-run covariability. *Econometrica*, *86*, 775–804.

Ng, S., & Perron, P. (1997). Estimation and inference in nearly unbalanced nearly cointegrated systems. *Journal of Econometrics*, *79*, 53–81.

Park, J. Y. (1992). Canonical cointegrating regressions. *Econometrica*, *60*, 119–143.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*, 95–131.

Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. *Journal of Econometrics*, *122*, 349–384.

Phillips, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica*, *59*, 283–306.

Phillips, P. C. B., & Lee, J. H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, *177*, 250–264.

Stock, J. H., & Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, *61*, 783–820.

Torous, W., & Valkanov, R. (2000). *Boundaries of predictability: Noisy predictive regressions*. UCLA. Unpublished manuscript.

Zivot, E. (2000). The power of single equation tests for cointegration when the cointegrating vector is prespecified. *Econometric Theory*, *16*, 407–439.

APPENDIX: PROOFS OF MAIN RESULTS

A.1. PRELIMINARY LEMMA

Lemma A1. *Under Assumptions A–E, we have that*

1. $\Omega_{11}^{-1/2}\frac{1}{T^2}\sum_{t=1}^{T}x_t x_t'\Omega_{11}^{-1/2'} \Rightarrow \int W_1 W_1',$ (1)

2. $\Omega_{11}^{-1/2}\omega_{2.1}^{-1/2}\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}x_t u_{y,t} \Rightarrow \lambda\int W_1 J_{12c},$ (2)

3. $\omega_{2.1}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{y,t}^2 \Rightarrow \lambda^2\int J_{12c}^2,$ (3)

4. $\omega_{2.1}^{-1/2}\frac{1}{T^{\frac{1}{2}}}\sum_{t=1}^{T}u_{y,t-1}(\Sigma^{-1/2}\varepsilon_t) \Rightarrow \lambda\int J_{12c}dW,$ (4)

5. $\Omega_{11}^{-1/2}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}(\Sigma^{-1/2}\varepsilon_t) \Rightarrow \int W_1 dW,$ (5)

where $J_{12c}(r) = W_{12}(r) + c\int_0^r e^{(r-s)c}W_{12}(s)ds$, $W_{12}(r) = \sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r) + W_2(r)$, and $\tilde{W}_1(r)$ *is a univariate standard Brownian motion.*

Proof. Under our assumptions, we have $\frac{1}{\sqrt{T}}\sum_{s=1}^{[Tr]}v_s \Rightarrow \Omega^{1/2}W(r)$ which implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}v_{x,t} \Rightarrow \Omega_{11}^{1/2}W_1(r)$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}v_{y,t} \Rightarrow \omega_{21}\Omega_{11}^{-1/2}W_1(r) + \omega_{2.1}^{1/2}W_2(r).$$

Recall that $\theta^2 = \delta'\delta$, where $\delta = \Omega_{11}^{-1/2}\omega_{12}\omega_{22}^{-1/2}$ is a vector containing the bivariate zero frequency correlations of each element of $v_{x,t}$ with $v_{y,t}$. Define $\bar{\delta}' = \omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}$ so that $\bar{\delta}'\bar{\delta} = \frac{\theta^2}{1-\theta^2}$. We can then see that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\bar{\delta}'\Omega_{11}^{-1/2}v_{x,t} \Rightarrow \sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r),$$

where $\tilde{W}_1$ is an univariate standard Brownian motion independent of W_2 and

$$\frac{1}{\sqrt{Tr}}\sum_{t=1}^{[Tr]}v_{y,t} \Rightarrow \omega_{2.1}^{1/2}\left[\omega_{2.1}^{-1/2}\omega_{21}\Omega_{11}^{-1/2}W_1(r) + W_2(r)\right] = \omega_{2.1}^{1/2}\left[\sqrt{\frac{\theta^2}{1-\theta^2}}\tilde{W}_1(r) + W_2(r)\right].$$

Using these limiting expressions, all results in Lemma A1 follow from FCLT and Continuous Mapping Theorem.

A.2. PROOF OF THEOREM 1

The results follow directly from Lemma A1 and the fact that

$$\sqrt{T}\left(\hat{\gamma}-\gamma\right)=\left(\frac{1}{T^2}\sum_{t=1}^{T}x_t x_t'\right)^{-1}\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T}x_t u_{y,t}\right)$$

and

$$T\cdot\mathrm{SE}(\hat{\gamma})=\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}x_t x_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{y,t}^2\right)\right]^{1/2},$$

where $\hat{u}_{y,t}=u_{yt}-\left(\hat{\gamma}-\gamma\right)'x_t$. Notice that

$$\omega_{2.1}^{-1/2}\hat{u}_{y,t}=\omega_{2.1}^{-1/2}u_{yt}-\omega_{2.1}^{-1/2}\sqrt{T}\left(\hat{\gamma}-\gamma\right)\frac{x_t}{\sqrt{T}}$$

and $\hat{u}_{y,t}$ is also $O_p(1)$, so that

$$\begin{aligned}\omega_{2.1}^{-1/2}\hat{u}_{y,t} &\Rightarrow \lambda J_{12c}-\omega_{2.1}^{-1/2}\lambda\left(\int W_1 J_{12c}\right)'\left(\int W_1 W_1'\right)^{-1}\omega_{2.1}^{1/2}\Omega_{11}^{-1/2}\Omega_{11}^{1/2}W_1\\ &= \lambda J_{12c}-\lambda\left(\int W_1 J_{12c}\right)'\left(\int W_1 W_1'\right)^{-1}W_1=\lambda\tilde{J}_{12c}^2.\end{aligned}$$

Therefore,

$$\omega_{2.1}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{y,t}^2\right)\Rightarrow\lambda^2\int\tilde{J}_{12c}^2.$$

A.3. PROOF OF THEOREM 2

To simplify the intuition of the proof, we assume no deterministic terms. Recall our conditional VECM (5)

$$\Delta y_t=\beta u_{y,t-1}+\varphi'\Delta x_t+\pi_1(L)'\Delta x_{t-1}+\pi_2\left(L\right)\Delta y_{t-1}+e_t.$$

This can be written in a compact form as

$$\Delta y_t=X_t'\Pi+e_t,$$

where

$$\Pi=\begin{pmatrix}\beta & \varphi' & \pi_{11}' & \cdots & \pi_{1p}' & \pi_{21} & \cdots & \pi_{2p}\end{pmatrix}',$$

$$X_t'=\begin{pmatrix}u_{y,t-1} & \Delta x_t' & \Delta x_{t-1}' & \cdots & \Delta x_{t-p}' & \Delta y_{t-1} & \cdots & \Delta y_{t-p}\end{pmatrix},$$

with $\beta = a_{22}(1)(\rho - 1)$, $\varphi' = \gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_t = \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$. Defining $\hat{\Pi}$ as the OLS estimator of Π, we have[11]

$$T\left(\hat{\Pi} - \Pi\right) = \left(\frac{1}{T}\sum X_t' X_t\right)^{-1}\left(\sum X_t' e_t\right).$$

Then, invoking the limiting results in Lemma A1, we have

$$\frac{1}{T}\sum u_{y,t-1}^2 \Rightarrow \omega_{2.1}\lambda^2 \int J_{12c}^2(r),$$

$$\begin{aligned}\sum u_{y,t-1}e_t &= \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\sum u_{y,t-1}\eta_{yt} \\ &\Rightarrow \omega_{2.1}^{1/2}\sigma_{2.1}^{1/2}\lambda^2 \int J_{12c}\,dW_2,\end{aligned}$$

$$\frac{1}{T}\sum u_{y,t-1}\Delta x_{t-i} \to 0$$

for $i = 0, 1, \ldots, p-1$ since $\frac{1}{\sqrt{T}}\sum u_{y,t-1}\Delta x_{t-1}$ is $O_p(1)$, and

$$\frac{1}{T}\sum u_{y,t-1}\Delta y_{t-i} = \gamma' \frac{1}{T}\sum u_{y,t-1}\Delta y_{t-i} + c\frac{1}{T^2}\sum u_{y,t-1}^2 + \frac{\lambda}{T^{3/2}}\sum u_{y,t-1}v_{yt-i} \to 0,$$

where $\to$ denotes convergence in probability. The result in (6) follows directly from the asymptotic block diagonality of $\left(\frac{1}{T}\sum X_t' X_t\right)^{-1}$. Furthermore, note that $\sum \Delta x_{t-i}e_t = \lambda\sigma_{2.1}^{1/2}\frac{1}{\sqrt{T}}\sum \Delta x_{t-i}\eta_{y,t}$ and $\sum \Delta y_{t-i}e_t = \lambda\sigma_{2.1}^{1/2}\frac{1}{\sqrt{T}}\sum \Delta y_{t-i}\eta_{y,t}$ are $O_p(1)$ so that $T(\tilde{\varphi} - \varphi)$, $T(\hat{\pi}_{2i} - \pi_{2i})$ and $T(\hat{\pi}_{1i} - \pi_{1i})$ will also be $O_p(1)$.

For the variance of the estimator $\tilde{\beta}$

$$T^2 \cdot \mathrm{Var}\left(\tilde{\beta}\right) = \left(\frac{1}{T^2}\sum u_{y,t-1}^2\right)^{-1}\left(\sum \hat{e}_t^2\right),$$

we have from Lemma A1 that

$$\frac{1}{T}\sum_{t=1}^{[T\cdot]} u_{y,t}^2 \Rightarrow \lambda^2\omega_{2.1}\int J_{12c}^2.$$

We can write

$$\hat{e}_t = \Delta y_t - X_t'\hat{\Pi} = e_t - \frac{X_t'}{T}T(\hat{\Pi} - \Pi).$$

Since $T(\hat{\Pi}-\Pi)$ are $O_p(1)$, $\hat{e}_t$ will converge in the limit to $e_t = \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$ and

$$\sum \hat{e}_t^2 \to \lambda^2\sigma_{2.1}$$

since $\frac{1}{T}\sum \eta_{y,t}^2 \to 1$. Therefore,

$$T\cdot\mathrm{SE}\left(\tilde{\beta}\right) \Rightarrow \omega_{2.1}^{-1/2}\sigma_{2.1}^{1/2}\left(\int J_{12c}^2\right)^{-1/2}.$$

Finally, using that $\beta = \rho - 1 = \frac{c}{T}$, the t-test for $H_0 : \beta = \beta_0$ has the following limiting representation

$$\tilde{t}_{\beta=\beta_0} = \frac{T\left(\tilde{\beta}-\beta_0\right)}{T\cdot\mathrm{SE}(\tilde{\beta})} \Rightarrow \frac{\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}dW_2\right)}{\left(\int J_{12c}^2\right)^{-1/2}}.$$

For the test of the hypothesis $H_0 : \beta = 0$, we have $c = 0$ and $J_{12c} = W_{12}$ in the above limiting expression.

A.4. PROOF OF THEOREM 3

The unconditional VECM is given by

$$\Delta y_t = \beta u_{y,t-1} + \pi_1(L)'\Delta x_{t-1} + \pi_2(L)\Delta y_{t-1} + \xi_t.$$

This can be written in a more compact form as $\Delta y_t = X_t'\Pi + \xi_t$, where the notation matches that of the proof of Theorem 2, except that

$$X_t' = \begin{pmatrix} u_{y,t-1} & \Delta x_{t-1}' & \cdots & \Delta x_{t-p}' & \Delta y_{t-1} & \cdots & \Delta y_{t-p} \end{pmatrix}$$

is redefined to omit Δx_t, $\Pi = \begin{pmatrix} \beta & \pi_{11}' & \cdots & \pi_{1p}' & \pi_{21} & \cdots & \pi_{2p} \end{pmatrix}'$, and the new error $\xi_t = \varphi'\Delta x_t + e_t$ thus includes the omitted $\varphi'\Delta x_t$. Substituting for $\varphi' = \gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}$ and $e_t = \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}$, we have

$$\xi_t = \left(\gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}\right)\Delta x_t + \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt}.$$

The OLS estimator now converges at rate $\sqrt{T}$ since

$$T\left(\hat{\Pi}-\Pi\right) = \left(\frac{1}{T}\sum X_t'X_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum X_t'\xi_t\right)$$

where

$$\frac{1}{\sqrt{T}}\sum u_{y,t-1}\xi_t = \frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma'\Delta x_t + \lambda\sigma_{21}\Sigma_{11}^{-1/2}\frac{1}{T}\sum u_{y,t-1}\Delta x_t + \sigma_{2.1}^{1/2}\lambda\frac{1}{T}\sum u_{y,t-1}\eta_{yt}.$$

The last two terms converge to zero while

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma'\Delta x_t = \frac{1}{\sqrt{T}}\sum u_{y,t-1}\gamma' v_{xt} \quad &\Rightarrow \quad \omega_{2.1}^{1/2}\lambda(\gamma'\Omega_{11}\gamma)^{1/2}\int J_{12c}\,d\tilde{W}_1 + \gamma'\Lambda_{y,x} \\ &= \quad \omega_{2.1}^{1/2}\lambda(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}\,d\tilde{W}_1 + \Lambda *\right).\end{aligned}$$

where $\Lambda_{y,x}$ and $\Lambda *$ are defined in Theorem 3. Noting that $\left(\frac{1}{T}\sum X_t' X_t\right)^{-1}$ is again asymptotically block diagonal and $\frac{1}{T}\sum u_{y,t-1}^2 \Rightarrow \omega_{2\cdot 1}\lambda^2\int J_{12c}^2(r)$, we have

$$\begin{aligned}\sqrt{T}\left(\hat{\beta}-\beta_0\right) \quad &= \quad \left(\frac{1}{T}\sum u_{y,t-1}^2\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum u_{y,t-1}\xi_t\right) \\ &\Rightarrow \quad \lambda^{-1}\omega_{2.1}^{-1/2}(\gamma'\Omega_{11}\gamma)^{1/2}\left(\int J_{12c}^2\right)^{-1}\left(\int J_{12c}\,d\tilde{W}_1 + \Lambda *\right).\end{aligned}$$

For the variance of the estimator, we proceed similarly as in the proof in Theorem 2. Since

$$\hat{\xi}_t = \Delta y_t - X_t'\hat{\Pi} = \xi_t - \frac{X_t'}{\sqrt{T}}\sqrt{T}\left(\hat{\Pi}-\Pi\right),$$

where $\xi_t = \left(\gamma' + \frac{\lambda}{\sqrt{T}}\sigma_{21}\Sigma_{11}^{-1/2}\right)\Delta x_t + \frac{\lambda}{\sqrt{T}}\sigma_{2.1}^{1/2}\eta_{yt} = \gamma'\Delta x_t + o_p(1)$, we have

$$\frac{1}{T}\hat{\xi}'\hat{\xi} = \frac{1}{T}\xi'\xi + o_p(1) = \frac{1}{T}\sum \gamma' v_{x,t}v_{x,t}'\gamma + o_p(1) \to \gamma'\Gamma_{0,xx}\gamma.$$

The variance of the estimator can then be expressed as

$$T\cdot\mathrm{Var}\left(\hat{\beta}\right) = \left(\frac{1}{T^2}\sum u_{y,t-1}^2\right)^{-1}\left(\frac{1}{T}\sum \hat{\xi}_t^2\right) \Rightarrow \gamma'\Gamma_{0,x}\gamma'\left(\omega_{2.1}\lambda^2\int J_{12c}^2\right)^{-1}.$$

Thus,

$$\hat{t}_{\beta=\beta_o} = \frac{\sqrt{T}\left(\hat{\beta}-\beta_0\right)}{\sqrt{T}\cdot\mathrm{SE}(\hat{\beta})} \Rightarrow \frac{(\gamma'\Omega_{11}\gamma)^{1/2}}{(\gamma'\Gamma_{0,x}\gamma)^{1/2}}\left(\int J_{12c}^2\right)^{-1/2}\left(\int J_{12c}\,d\tilde{W}_1 + \Lambda *\right).$$

As before, when $\beta = 0$, $c = 0$ and $J_{12c} = W_{12}$.

CHAPTER 11

SOME EXTENSIONS OF ASYMPTOTIC *F* AND *t* THEORY IN NONSTATIONARY REGRESSIONS

Yixiao Sun[a]

[a]*Department of Economics, University of California, San Diego, CA, USA*

ABSTRACT

The author develops and extends the asymptotic F- and t-test theory in linear regression models where the regressors could be deterministic trends, unit-root processes, near-unit-root processes, among others. The author considers both the exogenous case where the regressors and the regression error are independent and the endogenous case where they are correlated. In the former case, the author designs a new set of basis functions that are invariant to the parameter estimation uncertainty and uses them to construct a new series long-run variance estimator. The author shows that the F-test version of the Wald statistic and the t-statistic are asymptotically F and t distributed, respectively. In the latter case, the author shows that the asymptotic F and t theory is still possible, but one has to develop it in a pseudo-frequency domain. The F and t approximations are more accurate than the more commonly used chi-squared and normal approximations. The resulting F and t tests are also easy to implement – they can be implemented in exactly the same way as the F and t tests in a classical normal linear regression.

Keywords: *F*-distribution; fixed-smoothing asymptotics; heteroscedasticity and autocorrelation; series long-run variance estimator; nonstationary process; *t*-distribution; transformed and augmented OLS; unit root

JEL classifications: C12; C13; C32

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 319–347

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A011

1. INTRODUCTION

This chapter considers time-series regressions in a nonstationary framework. The regressors can be deterministic trends, unit-root processes, or near-unit-root processes while the regression error is stationary but with an unknown autocorrelation function. The regressions we consider include trend regressions, cointegration regressions, and predictive regressions as special cases. For all these regressions, a main challenge for statistical inference is to account for the nonparametric autocorrelation in the regression error and the possible nonparametric correlation between the innovation of the regressor process and the regressor error. Standard practice is to estimate the unknown autocorrelation and correlation using a nonparametric kernel method but ignore the nonparametric estimation error for the convenience of statistical inference. However, by its nonparametric nature, the nonparametric estimation error can be large in finite samples. As a result, ignoring the estimation error can lead to highly unreliable inferences.

Recent literature has developed the fixed-smoothing asymptotics, an alternative type of asymptotics, to capture the nonparametric estimation error. The fixed-smoothing asymptotic theory, which includes, as a special case, the fixed-b asymptotic theory of Kiefer and Vogelsang (2002a, 2002b, 2005), was originally developed in the time-series setting. It has been extended to accommodate spatial data, spatial and temporal data, panel data, and clustered data. See Sun (2018) for a recent discussion. However, the fixed-smoothing asymptotic distributions are often nonstandard and thus not very convenient to use.

To obtain more accurate but at the same time more convenient fixed-smoothing approximations, we can use the series approach to correct the potential endogenous bias and robustify the inference. The underlying series variance estimator has a long history. A classical example is the average periodograms estimator, which involves taking a simple average of the first few periodograms. An appealing feature of the series approach is that we have the freedom to choose and design the basis functions so that the fixed-smoothing asymptotic distributions become standard F- and t-distributions. See, for example, Sun (2011) for trend regression, Müller (2007), Sun (2013, 2014a), and Lazarus et al. (2021) in the first-step Generalized Methods of Moments (GMM) or the Ordinary Least Squares (OLS) setting, and Sun (2014b), Hwang and Sun (2017), and Martínez-Iriarte et al. (2022) in the two-step GMM setting, Sun and Kim (2015) and Liu and Sun (2019) for spatial and panel data settings (difference-in-differences regressions). More recently, adopting the framework of Chang et al. (2018), Pellatt and Sun (2022) develops the asymptotic F theory in a continuous time setting. Earlier papers along this line of research include Phillips (2005) and Sun (2006).

The aim of this chapter is to review the asymptotic F and t theory in the nonstationary framework and extend it to cover the nonstationary cases that the theory is currently lacking. Section 2 considers the exogeneous case where the regressors follow either a deterministic trend function or a stochastic trend. It introduces a new idea to design the bases for series variance estimation. This approach involves projecting any given candidate bases onto the orthogonal complement of the column space spanned by the regressors. The projection ensures that the new series variance estimator is invariant to the parameter estimation error. After proper normalization, the projected bases are orthonormal. This enables us to establish that the associated Wald and t statistics are asymptotically F and t distributed, respectively. Section 3 examines the case with endogenous stochastic trends. The regressors can be unit-root or near-unit-root processes but are correlated with the regression error. Here we follow Phillips (2014) and Hwang and Sun (2018) and cast the regression as a low-frequency instrumental variable regression. Effectively, we convert a highly nonstandard inference problem into a standard inference problem in a classical normal linear regression. The asymptotic F and t theory then follows naturally. Section 4 presents a simulation study that demonstrates the higher size accuracy of the proposed F-test in a cointegration regression with exogenous regressors similar to that considered in Phillips and Park (1988). The last section concludes.

2. DETERMINISTIC AND EXOGENOUS CASES

Consider the regression model:

$$Y_t = X_t'\beta_0 + u_t, \quad t = 1,2,\ldots,T,$$

where $X_t \in \mathbb{R}^d$ for $d \geq 1$ is either a deterministic trend process, a unit-root process, or a near-unit-root process, and $\{u_t\}$ is a stationary zero-mean process that is independent of $\{X_t\}$. An intercept can be included as the first element of X_t.

For some $p \times d$ matrix $R = (R(i,j))$ and $p \times 1$ vector r, we are interested in testing

$$H_0 : R\beta_0 = r \text{ against } H_1 : R\beta_0 \neq r.$$

When $p = 1$, we may be interested in testing the null against a one-sided alternative, say, testing $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 > r$.

Based on the observations $\{X_t, Y_t\}_{t=1}^T$, we estimate β_0 by the OLS estimator:

$$\hat{\beta} = \left(\sum_{t=1}^T X_t X_t'\right)^{-1} \sum_{t=1}^T X_t Y_t.$$

For many regression models we consider here, the OLS estimator is asymptotically as efficient as the Generalized Least Squares (GLS) estimator. These include polynomial trend regressions and cointegration regressions. For the former case, see Grenander and Rosenblatt (1957), and for the latter case, see Phillips and Park (1988), which also provides general discussions on the reason for the asymptotic equivalence between OLS and GLS.

Assumption 1. (i) *For some diagonal matrix D_T, we have*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[\mathrm{Tr}]} u_t \\ D_T^{-1} X_{[\mathrm{Tr}]} \end{pmatrix} \Rightarrow \begin{pmatrix} B_u(r) \\ B_x(r) \end{pmatrix} := \begin{pmatrix} \omega_u W_u(r) \\ B_x(r) \end{pmatrix},$$

where $B_x(\cdot)$ is either a deterministic function or a stochastic process and $W_u(\cdot)$ is a standard Brownian motion that is independent of $B_x(\cdot)$.
(ii) $\int_0^1 B_x(r) B_x(r)' dr$ is of full rank d almost surely.

Assumption 1 requires that $B_x(r)$ and $W_u(r)$ be independent. When $B_x(r)$ is not random, this holds trivially. When $B_x(r)$ is random, this assumption amounts to assuming that the long-run correlation between $\{u_t\}$ and $\{\sqrt{T} D_T^{-1} \Delta X_t\}$ converges to zero. The latter condition holds if X_t is strictly exogenous in the sense that $\{X_t : t = 1, \ldots, T\}$ is independent of $\{u_t : t = 1, \ldots, T\}$. We consider the strictly exogenous case in this section and defer the endogenous case to the next section.

We now provide a few examples where Assumption 1 holds. These examples also show that our framework accommodates different types of regressions.

Example 1. ***Deterministic trend.*** *Let $X_t = (1, t, t^2)'$, in which case $D_T = \text{diag}(1, T, T^2)$ and $B_x(r) = (1, r, r^2)'$. One can also consider other types of trend functions.*

Example 2. ***Unit-root process.*** *Let $X_t = (1, \tilde{X}_t')'$ and $\tilde{X}_t$ be a unit-root process:*

$$\tilde{X}_t = \tilde{X}_{t-1} + u_{x,t} \quad \text{for } t = 1, 2, \ldots, T,$$

where $\tilde{X}_0 = o_p(\sqrt{T})$ and $u_{x,t}$ is a stationary zero-mean process. In this case,

$$D_T = diag\left(1, \sqrt{T}, \ldots, \sqrt{T}\right) \quad and \quad B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} W_x(r) \end{pmatrix},$$

where $W_x(\cdot)$ is a standard Brownian motion process, Ω_{xx} is the long-run variance of $\{u_{x,t} := \Delta \tilde{X}_t\}$ and $\Omega_{xx}^{1/2}$ is the unique and symmetric matrix square root of Ω_{xx} such that $\Omega_{xx}^{1/2} \left(\Omega_{xx}^{1/2}\right)' = \Omega_{xx}$.

Example 3. ***Near-unit-root process.*** *Let $X_t = (1, \tilde{X}_t')'$ and $\tilde{X}_t$ be a near-unit-root process:*

$$\tilde{X}_t = \left(1 - \frac{c}{T}\right) \tilde{X}_{t-1} + u_{x,t} \quad for \quad t = 1, 2, \ldots, T,$$

where $c > 0$, $\tilde{X}_0 = o_p(\sqrt{T})$, and $u_{x,t}$ is a stationary zero-mean process. In this case,

$$D_T = diag\left(1, \sqrt{T}, \ldots, \sqrt{T}\right) and \ B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} J_{c,x}(r) \end{pmatrix},$$

where $J_{c,x}(\cdot)$ is the Ornstein–Uhlenbeck (OU) process defined by

$$dJ_{c,x}(r) = -cJ_{c,x}(r) dr + dW_x(r)$$

with $J_{c,x}(0) = 0$. That is, $J_{c,x}(r) = \int_0^r e^{-c(r-s)} dW_x(s)$.

Example 4. ***Structural break.*** *Let* $X_t' = \left[X_t^{\circ\prime} 1(t \le \lambda T), X_t^{\circ\prime} 1(t > \lambda T) \right]$ *and* $\beta_0 = \left(\beta_{10}', \beta_{20}'\right)'$ *so that*

$$Y = X_t'\beta_0 + u_t = X_t^{\circ\prime} 1(t \le \lambda T)\beta_{10} + X_t^{\circ\prime} 1(t > \lambda T)\beta_{20} + u_t.$$

This model allows for a structural break in the linear relationship between Y_t *and* X_t°. *The possible break takes place at time* $t = \lambda T$, *where, for convenience,* λT *is assumed to be an integer. Assuming that*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[\mathrm{Tr}]} u_t \\ D_T^{-1} X_{[\mathrm{Tr}]}^{\circ} \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_u W_u(r) \\ B_x^{\circ}(r) \end{pmatrix}$$

for some D_T *and stochastic processes* $W_u(\cdot)$ *and* $B_x^{\circ}(\cdot)$, *we have*

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[\mathrm{Tr}]} u_t \\ D_T^{-1} X_{[\mathrm{Tr}]} \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_u W_u(r) \\ B_x(r) \end{pmatrix},$$

where

$$B_x(r) = \begin{pmatrix} B_x^{\circ}(r) 1\{r \le \lambda\} \\ B_x^{\circ}(r) 1\{r > \lambda\} \end{pmatrix}.$$

Example 5. ***High-order integrated process.*** *Let* $X_t = (1, \tilde{X}_t')'$ *and* $\tilde{X}_t$ *be an I(2) process:*

$$(1-L)^2 \tilde{X}_t = u_{x,t} \quad \text{for} \quad t = 1, 2, \ldots, T,$$

where L is the lag operator, $\tilde{X}_0 = o_p\left(\sqrt{T}\right)$, *and* $u_{x,t}$ *is a stationary zero-mean process. In this case,*

$$D_T = diag(1, T^{3/2}, \ldots, T^{3/2}) \;\; and \;\; B_x(r) = \begin{pmatrix} 1 \\ \Omega_{xx}^{1/2} \int_0^r W_x(s)\, ds \end{pmatrix},$$

where $W_x(\cdot)$ *is a standard Brownian motion process.*

In the above examples, the nonstationary regressors are of the same type. In principle, our setting can also accommodate regressors of different types given in these examples. The asymptotic *F* and *t* theory holds as long as Assumption 1 is satisfied.

Under Assumption 1, we have

$$\begin{aligned}\sqrt{T}D_T(\hat{\beta}-\beta) &= \left(\frac{1}{T}\sum_{t=1}^{T}(D_T^{-1}X_t)(D_T^{-1}X_t)'\right)^{-1}\left(\sum_{t=1}^{T}(D_T^{-1}X_t)\frac{u_t}{\sqrt{T}}\right)\\ &\Rightarrow \omega_u\left[\int_0^1 B_x(r)B_x(r)'\,dr\right]^{-1}\int_0^1 B_x(r)dW_u(r).\end{aligned}$$

Different elements of $\hat{\beta}$ may have different rates of convergence. To find the asymptotic distribution of $R(\hat{\beta}-\beta_0)$, we need to find the slowest rate of convergence among the elements of $\hat{\beta}$ that are involved in each restriction. More specifically, for the i-th restriction $R(i,\cdot)\beta = r_i$ where $R(i,\cdot)$ is the i-th row of R and r_i is the i-th element of r, we define the set

$$\mathcal{S}_i := \{j: \text{ for } j\in\{1,2,\ldots,d\} \text{ such that } R(i,j)\neq 0\},$$

which consists of the indices of the coefficients that appear in the i-th restriction. The rate of convergence of $R(i,\cdot)\hat{\beta}$ is given by $\sqrt{T}\min_{j\in\mathcal{S}_i} D_T(j,j)$. Let

$$\tilde{D}_T = \operatorname{diag}\left(\min_{j\in\mathcal{S}_1} D_T(j,j),\ldots,\min_{j\in\mathcal{S}_p} D_T(j,j)\right),$$

which is a $p\times p$ diagonal matrix. Then $\lim_{T\to\infty}\tilde{D}_T R D_T^{-1} = A$ for a matrix $A\in\mathbb{R}^{p\times d}$ whose (i,j)-th element $A(i,j)$ is equal to

$$A(i,j) = \lim_{T\to\infty}\tilde{D}_T(i,i)R(i,j)/D_T(j,j).$$

That is, A is the same as R after we zero out the elements in each row for which the corresponding coefficients can be estimated at a faster rate. So, under the null H_0,

$$\begin{aligned}\tilde{D}_T\sqrt{T}\left(R\hat{\beta}-r\right) &= \tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\\ &= \left(\tilde{D}_T R D_T^{-1}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left(\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\frac{u_t}{\sqrt{T}}\right)\\ &\Rightarrow \omega_u A\left[\int_0^1 B_x(r)B_x(r)'\,dr\right]^{-1}\int_0^1 B_x(r)dW_u(r)\\ &:= \omega_u\int_0^1 B_x^*(r)dW_u(r),\end{aligned} \tag{1}$$

where

$$B_x^*(r) = A\left[\int_0^1 B_x(r)B_x(r)'\,dr\right]^{-1}B_x(r).$$

The above asymptotic theory forms the basis for testing H_0 against H_1, but we still have to estimate the long-run variance ω_u^2. Here, we employ the series approach. Let $\{\phi_i(\cdot), i=1,2,\ldots,K\}$ be some basis functions on $L^2[0,1]$. The series estimator of ω_u^2 is

$$\hat{\omega}_u^2 = \frac{1}{K}\sum_{i=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\phi_i\left(\frac{t}{T}\right)\hat{u}_t\right]^2, \tag{2}$$

where $\hat{u}_t = Y_t - X_t'\hat{\beta}$. As a rule of thumb, we can use the formula developed by Phillips (2005) to choose K. See Section 4 for more details.

Based on $\hat{\omega}_u^2$, we construct the Wald statistic

$$F_T = \frac{\left(R\hat{\beta}-r\right)'\left[R\left(\sum_{t=1}^{T}X_tX_t'\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)}{p\hat{\omega}_u^2}. \tag{3}$$

When $p=1$, we construct the t-statistic

$$t_T = \frac{R\hat{\beta}-r}{\hat{\omega}_u\sqrt{R\left(\sum_{t=1}^{T}X_tX_t'\right)^{-1}R}}.$$

Let

$$\tilde{\phi}_i(r) = \phi_i(r) - \left[\int_0^1 \phi_i(s)B_x(s)'ds\right]\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1}B_x(r), \tag{4}$$

which is the projection of $\phi_i(r)$ onto the orthogonal complement of the space spanned by $B_x(r)$. By construction, $\int_0^1 \tilde{\phi}_i(r)B_x'(r)dr = 0.$[1]

Denote $\tilde{\phi}(r) = (\tilde{\phi}_1(r),\ldots,\tilde{\phi}_K(r))'$. To obtain the weak limits of $\hat{\omega}_u^2$, F_T, and t_T, we make the following assumption on the basis functions.

Assumption 2. *(i) For each $i=1,\ldots,K$, $\phi_i(\cdot)$ is continuously differentiable. (ii) $\int_0^1 \tilde{\phi}(r)\tilde{\phi}(r)'dr$ is of rank K almost surely.*

Using Assumption 1, we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\hat{u}_t \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(Y_t - X_t'\hat{\beta}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left(Y_t - X_t'\beta_0 - X_t'(\hat{\beta}-\beta_0)\right) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}u_t - \frac{1}{T}\sum_{t=1}^{[Tr]}D_T^{-1}X_t'\left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left[\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\frac{u_t}{\sqrt{T}}\right] \\
&\Rightarrow B_u(r) - \left(\int_0^r B_x(s)'ds\right)\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1}\int_0^1 B_x(\tau)dB_u(\tau).
\end{aligned}$$

Combining this with Assumption 2(i), we have

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i\left(\frac{t}{T}\right)\hat{u}_t \\
&\Rightarrow \int_0^1 \phi_i(r)\,dB_u(r) - \left[\int_0^1 \phi_i(s)B_x(s)'ds\right]\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1}\int_0^1 B_x(\tau)dB_u(\tau) \\
&= \int_0^1 \left\{\phi_i(r) - \left[\int_0^1 \phi_i(s)B_x(s)'ds\right]\left[\int_0^1 B_x(\tau)B_x(\tau)'d\tau\right]^{-1} B_x(r)\right\}dB_u(r) \\
&= \omega_u \int_0^1 \tilde{\phi}_i(r)\,dW_u(r).
\end{aligned}$$

Hence, for a fixed-K,

$$\hat{\omega}_u^2 \Rightarrow \omega_u^2 \frac{1}{K}\sum_{i=1}^{K}\left[\int_0^1 \tilde{\phi}_i(r)\,dW_u(r)\right]^2.$$

Under the fixed-K asymptotics, $\hat{\omega}_u^2$ converges weakly to a random variable that is proportional to ω_u^2. This is sufficient for asymptotically pivotal inference.

For the variance term in the test statistic F_T, we have

$$\begin{aligned}
&\tilde{D}_T R\left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} R\tilde{D}_T \\
&= \tilde{D}_T R D_T^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\left(D_T^{-1}X_t\right)\left(D_T^{-1}X_t\right)'\right)^{-1}\left(\tilde{D}_T R D_T^{-1}\right)' \\
&\Rightarrow A\left[\int_0^1 B_x(r)B_x(r)'\right]^{-1} A' = \int_0^1 B_x^*(r)B_x^*(r)'\,dr.
\end{aligned}$$

Using the above weak convergence results and that for $R\hat{\beta}-r$ in (1), we obtain, for a fixed-K,

$$\begin{aligned}
F_T &= \frac{1}{p\hat{\omega}_u^2}\left\{\tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\right\}'\left[\tilde{D}_T R\left(\sum_{t=1}^{T}X_tX_t'\right)^{-1}R\tilde{D}_T\right]^{-1}\left\{\tilde{D}_T\sqrt{T}R\left(\hat{\beta}-\beta_0\right)\right\} \\
&\Rightarrow \frac{\left\{\int_0^1 B_x^*(r)\,dW_u(r)\right\}'\left\{\int_0^1 B_x^*(r)B_x^*(r)'\,dr\right\}^{-1}\left\{\int_0^1 B_x^*(r)\,dW_u(r)\right\}/p}{\sum_{i=1}^{K}\left[\int_0^1 \tilde{\phi}_i(r)\,dW_u(r)\right]^2/K} \qquad (5) \\
&:= \frac{\eta_0'\eta_0/p}{\sum_{i=1}^{K}\eta_i^2/K} := F_\infty,
\end{aligned}$$

where

$$\eta_0 = \left\{ \int_0^1 B_x^*(r) B_x^*(r)' \, dr \right\}^{-1/2} \int_0^1 B_x^*(r) \, dW_u(r),$$

$$\eta_i = \int_0^1 \tilde{\phi}_i(r) \, dW_u(r) \text{ for } i = 1, 2, \ldots, K.$$

In general, $\int_0^1 B_x^*(r) B_x^*(r)' \, dr$ may not be invertible. Here we assume that $A = \lim_{T\to\infty} \tilde{D}_T R D_T^{-1}$ is of full (row) rank p so that $\int_0^1 B_x^*(r) B_x^*(r)' \, dr$ is invertible almost surely. The above asymptotic theory allows us to make asymptotically valid inferences, but the limiting distribution F_∞ is nonstandard, and so critical values have to be simulated.

We proceed to study how we may obtain a standard fixed-K asymptotic distribution. Note that B_x^* is a function of B_x. Conditional on B_x, η_0 is a standard normal vector. So $\eta_0'\eta_0 \sim \chi_p^2$ conditional on B_x. Moreover, conditional on B_x, both η_0 and η_i are normal, and their conditional covariance given B_x is

$$\operatorname{cov}(\eta_0, \eta_i) = \left\{ \int_0^1 B_x^*(r) B_x^*(r)' \, dr \right\}^{-1/2} \left(\int_0^1 B_x^*(r) \tilde{\phi}_i(r) \, dr \right) = 0 \text{ for } i = 1, \ldots, K.$$

Here, the second equality holds because

$$\int_0^1 B_x^*(r) \tilde{\phi}_i(r) \, dr = A \left[\int_0^1 B_x(r) B_x(r)' \, dr \right]^{-1} \int_0^1 B_x(r) \tilde{\phi}_i(r) \, dr = 0,$$

where we have used $\int_0^1 B_x(r) \tilde{\phi}_i(r) \, dr = 0$. Therefore, conditional on B_x, η_0 and $\{\eta_i, i = 1, \ldots, K\}$ are independent.

To reduce the asymptotic distribution F_∞ to a standard F distribution, we hope that η_i is i.i.d. $N(0, 1)$ conditional on B_x. For this, we require $\tilde{\phi}_i(r)$ to be orthonormal (conditional on B_x). But

$$\begin{aligned}
&\int_0^1 \tilde{\phi}_{i_1}(r) \tilde{\phi}_{i_2}(r) \, dr \\
&= \int_0^1 \phi_{i_1}(r) \phi_{i_2}(r) \, dr - \left[\int_0^1 \phi_{i_1}(r) B_x(r)' dr \right] \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} \int_0^1 B_x(s) \phi_{i_2}(s) \, ds \\
&= \int_0^1 \int_0^1 \phi_{i_1}(r) \left\{ \delta(r-s) - B_x(r)' \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} B_x(s) \right\} \phi_{i_2}(s) \, dr ds \\
&:= \int_0^1 \int_0^1 \phi_{i_1}(r) C(r,s) \phi_{i_2}(s) \, dr ds,
\end{aligned}$$

where

$$C(r,s) = \delta(r-s) - B_x(s)' \left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau \right]^{-1} B_x(r),$$

and $\delta(\cdot)$ is the Dirac delta function such that

$$\int_0^1\int_0^1 \phi_{i_1}(s)\delta(r-s)\phi_{i_2}(r)drds = \int_0^1 \phi_{i_1}(r)\phi_{i_2}(r)dr.$$

To design the basis functions $\{\phi_i(r)\}$ such that $\{\tilde{\phi}_i(r)\}$ are orthonormal on $L^2[0,1]$, we require that $\{\phi_i(r)\}$ be orthonormal with respect to the weighting function $C(r,s)$, that is,

$$\int_0^1\int_0^1 C(r,s)\phi_{i_1}(r)\phi_{i_2}(s)drds = 1\{i_1 = i_2\} \tag{6}$$

almost surely. Because of the general form of $C(r,s)$ and its randomness, commonly used basis functions do not satisfy the above condition.

Instead of searching for the basis functions that satisfy (6), we search for their discrete versions: the basis vectors. For each basis function $\phi_i(r)$, the corresponding basis vector is defined as

$$\phi_i = \left(\phi_i\left(\frac{1}{T}\right), \phi_i\left(\frac{2}{T}\right), \ldots, \phi_i\left(\frac{T}{T}\right)\right)'.$$

We note that it is the basis vectors that are used in the variance estimator. Basis functions only appear in the limiting distribution when $T \to \infty$.

Let

$$\mathbf{C}_T = T \cdot M_X \text{ for } M_X = I_T - X\left(X'X\right)^{-1} X',$$

where $X = \left(X_1, \ldots, X_T\right)' \in \mathbb{R}^{T\times d}$. By definition, $\mathbf{C}_T$ is symmetric and positive semidefinite. It is the discrete version of $C(r,s)$. For any two vectors $r_1, r_2 \in \mathbb{R}^T$, we define their inner product as

$$\langle r_1, r_2 \rangle = r_1' \mathbf{C}_T r_2 / T^2. \tag{7}$$

The discrete analogue of (6) is

$$\left\langle \phi_{i_1}, \phi_{i_2} \right\rangle = 1\{i_1 = i_2\} \text{ for } i_1, i_2 = 1, \ldots, K. \tag{8}$$

Given any set of basis vectors $\phi_1, \ldots, \phi_K$, let $\phi = (\phi_1, \ldots, \phi_K)$ be the $T \times K$ matrix of these basis vectors. Define

$$\phi^M = \sqrt{T}\left(M_X\phi\right)\left[\left(M_X\phi\right)' M_X\phi\right]^{-1/2},$$

where the superscript "M" signifies that the basis vectors in ϕ^M are obtained via a transformation involving M_X. We have

$$
\begin{aligned}
&T^{-2}\left(\phi^{M}\right)'\mathbf{C}_T\phi^{M}\\
&=TT^{-2}\left[\left(M_X\phi\right)'M_X\phi\right]^{-1/2}\phi'M_X\cdot\mathbf{C}_T\cdot M_X\phi\left[\left(M_X\phi\right)'M_X\phi\right]^{-1/2}\\
&=\left[\phi'M_X\phi\right]^{-1/2}\left(\phi'M_X\phi\right)\left[\phi'M_X\phi\right]^{-1/2}=I_K.
\end{aligned}
$$

That is, the columns of the matrix ϕ^M satisfy the conditions in (8).

If we use $\left\{\phi_i^M=\left(\phi_{i,1}^M,\ldots,\phi_{i,T}^M\right)',\quad i=1,\ldots,M\right\}$, the columns of ϕ^M, in constructing the variance estimator, that is, we estimate ω_u^2 by

$$
\hat{\omega}_{u,M}^2=\frac{1}{K}\sum_{i=1}^{K}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_{i,t}^M\hat{u}_t\right]^2,
$$

then

$$
\begin{aligned}
\hat{\omega}_{u,M}^2&=\frac{1}{TK}\hat{u}'\phi^M\left(\phi^M\right)'\hat{u}\\
&=\frac{1}{K}\left(\hat{u}'M_X\phi\right)\left(\phi'M_X\phi\right)^{-1}\left(\phi'M_X\hat{u}\right)=\frac{1}{K}\left(\hat{u}'\phi\right)\left(\phi'M_X\phi\right)^{-1}\left(\phi'\hat{u}\right) \qquad (9)\\
&\Rightarrow\omega_u^2\frac{1}{K}\left[\int_0^1\tilde{\phi}(r)'\,dW_u(r)\right]\left[\int_0^1\tilde{\phi}(r)\tilde{\phi}(r)'\,dr\right]^{-1}\left[\int_0^1\tilde{\phi}(r)\,dW_u(r)\right].
\end{aligned}
$$

In the above, the third equality holds because $\hat{u}=M_Xu$ and so $\hat{u}'M_X\phi=u'M_X'M_X\phi=u'M_XM_X\phi=u'M_X\phi=\hat{u}'\phi$. As a result,

$$
F_T^*:=\frac{\left(R\hat{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)/p}{\left(\hat{u}'M_X\phi\right)\left(\phi'M_X\phi\right)^{-1}\left(\phi'M_X\hat{u}\right)/K}\Rightarrow\frac{\eta_0'\eta_0/p}{\eta'\eta/K},
$$

where, as before

$$
\eta_0=\left\{\int_0^1B_x^*(r)B_x^*(r)'\,dr\right\}^{-1/2}\int_0^1B_x^*(r)\,dW_u(r),
$$

but now

$$
\eta:=\left(\eta_1,\eta_2,\ldots,\eta_K\right)'=\left[\int_0^1\tilde{\phi}(r)\tilde{\phi}(r)'\,dr\right]^{-1/2}\left[\int_0^1\tilde{\phi}(r)\,dW_u(r)\right].
$$

Under Assumption 2(ii), $\eta\sim N(0,I_K)$ conditional on B_x. Also, as we have shown before, η is independent of η_0 conditional on B_x. So, conditional on B_x,

$$
\frac{\eta_0'\eta_0/p}{\eta'\eta/K}\sim F_{p,K}.
$$

Given that the conditional distribution $F_{p,K}$ does not depend on the conditioning variable (i.e., B_x), $\frac{\eta_0'\eta_0 / p}{\eta'\eta / K} \sim F_{p,K}$ unconditionally. Therefore, $F_T^* \Rightarrow F_{p,K}$. We formalize this result and a similar result for a t-statistic in the theorem below.

Theorem 1. *Let Assumptions 1 and 2 hold. Assume further that* $\lim_{T\to\infty} \tilde{D}_T R D_T^{-1}$ *is of full row rank p. Then, for a fixed-K as* $T \to \infty$,

$$F_T^* := \frac{\left(R\hat{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)/p}{\left(\hat{u}'M_X\phi\right)\left(\phi'M_X\phi\right)^{-1}\left(\phi'M_X\hat{u}\right)/K} \Rightarrow F_{p,K}$$

$$t_T^* := \frac{R\hat{\beta}-r}{\sqrt{\left(\hat{u}'M_X\phi\right)\left(\phi'M_X\phi\right)^{-1}\left(\phi'M_X\hat{u}\right)/K}\sqrt{R\left(\sum_{t=1}^T X_tX_t'\right)^{-1}R'}} \Rightarrow t_K$$

where $M_X = I_T - X\left(X'X\right)^{-1}X'$, $F_{p,K}$ *is the standard F distribution with degrees of freedom* (p,K), *and* t_K *is the standard t distribution with degrees of freedom K.*

Remark 1. *In order to develop the asymptotic F and t theory, we use the novel series variance estimator* $(TK)^{-1}\hat{u}'\phi^M\left(\phi^M\right)'\hat{u}$ *instead of the usual series variance estimator* $(TK)^{-1}\left(\hat{u}'\phi\right)\left(\phi'\hat{u}\right)$. *To obtain* ϕ^M, *we first project* ϕ *on the orthogonal complement of the column space of X to obtain* $M_X\phi$ *and then orthonormalize it into* $\phi^M = \sqrt{T}\left(M_X\phi\right)\left[\left(M_X\phi\right)'M_X\phi\right]^{-1/2}$. *Note that*

$$(TK)^{-1}\hat{u}'\phi^M\left(\phi^M\right)'\hat{u} = (TK)^{-1}u'\phi^M\left(\phi^M\right)'u.$$

The series variance estimator is the same regardless of whether $\hat{u}$ *or the true u is used. The projection, therefore, ensures that the series variance estimator* $(TK)^{-1}\hat{u}'\phi^M\left(\phi^M\right)'\hat{u}$ *is invariant to the parameter estimation error. When X is random, the projection is onto a random subspace. For series variance estimation, the idea of using data-dependent and randomly orthonormalized basis functions is new in the literature.*

Remark 2. *From a theoretical point of view, the implied basis functions we use in the series variance estimation are the elements of the following vector:*

$$\tilde{\phi}^*(r) = \left[\int_0^1 \tilde{\phi}(\tau)\tilde{\phi}(\tau)'\,d\tau\right]^{-1/2}\tilde{\phi}(r).$$

From any given vector of basis functions $\phi(r) = \left(\phi_1(r),\ldots,\phi_K(r)\right)'$, *we first use the projection operation given in (4) to obtain* $\tilde{\phi}(r) = \left(\tilde{\phi}_1(r),\ldots,\tilde{\phi}_K(r)\right)'$ *and then use the orthonormalization (i.e., pre-multiplying* $\tilde{\phi}(r)$ *by* $\left[\int_0^1 \tilde{\phi}(\tau)\tilde{\phi}(\tau)'\,d\tau\right]^{-1/2}$*) to obtain* $\tilde{\phi}^*(r) := \left[\int_0^1 \tilde{\phi}(\tau)\tilde{\phi}(\tau)'\,d\tau\right]^{-1/2}\tilde{\phi}(r)$. *For practical implementation, we do* *not need to find the implied basis functions, as we only need to use the corresponding basis vectors* $\left(\phi' M_X \phi\right)^{-1/2}\left(\phi' M_X\right)$ *to compute* $\hat{\omega}_u^2$. *The basis functions in* $\tilde{\phi}^*(r)$ *appear only in the asymptotic distributions of the long-run variance estimator* $\hat{\omega}_u^2$ *and the test statistics* F_T^* *and* t_T^*. *More specifically, we can rewrite (9) as*

$$\hat{\omega}_{u,M}^2 \Rightarrow \omega_u^2 \frac{1}{K}\left[\int_0^1 \tilde{\phi}^*(r)\, dW_u(r)\right]'\left[\int_0^1 \tilde{\phi}^*(r)\, dW_u(r)\right].$$

The basis functions in $\tilde{\phi}^*(r)$ *appear in our asymptotic distributions only via the above weak limit of* $\hat{\omega}_{u,M}^2$.

Remark 3. *Theorem 1 extends Sun (*2011*) to allow for more general trend functions. For trend regressions, Sun (2011) considers a linear trend with* $X_t = (1,t)'$ *so that* $B_x(r) = (1,r)'$ *and employs the cosine basis functions* $\phi_i(r) = \sqrt{2}\cos(2\pi i r)$ *for* $i = 1,2,\ldots,K$. *These functions are special in that they are orthonormal on* $L^2[0,1]$ *and satisfy*

$$\int_0^1 \phi_i(s) B_x(s)\, ds = \begin{pmatrix} \int_0^1 \sqrt{2}\cos(2\pi s)\, ds \\ \int_0^1 s\sqrt{2}\cos(2\pi s)\, ds \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

As a result, $\tilde{\phi}_i(r) = \phi_i(r) - \left[\int_0^1 \phi_i(s) B_x(s)' ds\right]\left[\int_0^1 B_x(\tau) B_x(\tau)' d\tau\right]^{-1} B_x(r) = \phi_i(r)$ *for all i, that is, the projection does not change the original basis functions. Hence,* $\eta_i := \int_0^1 \tilde{\phi}_i(r)\, dW_u(r) = \int_0^1 \phi_i(r)\, dW_u(r)$ *is i.i.d.* $N(0,1)$. *The asymptotic F and t theory can then be established directly with the cosine basis functions. However, the requirements that* $\tilde{\phi}_i(r) = \phi_i(r)$ *for all i and* $\{\phi_i(r)\}$ *are orthonormal on* $L^2[0,1]$ *severely limit the set of basis functions we can use. For example, Sun (2011) has to rule out the sine functions* $\left\{\sqrt{2}\sin(2\pi i r)\right\}$ *with consequential adverse effect on the power of the resulting test. In contrast, Theorem 1 allows us to use any basis functions that satisfy Assumption 2. We can do so because the projection step preemptively purges the effect of the parameter estimation uncertainty and the orthonormalization step ensures the orthonormality of the implied basis functions* $\{\tilde{\phi}_i^*(r)\}$.

Remark 4. *For regressions with strictly exogenous integrated regressors, Theorem 1 can be regarded as an F-test version of Park and Phillips (1988) (Theorem 5.4) where the asymptotic chi-square test theory was developed. We note that an asymptotic F theory cannot be established for the usual test statistic constructed based on a kernel (long-run) variance estimator. A series variance estimator with carefully crafted basis functions/vectors appears to be indispensable for the asymptotic F theory.*

Remark 5. *For cointegration regressions, Bunzel (2006) and Jin et al. (2006) develop the fixed-b asymptotic theory for studentized test statistics. The asymptotic distributions in these two papers are nonstandard. In contrast, the asymptotic distributions in Theorem 1 are standard F- and t-distributions and are thus more* *convenient for practical use. In the presence of endogeneity, the case considered in the next section, we may employ the series long-run and half long-run variance estimators to obtain the fully modified OLS estimator of Phillips and Hansen (1990). Suppose we ignore the estimation error in the half long-run variance estimator, which effectively reduces the problem back to the exogenous case, then we can use the F and t approximations for inference. This may not be completely satisfactory, because the F- and t-distributions are not the exact asymptotic distributions of the F- and t-statistics under the fixed-K asymptotics. We use the F and t approximations only because they are expected to be more accurate than the chi-square and normal approximations. For the endogenous case, exact and asymptotically valid F and t tests are developed in the next section.*

Remark 6. *Theorem 1 allows for near-unit-root processes. It appears to be the first time that an asymptotic F and t theory is established in this setting. However, see Sun (2014c) for the F and t limit theory in a different setting where the regressor error is a near-unit-root process and Guo et al. (2018) for the asymptotic t theory in an autoregression where the process is moderately explosive. Theorem 1 also allows for I(2) processes. To the best of our knowledge, this has not been considered in the literature before.*

Remark 7. *The asymptotic F and t theory in the structural break setting given above appears to be new. Sun and Wang (2021) establish the asymptotic F and t theory in a structural break model, but they consider only the case when X_t° is stationary. Here we allow X_t° to be nonstationary.*

Remark 8. *Theorem 1 provides a unified framework that accommodates various non-stationary regressors. The idea may be extended under suitable conditions to allow for fractionally integrated processes, slowing-varying trend regressors, and nonlinear trends. See Phillips (2007) for slowing-varying trend regressions.*

To conclude this section, we outline the steps in conducting the asymptotic F and t tests:

(i) Estimate β_0 by the OLS estimator $\hat{\beta}$ and calculate the residual $\hat{u} = Y - X\hat{\beta}$.
(ii) Construct the $T \times K$ matrix $\phi = (\phi_1, \ldots, \phi_K)$ of the basis vectors. We recommend using the following Fourier series as the basis vectors:

$$\begin{aligned}
\phi_{2i-1} &= \left(\sqrt{2}\cos(2\pi i \frac{t}{T})\right)_{t=1}^{T} = \sqrt{2}\left(\cos(2\pi i \frac{1}{T}), \cos(2\pi i \frac{2}{T}), \ldots, \cos(2\pi i \frac{T}{T})\right)' \\
\phi_{2i} &= \left(\sqrt{2}\sin(2\pi i \frac{t}{T})\right)_{t=1}^{T} = \sqrt{2}\left(\sin(2\pi i \frac{1}{T}), \sin(2\pi i \frac{2}{T}), \ldots, \sin(2\pi i \frac{T}{T})\right)'
\end{aligned} \tag{10}$$

for $i = 1, 2, \ldots, K/2$ assuming that K is even.
(iii) For $M_X = I_T - X\left(X'X\right)^{-1} X'$, compute the test statistics F_T^* and t_T^* defined in Theorem 1.
(iv) For the F-test, compare F_T^* with critical values from the standard F- distribution $F_{p,K}$. For the t-test, compare t_T^* with critical values from the standard t-distribution t_K.

3. ENDOGENOUS CASE

3.1. Cointegration Regression

We consider the model

$$
\begin{aligned}
Y_t &= \alpha_0 + X_t'\beta_0 + u_{0t} \\
X_t &= \left(1 - \frac{c}{T}\right) X_{t-1} + u_{xt}
\end{aligned}
\tag{11}
$$

for $t = 1, \ldots, T$, where Y_t is a scalar time series and X_t is a $d_x \times 1$ vector of time series with $X_0 = o_p\left(\sqrt{T}\right)$. Here we single out the intercept because we allow the rest of the regressors X_t to be endogenous: $\{u_{xt}\}$ and $\{u_{0t}\}$ can be arbitrarily correlated. We assume that $c \geq 0$ so that both unit-root and near-unit-root processes can be accommodated.

We are interested in constructing a confidence interval for $R\beta_0$. The confidence interval can then be used in testing whether $R\beta_0 = r$ for some $r \in \mathbb{R}^p$.

We maintain the following assumption on $u_t = \left(u_{0t}, u_{xt}'\right)'$. Note that the definition of u_t is different from that in the previous section.

Assumption 3. *The functional central limit theorem holds*:

$$
T^{-1/2} \sum_{t=1}^{[T\cdot]} u_t \Rightarrow B(\cdot) = \Omega^{1/2} W(\cdot),
\tag{12}
$$

where $W(\cdot) := (W_0(\cdot), W_x'(\cdot))'$ is a $\left(d_x + 1\right)$-dimensional standard Brownian process,

$$
\Omega = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}' = \begin{pmatrix} \underset{1\times 1}{\omega_0^2} & \underset{1\times d_x}{\omega_{0x}} \\ \underset{d_x\times 1}{\omega_{x0}} & \underset{d_x\times d_x}{\Omega_{xx}} \end{pmatrix},
\tag{13}
$$

and Ω is positive-definite.

Using the Cholesky form of $\Omega^{1/2}$, we can write $B(\cdot)$ as

$$
B(\cdot) = \begin{pmatrix} B_0(\cdot) \\ B_x(\cdot) \end{pmatrix} = \begin{pmatrix} \omega_{0\cdot x} W_0(\cdot) + \omega_{0x} \Omega_{xx}^{-1/2} W_x(\cdot) \\ \Omega_{xx}^{1/2} W_x(\cdot) \end{pmatrix},
\tag{14}
$$

where $\omega_{0\cdot x}^2 = \omega_0^2 - \omega_{0x} \Omega_{xx}^{-1} \omega_{x0}$ and $\Omega_{xx}^{1/2}$ is the symmetric and positive-definite matrix square root of Ω_{xx}.

In the presence of endogeneity, $B_0(\cdot)$ and $B_x(\cdot)$ will be dependent, and the OLS estimator $\hat{\beta}$ of β_0 will have a second-order endogeneity bias and a complicated asymptotic distribution. To remove the endogeneity bias and restore the asymptotic (mixed) normality of $\hat{\beta}$, we may use the fully modified OLS estimator of Phillips and Hansen (1990). This estimator involves using a long-run variance and a half long-run variance to remove the dependence between $B_0(\cdot)$ and

$B_x(\cdot)$ and the endogeneity bias. Both the long-run variance and the half long-run variance are estimated nonparametrically. However, the estimation uncertainty, which is potentially very high, is ignored in the asymptotic chi-square and normal approximations. For this reason, the chi-square and normal tests often have large size distortion; see, for example, Vogelsang and Wagner (2014).

To confront the size distortion problem, we follow Hwang and Sun (2018)[2] and consider a different estimation approach. We assume that c is known. If X_t is a unit-root process, we know that $c=0$. So the assumption is not restrictive. However, when X_t is a near-unit-root process, we do not know c in general, and the assumption becomes restrictive. In this case, we can first construct a confidence interval for c and then use Bonferroni's method to construct the confidence bound for $R\beta_0$. In this case, our asymptotic theory below forms the basis for the Bonferroni's method. For more details, see, for example, Phillips (2015) for such a practice in predictive regressions. The same method can be used here.

Define the quasi-differenced process $\Delta_c X_t$ as

$$\Delta_c X_t = X_t - \left(1-\frac{c}{T}\right)X_{t-1}.$$

Obviously, $\Delta_c X_t = u_{x,t}$. Let $\delta_0 = \Omega_{xx}^{-1}\omega_{x0}$ be the long-run regression coefficient when u_{0t} is regressed on u_{xt} and

$$u_{0\cdot x,t} = u_{0t} - (\Delta_c X_t)'\delta_0 = u_{0t} - u_{x,t}'\delta_0$$

be the corresponding long-run regression error. Then, we obtain the augmented regression

$$Y_t = \alpha_0 + X_t'\beta_0 + (\Delta_c X_t)'\delta_0 + u_{0\cdot x,t}, \tag{15}$$

where, by definition, the long-run correlation between $u_{0\cdot x,t}$ and $u_{x,t}$ is zero. The augmentation is designed to purge the dependence between $B_0(\cdot)$ and $B_x(\cdot)$.

Note that the zero long-run correlation between $u_{0\cdot x,t}$ and $u_{x,t}$ does not rule out that $u_{0\cdot x,t}$ may be still correlated with X_t and $\Delta_c X_t$. Hence, the (augmented) OLS estimator (denoted by $\hat{\beta}_{\text{AOLS}}$) of β_0 based on the augmented regression can still have a second-order endogeneity bias. More precisely, the mean of the asymptotic distribution of $T(\hat{\beta}_{\text{AOLS}} - \beta_0)$ may not be zero. Ignoring the nonzero mean leads to invalid and unreliable statistical inferences.

To remove the endogeneity bias, we follow Phillips (2014) and Hwang and Sun (2018) and run the regression in a different domain, which resembles the frequency domain, but any set of orthonormal basis functions in $L^2[0,1]$ can be used. For convenience, we refer to this domain as the pseudo-frequency domain. Let $\{\phi_i\}_{i=1}^K$ be a set of K such basis functions on $L^2[0,1]$. For each $i=1,\ldots,K$, we transform all variables in the augmented regression into

$$\mathbb{W}_i^{\alpha} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_i\left(\frac{t}{T}\right),$$
$$\mathbb{W}_i^{y} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_t\phi_i\left(\frac{t}{T}\right), \mathbb{W}_i^{x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\phi_i\left(\frac{t}{T}\right), \tag{16}$$
$$\mathbb{W}_i^{\Delta_c x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\Delta_c X_t)\phi_i\left(\frac{t}{T}\right), \mathbb{W}_i^{0\cdot x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{0\cdot x,t}\phi_i\left(\frac{t}{T}\right).$$

The basis transformation is designed to extract the long-run component in the time-series data.

Based on the augmented regression and the transformed data, we have

$$\mathbb{W}_i^{y} = \alpha_0\mathbb{W}_i^{\alpha} + \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta_c x\prime}\delta_0 + \mathbb{W}_i^{0\cdot x} \text{ for } i = 1,\ldots,K. \tag{17}$$

Under the assumption that each function $\phi_i(\cdot)$ is continuously differentiable and satisfies $\int_0^1 \phi_i(r)dr = 0$, which we will maintain, we have

$$\mathbb{W}_i^{\alpha} = \sqrt{T}\int_0^1 \phi_i(r)dr + \sqrt{T}O(1/T) = O\left(1/\sqrt{T}\right) = o(1). \tag{18}$$

So

$$\mathbb{W}_i^{y} = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta_c x\prime}\delta_0 + \mathbb{W}_{\alpha,i}^{0\cdot x} \text{ for } i = 1,\ldots,K, \tag{19}$$

where

$$\mathbb{W}_{\alpha,i}^{0\cdot x} = \mathbb{W}_i^{0\cdot x} + \mathbb{W}_i^{\alpha} = \mathbb{W}_i^{0\cdot x} + o(1).$$

Our estimation and inference will be based on equation (19), which can be regarded as a low-frequency regression.

Putting (19) in a vector form, we have

$$\mathbb{W}^{y} = \mathbb{W}^{x}\beta_0 + \mathbb{W}^{\Delta_c x}\delta_0 + \mathbb{W}_{\alpha}^{0\cdot x}, \tag{20}$$

where $\mathbb{W}^{y} = (\mathbb{W}_1^{y},\ldots,\mathbb{W}_K^{y})'$ and $\mathbb{W}^{x}$, $\mathbb{W}^{\Delta x}$, and $\mathbb{W}_{\alpha}^{0\cdot x}$ are defined similarly. Running OLS based on the above equation leads to the transformed and augmented OLS (TAOLS) estimator of $\gamma_0 = (\beta_0', \delta_0')'$:

$$\hat{\gamma}_{\text{TAOLS}} = (\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\mathbb{W}^{y},$$

where $\tilde{\mathbb{W}} = \left(\mathbb{W}^{x}, \mathbb{W}^{\Delta x}\right)$. See Hwang and Sun (2018) for discussions on the efficiency and robustness of this estimator. The TAOLS estimator can be regarded as an IV estimator based on the augmented equation in (15) using the basis vectors $\phi_1,\ldots,\phi_K$ as the instruments. See Phillips (2014) for more details.

Let

$$P_x = \mathbb{W}^{x}(\mathbb{W}^{x\prime}\mathbb{W}^{x})^{-1}\mathbb{W}^{x\prime}, P_{\Delta_c x} = \mathbb{W}^{\Delta_c x}(\mathbb{W}^{\Delta_c x\prime}\mathbb{W}^{\Delta_c x})^{-1}\mathbb{W}^{\Delta_c x\prime},$$

and $M_x = I_K - P_x$, $M_{\Delta_c x} = I_K - P_{\Delta_c x}$. Then we can represent $\hat{\gamma}_{\text{TAOLS}}$ as

$$\hat{\gamma}_{\text{TAOLS}} = \begin{pmatrix} \hat{\beta}_{\text{TAOLS}} \\ \hat{\delta}_{\text{TAOLS}} \end{pmatrix} = \begin{pmatrix} (\mathbb{W}^{x\prime} M_{\Delta_c x} \mathbb{W}^x)^{-1} (\mathbb{W}^{x\prime} M_{\Delta_c x} \mathbb{W}^y) \\ (\mathbb{W}^{\Delta_c x\prime} M_x \mathbb{W}^{\Delta_c x})^{-1} (\mathbb{W}^{\Delta_c x\prime} M_x \mathbb{W}^y) \end{pmatrix}. \tag{21}$$

To establish the asymptotic properties of $\hat{\gamma}_{\text{TAOLS}}$, we make the following assumption.

Assumption 4. (*i*) *For every* $i = 1, \ldots, K$, $\phi_i(\cdot)$ *is continuously differentiable; (ii) for every* $i = 1, \ldots, K, \phi_i(\cdot)$ *satisfies* $\int_0^1 \phi_i(r)dr = 0$; *(iii) the functions* $\{\phi_i(\cdot)\}_{i=1}^K$ *are orthonormal in* $L^2[0,1]$.

Under Assumptions 3 and 4(i and ii), we can use summation by parts, the continuous mapping theorem, and integration by parts to obtain

$$\begin{aligned}
\mathbb{W}_{\alpha,i}^{0\cdot x} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_i\left(\frac{t}{T}\right) u_{0\cdot xt} + \mathbb{W}_i^{\alpha} \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_i\left(\frac{t}{T}\right) \left(u_{0t} - u_{xt}'\delta_0\right) + o(1) \\
&\Rightarrow \int_0^1 \phi_i(r)\, d\left[B_0(r) - B_x(r)'\delta_0\right] \\
&= \omega_{0\cdot x} \int_0^1 \phi_i(r)\, dW_0(r) := \omega_{0\cdot x} \nu_i.
\end{aligned}$$

Similarly,

$$\frac{\mathbb{W}_i^x}{T} = \frac{1}{T^{3/2}} \sum_{t=1}^{T} \phi_i\left(\frac{t}{T}\right) X_t \Rightarrow \Omega_{xx}^{1/2} \int_0^1 \phi_i(r) J_{c,x}(r) dr := \xi_i,$$

and

$$\mathbb{W}_i^{\Delta_c x} \Rightarrow \int_0^1 \phi_i(r)\, dB_x(r) = \Omega_{xx}^{1/2} \int_0^1 \phi_i(r)\, dW_x(r) := \eta_i.$$

Let

$$\begin{aligned}
\nu &\equiv \left(\nu_1, \nu_2, \ldots, \nu_K\right)' \in \mathbb{R}^{K\times 1}, \\
\xi &\equiv \left(\xi_1, \xi_2, \ldots, \xi_K\right)' \in \mathbb{R}^{K\times d_x}, \\
\eta &\equiv (\eta_1, \eta_2, \ldots, \eta_K)' \in \mathbb{R}^{K\times d_x},
\end{aligned}$$

and $\zeta = (\xi, \eta)$. Then

$$\left(\mathbb{W}^x / T, \mathbb{W}^{\Delta x}, \mathbb{W}_{\alpha}^{0\cdot x}\right) \Rightarrow \left(\xi, \eta, \omega_{0\cdot x}\nu\right), \tag{22}$$

where $\zeta \perp \nu$. Also, it follows from Assumption 4(iii) that $\nu \sim N(0, I_K)$. In particular, for

$$\Upsilon_T = \begin{pmatrix} T \cdot I_{d_x} & \underset{d_x \times d_x}{O} \\ \underset{d_x \times d_x}{O} & I_{d_x} \end{pmatrix},$$

where O is a matrix of zeros whose dimension may be different at different occurrences, we have $\tilde{\mathbb{W}}\Upsilon_T^{-1} \Rightarrow \zeta$. It then follows that

$$\begin{aligned} \Upsilon_T \left(\hat{\gamma}_{\text{TAOLS}} - \gamma_0\right) &= (\Upsilon_T^{-1}\tilde{\mathbb{W}}'\tilde{\mathbb{W}}\Upsilon_T^{-1})^{-1}(\tilde{\mathbb{W}}\Upsilon_T^{-1})'\mathbb{W}_\alpha^{0\cdot x} \\ &\Rightarrow \omega_{0\cdot x}(\zeta'\zeta)^{-1}\zeta'\nu = \omega_{0\cdot x} \begin{pmatrix} \left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu \\ \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu \end{pmatrix}, \end{aligned}$$

where

$$M_\eta = I_{d_x} - \eta\left(\eta'\eta\right)^{-1}\eta' \text{ and } M_\xi = I_{d_x} - \xi\left(\xi'\xi\right)^{-1}\xi'.$$

We formalize the above asymptotic result in the theorem below.

Theorem 2. *Let Assumptions 3 and 4 hold. Then under the fixed-K asymptotics where K is held fixed as $T \to \infty$, we have*

$$T(\hat{\beta}_{\text{TAOLS}} - \beta_0) \Rightarrow \omega_{0\cdot x}\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu, \tag{23}$$

$$\hat{\delta}_{\text{TAOLS}} - \delta_0 \Rightarrow \omega_{0\cdot x}\left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \nu, \tag{24}$$

jointly.

Except for the difference in the definitions and the distributions of $\zeta = (\xi, \eta)$, Theorem 2 is identical to Theorem 1 of Hwang and Sun (2018).

Conditional on (ξ, η), both limiting distributions in Theorem 2 are normal with mean zero. There is no second-order endogeneity bias in the TAOLS estimator. As in Hwang and Sun (2018), the TAOLS approach successfully removes the two problems that plague the usual OLS estimator. It paves the way for developing standard inference procedures.

To make inferences on $R\beta_0$, we estimate $\omega_{0\cdot x}^2$ by

$$\begin{aligned} \hat{\omega}_{0\cdot x}^2 &= \frac{1}{K}\sum_{i=1}^{K}\left(\mathbb{W}_i^{y} - \mathbb{W}_i^{x'}\hat{\beta}_{\text{TAOLS}} - \mathbb{W}_i^{\Delta_c x'}\hat{\delta}_{\text{TAOLS}}\right)^2 \\ &= \frac{1}{K}\mathbb{W}_\alpha^{0\cdot x'}\left[I_K - \tilde{\mathbb{W}}(\tilde{\mathbb{W}}'\tilde{\mathbb{W}})^{-1}\tilde{\mathbb{W}}'\right]\mathbb{W}_\alpha^{0\cdot x}. \end{aligned}$$

We can then construct the test statistic

$$F(\hat{\beta}_{\text{TAOLS}}) = \frac{1}{\hat{\omega}_{0\cdot x}^2}\left[R\left(\hat{\beta}_{\text{TAOLS}} - \beta_0\right)\right]'\left[R(\mathbb{W}^{x'}M_{\Delta_c x}\mathbb{W}^{x})^{-1}R'\right]^{-1}\left[R\left(\hat{\beta}_{\text{TAOLS}} - \beta_0\right)\right]/\, p, \tag{25}$$

and for $p=1$,

$$t(\hat{\beta}_{\text{TAOLS}}) = \frac{R\left(\hat{\beta}_{\text{TAOLS}} - \beta_0\right)}{\sqrt{\hat{\omega}_{0\cdot x}^2 R(\mathbb{W}^{x\prime} M_{\Delta_c x} \mathbb{W}^x)^{-1} R'}}.$$

Using (22), we have

$$F(\hat{\beta}_{\text{TAOLS}}) \Rightarrow \frac{\left[R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu\right]' \left[R\left(\xi' M_\eta \xi\right)^{-1} R\right]^{-1} \left[R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu\right] / p}{K^{-1} \nu' \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right]\nu}$$

$$= \frac{\left\| \left[R\left(\xi' M_\eta \xi\right)^{-1} R\right]^{-1/2} R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu \right\|^2 / p}{\left\| \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right]\nu \right\|^2 / K}.$$

Conditional on $\zeta = (\xi, \eta)$, we have

$$\left\| \left[R\left(\xi' M_\eta \xi\right)^{-1} R\right]^{-1/2} R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu \right\|^2 \sim \chi_p^2$$

$$\left\| \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right]\nu \right\|^2 \sim \chi_{K-2d_x}^2$$

and conditional on ζ,

$$\begin{aligned}
&\operatorname{cov}\left(R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu, \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right]\nu\right) \\
&= R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right] \\
&= R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta - R\left(\xi' M_\eta \xi\right)^{-1} \xi' \left[M_\eta \xi, O\right](\zeta'\zeta)^{-1}\zeta \\
&= R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta - [R, O] \begin{pmatrix} \left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \\ \left(\eta' M_\xi \eta\right)^{-1} \eta' M_\xi \end{pmatrix} = O.
\end{aligned}$$

So, conditional on ζ, the numerator and the denominator of the limiting distribution of $F(\hat{\beta}_{\text{TAOLS}})$ follow independent chi-square distributions. Hence,

$$\frac{\left\| \left[R\left(\xi' M_\eta \xi\right)^{-1} R\right]^{-1/2} R\left(\xi' M_\eta \xi\right)^{-1} \xi' M_\eta \nu \right\|^2 / p}{\left\| \left[I_K - \zeta(\zeta'\zeta)^{-1}\zeta\right]\nu \right\|^2 / \left(K - 2d_x\right)} \sim F_{p, K-2d_x}.$$

But $F_{p,K-2d_x}$ does not depend on the conditioning variable ζ, thus, it is also the unconditional distribution. We have, therefore, shown that

$F(\hat{\beta}_{\text{TAOLS}}) \Rightarrow \dfrac{K}{K-2d_x} \cdot F_{p,K-2d_x}$. We collect this and the result on the t-statistic in the theorem below.

Theorem 3. *Let Assumptions 3 and 4 hold. Assume that $K > 2d_x$. Under the fixed-K asymptotics, we have*

$$F^*(\hat{\beta}_{\text{TAOLS}}) := \frac{K-2d_x}{K} F(\hat{\beta}_{\text{TAOLS}}) \Rightarrow F_{p,K-2d_x} \quad and$$
$$t^*(\hat{\beta}_{\text{TAOLS}}) := \sqrt{\frac{K-2d_x}{K}} t(\hat{\beta}_{\text{TAOLS}}) \Rightarrow t_{K-2d_x} \quad for \quad p=1,$$

where $F_{p,K-2d_x}$ is the standard F distribution with degrees of freedom p and $K-2d_x$, and t_{K-2d_x} is the standard t distribution with degrees of freedom $K-2d_x$.

Remark 9. *If we pretend that all variables in the regression (20) are distributed exactly as their respective asymptotic normal distributions, then we obtain a classical normal linear regression model (CNLRM). The asymptotic F and t theory in Theorem 3 is the same as the exact F and t theory in a CNLRM. The test statistic $F^*(\hat{\beta}_{\text{TAOLS}})$ can be equivalently computed by the classical formula that compares the sums of squared residuals for restricted and unrestricted regressions.*

Remark 10. *To establish the asymptotic F and t theory, we employ a conditioning argument by conditioning on ζ. The exact form of the distribution of ζ is not essential. The asymptotic F and t theory holds regardless of the distribution of ζ. While the distribution of ζ in Hwang and Sun (2018) is different from what we have here, the asymptotic F- and t-distributions in Theorem 3 are the same as those in Theorem 3 of Hwang and Sun (2018). There is an opportunity to extend the asymptotic F and t theory further to allow for other distributions of ζ; see, for example, Pellatt and Sun (2022).*

3.2. Predictive Regression

As a variant of the model in the previous subsection, we consider the predictive regression:

$$\begin{aligned} Y_t &= \alpha_0 + X_{t-1}'\beta_0 + u_{0t}, \\ X_t &= \left(1 - \frac{c}{T}\right) X_{t-1} + u_{xt}. \end{aligned} \tag{26}$$

To describe the information filtration, we let $\left(u_{0t}', \varepsilon_{xt}'\right)'$ be a martingale difference sequence. We assume that $u_{xt} = g(\varepsilon_{xt}, \varepsilon_{x,t-1}, \ldots)$ for some measurable function g.

There is a large econometric and finance literature on this type of regression; see Phillips (2015) for a recent review. Here we do not restrict $\{u_{xt}\}$ to be a martingale difference sequence, but we still maintain Assumption 3. Our assumption is less restrictive than most of the existing literature where the martingale difference assumption is maintained. However, see Hjalmarsson (2007), which allows u_{xt} to be a linear process driven by $\{\varepsilon_{xt}\}$.

Define $\mathbb{W}_i^{\alpha}, \mathbb{W}_i^{y}, \mathbb{W}_i^{\Delta_c x}, \mathbb{W}_i^{0\cdot x}$ in the same way as before, but define $\mathbb{W}_{i,-1}^{x}$ as

$$\mathbb{W}_{i,-1}^{x} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{t-1}\phi_i\left(\frac{t}{T}\right).$$

The transformed and augmented regression model becomes

$$\mathbb{W}_i^{y} = \mathbb{W}_{i,-1}^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta_c x\prime}\delta_0 + \mathbb{W}_{\alpha,i}^{0\cdot x} \quad \text{for } i = 1,\ldots,K, \tag{27}$$

which is the same as (19) but with $\mathbb{W}_{i,-1}^{x}$ replacing $\mathbb{W}_i^{x}$. Noting that

$$\frac{\mathbb{W}_{i,-1}^{x}}{T} = \frac{1}{T^{3/2}}\sum_{s=1}^{T}\phi_i\left(\frac{s}{T}\right)X_{s-1} \Rightarrow \Omega_{xx}^{1/2}\int_0^1 \phi_i(r) J_{c,x}(r)dr = \xi_i,$$

the asymptotic distribution of $\mathbb{W}_{i,-1}^{x}/T$ is the same as that of $\mathbb{W}_i^{x}/T$. Therefore, all variables in (27) have the same asymptotic distributions as those in (19). It then follows that the asymptotic F and t theory in Theorem 3 continues to hold for the predictive regression.

To conclude this section, we outline the steps in conducting the asymptotic F and t tests[3]:

(i) Choose a sequence of orthonormal basis functions on $L^2[0,1]$. For example, we can use $\left\{\sqrt{2}\cos(2\pi r i), \sqrt{2}\sin(2\pi r i)\right\}$ for $i = 1,2,\ldots,K/2$.

(ii) Project the time-series data Y_t, X_t, and $\Delta_c X_t$ onto the space spanned by the basis vectors to obtain $\mathbb{W}_i^{y}, \mathbb{W}_i^{x}$, and $\mathbb{W}_i^{\Delta_c x}$.

(iii) Estimate

$$\mathbb{W}_i^{y} = \mathbb{W}_i^{x\prime}\beta_0 + \mathbb{W}_i^{\Delta_c x\prime}\delta_0 + \mathbb{W}_{\alpha,i}^{0\cdot x} \quad \text{for } i = 1,\ldots,K. \tag{28}$$

by OLS with $\mathbb{W}_{\alpha,i}^{0\cdot x}$ as the regression error.

(iv) Conduct inferences in the usual way, treating the above as the CNLRM. More specifically, use the F- and t-distributions to construct confidence intervals and perform hypothesis testing.

4. A SIMULATION STUDY

In this section, we compare the finite sample performances of the proposed F-test with those of the conventional chi-squared tests. There has already been some simulation evidence that the F-test (or the t-test) has more accurate size than the chi-squared tests (or the normal tests) in some nonstationary regressions, such as the linear trend regression in Sun (2011) and the cointegration regression with endogeneity in Hwang and Sun (2018). We have covered a few other nonstationary regressions in this chapter. In light of the space constraint and given that this chapter is a tribute to Joon Park's seminal contributions to econometrics, we consider only the cointegration regression of Phillips and Park (1988) where the regressors are strictly exogeneous. More specifically, we consider the following data-generating process with three regressors $X_t = (1, \tilde{X}_t')' \in \mathbb{R}^3$ and $\tilde{X}_t = \left(\tilde{X}_{t1}, \tilde{X}_{t2}\right)' \in \mathbb{R}^2$:

$$\begin{aligned} Y_t &= X_t'\beta_0 + u_t, \\ \tilde{X}_t &= \tilde{X}_{t-1} + u_{x,t}, \\ u_t &= \rho u_{t-1} + \epsilon_{yt}, \\ u_{x,t} &= \rho u_{x,t-1} + \epsilon_{xt}, \end{aligned}$$

where $\left(\epsilon_{yt}, \epsilon_{xt}'\right)'$ is i.i.d. $N(0, I_3)$.[4] Since $\left\{\epsilon_{yt}\right\}_{t=1}^T$ and $\left\{\epsilon_{xt}\right\}_{t=1}^T$ are independent, $\tilde{X}_t$ is strictly exogeneous. The parameter ρ controls the persistence of u_t and each component of u_{xt}. We consider the following values of ρ:

$$\rho \in \{0.05, 0.20, 0.35, 0.50, 0.75, 0.90\}.$$

Without loss of generality, we set the true coefficient vector to be $\beta_0 = \left(\beta_{10}, \beta_{20}, \beta_{30}\right)' = (1,1,1)'$.

We test $H_0 : R\beta_0 = r$ against $H_1 : R\beta_0 \neq r$ where

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } r = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

so that $p = 2$, that is, we perform a test on whether the coefficients β_{20} and β_{30} on the nonconstant regressors are jointly one. We consider two sample sizes $T = 100$ and $T = 200$. The number of simulation replications is 10,000.

In Phillips and Park (1988), $\{u_t\}$ is known to follow an autoregressive (AR) process with a known AR order, and hence the long-run variance of $\{u_t\}$ can be estimated by running an AR regression based on the OLS residuals. In contrast, here we assume that we do not know the true data-generating process for $\{u_t\}$, and we estimate the long-run variance of $\{u_t\}$ nonparametrically.

Depending on the long-run variance estimator and the critical value used, we consider three groups of 5% tests. The first group consists of two tests that are based on the series long-run variance estimator. Using the Fourier basis functions given in (10), we compute the test statistic

$$F_{T,\text{Fourier}} = \left(R\hat{\beta} - r\right)' \left[R\left(X'X\right)^{-1} R\right]^{-1} \left(R\hat{\beta} - r\right) / \left(p\hat{\omega}_u^2\right),$$

where $\hat{\omega}_u^2 = (KT)^{-1}\left(\phi'\hat{u}\right)'\left(\phi'\hat{u}\right)$. See equations (2) and (3). Both tests in this group are based on the test statistic $F_{T,\text{Fourier}}$. The first test uses the simulated critical value from the nonstandard distribution F_∞ given in (5) and is referred to as "Fourier-F_∞." The second test uses the 5% critical value from $\chi_2^2/2$, a (normalized) chi-squared distribution[5] and is referred to as "Fourier-χ_2^2."

The second group of tests is similar to the first group but is based on the transformed Fourier series. The first test in this group is the F-test detailed at the end of Section 2. The second test in this group uses the same test statistic but employs the 5% critical value from $\chi_2^2/2$. We refer to these two tests as "Transformed-Fourier-$F_{2,K}$" and "Transformed-Fourier-χ_2^2," respectively.

The third group consists of three chi-squared tests that use kernel estimators of the long-run variance ω_u^2 of $\{u_t\}$. We include the Bartlett, Parzen, or Quadratic Spectral (QS) kernels in our simulation study, but here we focus on the case with the QS kernel, as the results for the other two cases are qualitatively similar. Given the QS kernel $k_{QS}(\cdot)$, we first construct

$$\tilde{\omega}_u^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{u}_t\hat{u}_s k_{QS}\left(\frac{t-s}{bT}\right),$$

where b is a smoothing parameter, and then compare the test statistic

$$F_{T,QS} = \left(R\hat{\beta}-r\right)'\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)/\left(p\tilde{\omega}_u^2\right)$$

with the critical value from $\chi_2^2/2$. We refer to the resulting test as "QS-χ_2^2."

For the series-based tests, we need to choose K, and for the kernel-based tests, we need to choose b. We consider both pre-specified values and data-driven choices. In the former case, we set $K = 8$ and 16. Hwang and Sun (2018) consider these two values of K in their simulation study and provide some justifications. Nevertheless, these two values should be regarded as rule of thumb choices. For each $K = 8$ or 16, we obtain the following comparable value of b:

$$b = \left(\int_{-\infty}^{\infty}k_{QS}^2(x)\,dx\cdot K\right)^{-1} = 1/K.$$

Under the above relationship between K and b, the Fourier series and QS kernel long run variance (LRV) estimators have the same asymptotic variance under the conventional asymptotics.

For the data-driven choices of K and b, we use the Mean Squared Error (MSE) based rules developed by Phillips (2005) and Andrews (1991), respectively. We employ the AR(1) plug-in implementation. After fitting an AR(1) model to the residual process $\{\hat{u}_t\}$ by OLS, we compute

$$\hat{K} = 0.7134\left[\hat{\alpha}(2)\right]^{-1/5}T^{4/5} \text{ and } \hat{b}_{QS} = 1.3221\left[\hat{\alpha}(2)T\right]^{1/5}/T,$$

where $\hat{\alpha}(2) = 4\hat{\rho}^2(1-\hat{\rho})^{-4}$ and $\hat{\rho}$ is the estimated AR coefficient. These data-driven choices of K and b can be justified in the asymptotic framework of Andrews

(1991), but they are not necessarily most suitable for testing problems. It is beyond the scope of this chapter to derive a testing-optimal choice of K or b along the lines of Sun et al. (2008).

Fig. 1 reports the empirical type I errors of the five tests when $K = 8$ and $b = 1/8$ for $T = 100$ and $T = 200$. The results for the case with $K = 16$ are qualitatively similar and are omitted here. Fig. 1 shows that the empirical type I errors of the chi-squared tests are substantially larger than the nominal significance level. Ignoring the estimation error in the long-run variance estimator, each chi-squared test over-rejects the null hypothesis. In contrast, the proposed F-test and the non-standard F_∞ test have very accurate size. The empirical null rejection probability for both tests is very close to the nominal significance level, except when ρ is large (i.e., when ρ is larger than 0.75). We note that the F-test and the nonstandard F_∞ test achieve similar size accuracy. Given that there is no need to simulate critical values, we recommend the more convenient F-test.

In Fig. 1, the size distortion of the chi-squared tests does not decrease significantly with the sample size. The reason is that the smoothing parameter K is fixed at $K = 8$ and the smoothing parameter b is fixed at $1/8$. Hence, the estimation uncertainty in the LRV estimator remains more or less the same across the two sample sizes $T = 100$ and $T = 200$. As a result, the size performance of the chi-squared tests does not improve much as the sample size increases.

Fig. 2 reports the empirical type I errors when K and b are data-driven. The F-test and the nonstandard F_∞ test still have more accurate size than any of the chi-squared tests, and they achieve more or less the same size accuracy. However, the F-test is more convenient to use and hence is recommended.

Another observation from Fig. 2 is that the size distortion of all tests decreases as the sample size increases from 100 to 200. This is a feature of data-driven choices of the smoothing parameters. Unreported results show that as the sample size increases, the average value of the data-driven K's increases and the average value of the data-driven b's decreases. Hence, a larger sample size leads to LRV estimators with smaller variability. As a result, all tests become less size-distorted as the sample size increases.

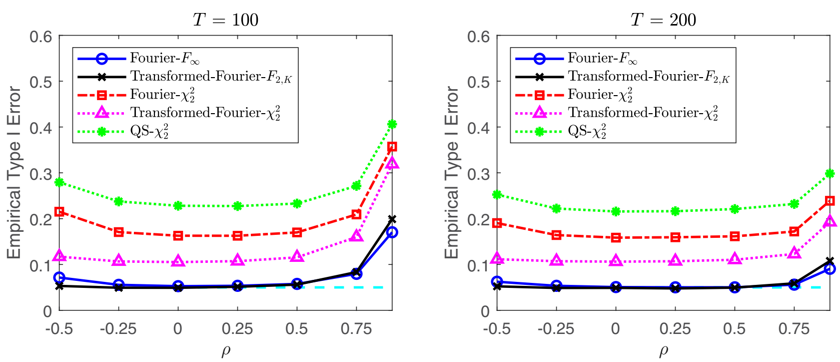


Fig. 1. Empirical Type I Error of Different 5% Tests With Smoothing Parameters $K = 8$ and $b = 1/K$.

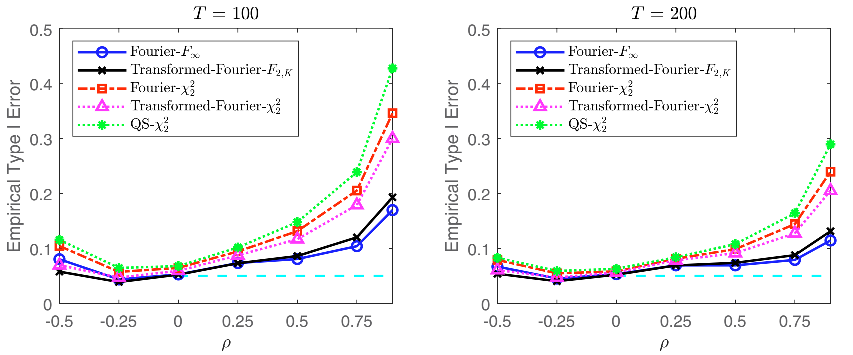


Fig. 2. Empirical Type I Error of Different 5% Tests With Data-driven Smoothing Parameters.

We have simulated the size-adjusted power curves. The two tests in the first group "Fourier-F_∞" and "Fourier-χ_2^2" have the same size-adjusted power, as they are based on the same test statistic. Similarly, the two tests in the second group "Transformed-Fourier-Transformed-Fourier-$F_{2,K}$" and "Transformed-Fourier-χ_2^2" have the same size-adjusted power. It suffices to consider only three tests for the size-adjusted power comparison, and we denote the three tests by "Fourier," "Transformed-Fourier," and "QS" in our power figure.

We consider the same data-generating processes as before but now $\beta_1 = \beta_{10}$ and

$$(\beta_2, \beta_3)' = (\beta_{20}, \beta_{30})' + \theta / T$$

for some θ. For each simulation replication, we draw a different value of θ uniformly over a circle. We plot the size-adjusted power as a function of the radius $\|\theta\|$ in Fig. 3 when $\rho = 0.5$ and $K = 8$. Fig. 3 shows that "Transformed-Fourier" is slightly more powerful than "Fourier" and "QS." For other parameter

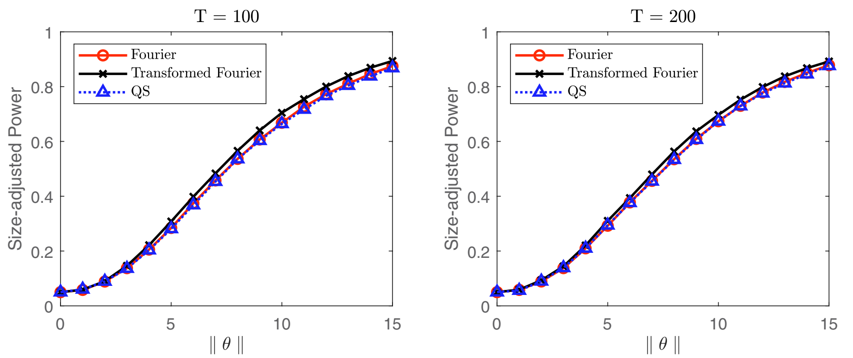


Fig. 3. Size-adjusted Power of Different 5% Tests With Smoothing Parameters $K = 8$ and $b = 1/K$.

configurations (i.e., different values of ρ) and smoothing parameter choices, "Transformed-Fourier" is as powerful as "Fourier" and "QS."

To sum up, we have found that for the cointegration regression considered here, the proposed F-test has more accurate size than the commonly used chi-squared tests, and it is as powerful as and sometimes more powerful than the chi-squared tests.

5. CONCLUSION

This chapter has developed the asymptotic F and t theory for regressions with nonstationary regressors of general forms. Linear and nonlinear trend functions, unit-root processes, near-unit-root processes with an additional complication from a structural break are all accommodated. While our asymptotic theory covers some existing results, it also covers many new cases that the asymptotic F and t theory is currently lacking in the literature. Depending on whether the regressors are endogenous or not, we develop the asymptotic F and t theory in different domains: the time domain or the pseudo-frequency domain. In both cases, statistical inferences are very easy to implement. In particular, in the latter case, we only need to transform our data using real matrix multiplications and then conduct the F and t tests as if we have a classical normal linear regression model. There is no need to use complex exponentials or explicitly estimate the long-run variance and half long-run variance.

As discussed before, it will be interesting to extend the theory further to allow the regressors to be fractionally integrated or follow a slow-varying trend. It will also be interesting to extend the theory to cover regressions with both nonstationary regressors, exogenous or not, and stationary regressors, such as the regressions considered by Park and Phillips (1989). The idea of using data-dependent and random basis functions in series variance estimation may be extended to functional data analysis.

NOTES

1. Here "0" stands for a vector of zeros, and its dimension may be different at different occurrences.
2. Hwang and Sun (2018) tackle the size distortion problem in a cointegration regression where $\{X_t\}$ are unit-root processes. Here, we generalize their asymptotic theory to allow $\{X_t\}$ to be near-unit-root processes. We employ the same argument as in Hwang and Sun (2018). To make the paper self-contained, we outline the main steps of the argument here.
3. We consider the cointegration regression here. For the predictive regression, we only need to change X_t into X_{t-1}.
4. Other distributions have been considered, but the simulation results are qualitatively close to what we report here.
5. More precisely, $\chi_2^2/2$ stands for the distribution of $\mathcal{Z}^2/2$ where $\mathcal{Z}^2 \sim \chi_2^2$.

REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*, 817–858.

Bunzel, H. (2006). Fixed-b asymptotics in single-equation cointegration models with endogenous regressors. *Econometric Theory*, *22*(4), 743–755.

Chang, Y., Lu, Y., & Park, J. Y. (2018). Understanding regressions with observations collected at high frequency over long span. Working Paper 2018-10. University of Sydney, School of Economics.

Grenander, U., & Rosenblatt, M. (1957). *Statistical analysis of stationary time series*. Wiley Publication in Mathematical Statistics. Wiley.

Guo, G., Sun, Y., & Wang, S. (2018). Testing for moderate explosiveness. *The Econometrics Journal*, *22*(1), 73–95.

Hjalmarsson, E. (2007). Fully modified estimation with nearly integrated regressors. *Finance Research Letters*, *4*(2):92–94.

Hwang, J., & Sun, Y. (2017). Asymptotic F and t tests in an efficient GMM setting. *Journal of Econometrics*, *198*, 277–295.

Hwang, J., & Sun, Y. (2018). Simple, robust, and accurate F and t tests in cointegrated systems. *Econometric Theory*, *34*(5), 949–984.

Jin, S., Phillips, P. C. B., & Sun, Y. (2006). A new approach to robust inference in cointegration. *Economics Letters*, *91*(2), 300–306.

Kiefer, N. M., & Vogelsang, T. J. (2002a). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory*, *18*, 1350–1366.

Kiefer, N. M., & Vogelsang, T. J. (2002b). Heteroskedasticity-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, *70*, 2093–2095.

Kiefer, N. M., & Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, *21*, 1130–1164.

Lazarus, E., Lewis, D. J., & Stock, J. H. (2021). The size-power tradeoff in HAR inference. *Econometrica*, *89*(5), 2497–2516.

Liu, C., & Sun, Y. (2019). A simple and trustworthy asymptotic t test in difference-in-differences regressions. *Journal of Econometrics*, *210*(2), 327–362.

Martínez-Iriarte, J., Sun, Y., & Wang, X. (2020). Asymptotic F tests under possibly weak identification. *Journal of Econometrics*, *218*(1), 140–177.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics*, *141*, 1331–1352.

Park, J. Y., & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*(3), 468–497.

Park, J. Y., & Phillips, P. C. B. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*(1), 95–131.

Pellatt, D., & Sun, Y. (2022). Asymptotic F test in regressions with observations collected at high frequency over long span. Forthcoming, *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2022.10.007.

Phillips, P. C. B. (2005). HAC estimation by automated regression. *Econometric Theory*, *21*, 116–142.

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, *23*(4), 557–614.

Phillips, P. C. B. (2014). Optimal estimation of cointegrated systems with irrelevant instruments. *Journal of Econometrics*, *178*, 210–224.

Phillips, P. C. B. (2015). Halbert White Jr. Memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, *13*, 521–555.

Phillips, P. C. B., & Park, J. Y. (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association*, *83*(401), 111–115.

Phillips, P. C. B., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I(1) processes. *The Review of Economic Studies*, *57*(1), 99–125.

Sun, Y. (2006). Best quadratic unbiased estimators of integrated variances in the presence of market microstructure noise. Working Paper. Department of Economics, UC San Diego.

Sun, Y. (2011). Robust trend inference with series variance estimator and testing-optimal smoothing parameter. *Journal of Econometrics*, *164*, 345–366.

Sun, Y. (2013). A heteroskedasticity and autocorrelation robust F test using orthonormal series variance estimator. *Econometrics Journal*, *16*, 1–26.

Sun, Y. (2014a). Let's fix it: Fixed-b asymptotics versus small-b asymptotics in heteroskedasticity and autocorrelation robust inference. *Journal of Econometrics*, *178*, 659–677.

Sun, Y. (2014b). Fixed-smoothing asymptotics in a two-step GMM framework. *Econometrica*, *82*, 2327–2370.

Sun, Y. (2014c). Fixed-smoothing asymptotics and asymptotic F and t tests in the presence of strong autocorrelation. *Advances in Econometrics*, *33*, 23–63.

Sun, Y. (2018). Comment on “HAR inference: Recommendations for practice” by Lazarus, Lewis, Stock, and Watson.*Journal of Business & Economic Statistics*, *36*(4), 565–568.

Sun, Y., & Kim, M. S. (2015). Asymptotic F test in a GMM framework with cross sectional dependence. *Review of Economics and Statistics*, *97*, 210–223.

Sun, Y., Philips, P. C. B., & Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica*, *76*, 175–194.

Sun, Y., & Wang, X. (2021). An asymptotically F-distributed Chow test in the presence of heteroscedasticity and autocorrelation. *Econometric Reviews*, *41*(2), 177–206.

Vogelsang, T. J., & Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics*, *178*(2), 741–760.

This page intentionally left blank

CHAPTER 12

NON-STATIONARY PARAMETRIC SINGLE-INDEX PREDICTIVE MODELS: SIMULATION AND EMPIRICAL STUDIES

Ying Zhou[a], Hsein Kew[a] and Jiti Gao[a]

[a]*Department of Econometrics and Business Statistics, Monash University, Clayton, VIC, Australia*

ABSTRACT

This chapter considers the estimation of a parametric single-index predictive regression model with integrated predictors. This model can handle a wide variety of non-linear relationships between the regressand and the single-index component containing either the cointegrated predictors or the non-cointegrated predictors. The authors introduce a new estimation procedure for the model and investigate its finite sample properties via Monte Carlo simulations. This model is then used to examine stock return predictability via various combinations of integrated lagged economic and financial variables.

Keywords: Non-linearity; non-stationarity; single-index models; stock return predictability; cointegration; constrained least squares estimator

JEL classifications: C13; C14; C32; C51

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 349–365

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A012

1. INTRODUCTION

Previous studies in the empirical literature have taken non-linearities into account when modeling financial and macroeconomic series. To account for this non-linearity feature, a wide variety of non-linear econometric models has been developed under the assumption that the regressors are stationary. However, financial and macroeconomics time-series data are usually non-stationary and cointegrated. Subsequent studies have developed estimation methods for the non-linear econometric models with non-stationary regressors. For instance, Wooldridge (1994) and Andrews and McDermott (1995) develop an asymptotic theory of estimation for stationary time series around a deterministic (but not stochastic) trend function. In a series of papers, Park and Phillips (1999), Park and Phillips (2000) and Park and Phillips (2001) develop an asymptotic theory for a class of non-linear regressions with an integrated scalar regressor. These papers continue the work of Park and Phillips (1988) and Park and Phillips (1989) on linear models in which the regressor is an integrated time series. Chang et al. (2001) and Chang and Park (2003) extend the non-linear models to allow for multiple integrated regressors. Also, Park (2002) and Chung and Park (2007) develop an asymptotic theory for non-linear heteroskedastic models.

Building on this growing literature, this chapter focuses on a parametric single-index predictive model with integrated regressors. This model: (a) accounts for the non-linearities in the data when forecasting next period's macroeconomic and financial variables; (b) helps ease the curse of dimensionality when it comes to parametric non-linear function estimation involving multivariate integrated regressors due to the spatial feature of a multivariate Brownian motion (Revuz & Yor, 2013); and (c) allows for the presence of cointegrated regressors or non-cointegrated regressors. Our work is related to the semiparametric single-index framework used by Dong et al. (2016) in which the integrated predictors are not cointegrated, and by Zhou et al. (2018) in which the integrated predictors are cointegrated.

In our parametric specification, we develop a new estimation procedure for the single-index model and show, via a Monte Carlo experiment, that it has better finite sample properties than the usual non-linear least squares estimator. To illustrate the use of the single-index model, we provide an application to predictability of US stock market returns. There is a large and growing empirical literature that attempts to predict stock returns using lagged macroeconomic and financial variables. Most empirical studies in this literature have focused on univariate linear predictive regression models, including, for example, the papers by Campbell (1987), Fama and French (1989), Pesaran and Timmermann (1995) and Choi et al. (2016). However, several subsequent papers argue that multiple predictors can explain the variation in stock returns better than a single predictor because multivariate predictive models are less prone to suffer from omitted predictors problem (Cochrane, 2011).

A linear specification for the bivariate/multivariate predictive regression is used in many empirical applications. However, Welch and Goyal (2008) find that the linear predictive models predict poorly out-of-sample than the standard historical average benchmark. We examine in this chapter whether the parametric single-index model, which explicitly allows for non-linearities in the predictive relations, can produce better out-of-sample fits than the historical average benchmark.

This chapter is organized as follows. Section 2 introduces the parametric single-index predictive model and presents a parametric estimator for this model. Section 3 examines the small-sample properties of the estimators by a Monte Carlo approach. The application to stock market return predictability is given in Section 4. Section 5 concludes.

2. PARAMETRIC SINGLE-INDEX PREDICTIVE MODEL

In this chapter, we study the estimation for a parametric single-index predictive model with non-stationary predictors

$$y_t = f\left(x'_{t-1}\theta_0, \gamma_0\right) + e_t, \quad t = 2, \ldots, T, \tag{2.1}$$

where $f(.,.)$ is a known univariate function, x_{t-1} is a d-dimensional integrated process of order one, θ_0 is a d-dimensional unknown true parameter vector that lies in the parameter set Θ, γ_0 is a m-dimensional unknown true parameter vector that lies in the parameter set Γ and e_t is a martingale difference process. The parameter sets Θ and Γ are assumed to be compact and convex subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively. In order to ensure that θ_0 is uniquely identifiable, we will need to impose $\theta'_0\theta_0 = 1$.

The linear combination $x'_{t-1}\theta_0$ in (2.1) is called the single-index component. Since x_{t-1} is a vector of $I(1)$ time series, we impose the following two assumptions on the single-index component. The first assumption rules out cointegration among the predictors, x_{t-1}, and θ_0 is the vector of single-index coefficients. By contrast, the second assumption permits cointegrated predictors via $x'_{t-1}\theta_0 \sim I(0)$ with θ_0 being the cointegrating vector.

Our single-index predictive model can be used to predict stock market returns without imposing linearity. The most commonly used predictors, such as dividend-price ratio, earning-price ratio, book-to-market ratio, term spread and dividend-payout ratio, have been found to be non-stationary and often contain an autoregressive unit root; see Table 7 of Kostakis et al. (2015) and Table 4 of Campbell and Yogo (2006). Campbell et al. (2004) estimate a linear trivariate predictive model containing earning-price ratio, term spread and book-to-market ratio as predictors. In the absence of a cointegrating relationship among these predictors, we could utilize this single-index model to reexamine the out-of-sample forecasting performance of these predictors. In the cointegrated predictors case, Lettau and Ludvigson (2001) consider $x_{t-1} = \left(c_{t-1}, a_{t-1}, y_{t-1}\right)'$ where c_t is log consumption, a_t is log asset wealth and y_t is log labor income. They show that, under the general household budget constraint framework, the predictors c_t, a_t and y_t should move together over the long run and are hence cointegrated. They demonstrate that the cointegrating residual (termed the "cay" variable) from regressing c_{t-1} on a_{t-1} and y_{t-1} is a strong predictor of stock returns when using a linear predictive model. Again, we could utilize this single-index model to re-assess the forecasting ability of the cointegrated cay predictor.

To illustrate the role played by the parameter vector $\gamma_0 = \left(\gamma_{1,0},\ldots,\gamma_{m,0}\right)'$ in (2.1), consider the case where y_t is related to the single-index component through a quadratic functional form

$$f\left(u_{t-1},\gamma_0\right) = \gamma_{1,0} + \gamma_{2,0}u_{t-1} + \gamma_{3,0}u_{t-1}^2,$$

where $u_{t-1} = x_{t-1}'\theta_0$. Here γ_0 is the vector of coefficients for the single-index components. Non-zero elements of γ_0 indicate that the single-index component is a useful predictor of y_t. The single-index model (2.1) includes as a special case the simple neural network model – with $f\left(u_{t-1},\gamma_0\right) = \gamma_{1,0} + \gamma_{2,0}G\left(\gamma_{3,0} + u_{t-1}\right)$ for a known function $G(.)$ – considered by Chang and Park (2003).

In the case where x_{t-1} is a univariate integrated predictor, Park and Phillips (2001) consider a parametric non-linear non-stationary regression model: $y_t = f\left(x_{t-1},\gamma_0\right) + e_t$ They use a non-linear least squares (NLS) estimator to estimate their model and show that the limiting distribution of this estimator is a functional of a univariate Brownian motion process. An extension of their univariate framework to include multivariate predictors suffers from the curse of dimensionality problem because the limiting distribution would naturally be a function of a d-dimensional vector Brownian motion processes. As the dimension of x_t increases, the vector Brownian motion processes is known to exhibit increasing spatial behavior; see Revuz and Yor (2013). The use of the single-index method avoids this spatial problem because it reduces the dimension of multivariate predictors to a univariate linear combination $x_{t-1}'\theta_0$, and thereby eliminates the necessity for vector Brownian motion processes in the development of the limit theory.

As in Chang and Park (2003), the single-index model (2.1) can be estimated by a NLS estimator. Define the sum-of-squared-errors by

$$Q_n\left(\theta,\gamma\right) = \sum_{t=1}^{T}\left(y_t - f\left(x_{t-1}'\theta,\gamma\right)\right)^2.$$

The NLS estimator $\hat{\theta}$ and $\hat{\gamma}$ is given by minimizing $Q_n\left(\theta,\gamma\right)$ over $\theta \in \Theta$ and $\gamma \in \Gamma$, that is

$$\left(\hat{\theta},\hat{\gamma}\right) = \arg\min_{\theta\in\Theta,\gamma\in\Gamma} Q_n\left(\theta,\gamma\right). \tag{2.2}$$

The solutions to (2.2) must be found numerically because there is no closed-form solution. The '*nloptr*' package in R with Augmented Lagrangian Algorithm can be used as an optimization routine to numerically solve (2.2).

In an attempt to improve the finite sample properties of the NLS estimator, we truncate the squared errors $\left(y_t - f\left(x_{t-1}'\theta,\gamma\right)\right)^2$ and we impose a constraint on the coefficient vector θ in the estimation procedure. To this end, we define the modified sum-of-squared-errors by

$$Q_{n,m}(\theta,\gamma)=\sum_{t=1}^{T}\left(y_t-f\left(x_{t-1}'\theta,\gamma\right)\right)^2 I\left(\|x_{t-1}\|\le M_T\right)+\lambda\left(\|\theta\|^2-1\right), \tag{2.3}$$

where $I(.)$ denotes the indicator function, $\|.\|$ is the Euclidean norm, M_T is a positive and increasing sequence satisfying $M_T\to\infty$ as $T\to\infty$ and λ is a Lagrange multiplier.

The reason for truncating the squared-errors $\left(y_t-f\left(x_{t-1}'\theta,\gamma\right)\right)^2$ in (2.3) is that the presence of integrated predictors will tend to produce far too few observations at distinct spatial locations, especially so when using relatively small-sample sizes and the parametric single-index function $f(u,\gamma)$ is non-bounded and non-integrable. These observations may cause a standard optimization routine to fail to converge when solving (2.2). This truncation method was originally used by Li et al. (2016) for the case when the univariate regressor x_{t-1} follows a null recurrent Markov process, which is known to exhibit spatial structure.

The constrained least squares (CLS) estimator $\tilde{\theta}$ and $\tilde{\gamma}$ is given by minimizing $Q_{n,m}(\theta,\gamma)$ over $\theta\in\Theta$ and $\gamma\in\Gamma$ such that the restriction $\|\theta\|^2=1$ holds; that is

$$\left(\tilde{\theta},\tilde{\gamma}\right)=\arg\min_{\theta\in\Theta,\gamma\in\Gamma,\|\theta\|^2=1} Q_{n,m}(\theta,\gamma). \tag{2.4}$$

A constrained optimization method is required to find the solutions to (2.4) and the '*nloptr*' package in R can still be used to numerically solve (2.4) subject to the constraint that $\|\theta\|^2=1$.

The constraint $\|\theta\|^2=1$ in (2.3) scales the CLS estimator to the surface of the unit ball and Zhou et al. (2018) demonstrate that, in their semi-parametric single-index model, this constraint on the estimation procedure causes the CLS estimator of θ_0 to converge to their true values at a faster rate than that for the case without constraints. In our parametric single-index model, the simulation results in the next section show significant finite sample gains from imposing this constraint on the estimation procedure when comparing the NLS estimator with the CLS estimator.

3. SIMULATION RESULTS

In this section, we investigate the finite sample properties of the NLS and the proposed CLS estimators in multivariate non-stationary settings. The predictors x_{t-1} is a two-vector integrated time series. Data were generated on the following models:

$$y_t = f\left(x'_{t-1}\theta_0, \gamma_0\right) + e_t, \quad e_t \sim i.i.N(0,1),$$
$$x_t = x_{t-1} + v_{1t}, \quad x_0 = (0,0)',$$
$$v_{1t} = \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix} \sim i.i.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right).$$

We consider the following non-linear functions:

$$\begin{aligned} f_1(u,\gamma_0) &= \sin\left(u + \gamma_{1,0}\right), \\ f_2(u,\gamma_0) &= \cos\left(u + \gamma_{1,0}\right), \\ f_3(u,\gamma_0) &= 1 - e^{-\gamma_{1,0}\left(u-\gamma_{2,0}\right)^2}, \\ f_4(u,\gamma_0) &= \gamma_{1,0} e^{-\gamma_{2,0}u^2}, \\ f_5(u,\gamma_0) &= \gamma_{1,0} + \gamma_{2,0}u + \gamma_{3,0}u^2. \end{aligned} \tag{3.1}$$

with $\gamma_{1,0} = 0.2$, $\gamma_{2,0} = 0.3$ and $\gamma_{3,0} = 0.3$. The first four functions are bounded on $\mathbb{R}$ and the last one is unbounded on $\mathbb{R}$.

We consider sample sizes $T = 100, 500, 1{,}000$ and we use $M = 1{,}000$ simulation replications. For the CLS estimation method, we follow Li et al. (2016) by choosing $M_T = \sqrt{T}$. Let $\theta_0 = \left(\theta_{1,0}, \theta_{2,0}\right)'$. To evaluate the finite sample performance of the NLS and proposed CLS estimators, we compute the bias and standard deviation of each element of $\hat{\theta}$ and $\tilde{\theta}$ defined in the previous section. For example, let $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2\right)'$ and $\hat{\theta}_i^{(j)}$ denote the j-th replication of the estimate $\hat{\theta}_i$ for $i = 1, 2$. Then, for the NLS estimator, we have

$$\text{bias} = \bar{\hat{\theta}}_i - \theta_{i,0},$$

where $\bar{\hat{\theta}}_i = M^{-1}\sum_{j=1}^{M}\hat{\theta}_i^{(j)}$; and

$$\text{standard deviation (s.d.)} = \sqrt{M^{-1}\sum_{j=1}^{M}\left(\hat{\theta}_i^{(j)} - \bar{\hat{\theta}}_i\right)^2}.$$

Since $\hat{\theta}_1$ and $\hat{\theta}_2$ are correlated, we also calculate a type of estimated covariance of the form:

$$\sigma_\theta = \frac{1}{M}\sum_{j=1}^{M}\left(\hat{\theta}_1^{(j)} - \bar{\hat{\theta}}_1\right)\left(\hat{\theta}_2^{(j)} - \bar{\hat{\theta}}_2\right), \quad \text{std}_\theta = |\,\sigma_\theta\,|. \tag{3.2}$$

We first consider the case in which the predictors are not cointegrated, so that the single-index component $x'_{t-1}\theta_0$ is purely non-stationary by setting $\theta_0 = (0.8, 0.6)'$. The simulation results are reported in Table 1. As we can see, the biases, standard deviations and σ_θ for the NLS and CLS estimators decrease as the sample size increases. These results are promising, and indicate that both

Table 1. Finite Sample Properties of NLS and CLS Estimators: Non-cointegrated Predictors.

			CLS			NLS		
			$T = 100$	$T = 500$	$T = 1{,}000$	$T = 100$	$T = 500$	$T = 1{,}000$
$f_1(u, \gamma_0)$	$\theta_{1,0}$	Bias	0.00422	0.00112	0.00038	0.01497	0.00532	0.00369
		s.d.	0.02191	0.00600	0.00365	0.08951	0.04345	0.03039
	$\theta_{2,0}$	Bias	−0.00068	−0.00013	−4.5E-05	−0.00227	−0.00250	0.00200
		s.d.	0.00271	0.00061	0.00037	0.15105	0.06067	0.04121
		std_θ	0.00845	0.00689	0.00137	0.10982	0.07392	0.03368
	$\gamma_{1,0}$	Bias	−0.00779	−0.00032	−0.00152	0.00582	0.00326	0.00266
		s.d.	0.17540	0.07832	0.05733	0.13945	0.05851	0.04129
$f_2(u, \gamma_0)$	$\theta_{1,0}$	Bias	0.00447	0.00139	0.00064	−0.00128	−0.00016	−0.00011
		s.d.	0.023040	0.00642	0.00336	0.03094	0.00755	0.00390
	$\theta_{2,0}$	Bias	−0.00073	−0.00016	−7E–05	−0.00058	−0.00003	−0.00010
		s.d.	0.00292	0.00066	0.00034	0.02912	0.00722	0.00386
		std_θ	0.00845	0.00689	0.00137	0.11256	0.29867	0.01399
	$\gamma_{1,0}$	Bias	−0.00211	−0.0057	−0.00161	0.04223	0.00609	−0.00229
		s.d.	0.16254	0.08333	0.05583	0.09154	0.04721	0.02063
$f_3(u, \gamma_0)$	$\theta_{1,0}$	Bias	−0.00011	−0.00005	0.00007	0.00634	0.00409	0.00380
		s.d.	0.00371	0.00312	0.00254	0.02708	0.01484	0.01076
	$\theta_{2,0}$	Bias	0.00007	0.00003	−0.00006	0.06275	0.01113	0.00459
		s.d.	0.00267	0.00225	0.00195	0.19590	0.07324	0.05620
		std_θ	0.00345	0.00279	0.00011	0.03922	0.00981	0.00136
	$\gamma_{1,0}$	Bias	−0.00083	0.00460	0.00861	0.01342	0.00349	0.00185
		s.d.	0.02793	0.14505	0.16913	0.03588	0.0.00877	0.00576
	$\gamma_{2,0}$	Bias	0.09702	0.07387	0.03148	0.03655	0.01197	0.00696
		s.d.	1.21163	0.90087	0.51446	0.05355	0.02329	0.01733
$f_4(u, \gamma_0)$	$\theta_{1,0}$	Bias	−0.00047	−0.00054	−0.00028	0.03116	0.03489	0.02974
		s.d.	0.03926	0.00974	0.00485	0.41504	0.15343	0.10855
	$\theta_{2,0}$	Bias	−0.00049	0.00012	4E-05	−0.00729	0.01939	0.01315
		s.d.	0.02934	0.00853	0.00444	0.37102	0.14480	0.08879
		std_θ	0.00552	0.00398	0.00275	0.00992	0.00584	0.00412
	$\gamma_{1,0}$	Bias	0.00288	−0.00023	−0.00011	0.01621	−0.00864	−0.00925
		s.d.	0.02392	0.00738	0.00336	0.09407	0.05596	0.03377
	$\gamma_{2,0}$	Bias	0.00171	−0.00087	−0.00009	−0.00210	0.00759	0.01689
		s.d.	0.03928	0.00664	0.00874	0.37198	0.16559	0.11484
$f_5(u, \gamma_0)$	$\theta_{1,0}$	Bias	−0.00077	−0.00001	0.00000	0.00237	0.00167	0.00094
		s.d.	0.00571	0.00090	0.00041	0.01033	0.00494	0.00326
	$\theta_{2,0}$	Bias	0.00006	0.00000	0.00000	0.00220	0.00139	0.00068
		s.d.	0.00053	0.00009	0.00004	0.01132	0.00483	0.00329
		std_θ	0.00053	0.00019	0.00008	0.00444	0.00192	0.00107
	$\gamma_{1,0}$	Bias	−0.00071	0.00167	0.00168	−0.00003	−0.00049	−0.00013
		s.d.	0.01609	0.00513	0.00322	0.00809	0.00296	0.00214
	$\gamma_{2,0}$	Bias	0.00008	0.00000	0.00000	−0.00174	−0.00111	−0.00062
		s.d.	0.00472	0.00060	0.00028	0.00446	0.00238	0.00148
	$\gamma_{3,0}$	Bias	0.00014	−0.00003	−0.00001	0.00270	0.00107	0.00045
		s.d.	0.00595	0.00084	0.00040	0.00581	0.00358	0.00233

NLS and CLS are consistent estimators of θ_0. Table 1 also shows that our proposed CLS estimator exhibits a much better finite sample performance than the NLS estimator, especially much smaller finite sample biases and standard deviations across all experiments. It is thus useful to incorporate the truncation method and impose a constraint on the coefficient vector θ in the estimation procedure.

We now allow for the possibility of cointegration among the predictors. Following Zhou et al. (2018), we set $\theta_0 = (0.6, -0.8)'$ with

$$\begin{aligned} x_t &= x_{t-1} + v_{2t}, \\ v_{2t} &= \epsilon_t + C\epsilon_{t-1}, \end{aligned}$$

where $\epsilon_t \sim i.i.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$ and $C = \begin{pmatrix} -1 & 4/3 \\ 0 & 0 \end{pmatrix}$.

We report simulation results in Table 2 only for the CLS estimator since in the non-cointegrated case we find that there are finite sample gains when resorting to a constrained estimation procedure. As can be seen, the performance of the CLS estimator tends to improve as the sample size increases across the different regression functions. Deriving the theoretical properties of the CLS estimator under both non-cointegrated and cointegrated predictors should be a good topic for future research.

4. EMPIRICAL ILLUSTRATION: STOCK MARKET RETURN PREDICTABILITY

In this section, we use the single-index model containing multivariate integrated regressors to predict future US stock market returns. We focus on out-of-sample forecasts because it involves forecasters making predictions in real time (Stock & Watson, 2007). The dataset used for this study is available from Amit Goyal's website. The examined period is quarterly data from 1956:Q1 to 2018:Q4. The dependent variable, y_t, is the equity premium defined as the S&P500 value-weighted log excess returns.

We select a combination of predictors x_{t-1} that has been found to be important, both theoretically and empirically, in prior studies of stock return predictability. In particular, the following combinations of non-cointegrated predictors are used in the single-index component: (a) dividend-price ratio and T-bill rate, see Ang and Bekaert (2007); (b) dividend-price ratio, T-bill rate, default yield spread and term spread, see Ferson and Schadt (1996); (c) dividend-price ratio and book-to-market ratio, see Kothari and Shanken (1997); (d) dividend-price ratio and dividend-payout ratio, see Lamont (1998); (e) earning-price ratio, term spread and book-to-market ratio, see Campbell et al. (2004). The following four combinations of cointegrated predictors are considered in Zhou et al. (2018); (f) dividend-price ratio and dividend yield; (g) T-bill rate and long-term yield; (h) dividend-price ratio and earning-price

Table 2. Finite Sample Properties of CLS Estimator: Cointegrated Predictors.

		$f_1(u,\gamma_0)$			$f_2(u,\gamma_0)$			$f_3(u,\gamma_0)$			
		$\gamma_{1,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\gamma_{2,0}$	$\theta_{1,0}$	$\theta_{2,0}$
$T = 100$	Bias	−0.00707	0.00814	−0.03127	−0.00511	0.00815	−0.02885	−0.01325	0.03478	0.02894	0.56838
	std	0.05976	0.08368	0.22101	0.05561	0.07984	0.19458	0.18462	0.25850	0.39893	0.86347
	std_θ		0.21981			0.19987				0.05635	
$T = 500$	Bias	−0.00049	0.00065	−0.00102	−0.00115	0.00124	−0.00589	−0.01193	−0.03216	0.02289	0.11934
	std	0.01789	0.02420	0.07311	0.01135	0.01536	0.05785	0.06733	0.08842	0.0.07056	0.21906
	std_θ		0.07272			0.06180				0.01955	
$T = 1000$	Bias	−0.00007	0.00083	−0.00120	−0.00066	0.00075	−0.00257	−0.00434	0.00793	0.00173	0.09899
	std	0.00812	0.01115	0.04286	0.00764	0.01072	0.03726	0.01899	0.03194	0.03956	0.09878
	std_θ			0.04471			0.03731				0.01266

		$f_4(u,\gamma_0)$				$f_5(u,\gamma_0)$				
		$\gamma_{1,0}$	$\gamma_{2,0}$	$\theta_{1,0}$	$\theta_{2,0}$	$\gamma_{1,0}$	$\gamma_{2,0}$	$\gamma_{3,0}$	$\theta_{1,0}$	$\theta_{2,0}$
$T = 100$	Bias	0.04778	−0.05835	0.42574	0.44352	0.00244	−0.00290	0.00176	−0.00725	−0.02559
	std	0.22502	30026	79131	1.35994	0.00982	0.01532	0.02586	0.09341	0.05667
	std_θ			0.09709					0.10301	
$T = 500$	Bias	−0.00359	0.00408	0.09030	17635	0.00041	−0.00047	0.00077	−0.00025	−0.00352
	std	0.07905	0.10786	0.32777	0.69730	0.00386	0.00609	0.00896	0.03894	0.02578
	std_θ				0.04742					0.05841
$T = 1000$	Bias	−0.00266	0.00241	0.05001	0.07098	4.37E-05	−0.00027	0.00008	−0.00049	−0.00103
	std	0.05233	0.07447	0.23882	0.37198	0.00266	0.00418	0.00532	0.02730	0.03349
	std_θ			0.008577438					0.03156	

ratio; (i) baa- and aaa-rated corporate bond yields; and (j) c_t, a_t and y_t, see Lettau and Ludvigson (2001).

We use the same non-linear regression functions (namely $f_1(u,\gamma), f_2(u,\gamma), f_3(u,\gamma), f_4(u,\gamma)$ and $f_5(u,\gamma)$) as in the Monte Carlo section. The single-index model is estimated using the CLS estimator with $M_T = \sqrt{T}$. We consider one-step ahead forecasting of the equity premium using all the 10 combinations of predictors and the five non-linear functions, for a total of 50 single-index predictive models. We generate pseudo out-of-sample forecasts from a sequence of recursive predictions. To examine the forecasting performance, we follow Campbell and Thompson (2008) and compute the out-of-sample R^2 statistic:

$$R^2_{\text{OOS}} = 1 - \frac{T_2^{-1}\sum_{s=1}^{T_2}\left(y_{T_1+s} - \hat{y}_{T_1+s}\right)^2}{T_2^{-1}\sum_{s=1}^{T_2}\left(y_{T_1+s} - \bar{y}_{T_1+s}\right)^2},$$

where $T_2 = T - T_1$ with T_1 is the number of observations in the (in-sample) estimation period, $\hat{y}_{t+1} = f\left(x_t'\tilde{\theta}, \tilde{\gamma}\right)$ and $\bar{y}_{t+1} = t^{-1}\sum_{s=1}^{t} y_s$ is the historical average benchmark forecast. The single-index model has a forecasting performance that beats the benchmark if $R^2_{\text{OOS}} > 0$. We report out-of-sample results in a set of figures that show the R^2_{OOS} statistic on the vertical axis and the beginning of the various out-of-sample evaluation periods on the horizontal axis. This will provide us with an assessment of the robustness of the out-of-sample forecasting results.

We present only the combinations of predictors and non-linear functions that are found to produce positive R^2_{OOS} values over a substantial number of out-of-sample periods. When using $f_3(u,\gamma)$ as a non-linear function, the following combinations of predictors generate out-of-sample gains: (a) earning-price ratio, book-to-market ratio and term spread (see Fig. 1); (b) dividend-price ratio and dividend-payout ratio (see Fig. 2); (c) c_t, a_t and y_t (see Fig. 3). In addition, Fig. 4 displays the results when we use pairs of cointegrated predictors considered in Zhou et al. (2018). As this Fig. 4 shows the pair of long-term yield and three-month T-bill rate produces $R^2_{\text{OOS}} > 0$ when using $f_1(u,\gamma)$, $f_4(u,\gamma)$ and $f_5(u,\gamma)$ as non-linear functions. We also find that when using $f_3(u,\gamma)$ as a non-linear function, the following three pairs of predictors generate positive R^2_{OOS} values: (a) dividend-price ratio and earning-price ratio; (b) dividend-price ratio and dividend yield; and (c) baa- and aaa-rated corporate bond yields. These findings indicate that exploiting non-linearities in the data can lead to improved forecast accuracy relative to the historical average. Overall, the single-index predictive model shows that the largest out-of-sample forecast gains in all the figures come from using the combination of aaa- and baa-rated corporate bond rates, and using the non-linear function $f_3(u,\gamma)$.

While we have shown that the single-index predictive model produces $R^2_{\text{OOS}} > 0$, it is also interesting to test $H_0 : R^2_{\text{OOS}} \le 0$ against $H_A : R^2_{\text{OOS}} > 0$. When comparing forecasts from nested models, Clark and West (2007) suggest the mean

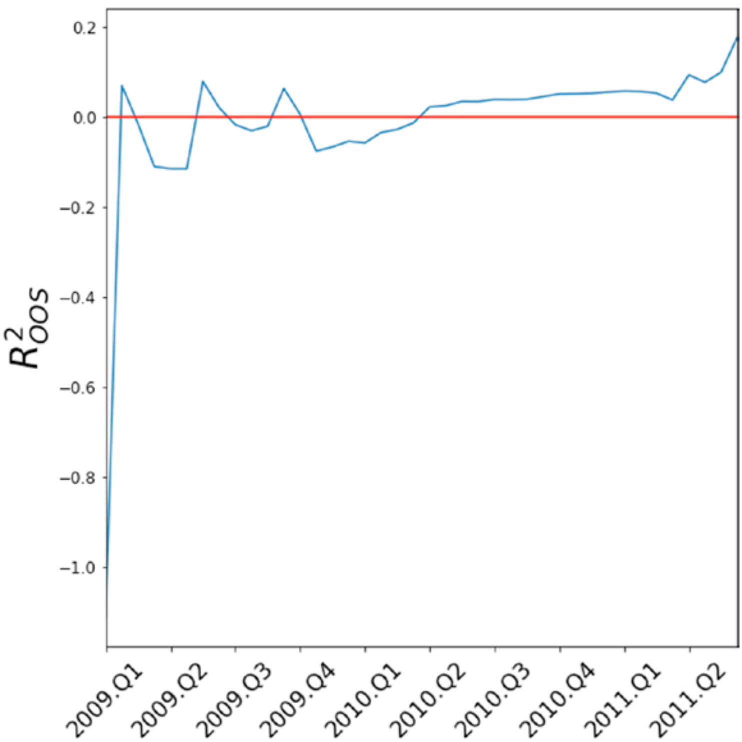


Fig. 1. Predictors: Earning–Price Ratio, Book-to-Market Ratio, Term Spread and $f_3(u,\gamma)$.

square prediction error (MSPE)-*adjusted* statistic, which is the *t*-statistic on the constant coefficient from the regression of f_{T_1+s} on a constant, where

$$f_{T_1+s} = \left(y_{T_1+s} - \overline{y}_{T_1+s}\right)^2 - \left[\left(y_{T_1+s} - \hat{y}_{T_1+s}\right)^2 - \left(\overline{y}_{T_1+s} - \hat{y}_{T_1+s}\right)^2\right],$$

for $s = 1,\dots,T_2$. The one-sided (upper-tail) critical value for this statistic can be obtained from the standard normal distribution. We report this statistic in the fourth and eighth columns of Table 3.

We use 1956:Q1 to 2001:Q4 as the initial estimation period and so the out-of-sample period is from 2002:Q1 to 2018:Q4. Table 3 reports the R^2_{OOS} statistics for all the five non-linear functional forms considered in Section 3 and using the five combinations of cointegrated predictors and two combinations of non-cointegrated predictors. Table 3 shows that a number of the R^2_{OOS} statistics are positive in the third column. Nevertheless, the three positive R^2_{OOS} statistics (for $f_3(u,\gamma)$ with the combination of dividend-price ratio (dp) and earning-price ratio (ep); $f_4(u,\gamma)$ with the combination of T-bill rate (tbl) and long-term yield (lty); and

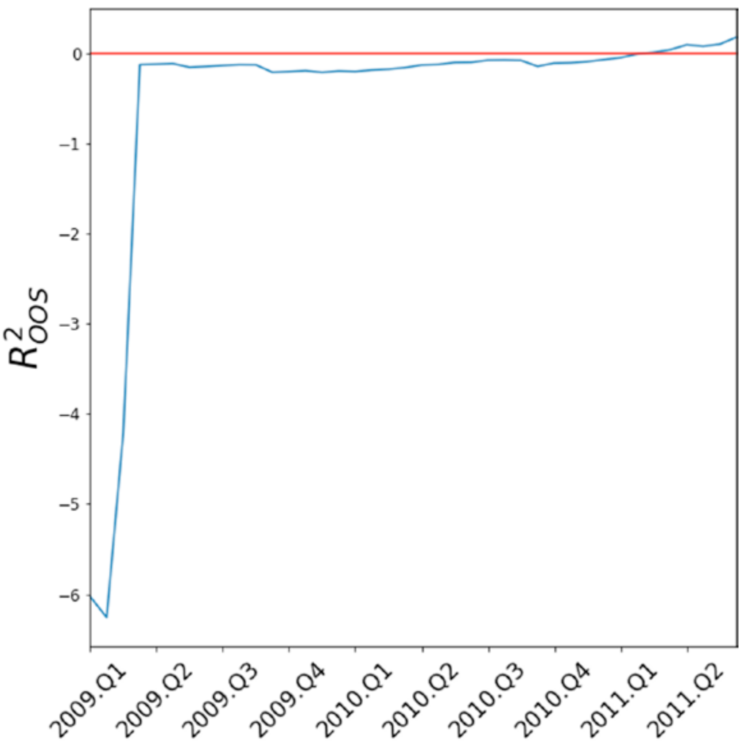


Fig. 2. Predictors: Dividend-price, Dividend–payout and $f_3(u,\gamma)$.

$f_4(u,\gamma)$ with the combination of earning-price ratio, book-to-market ratio (bm) and term spread (tms)) indicate that their MSPE-*adjusted* statistics are significantly greater than zero at the 10% level.[1] These tests, thus, suggest that allowing for non-linearities in the $f\left(x'_{t-1}\theta,\gamma\right)$ specification displays statistically significant out-of-sample predictive value for stock returns. As a comparison, we consider a linear function $f_6\left(u,\gamma_0\right)=\gamma_{1,0}+\gamma_{2,0}u$ for the predictive model, and report the results in Table 3. All seven of the R^2_{OOS} statistics are negative and thus suggesting that the linear predictive regression fails to outperform the historical average benchmark.

We next calculate the realized utility gain (or certainty equivalent return (CER), gain) for a mean-variance optimizing investor over the out-of-sample period. It measures the economic gains from using forecasts obtained from the single-index model relative to the historical average model for the risk-averse investor. Following Campbell and Thompson (2008), the investor, at the end of quarter t, allocates the share of the portfolio to equities during quarter $t+1$ as $\omega_t=\gamma^{-1}\hat{r}_{t+1}\hat{\sigma}^{-2}_{t+1}$ where $\hat{r}_{t+1}$ is a forecast of the simple equity premium and $\hat{\sigma}^2_{t+1}$ is variance of forecasts. The investor allocates the remaining share $1-\omega_t$ to risk-free bills and hence the quarter-$(t+1)$ portfolio return is $R_{p,t+1}=\omega_t r_{t+1}+R_{f,t+1}$,

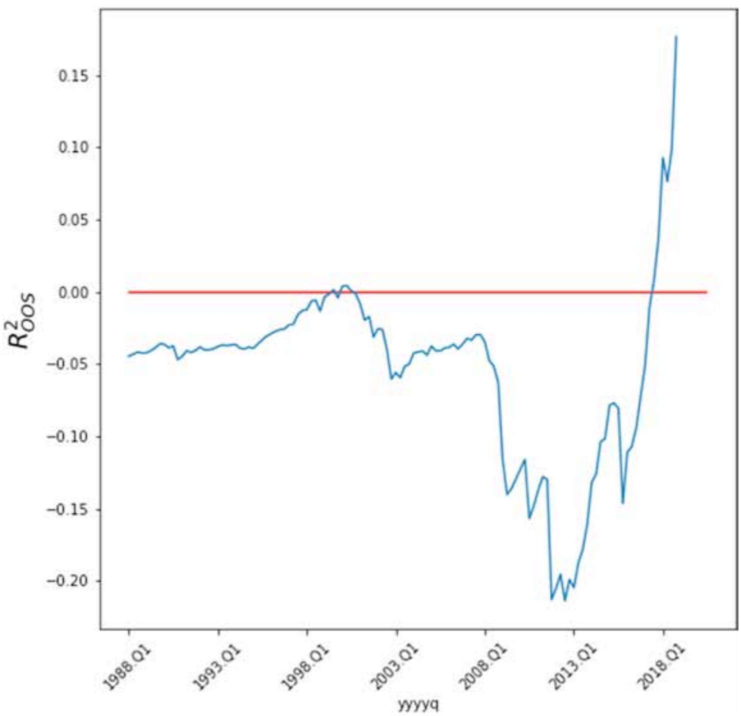


Fig. 3. Predictors: Log Consumption, Log Asset Wealth and Log Labor Income and $f_3(u,\gamma)$.

where r_{t+1} is the equity premium and $R_{f,t+1}$ is the risk-free rate. Following Neely et al. (2014) and Campbell and Thompson (2008), the investor estimates the variance of the equity premium using a five-year rolling window of past returns and imposes the constraint that the portfolio weight ω_t lies between 0 and 1.5 in each quarter.

The CER for the portfolio is $\text{CER}_p = \hat{\mu}_p + 0.5\gamma\hat{\sigma}_p^2$ where $\hat{\mu}_p$ is the mean for the portfolio and $\hat{\sigma}_p^2$ is the variance for the portfolio over the out-of-sample period. The CER gain is the difference between the CER for the investor who uses forecasts based on the proposed single-index model (2.1) and the CER for the investor who uses the historical average forecast. We express, as is common, this CER gain in annualized percentage return (by multiplying it by 400).[2] The fifth and last columns in Table 3 give the CER gains for the relative risk coefficient $\gamma = 5$. There are quite a number of cases where the CER gains are positive, which suggest that the mean-variance investors enjoy economic benefits from using single-index predictive regression forecasts relative to the historical average. In particular, the non-linear specification $f_2(u,\gamma)$ and the cointegrated pair of T-bill rate and long-term yield produce the largest CER gain of 298 basis points. When using $f_1(u,\gamma)$ and $f_2(u,\gamma)$ with T-bill rate and long-term yield as predictors and using $f_2(u,\gamma)$ with baa- and

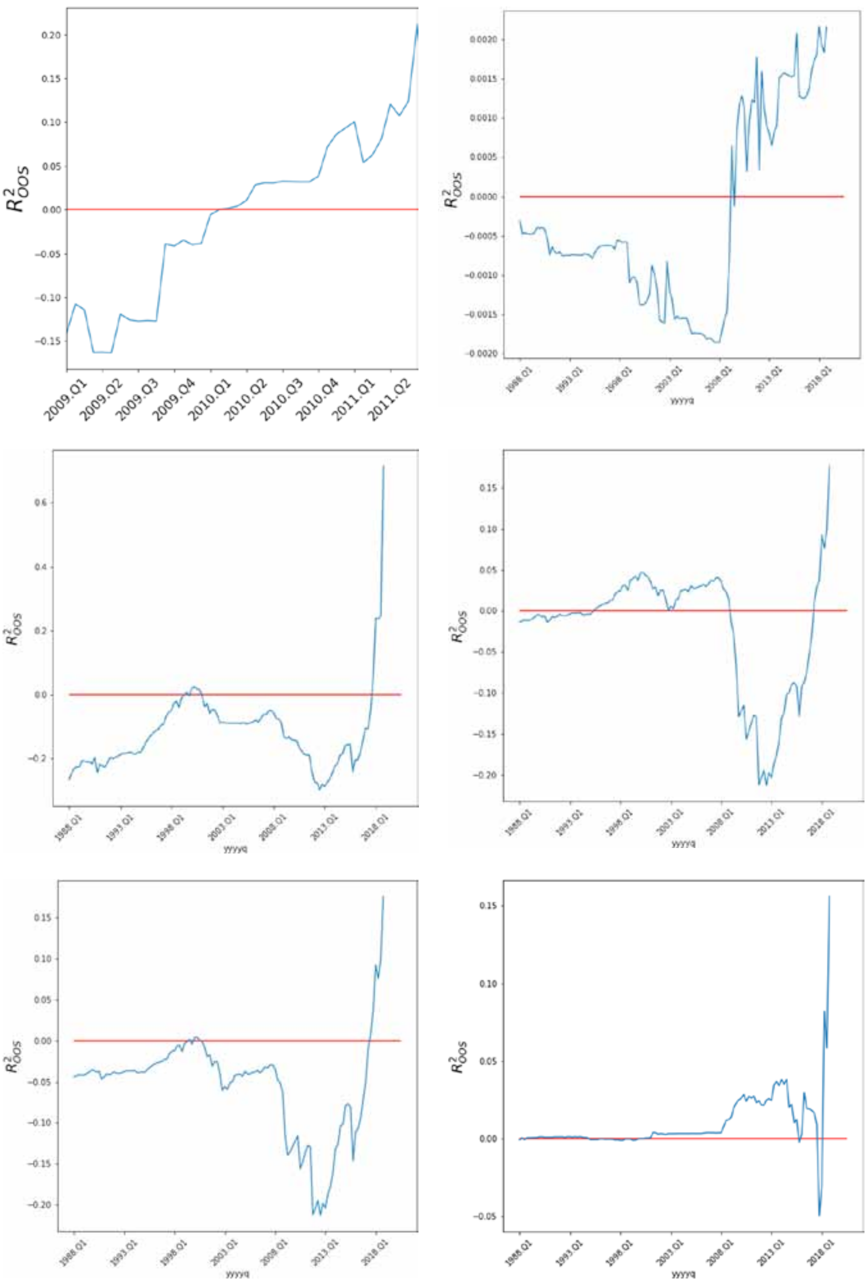


Fig. 4. Forecasting Results Using Cointegrated Predictors. (a) Predictors: Long-term Yield and T-bill Rate and $f_1(u,\gamma)$. (b) Predictors: Long-term Yield and T-bill Rate and $f_4(u,\gamma)$. (c) Predictors: Long-term Yield and T-bill Rate and $f_5(u,\gamma)$. (d) Predictors: Dividend-price and Earning-price and $f_3(u,\gamma)$. (e) Predictors: Dividend-price and Dividend Yield and $f_3(u,\gamma)$. (f) Predictors: BAA and AAA and $f_3(u,\gamma)$.

Table 3. Out-of-Sample Forecasting Period: 2002:Q1–2018:Q4.

Function	Predictors	R^2_{OOS}	MSPE-adj	CER Gains	Predictors	R^2_{OOS}	MSPE-adj	CER Gains
$f_1(u,\gamma)$	dp, dy	0.291	−0.1852	−1.36	ep, bm, tms	−0.165	1.3373*	−0.01
	tbl, lty	−0.020	−0.4700	2.06	dp, de	−0.221	1.0108	−10.59
	dp, ep	−0.151	0.0606	−2.81	c, a, y	−0.049	0.9668	−1.92
	baa, aaa	0.044	−0.0195	−0.32				
$f_2(u,\gamma)$	dp, dy	−0.147	0.1763	−5.50	ep, bm, tms	−0.167	1.4923*	−0.01
	tbl, lty	−0.016	1.0662	2.98	dp, de	−0.196	0.9960	0.19
	dp, ep	−0.178	2.3239**	2.33	c, a, y	−0.079	1.8943*	−10.34
	baa, aaa	0.051	0.8020	1.13				
$f_3(u,\gamma)$	dp, dy	0.154	−1.7763	−2.97	ep, bm, tms	0.016	1.0831	−10.06
	tbl, lty	0.00	0.9345	0.55	dp, de	−0.073	1.5312*	−10.59
	dp, ep	0.153	1.5763*	−2.97	c, a, y	−0.046	1.5310*	−10.59
	baa, aaa	0.153	0.7625	0.27				
$f_4(u,\gamma)$	dp, dy	−0.160	1.4830*	0.98	ep, bm, tms	0.092	1.3319*	−0.81
	tbl, lty	0.030	2.3102**	0.29	dp, de	−0.410	0.5125	−1.36
	dp, ep	−0.160	1.4830*	0.98	c, a, y	0.173	−1.4887	0.19
	baa, aaa	−0.014	1.2928*	−1.90				
$f_5(u,\gamma)$	dp, dy	−0.066	−0.6553	−2.02	ep, bm, tms	−0.567	1.4758*	0.04
	tbl, lty	0.656	0.4736	0.07	dp, de	−0.444	0.5302	−1.99
	dp, ep	0.071	−0.6178	0.93	c, a, y	−0.034	0.8112	−2.84
	baa, aaa	0.188	−1.8653	−1.70				
$f_6(u,\gamma)$	dp, dy	−0.176	1.0860	−0.03	ep, bm, tms	−0.023	1.2496	0.01
	tbl, lty	−0.000	0.2327	−0.00	dp, de	−0.505	0.3734	−0.03
	dp, ep	−0.583	0.3434	−0.02	c, a, y	−0.007	1.9442*	−0.02
	baa, aaa	−0.140	0.6921	−0.02				

Notes: *, ** and *** indicate 10%, 5%, and 1% significance levels.
dp, is dividend-price ratio; dy, dividend yield; tbl, T-bill rate; lty, long-term yield; ep, earning-price ratio; baa and aaa, bond yields; bm, book-to-market ratio; de, dividend-payout ratio; c, a and y are consumption, asset wealth and labor income (all in logs).

aaa-rated corporate bond yields as predictors, the CER gains are more than 100 basis points. Our results suggest that a risk-adverse investor who optimally allocates their resources between equities and risk-free bills can receive sizable utility gains by using the single-index *vis-á-vis* the historical average forecasts. Our results complement existing empirical examinations of stock return predictability, which typically focus on linear predictive models.

5. CONCLUSION

This chapter considers a parametric single-index predictive model with integrated predictors. We propose a new constrained estimation procedure to estimate this model and show that it has better finite sample properties than the usual NLS estimator. We apply the model to examine the predictability of stock returns using cointegrated and non-cointegrated predictors. We find that several combinations of the predictors used in prior studies deliver out-of-sample forecasting gains relative to the standard historical average benchmark over a large number of evaluation periods when using the single-index predictive model that accounts for the non-linearities in the time-series data.

We leave the development of an associated estimation theory and asymptotic properties for future research, because considerable efforts are needed for addressing challenging issues when considering several different scenarios: (i) estimation of non-linearly cointegrated models with constraints or without constraints; and (ii) estimation of non-cointegrated models with constraints or without constraints.

NOTES

1. Neely et al. (2014) points out that the nested MSPE-*adjusted* statistic can be statistically significant even if the R^2_{OOS} statistic is negative.
2. The calculations were performed via a Matlab program available on Dave Rapach's website.

ACKNOWLEDGMENTS

The authors thank Yoosoon Chang, two anonymous referees and Gael Martin for helpful suggestions on an earlier draft. This chapter is based on the second chapter of the first author's doctoral dissertation at Monash University. The third author also acknowledges the financial support by the Australian Research Council Discovery Grants Program under Grant Number: DP170104421.

REFERENCES

Andrews, D. W., & McDermott, C. J. (1995). Nonlinear econometric models with deterministically trending variables. *The Review of Economic Studies*, *62*(3), 343–360.

Ang, A., & Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, *20*(3), 651–707.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, *18*(2), 373–399.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Campbell, J. Y., Vuolteenaho, T., & Ramadorai, T. (2004). *Caught on tape: Predicting institutional ownership with order flow*. Harvard Institute of Economic Research Discussion Paper (2046).

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, *81*(1), 27–60.

Chang, Y., & Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, *114*(1), 73–106.

Chang, Y., Park, J. Y., & Phillips, P. C. (2001), Nonlinear econometric models with cointegrated and deterministically trending regressors. *The Econometrics Journal*, *4*(1), 1–36.

Choi, Y., Jacewitz, S., & Park, J. Y. (2016). A reexamination of stock return predictability. *Journal of Econometrics*, *192*(1), 168–189.

Chung, H., & Park, J. Y. (2007). Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics*, *137*(1), 230–259.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291–311.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Dong, C., Gao, J., & Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *The Annals of Statistics*, *44*(1), 425–453.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, *25*(1), 23–49.

Ferson, W. E., & Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *The Journal of Finance*, *51*(2), 425–461.

Kostakis, A., Magdalinos, T., & Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, *28*(5), 1506–1553.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics*,*44*(2), 169–203.

Lamont, O. (1998). Earnings and expected returns. *The Journal of Finance*, *53*(5), 1563–1587.

Lettau, M., & Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, *56*(3), 815–849.

Li, D., Tjøstheim, D., & Gao, J. (2016). Estimation in nonlinear regression with Harris recurrent Markov chains. *The Annals of Statistics*, *44*(5), 1957–1987.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science*, *60*(7), 1772–1791.

Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics*, *110*(2), 383–415.

Park, J. Y., & Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, *4*(3), 468–497.

Park, J. Y., & Phillips, P. C. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, *5*(1),95–131.

Park, J. Y., & Phillips, P. C. (1999), 'Asymptotics for nonlinear transformations of integrated time series'. *Econometric Theory*, *15*(3), 269–298.

Park, J. Y., & Phillips, P. C. (2000). Nonstationary binary choice. *Econometrica*, *68*(5), 1249–1280.

Park, J. Y., & Phillips, P. C. (2001). Nonlinear regressions with integrated time series. *Econometrica*, *69*(1), 117–161.

Pesaran, M. H., & Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance*, *50*(4), 1201–1228.

Revuz, D., & Yor, M. (2013). *Continuous martingales and Brownian motion* (Vol. 293). Springer Science & Business Media.

Stock, J. H., & Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*, 3–33.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. *Handbook of Econometrics*, *4*, 2639–2738.

Zhou, W., Gao, J., Kew, H., & Harris, D. (2018). Semiparametric single-index predictive regression. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3214042

This page intentionally left blank

CHAPTER 13

BEST LINEAR PREDICTION IN COINTEGRATED SYSTEMS

Yun-Yeong Kim

ABSTRACT

This chapter introduces the best linear predictor (BLP) with the asymptotic minimum mean squared forecasting error (MSFE) among linear predictors of variables in cointegrated systems. Accordingly, the authors show that (i) if the autocorrelation coefficient of the cointegration error between the prediction time and the predicted targeting time is larger than ½ (representing a short prediction period), then the BLP is deduced from the random walk model; and (ii) in other cases (representing a long prediction period), the BLP is deduced from the cointegration model. Under this scheme, we suggest a switching predictor that automatically selects the random walk or cointegration model according to the size of the estimated autocorrelation coefficient. These results effectively explain the superiority reversal in the short- and long-term prediction of the exchange rate between the random walk and the structural/cointegration model (known as the Meese–Rogoff or disconnect puzzle).

Keywords: Best linear prediction; random walk model; cointegration model; cointegration error; autocorrelation coefficient; switching predictor

JEL classification: C3

Essays in Honor of Joon Y. Park: Econometric Theory
Advances in Econometrics, Volume 45A, 367–391

Published under exclusive licence by Emerald Publishing Limited
ISSN: 0731-9053/doi:10.1108/S0731-90532023000045A013

1. INTRODUCTION

"*And, behold, I come quickly; and My reward is with Me, to give every man according as his work shall be*." (Revelation Chapter 22:12-14)

The random walk model is known to perform effectively in the short-run forecasting of economic variables, including the exchange rate and the stock price, while the structural models based on fundamentals performs relatively well in long-term forecasting. In particular, Meese and Rogoff (1983a) found that no model using macro-fundamentals (e.g., structural or VAR models based on the monetary approach) could beat the random walk model in out-of-sample short-term forecasts. However, Mark (1995) and Meese and Rogoff (1983b) found that the random walk model no longer consistently has the lowest MSFE for long-term forecast horizons. In particular, the cointegration model has the potential to be superior to the random walk model in long-term prediction because it explains the long-run equilibrium of the exchange rate. This is known as the Meese–Rogoff or disconnect puzzle in exchange rate theory.

The existing approach to solving the disconnect puzzle focuses on improving the prediction model using fundamentals, presupposing that this model does not beat the random walk model because of modeling errors, nonlinearity, endogeneity, and parameter instability. Some studies have tried to improve the prediction performance of structural models by using different datasets, more sophisticated techniques, or new variables (Chinn & Meese, 1995; MacDonald & Marsh, 1997; Mark, 1995; Mark & Sul, 2001; Meese & Rogoff, 1983b). Others have attempted to apply economy-wide macro-econometric models (Gandolfo & Padoan, 1990). Some researchers have introduced nonlinearity into the exchange rate models (Kilian & Taylor, 2003; Taylor & Peel, 2000). The relative superiority of long-term prediction using cointegration information has also been proved by Engle and Yoo (1987), Lin and Tsay (1996), and Reinsel and Ahn (1992).

In this regard, this chapter introduces the BLP with the asymptotic minimum MSFE among linear predictors of variables in cointegrated systems. Accordingly, we show that (i) if the autocorrelation coefficient of the cointegration error between the prediction time and the predicted targeting time is larger than ½ (representing a short prediction period), then the BLP is deduced from the random walk model; and (ii) in other cases (representing a long prediction period), the BLP is deduced from the cointegration model.[1] Under this scheme, we suggest a switching predictor that automatically selects the random walk or cointegration model according to the size of the estimated autocorrelation coefficient. These results effectively explain the superiority reversal in the short- and long-term prediction of the exchange rate between the random walk and the structural/cointegration model (known as the Meese–Rogoff or disconnect puzzle). The results imply that the difference in predictive performance between the random walk and the structural/cointegration model is not directly due to the insufficient use or misuse of economic theory but rather a result of the difference in the information used (especially the memory of cointegration errors) in the prediction.

The rest of the chapter proceeds as follows. Section 2 introduces the switching predictor of the random walk and cointegration models and Section 3 presents the prediction with additional cointegration error dynamics information. Section 4 reports the Monte Carlo simulation results and Section 5 presents the application results. Finally, Section 6 concludes the study.

2. SWITCHING PREDICTOR OF THE RANDOM WALK AND COINTEGRATION MODELS

Consider Phillips' (1991) triangular representation of a system with a cointegrating relationship of I(1) variables y_{1t} and y_{2t}:

$$\underset{(1\times1)}{y_{1t}} = \underset{(1\times1)}{\delta} + \underset{(1\times g)}{\Gamma'}\underset{(g\times1)}{y_{2t}} + \underset{(1\times1)}{z_t} \tag{2.1}$$

and

$$\underset{(g\times1)}{\Delta y_{2t}} = \underset{(g\times1)}{\mu} + \underset{(g\times1)}{u_t} \tag{2.2}$$

for $t = 1, 2, \ldots, T$, where δ and μ are constant terms, Γ is a cointegration coefficient vector, z_t is the cointegration error, $\Delta y_{2t} \equiv y_{2t} - y_{2t-1}$ is a time difference, and $u_t \equiv (u_{1t}, u_{2t}, \ldots, u_{gt})'$ are mean zero stationary processes, respectively.

For instance, if $z_t = \phi z_{t-1} + \varepsilon_t$ and $\Gamma = 1$ in equations (2.1) and (2.2), then the above AR(1) process of z_t is in line with Kaul (1996, p. 276), Poterba and Summers (1988), and Summers (1986) and is often referred to as a "fad" model of the stock price. Poterba and Summers (1988) and Summers (1986) argued that the long temporary price swings assumed in models of an inefficient market imply a slowly decaying stationary price component z_t.

We aim to predict a variable y_{1t+h} for $h \in Z^+$, where Z^+ denotes a set of positive integers. For example, y_{1t} could denote the exchange rate and y_{2t} the relative price representing the purchasing power parity. Frankel's (1979) model of exchange rate determination includes income, monetary aggregates, and short- and long-term interest rates (Meese & Rogoff, 1983a, p. 5).

We consider a predictor of y_{1t+h} at time t using information set $\Omega_t \equiv (y_{11}, y_{21}', y_{12}, y_{22}', \ldots, y_{1t}, y_{2t}')'$. In particular, we consider the following class of linear form predictor:

$$F_{<x_t>} \equiv \theta + \Theta' x_t \tag{2.3}$$

which is a linear combination of $x_t \equiv (y_{1t}, y_{2t}', y_{1t-1}, y_{2t-1}', \ldots, y_{1t-p+1}, y_{2t-p+1}')'$, where θ is a constant and Θ is a $p(g+1)\times 1$ vector of the coefficient and $p(g+1) < T$. Thus, equation (2.3) includes predictions using the vector-autoregressive (VAR) or autoregressive (AR) model with variables of x_t.

Before proceeding, we define the estimator of cointegration error $\hat{z}_t \equiv y_{1t} - \hat{\delta} - \hat{\Gamma}' y_{2t}$, where

$$(\hat{\delta}, \hat{\Gamma}')' \equiv \left(\sum_{i=1}^{T} \begin{bmatrix} 1 & y_{2i} \\ y_{2i}' & y_{2i} y_{2i}' \end{bmatrix} \right)^{-1} \sum_{i=1}^{T} \begin{bmatrix} y_{1i} \\ y_{2i} y_{1i} \end{bmatrix} \tag{2.4}$$

is an ordinary least square (OLS) estimator of the cointegration coefficient $(\delta, \Gamma')'$ of equation (2.1) at time t using a given sample of size T.

Now, we assume the following standard regularity conditions:

Assumption 2.1 *Suppose that*[2]

(a) $T^{-1}\sum_{t=1}^{T} z_{t+l}u_{t+k} = o_p(1)$ *for any integer l or k*;

(b) $T^{-1}\sum_{t=1}^{T} z_t = O_p(1)$;

(c) $T^{-2}\sum_{t=1}^{T} y_{2t} = O_p(1)$;

(d) $T^{-3/2}\sum_{t=1}^{T} y_{2t}z_{t+h} = O_p(1)$;

(e) $T^{-3/2}\sum_{t=1}^{T} (y_{2t+h} - y_{2t})y_{2t}' = O_p(1)$;

(f) $T^{-3}\sum_{t=1}^{T} y_{2t}y_{2t}' = O_p(1)$;

(g) $T^{-1}\sum_{t=1}^{T} z_{t+h}z_t \rightarrow_p E(z_{t+h}z_t)$;

(h) $T^{-1}\sum_{t=1}^{T} u_t = o_p(1)$;

(i) $T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{h} u_{t+i}\right)\left(\sum_{i=1}^{h} u_{t+i}\right)' \rightarrow_p \Sigma_h < +\infty$.

(j) $\left[\begin{array}{c} T^{1/2}(\hat{\delta}-\delta) \\ T^{3/2}(\hat{\Gamma}-\Gamma) \end{array}\right] = O_p(1)$.

Remark 2.2 (i) Assumption 2.1(a) implies that a law of large numbers holds where z_t and u_t are independent from each other[3]; and $E(u_t) = 0$[4]; (ii) Assumption 2.1(c)–(f) hold because $\mu \neq 0$ and then y_{2t} has a drift term. For this, see Hamilton (1994, Proposition 17.3); and (iii) see Hamilton (1994, 7.2.15) where Assumption 2.1(g) holds in a certain stationarity assumption. ■

Further, we define $\hat{\rho}_h \equiv \hat{\gamma}_h / \hat{\gamma}_0$ as an estimator of the h-th autocorrelation of cointegration error z_t, $\rho_h \equiv \gamma_h / \gamma_0$, where $\gamma_h \equiv E(z_{t+h}z_t)$ and $\hat{\gamma}_h \equiv T^{-1}\sum_{i=1}^{T} \hat{z}_{i+h}\hat{z}_i$.

Consider a linear combination of random walk predictor $F_1(\equiv h\hat{\Gamma}'\hat{\mu} + y_{1t})$ and cointegration model predictor $F_0(\equiv h\hat{\Gamma}'\hat{\mu} + \hat{\delta} + \hat{\Gamma}'y_{2t})$ as follows:

$$F_w \equiv wF_1 + (1-w)F_0 \ [\equiv F_0 + w\hat{z}_t] \tag{2.5}$$

where w is a real constant and $\hat{\mu} \equiv T^{-1}\sum_{i=1}^{T} y_{2i}$ is an estimator of μ. Note that (2.5) is general enough to cover various important predictors. For instance, if $w = 0$, then F_0 is a cointegration model, and, if $w = 1$, then F_1 is a random walk model, respectively. Further, if $w = 1_{[\hat{\rho}_h \geq 1/2]}$, then

$$F_{1_{[\hat{\rho}_h \geq 1/2]}} \equiv 1_{[\hat{\rho}_h \geq 1/2]}F_1 + (1 - 1_{[\hat{\rho}_h \geq 1/2]})F_0 \tag{2.6}$$

is a switching predictor that automatically selects the random walk or cointegration model according to the size of the estimated autocorrelation coefficient,[5] where $1_{[\hat{\rho}_h \geq 1/2]}$ is an index function such that

$$1_{[\hat{\rho}_h \geq 1/2]} = \begin{cases} 0 \text{ if } \hat{\rho}_h < 1/2 \\ 1 \text{ if } \hat{\rho}_h \geq 1/2 \end{cases}$$

and, by definition,

$$F_{1_{[\hat{\rho}_h \geq 1/2]}} = \begin{cases} F_0 \text{ } if \text{ } \hat{\rho}_h < 1/2 \\ F_1 \text{ } if \text{ } \hat{\rho}_h \geq 1/2. \end{cases} \tag{2.7}$$

Remark 2.3. (a) In a certain case, Phillips' (1991) triangular representation might be interpreted as a random walk model since we may derive the following equation from (2.1) and (2.2):

$$\Delta y_{1t} = \Gamma'\mu + \Gamma' u_t + \Delta z_t. \tag{2.8}$$

If $\{\Gamma' u_t + \Delta z_t\}$ is an independent and identically distributed (IID) process, then (2.8) is a random walk process with a drift.
(b) For long-term prediction, switching predictor (2.6) eventually converges to the cointegration model predictor in general. For instance, suppose that z_t is an autoregressive-moving average [ARMA (p, q)] process and that the roots of the characteristic equation lie outside the unit circle. Then, we note that

$$\lim_{h\to\infty} \gamma_h = \lim_{h\to\infty} \rho_h = 0$$

from Hamilton (1994, p. 60). In this case, we obtain

$$p\lim_{h\to\infty}(F_{1_{[\hat{\rho}_h \geq 1/2]}} - F_0) = 0 \tag{2.9}$$

for a given y_{2t} if $\hat{\delta} \to_p \delta, \hat{\Gamma} \to_p \Gamma$ and $\hat{\rho}_h \to_p \rho_h (\equiv \gamma_h / \gamma_0)$ under Assumption 2.1.[6] ■

Next, the *MSFE* of a predictor F_b is defined in the same way as follows by subscript b, which distinguishes the predictor[7]:

$$MSFE_b \equiv T^{-1} \sum\nolimits_{t=1}^{T} (y_{1t+h} - F_b)^2. \tag{2.10}$$

Theorem 2.4 *Suppose that* Assumption 2.1 *holds. Then*

(i) $MSFE_0 - MSFE_1 \to_p 2\gamma_0(\rho_h - 1/2) \begin{cases} < 0 \text{ } if \text{ } \rho_h < 1/2 \\ \geq 0 \text{ } if \text{ } \rho_h \geq 1/2, \end{cases}$

(ii) $MSFE_{1_{[\hat{\rho}_h \geq 1/2]}} = \begin{cases} MSFE_0 + o_p(1) \text{ } if \text{ } \rho_h < 1/2 \\ MSFE_1 + o_p(1) \text{ } if \text{ } \rho_h \geq 1/2. \end{cases}$

The proofs of all theorems are in the Appendix A.

Theorem 2.4 shows that (i) if the autocorrelation coefficient of the cointegration error between the prediction time and the predicted targeting time is larger than ½ (representing a short prediction period), then a predictor with the minimum *MSFE* is deduced from the random walk model; and (ii) in other cases (representing a long prediction period), a predictor with the minimum *MSFE* is deduced from the cointegration model.

These results provide a solution for the Meese–Rogoff puzzle, that is, the random walk model performs well in the short-run forecasting of exchange rates, while the structural/cointegration model based on fundamentals is relatively efficient in long-term forecasting.

Example 2.5 Suppose that $z_t = \phi z_{t-1} + \varepsilon_t$, where $|\phi| < 1$ and $\{\varepsilon_t\}$ is a mean zero IID process with a variance $\sigma_\varepsilon^2 (> 0)$. Then, we obtain

$$MSFE_0 - MSFE_1 \rightarrow_p \frac{2\phi^h - 1}{1-\phi^2}\sigma_\varepsilon^2 \begin{cases} < 0 \text{ if } \phi^h < 1/2 \\ \geq 0 \text{ if } \phi^h \geq 1/2. \end{cases} \tag{2.11}$$

where $\rho_h \equiv \phi^h$.

The above relationship (2.11) means that, when ϕ is small and the external shock for the cointegration error rapidly disappears over time, the cointegration model becomes better at prediction than the random walk model. For instance, suppose that $\phi = 0$. Then, note that

$$MSFE_0 - MSFE_1 \rightarrow_p -\sigma_\varepsilon^2 < 0$$

for any $h \in Z^+$. Further, if h becomes large, the conformable *MSFE* difference ultimately becomes

$$\lim_{h\to\infty}(MSFE_0 - MSFE_1) \rightarrow_p -\frac{1}{1-\phi^2}\sigma_\varepsilon^2 < 0. \qquad \blacksquare$$

Theorem 2.6. *Suppose that* Assumption 2.1 *holds. Then*

$$MSFE_{<x_t>} = MSFE_w + \underbrace{T^{-1}\sum\nolimits_{t=1}^{T}\tilde{\Theta}' x_t x_t{}'\tilde{\Theta}}_{O_p(T^2)} + O_p(T) \textit{ for any } w \textit{ and}$$

where $\tilde{\Theta} \equiv [0, \Gamma', 0_{(p-1)(g+1)}{}']' - \Theta\ (\neq 0)$, *with* 0_a *as an* $a \times 1$ *vector of zeroes and* $T^{-1}\tilde{\Theta}'\sum\nolimits_{t=1}^{T} x_t x_t{}'\tilde{\Theta} \geq 0$.

Theorem 2.6 shows that F_w in (2.5) is the BLP with the minimum *MSFE* among the linear predictors (*cf*. defined in (2.3)) of variables in cointegrated systems.

We may apply the above results to the case in which there are m different cointegration relationships as Γ^i (i = 1, 2, …, m). We define $\hat{z}_t^i \equiv y_{1t} - \hat{\delta}^i - \hat{\Gamma}^j y_{2t}$ and $\hat{\gamma}_h^i \equiv T^{-1}\sum\nolimits_{i=1}^{T} \hat{z}_{t+h}^i \hat{z}_t^i$, where $\hat{\delta}^i$ and $\hat{\Gamma}^i$ are OLS estimators of δ^i and Γ^i. For instance, a monetary approach model of exchange rate determination may include income, monetary aggregates, and short- and long-term interest rates (or some of them).

Example 2.7 The long-run equilibrium variables to define the cointegration of the exchange rate (y_{1t}) may be determined by different models, such as the following: (i) the basic model by Lucas (1982); (ii) the model by Bilson (1978), including short-term interest rates; (iii) Frankel's (1979) model, including the short- and long-term interest rates from the basic model; and (iv) a portfolio equilibrium model that adds stock prices to the Frankel model. In the following, * indicates the economic variables of other countries.

Basic model:

$$y_{1t} = \alpha_1(m_t - m_t^*) + \alpha_2(n_t - n_t^*) + z_t^1$$

Bilson model:

$$y_{1t} = \beta_1(m_t - m_t^*) + \beta_2(n_t - n_t^*) + \beta_3(i_t - i_t^*) + z_t^2$$

Frankel model:

$$y_{1t} = \gamma_1(m_t - m_t^*) + \gamma_2(n_t - n_t^*) + \gamma_3(i_t - i_t^*) + \gamma_4(i_{Lt} - i_{Lt}^*) + z_t^3$$

Extended model:

$$y_{1t} = \kappa_1(m_t - m_t^*) + \kappa_2(n_t - n_t^*) + \kappa_3(i_t - i_t^*) + \kappa_4(i_{Lt} - i_{Lt}^*) + \kappa_5(s_t - s_t^*) + z_t^4$$

where the income (n_t, n_t^*), monetary amount (m_t, m_t^*), short-term interest rate (i_t, i_t^*), long-term interest rate (i_{Lt}, i_{Lt}^*), and stock price (s_t, s_t^*) are the included variables. In this case, we may let $y_{2t} \equiv (m_t - m_t^*, n_t - n_t^*, i_t - i_t^*, i_{Lt} - i_{Lt}^*, s_t - s_t^*)'$, $\Gamma^1 \equiv (\alpha_1, \alpha_2, 0, 0, 0)'$, $\Gamma^2 \equiv (\beta_1, \beta_2, \beta_3, 0, 0)'$, $\Gamma^3 \equiv (\gamma_1, \gamma_2, \gamma_3, \gamma_4, 0)'$, and $\Gamma^4 \equiv (\kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5)'$. ■

Then, the optimal BLP among the candidate cointegration model predictors to minimize the *MSFE* may be derived by solving the following minimization problem:

$$\mathrm{Min}_i\, 2\hat{\gamma}_0^i \left| \hat{\rho}_h^i - \frac{1}{2} \right|, \tag{2.12}$$

for $i = 1, 2, \ldots, m$ because the following approximation from Theorem 2.4(i) holds:

$$MSFE_{0,i} = MSFE_1 + 2\hat{\gamma}_0^i \left| \hat{\rho}_h^i - \frac{1}{2} \right| + o_p(1),$$

whereas $MSFE_1 \equiv T^{-1}\sum_{t=1}^{T}(y_{1t+h} - y_{1t})^2$ is not changed by the selection of the cointegration explanatory variable y_{2t}^i.

Next, define a cointegration explanatory variable as $y_{2t}^\dagger$ corresponding to the solution to the above problem (2.12) to get the optimal BLP. Then (given multiple cointegration vectors), the conformable optimal BLP is defined as follows:

$$F^\dagger_{[\hat{\rho}_h^\dagger \geq 1/2]} \equiv 1_{[\hat{\rho}_h^\dagger \geq 1/2]} F_1 + (1 - 1_{[\hat{\rho}_h^\dagger \geq 1/2]}) F_0^\dagger, \tag{2.13}$$

where $\hat{\rho}_h^\dagger \equiv \hat{\gamma}_h^\dagger / \hat{\gamma}_0^\dagger$ and $F_0^\dagger \equiv \hat{\delta}^\dagger + \hat{\Gamma}^\dagger{}' y_{2t}^\dagger$.

3. PREDICTION WITH NON-IID COINTEGRATION ERROR DYNAMICS[8]

If the dynamic structure of cointegration error z_t is not an IID process (e.g., ARMA process), it may be considered for prediction.[9] In this case, we want to determine whether the forecasting performance of the random walk/cointegration model switching predictor presented above can be improved.

To illustrate this, we assume that $y_{1t+h}(=\Gamma' y_{2t+h}+z_{t+h})$ is predicted at time t, where $\Delta y_{2t}=u_t$[10] holds.[11] We further assume that (i) $\{\varepsilon_t, u_t\}$ is an IID process with mean zero and finite variance, independent from each other, and $E(\varepsilon_t u_t)=0$ holds and that (ii) the Γ and coefficients of the following AR(1) and MA(1) models are known.

First, suppose that z_t follows the AR(1) process below:

$$z_t=\phi z_{t-1}+\varepsilon_t, \tag{3.1}$$

where $|\phi|<1$.

In this case, the following linear predictor for y_{1t+h} can be considered:

$$F_{\varphi_h}\equiv\Gamma' y_{2t}+\varphi_h z_t, \tag{3.2}$$

where φ_h is an arbitrary constant that may depend on h.

If the *MSFE* is to be minimized, it is well known that the following linear projection predictor is optimal:

$$F_{\bar{\varphi}_h}=\Gamma' y_{2t}+\frac{\phi^h}{1+\phi^2}z_t, \tag{3.3}$$

where $\bar{\varphi}_h\equiv\dfrac{E(z_t z_{t+h})}{E(z_t^2)}=\dfrac{\phi^h}{1+\phi^2}$.

Finally, by using the property of linear projection, the following result can be obtained:

$$MSFE_{\bar{\varphi}_h}\le MSFE_{\varphi_h} \tag{3.4}$$

for any φ_h that includes 0 or 1.[12]

Second, suppose that z_t follows an MA(1) process as

$$z_t=\varepsilon_t+\theta\varepsilon_{t-1}. \tag{3.5}$$

In this case, the linear projection predictor for y_{1t+1} corresponding to note (3.3) is given by

$$F_{\bar{\varphi}_h}=\begin{cases}\Gamma' y_{2t}+\dfrac{\theta}{1+\theta^2}z_t & \text{if } h=1\\ \Gamma' y_{2t} & \text{if } h>1,\end{cases} \tag{3.6}$$

where $\bar{\varphi}_1\equiv\dfrac{E(z_t z_{t+1})}{E(z_t^2)}=\dfrac{\theta}{1+\theta^2}$.

An alternative univariate forecast (which is also the conditional expectation) for y_{1t+1} was suggested by Christoffersen and Diebold (1998, p. 13) as follows:

$$y_{1t}^{CD} \equiv \Gamma' y_{2t} + \theta\varepsilon_t \; [\equiv E(y_{1t+1} \mid I_t)], \tag{3.7}$$

assuming $\varepsilon_t \in I_t$, where I_t is an information set at time t. We may readily show that

$$MSFE_{\varphi_1} \geq MSFE_{\bar{\varphi}_1} = MSFE_{y_{1t}^{CD}} + \frac{\theta^6 + \theta^4}{1+\theta^2}\sigma_\varepsilon^2 \tag{3.8}$$

for any φ_1 that includes 0 or 1 because $MSFE_{y_{1t}^{CD}} = \sigma_\varepsilon^2$ and $MSFE_{\bar{\varphi}_1} = \left(1 + \frac{\theta^6 + \theta^4}{1+\theta^2}\right)\sigma_\varepsilon^2$.[13]

Therefore, when the constituent variables and coefficients of the AR(1) and MA(1) models are known, the AR(1) and MA(1) model-based predictors have lower MSFEs than in the random walk and cointegration models, as shown in (3.4) and (3.8). However, it seems difficult to realize this superiority in prediction using a finite sample due to the following two problems.

First, compared with the random walk and cointegration models, the AR(1) and MA(1) model predictions use one or more additional variables and coefficients, and, accordingly, additional estimation errors are generated. More seriously, in the case of the MA(1) model, it can be written as $\varepsilon_t = \sum_{i=0}^{\infty} (-\theta)^i z_{t-i}$ in Equation (3.5), so it is necessary to estimate the infinite lag term of the cointegration error z_t to estimate ε_t consistently. Second, if exact information about the dynamic structure of cointegration error z_t (e.g., the ARMA(p,q) process structure) is not known, it will increase the prediction error if it is used for prediction even though this structure is not accurate.

On the other hand, according to the above results, in the long run, the optimal predictions given by the AR(1) and MA(1) models all converge to the cointegration model, as shown in (3.3) and (3.6). In general, in the long run, a prediction obtained through any ARMA(p,q) model converges to the cointegration model (*cf*, Remark 2.3(b)), so the advantage of prediction efficiency gained through the information of the known ARMA(p,q) model disappears.

4. MONTE CARLO SIMULATION RESULTS

In this section, we conduct a Monte Carlo experiment[14] to check the small-sample properties of the suggested predictors. The basic simulation model used is of the form[15]

$$y_{1t} = \delta + \Gamma y_{2t} + z_t, \tag{4.1}$$

$$y_{2t} = \mu + \phi y_{2t-1} + u_t, \tag{4.2}$$

and

$$z_t = \psi z_{t-1} + \varepsilon_t \tag{4.3}$$

for $t = 1, 2, \ldots, 100$ and $(u_t, \varepsilon_t)' \sim \text{IID}N(0, \text{I}_2)$. The parameters are set as $\Gamma = 1$, $\phi = 1,\ 0.95$, $\psi = \pm 0.9,\ \pm 0.3$, $\delta = 0,\ 0.5$, and $\mu = 0,\ 0.5$, respectively.

Then, y_{1t+h} is predicted at time t with 100 samples, where $h = 1, 2, \ldots, 20$. The *MSFEs* of the four predictors in (4.1), (4.2), and (4.3) using a VAR (1) model (a cointegration relationship is not imposed)[16] representing a predictor $F_{<x_t>}$ are compared with one another. Subsequently, the *MSFEs* are computed by the mean of the samples from 10,000 repetitions of the above experiments.

The simulation results in Appendix B confirm the theoretical expectation, that is, when the prediction period h is small, F_1 has the smallest *MSFE*, whereas, when the prediction period h is large, F_0 has the smallest *MSFE*. On the other hand, the switching predictor $F_{[\hat{\rho}_h \geq 1/2]}$ has the smallest *MSFE* when selecting the smallest *MSFE* between F_1 and F_0. Similar results are obtained when y_{1t} is near the I(1) process (i.e., $\phi = 0.95$). Finally, the autocorrelation coefficient roughly represents the quality of prediction using the cointegration vector.

5. APPLICATION TO THE US' ECONOMIC VARIABLES

In this section, out-of-sample forecasts for the exchange rate and other economic variables are obtained using the suggested predictors in Section 3. The forecast performance is compared with the *MSFE* calculated using an h-period (10, 20, …, 180) ahead of forecast errors and an n-repeated rolling regression with a same-data number.

The data have a monthly frequency and extend from January, 1971 to July, 2019. The data period considers the timing with which the foreign exchange rate regime shifted to the floating rates from the pegged rates under the Bretton Woods system. The data source is the FRB St. Louis FRED, except for the stock price (S&P index), which is taken from R. Shiller's website (http://www.econ.yale.edu/∼shiller/data.htm).

The cointegration fundamentals (y_{2t}) are as follows: the exchange rate (i.e., yen/US dollar for Japan and US dollar/foreign currency for the other countries) uses the relative consumer price index (CPI) ratio of the two countries (Japanese CPI/US CPI for Japan and US CPI/foreign CPI for the other countries); the stock price index uses the currency portion of M1; the unemployment rate, oil price, and long-term interest rate all use the industrial production index (IPI); the short-term interest rate uses inflation (computed by the log time difference of CPIs); inflation uses M1; and, finally, the IPI uses the currency portion of M1.

First, augmented Dickey–Fuller tests are conducted to check the unit root of the variables considered. If no unit root exists, such as for Switzerland's exchange rate against the US dollar, the prediction superiority of the random walk model or the cointegration model presented in Chapter 2 is not guaranteed. See Table 1 for the unit root test results.

Table 1. Unit Root Test Results (p-Values).

	Included Terms	None	Intercept	Trend and Intercept
Exchange rate against the US dollar	UK	0.185	0.110	0.135
	Japan	0.008*	0.070	0.339
	Canada	0.760	0.311	0.651
	Switzerland	0.000*	0.001*	0.010*
	Sweden	0.805	0.545	0.319
	Norway	0.648	0.190	0.245
	Denmark	0.471	0.163	0.419
Relative price against the US CPI	UK	0.055	0.000*	0.008*
	Japan	0.189	0.789	0.001*
	Canada	0.646	0.801	0.651
	Switzerland	0.004	0.064	0.053
	Sweden	0.741	0.428	0.902
	Norway	0.840	0.262	0.614
	Denmark	0.775	0.150	0.233
Other US economic variables	Oil price	0.235	0.131	0.039
	Unemployment Rate	0.374	0.030*	0.071
	Inflation rate	0.124	0.119	0.000*
	S&P index	0.999	1.000	0.992
	Short-term interest rate	0.251	0.517	0.118
	Long-term interest rate	0.360	0.800	0.165
	Industrial production index	0.986	0.801	0.227
	M1	0.999	0.999	0.998
	Currency portion of M1	0.999	1.000	1.000

Note: For the p-value of the null hypothesis, the variable has a unit root, and the lag length is selected using the Schwarz information criterion.
*Denotes the rejection of the null hypothesis at the 5% level.

The Johansen cointegration tests using a VAR model of $(y_{1t}, y_{2t})'$ are then conducted to check whether a cointegration vector exists in the model. The null hypothesis that "the hypothesized number of the cointegration is 0" is rejected at the 5% level, except for the exchange rates of Canada, Sweden and Norway and the US long-term interest rate, using the Johansen tests.

The prediction experiments begin in April 1979, which is 100 months after the beginning of the data because 100 months' worth of data are used in the estimation of the cointegration coefficients. The estimation results in Table 3 show the relative efficiency ratio of the random walk model prediction compared with the cointegration model prediction,[17] computed as $MSFE_0 / MSFE_1$ [18]

They show that the random walk and cointegration models have relatively high levels of forecasting performance in the short and the long run, respectively. For instance, for the exchange rate of the British pound against the US dollar, the cointegration model gives better predictions than the random walk model even in the short-term prediction (e.g., over 20 months). In the case of US inflation

Table 2. Cointegration Test Results (p-Values).

	Cointegration Test Method	Trace	Max-Eigen.
Exchange rate against US dollar	UK	0.000*	0.000*
	Japan	0.000*	0.000*
	Canada	0.634	0.575
	Switzerland	0.000*	0.000*
	Sweden	0.522	0.448
	Norway	0.038	0.037
	Denmark	0.003*	0.001*
Other US economic variables	Oil price	0.010*	0.017*
	Unemployment rate	0.005*	0.006*
	Inflation	0.000*	0.000*
	S&P index	0.000*	0.000*
	Short-term interest rate	0.000*	0.000*
	Long-term interest rate	0.469	0.388
	Industrial production index	0.000*	0.000*

Note: The null hypothesis is that the hypothesized number of the cointegration is 0.
*denotes the rejection of the null hypothesis at the 5% level.

prediction, the cointegration model provides a better prediction than the random walk model even in short-term forecasts over a period of 10 months or less. For the stock price index, unemployment rate, and IPI, the relative forecast efficiency of the cointegration model over that of the random walk model starts from between 30 and 40 months, and, for the oil price and short-term interest rate, it starts from 100 months. However, if no cointegration relationship is observed (see Table 2 for the long-term interest rate), in all the cases, the random walk model's prediction is superior to that of the cointegration model (see Table 3).

The number h^c in Table 3 indicates the first switching time h in the prediction target period when the autocorrelation coefficient of the estimated cointegration error is less than ½; that is, $h^c \equiv \min\{h \mid \hat{\rho}_h < 1/2\}$. This indicates the starting time that we theoretically expect, for which the cointegration model prediction is superior to that found using the random walk model. The numbers $\hat{h}^c$ in the last row of Table 3 represent the first observed switching time of the MSFEs' size between the random walk and the cointegration model; it appears to be generally proportional to the theoretically expected switching time h^c. In this regard, the autocorrelation coefficient of the cointegration error seems to be an efficient indicator of the quality of the cointegration model's predictive performance.

6. CONCLUSION

This chapter attempted to explain why the random walk model is superior for short-term exchange rate prediction and inferior for long-term exchange rate prediction to the cointegration model using fundamentals. For this purpose, the random walk model was shown to be relatively efficient by using the cointegration error as a predictor variable in predicting the short-term exchange rate compared with the cointegration model using fundamentals.

Table 3. Relative Efficiency Ratio: $MSFE_0 / MSFE_1$.

Exchange Rates Against the US Dollar							
h	UK	Japan	Canada	Switzerland	Sweden	Norway	Denmark
10	1.916	2.866	5.630	2.736	3.263	3.114	3.483
20	**0.916**	1.569	3.171	1.296	1.757	1.930	1.731
30	**0.571**	1.182	2.191	1.002	1.235	1.416	1.276
40	**0.459**	**0.967**	1.747	**0.840**	1.070	1.206	1.074
50	**0.401**	**0.954**	1.434	**0.743**	**0.920**	1.030	**0.936**
60	**0.383**	1.009	1.215	**0.616**	**0.780**	**0.892**	**0.794**
70	**0.391**	1.042	1.059	**0.595**	**0.686**	**0.784**	**0.690**
80	**0.435**	1.117	**0.954**	**0.703**	**0.620**	**0.721**	**0.621**
90	**0.486**	1.141	**0.821**	**0.765**	**0.572**	**0.633**	**0.557**
100	**0.507**	1.106	**0.701**	**0.770**	**0.540**	**0.576**	**0.509**
110	**0.449**	1.055	**0.589**	**0.673**	**0.467**	**0.477**	**0.426**
120	**0.439**	1.008	**0.521**	**0.617**	**0.419**	**0.396**	**0.351**
130	**0.419**	**0.999**	**0.465**	**0.632**	**0.396**	**0.323**	**0.302**
140	**0.422**	1.030	**0.457**	**0.643**	**0.439**	**0.314**	**0.282**
150	**0.428**	1.062	**0.483**	**0.671**	**0.502**	**0.324**	**0.289**
160	**0.439**	1.093	**0.534**	**0.695**	**0.588**	**0.392**	**0.317**
170	**0.496**	1.112	**0.611**	**0.750**	**0.724**	**0.577**	**0.403**
180	**0.634**	1.127	**0.686**	**0.880**	**0.999**	1.153	**0.656**
h^c	17	24	34	18	26	20	21
$\hat{h}^c$	20	40	80	40	50	60	50

Other US Economic Variables							
h	Oil Price	Unemployment Rate	Inflation	Stock Price	Short-term Interest Rate	Long-term Interest Rate	Industrial Production Index
10	2.082	4.429	**0.793**	2.507	6.081	4.395	24.232
20	1.728	1.995	**0.593**	1.322	2.840	2.733	7.913
30	1.529	1.328	**0.666**	1.055	2.179	2.821	4.234
40	1.381	**0.982**	**0.686**	**0.951**	1.734	2.967	2.552
50	1.238	**0.833**	**0.765**	**0.889**	1.437	2.417	1.626
60	1.185	**0.775**	1.033	**0.835**	1.298	2.125	1.083
70	1.218	**0.718**	**0.752**	**0.833**	1.229	2.163	**0.783**
80	1.109	**0.733**	**0.668**	**0.857**	1.148	2.185	**0.610**
90	1.041	**0.785**	**0.696**	**0.867**	1.048	2.113	**0.482**
100	1.023	**0.833**	**0.777**	**0.855**	**0.980**	1.961	**0.398**
110	**0.964**	**0.923**	**0.838**	**0.847**	**0.897**	1.942	**0.362**
120	**0.970**	1.086	1.145	**0.866**	**0.809**	1.875	**0.347**
130	**0.949**	1.031	**0.937**	**0.933**	**0.747**	1.777	**0.335**
140	**0.990**	**0.877**	**0.954**	**0.923**	**0.767**	1.702	**0.324**
150	**0.969**	**0.787**	**0.976**	**0.912**	**0.790**	1.699	**0.321**
160	**0.956**	**0.722**	**0.977**	**0.931**	**0.783**	1.681	**0.329**
170	**0.927**	**0.702**	1.055	**0.960**	**0.750**	1.657	**0.342**
180	**0.901**	**0.760**	1.422	**0.991**	**0.697**	1.662	**0.355**
h^c	40	37	3	20	21	112	58
$\hat{h}^c$	110	40	<10	40	100	>100	70

The shaded areas indicate that the cointegration model's prediction is better than the random walk model's prediction.

Empirically, it was shown that the same phenomenon occurs not only with the exchange rate but also with other economic variables, such as the stock price, unemployment rate, oil price, long-term and short-term interest rates, IPI, oil price, inflation, and M1. These results imply that the random walk model might not be an efficient approximation method to describe the time-series dynamics of exchange rates and economic variables when the data frequency is low. In this sense, the disconnect puzzle (or the Meese–Rogoff puzzle) addresses not the problem of "economic theory" but the statistical/econometric issue.

Finally, the problem of expanding the case in which y_{1t} is a vector to find the vector BLP will be discussed in the next study.

DATA DESCRIPTION

Spot Crude Oil Price: West Texas Intermediate (WTI) dollars per barrel, not seasonally adjusted

Unemployment Rate: Not seasonally adjusted

Consumer Price Index: All items in the US City Average, All Urban Consumers Index 1982–1984 = 100, monthly, seasonally adjusted

S&P Composite Index: From R. Shiller's website

Short-term Interest Rate: 1-Year Treasury Constant Maturity Rate, not seasonally adjusted

Long-term Interest Rate: 10-Year Treasury Constant Maturity Rate, not seasonally adjusted

Industrial Production Index: Index 2012 = 100, seasonally adjusted

Currency Component of M1: Not seasonally adjusted

NOTES

1. Here, the autocorrelation coefficient of the cointegration error represents the benefit/cost ratio that arises because the random walk model additionally incurs the cointegration error term at the time of forecasting compared with the cointegration model. Here, the cost is induced by the increase of the predictor's own instability because of the additionally incurred cointegration error term, while the benefit is induced by its correlation with the cointegration error of the forecasting target time included in the forecasting target variable.

2. Note that μ in (2.2) cannot be zero for Assumption 2.1 to hold. So, it is assumed that μ is nonzero throughout the chapter, and then the reason will be explained when it is restricted as $\mu = 0$.

3. Independence from fundamental shock (i.e., u_t) can be justified if the shock of cointegration error z_t reflects the fad, as in Poterba and Summers (1988) and Summers (1986).

4. E denotes an expectation operator.

5. This predictor asymptotically has an *MSFE* based on F_0 or F_1 (*cf.* defined in (2.6)) and can be said to be a predictor of class defined in (2.5) asymptotically.

6. In this chapter, $p$lim is assumed to be defined when the sample size T approaches infinity.

7. For the convenience of analysis, it is assumed that the sample sizes for the model coefficient and MSFE estimation are all equal to T.

8. The further discussion here follows editor Zack Miller's suggestion, and I am deeply grateful for his kind advice, which made this important discussion possible.

9. If z_t follows a mean zero IID process, the prediction efficiency of the aforementioned cointegration or random walk model cannot be improved.

10. So, we assume that $\mu = 0$ while it does not make a difference to the result of comparing the differences in prediction performance.

11. The results obtained using the more general ARMA(p,q) model are also expected to be similar.

12. The predictor presented in (3.2) is a predictor of the class presented in (2.5), so the result of (3.4) does not contradict the contents of Theorem 2.6.

13. Therefore, the method of (3.7) can show better predictive performance than the random walk or cointegration model under the assumed conditions.

14. GAUSS 20 and EXCEL 2016 were used for the simulation.

15. Equations (4.1)–(4.3) can be expressed in a state space form as follows (and this is exploited in the simulation):

$$y_{1t} = \delta + (\Gamma, 1)(y_{2t}, z_t)'$$

$$\begin{pmatrix} y_{2t} \\ z_t \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} + \begin{pmatrix} \phi & 0 \\ 0 & \psi \end{pmatrix} \begin{pmatrix} y_{2t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix}.$$

16. It represents the unrestricted vector autoregression of Engle and Yoo (1987, Table 1, p. 151).

17. A constant term is not included in either prediction while it makes little difference to the result of comparing the differences in prediction performance.

18. If the ratio is greater than 1, the random walk model's prediction is better than that of the cointegration model.

ACKNOWLEDGMENTS

I thank God for knowing everything and leading me to the Navotas Charity Foundation (http://navotas.or.kr/) and confess that I have received all grace through it. I appreciate the comments of the editor, Zack Miller, and an anonymous referee. However, all remaining errors belong to me as the author. Lastly, I am especially grateful for the loving guidance of Professor Joon Y. Park, who was a member of my PhD committee.

REFERENCES

Bilson, J. (1978). The current experience with floating exchange rates: An appraisal of the monetary approach. *American Economic Review: Papers and Proceedings*, *68*(2), 392–397.

Chinn, M., & Meese, R. (1995). Banking on currency forecasts: How predictable is change in money? *Journal of International Economics*, *38*, 161–178.

Christoffersen, P., & Diebold, F. (1998). Cointegration and long-horizon forecasting. *Journal of Business & Economic Statistics*, *16*(4), 450–458.

Engle, R. F., & Yoo, B. S. (1987). Forecasting and testing in cointegrated systems. *Journal of Econometrics*, *35*, 143–159.

Frankel, J. A. (1979). On the mark: A theory of floating exchange rates based on real interest differentials. *American Economic Review*, *69*, 610–622.

Gandolfo, G., & Padoan, C. (1990). A presentation of the Italian continuous time model. In D. M. Christodoulakis (Ed.), *Dynamic modelling and control of national economics*. Pergamon Press.

Hamilton, J. (1994). *Time series analysis*. Princeton University Press.

Kaul, G. (1996). Predictable components in stock returns. In G. Maddala & C. Rao (Eds.), *Statistical methods in finance* (Handbook of Statistics, Vol. 14, pp. 269–296). Elsevier Science.

Kilian, L., & Taylor, M. (2003). Why is it so difficult to beat random walk forecasts of exchange rates? *Journal of International Economics*, *53*(1), 85–107.

Lin, J.-L., & Tsay, R. S. (1996). Cointegration constraints and forecasting: An empirical examination. *Journal of Applied Econometrics*, *11*, 519–538.

Lucas, R. (1982). Interest rates and currency prices in a two-country world. *Journal of Monetary Economics*, *10*(3), 335–359.

MacDonald, R., & Marsh, T. (1997). On fundamentals and exchange rates: A Casselian perspective. *Review of Economics and Statistics*, *79*, 655–666.

Mark, N. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, *85*, 201–218.

Mark, N., & Sul, D. (2001). Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel. *Journal of International Economics*, *53*(1), 29–52.

Meese, R., & Rogoff, K. (1983a). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*, 3–24.

Meese, R., & Rogoff, K. (1983b). The out-of-sample failure of empirical exchange rate models: Sampling error or misspecification? In J. A. Frenkel (Ed.), *Exchange rates and international macroeconomics* (pp. 67–112). Chicago University Press and National Bureau of Economic Research.

Phillips, P. (1991). Optimal inference in cointegrated systems. *Econometrica*, *59*, 283–306.

Poterba, J., & Summers, L. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, *22*, 27–59.

Reinsel, G. C., & Ahn, S. K. (1992). Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis*, *13*, 353–375.

Summers, L. (1986). Does the stock market rationally reflect fundamental values? *Journal of Finance*, *41*, 591–601.

Taylor, M., & Peel, M. (2000). Non-linear adjustment, long-term equilibrium and exchange rate fundamentals. *Journal of International Money and Finance*, *19*, 33–53.

APPENDIX A: PROOFS OF THEOREMS

Proposition A.1 $T^{-1}\sum_{t=1}^{T}\hat{z}_{t+h}\hat{z}_t \to_p \gamma_h \equiv E(z_{t+h}z_t)$.

Proof: Note that

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\hat{z}_{t+h}\hat{z}_t - T^{-1}\sum_{t=1}^{T}z_{t+h}z_t \\
&= T^{-1}\sum_{t=1}^{T}(\hat{z}_{t+h}-z_{t+h})(\hat{z}_t - z_t) + T^{-1}\sum_{t=1}^{T}(\hat{z}_{t+h}-z_{t+h})z_t + T^{-1}\sum_{t=1}^{T}(\hat{z}_t - z_t)z_{t+h} \\
&= T^{-1}\sum_{t=1}^{T}[(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t+h}][(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}] \\
&\quad + T^{-1}\sum_{t=1}^{T}[(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t+h}]z_t \\
&\quad + T^{-1}\sum_{t=1}^{T}[(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}]z_{t+h} \qquad \text{(A.1)}
\end{aligned}$$

[using $\hat{z}_t - z_t = (\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}$ for the second equality]

$$\begin{aligned}
&= (\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})^2 + T^{-1}(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})})'\underbrace{\sum_{t=1}^{T}y_{2t}}_{O_p(T^2)} \\
&\quad + T^{-1}(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})})'\underbrace{\sum_{t=1}^{T}y_{2t+h}}_{O_p(T^2)} + T^{-1}(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})})'(\underbrace{\sum_{t=1}^{T}y_{2t+h}y_{2t}}_{O_p(T^3)})(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})}) \\
&\quad + T^{-1}(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})\underbrace{\sum_{t=1}^{T}z_t}_{O_p(T)} + T^{-1}(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})})'\underbrace{\sum_{t=1}^{T}y_{2t+h}z_t}_{O_p(T^{3/2})} \\
&\quad + T^{-1}(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})\underbrace{\sum_{t=1}^{T}z_{t+h}}_{O_p(T)} + T^{-1}(\underbrace{\hat{\Gamma}-\Gamma}_{O_p(T^{-3/2})})'\underbrace{\sum_{t=1}^{T}y_{2t}z_{t+h}}_{O_p(T^{3/2})} = o_p(1)
\end{aligned}$$

Finally, the claimed result holds from (A.1) and Assumption 2.1(g). Q.E.D.

Proposition A.2 $MSFE_w = T^{-1}\sum_{t=1}^{T}(y_{1t+h} - h\Gamma'\mu - \delta - \Gamma'y_{2t} - wz_t)^2 + o_p(1)$.

Proof: Note that

$$\begin{aligned}
MSFE_w &\equiv T^{-1}\sum_{t=1}^{T}(y_{1t+h} - h\hat{\Gamma}'\hat{\mu} - \hat{\delta} - \hat{\Gamma}'y_{2t} - w\hat{z}_t)^2 \\
&= T^{-1}\sum_{t=1}^{T}[y_{1t+h} - h\Gamma'\mu - \delta - \Gamma'y_{2t} - wz_t + (1-w)(\hat{z}_t - z_t) + h(\Gamma'\mu - \hat{\Gamma}'\hat{\mu})]^2 \\
&= T^{-1}\sum_{t=1}^{T}(y_{1t+h} - h\Gamma'\mu - \delta - \Gamma'y_{2t} - wz_t)^2 \\
&\quad + (1-w)^2\underbrace{T^{-1}\sum_{t=1}^{T}(\hat{z}_t - z_t)^2}_{A} + h^2(\underbrace{\Gamma'\mu - \hat{\Gamma}'\hat{\mu}}_{B})^2 \\
&\quad + 2(1-w)\underbrace{T^{-1}\sum_{t=1}^{T}(y_{1t+h} - h\Gamma'\mu - \delta - \Gamma'y_{2t} - wz_t)(\hat{z}_t - z_t)}_{C}
\end{aligned}$$

$$+2h(\Gamma'\mu-\hat{\Gamma}'\hat{\mu})\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma'y_{2t}-wz_t)}_{D}$$
$$+2(1-w)h(\Gamma'\mu-\hat{\Gamma}'\hat{\mu})\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}(\hat{z}_t-z_t)}_{E}$$
$$=T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma'y_{2t}-wz_t)^2+o_p(1)$$

because

$$A\equiv T^{-1}\sum\nolimits_{t=1}^{T}(\hat{z}_t-z_t)^2$$
$$=T^{-1}\sum\nolimits_{t=1}^{T}[(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}]^2$$
$$=\underbrace{(\delta-\hat{\delta})^2}_{O_p(T^{-1})}+T^{-1}(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'\underbrace{\sum\nolimits_{t=1}^{T}y_{2t}y_{2t}{}'}_{O_p(T^3)}(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})$$
$$+2T^{-1}(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'\underbrace{\sum\nolimits_{t=1}^{T}y_{2t}}_{O_p(T^2)}=o_p(1),$$

$$B\equiv\Gamma'\mu-\hat{\Gamma}'\hat{\mu}$$
$$=(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'\mu-\hat{\Gamma}'(\underbrace{\mu-\hat{\mu}}_{O_p(T^{-1/2})})=o_p(1),$$

$$C\equiv T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma'y_{2t}-wz_t)(\hat{z}_t-z_t)$$
$$=T^{-1}\sum\nolimits_{t=1}^{T}[\Gamma'(y_{2t+h}-y_{2t})-h\Gamma'\mu+z_{t+h}-wz_t][(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}]$$
$$=T^{-1}\sum\nolimits_{t=1}^{T}[\Gamma'\sum\nolimits_{i=1}^{h}u_{t+i}+z_{t+h}-wz_t][(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}]$$
$$=(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})\Gamma'\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{h}u_{t+i}}_{o_p(1)}+\Gamma'T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{h}u_{t+i}y_{2t}{}'}_{O_p(T^{3/2})}(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})$$
$$+(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}z_{t+h}}_{O_p(1)}+T^{-1}(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'\underbrace{\sum\nolimits_{t=1}^{T}y_{2t}z_{t+h}}_{O_p(T^{3/2})}$$
$$-w(\underbrace{\delta-\hat{\delta}}_{O_p(T^{-1/2})})\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}z_t}_{O_p(1)}-wT^{-1}(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'\underbrace{\sum\nolimits_{t=1}^{T}y_{2t}z_t}_{O_p(T^{3/2})}=o_p(1),$$

$$D\equiv T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma'y_{2t}-wz_t)$$
$$=T^{-1}\sum\nolimits_{t=1}^{T}[\Gamma'\sum\nolimits_{i=1}^{h}u_{t+i}+z_{t+h}-wz_t]$$
$$=\Gamma'\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{h}u_{t+i}}_{o_p(1)}+\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}(z_{t+h}-wz_t)}_{O_p(1)}=O_p(1)$$

and

$$\begin{aligned} E &\equiv T^{-1}\sum_{t=1}^{T}(\hat{z}_t - z_t) \\ &= T^{-1}\sum_{t=1}^{T}\left[(\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}\right] \\ &= \delta-\hat{\delta}+(\underbrace{\Gamma-\hat{\Gamma}}_{O_p(T^{-3/2})})'T^{-1}\underbrace{\sum_{t=1}^{T}y_{2t}}_{O_p(T^2)} = o_p(1), \end{aligned}$$

using Assumption 2.1 (h), (j),

$$\begin{aligned} y_{1t+h} - h\Gamma'\mu - \delta - \Gamma' y_{2t} &= \Gamma'(y_{2t+h} - y_{2t}) - h\Gamma'\mu + z_{t+h} \\ &= \Gamma'\sum_{i=1}^{h} u_{t+i} + z_{t+h} \end{aligned} \tag{A.2}$$

and $\hat{z}_t - z_t = (\delta-\hat{\delta})+(\Gamma-\hat{\Gamma})'y_{2t}$ under Assumption 2.1. Q.E.D.

Theorem 2.4 (i) Note that

$$\begin{aligned} MSFE_w &= T^{-1}\sum_{t=1}^{T}(y_{1t+h} - h\Gamma'\mu - \delta - \Gamma' y_{2t} - wz_t)^2 + o_p(1) \\ &= T^{-1}\sum_{t=1}^{T}(\Gamma'\sum_{i=1}^{h}u_{t+i} + z_{t+h} - wz_t)^2 + o_p(1) \\ &= \Gamma' T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{h}u_{t+i}\right)\left(\sum_{i=1}^{h}u_{t+i}\right)'\Gamma \\ &\quad + T^{-1}\sum_{t=1}^{T}z_{t+h}^2 + w^2 T^{-1}\sum_{t=1}^{T}z_t^2 - 2wT^{-1}\sum_{t=1}^{T}z_{t+h}z_t \\ &\quad + 2\Gamma'\underbrace{T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{h}u_{t+i}\right)\left(z_{t+h} - wz_t\right)}_{o_p(1)} + o_p(1) \\ &= \Gamma'\Sigma_h\Gamma + (1+w^2)\gamma_0 - 2w\gamma_h + o_p(1). \end{aligned} \tag{A.3}$$

from Proposition A.2 for the first equality; and from Assumption 2.1(a), (g) and (i) for the final equality.

Accordingly, the asymptotic difference between the MSFEs is given as follows:

$$\begin{aligned} MSFE_0 - MSFE_1 &= 2\gamma_h - \gamma_0 + o_p(1) \\ &\equiv 2\gamma_0(\rho_h - 1/2) + o_p(1) \end{aligned}$$

from (A.3) as claimed.
(ii) First, suppose that $\rho_h \in [-1, 1/2)$. Note

$$MSFE_{1_{[\hat{\rho}_h \geq 1/2]}} - MSFE_0 \to_p 0 \tag{A.4}$$

because we may write that

$$\begin{aligned} MSFE_{1_{[\hat{\rho}_h \geq 1/2]}} = MSFE_0 + \left(1_{[\hat{\rho}_h \geq 1/2]}\right)^2 T^{-1}\sum_{t=1}^{T}\hat{z}_t^2 - 2\times 1_{[\hat{\rho}_h \geq 1/2]} \\ T^{-1}\sum_{t=1}^{T}\hat{z}_t(y_{1t+h} - y_{1t} + \hat{z}_t - h\hat{\Gamma}'\hat{\mu}) \end{aligned}$$

while $1_{[\hat{\rho}_h \geq 1/2]} \to_p 0$ from Slutsky's theorem because $1_{[\rho_h \geq 1/2]}$ is a continuous function for a $\rho_h \in [-1, 1/2)$ and $\hat{\rho}_h \to_p \rho_h$ where $T^{-1}\sum_{t=1}^{T} \hat{z}_t^2$ and $T^{-1}\sum_{t=1}^{T} \hat{z}_t (y_{1t+h} - y_{1t} + \hat{z}_t - h\hat{\Gamma}'\hat{\mu})$ are $O_p(1)$.

Second, suppose that $\rho_h \in (1/2, 1]$. Note

$$MSFE_{1_{[\hat{\rho}_h \geq 1/2]}} - MSFE_1 \to_p 0 \tag{A.5}$$

because we may write that

$$MSFE_{1_{[\hat{\rho}_h \geq 1/2]}} = MSFE_1 + \left(1 - 1_{[\hat{\rho}_h \geq 1/2]}\right)^2 T^{-1}\sum_{t=1}^{T} \hat{z}_t^2 + 2 \times \left(1 - 1_{[\hat{\rho}_h \geq 1/2]}\right) \\ T^{-1}\sum_{t=1}^{T} \hat{z}_t (y_{1t+h} - y_{1t} - h\hat{\Gamma}'\hat{\mu})$$

while $1_{[\hat{\rho}_h \geq 1/2]} \to_p 1$ from Slutsky's theorem because $1_{[\rho_h \geq 1/2]}$ is a continuous function for a $\rho_h \in (1/2, 1]$ and $\hat{\rho}_h \to_p \rho_h$ where $T^{-1}\sum_{t=1}^{T} \hat{z}_t^2$ and $T^{-1}\sum_{t=1}^{T} \hat{z}_t (y_{1t+h} - y_{1t} - h\hat{\Gamma}'\hat{\mu})$ are $O_p(1)$.

Third, suppose that $\rho_h = 1/2$. Define $\{\hat{\rho}_{h,T^+}\}$ and $\{\hat{\rho}_{h,T^-}\}$ as the T-dependent subsequences of $\{\hat{\rho}_h \equiv \hat{\rho}_{h,T}\} = \{\hat{\rho}_{h,T^+}\} \cup \{\hat{\rho}_{h,T^-}\}$ where $\hat{\rho}_{h,T^+} \geq 1/2$ and $\hat{\rho}_{h,T^-} < 1/2$ with $\{T\} = \{T^+\} \cup \{T^-\}$. Then note that

$$MSFE_{1_{[\hat{\rho}_{h,T^+} \geq 1/2]}} - MSFE_1 \to_p 0 \tag{A.6}$$

because $1_{[\hat{\rho}_{h,T^+} \geq 1/2]} = 1$ and that

$$MSFE_{1_{[\hat{\rho}_{h,T^-} \geq 1/2]}} - MSFE_0 \to_p 0 \tag{A.7}$$

because $1_{[\hat{\rho}_{h,T^-} \geq 1/2]} = 0$. Further, note that

$$MSFE_0 - MSFE_1 \to_p 0 \tag{A.8}$$

from Theorem 2.4(i) because $\rho_h = 1/2$.

Consequently, we get

$$MSFE_{1_{[\hat{\rho}_{h,T} \geq 1/2]}} - MSFE_1 \to_p 0 \tag{A.9}$$

from (A.6), (A.7), and (A.8).

Therefore, the claimed result holds from (A.4), (A.5), and (A.9). Q.E.D.

Theorem 2.6 First, from definition, note that

$$
\begin{aligned}
FMSE_{<x_t>} &= T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-\theta-\Theta' x_t)^2 \\
&= T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma' y_{2t}-wz_t+h\Gamma'\mu+\delta-\theta \\
&\quad +\Gamma' y_{2t}-\Theta' x_t+wz_t)^2 \\
&= T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma' y_{2t}-wz_t)^2 \\
&\quad +2T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma' y_{2t}-wz_t)(h\Gamma'\mu+\delta-\theta \\
&\quad +\Gamma' y_{2t}-\Theta' x_t+wz_t) \\
&\quad +T^{-1}\sum\nolimits_{t=1}^{T}(h\Gamma'\mu+\delta-\theta+\Gamma' y_{2t}-\Theta' x_t+wz_t)^2 \\
&= MSFE_w+2T^{-1}\sum\nolimits_{t=1}^{T}(y_{1t+h}-h\Gamma'\mu-\delta-\Gamma' y_{2t}-wz_t) \\
&\quad (h\Gamma'\mu+\delta-\theta+\tilde{\Theta}' x_t+wz_t) \\
&\quad +T^{-1}\sum\nolimits_{t=1}^{T}(h\Gamma'\mu+\delta-\theta+\tilde{\Theta}' x_t+wz_t)^2+o_p(1) \\
&= MSFE_w+\underbrace{2T^{-1}\sum\nolimits_{t=1}^{T}\left[\Gamma'\left(\sum\nolimits_{i=1}^{h}u_{t+i}\right)+z_{t+h}-wz_t\right](h\Gamma'\mu+\delta-\theta+\tilde{\Theta}' x_t+wz_t)}_{A} \\
&\quad +\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}(h\Gamma'\mu+\delta-\theta+\tilde{\Theta}' x_t+wz_t)^2}_{B}+o_p(1)
\end{aligned}
\tag{A.10}
$$

using Proposition A.2 and (A.2) for the fourth and fifth equality.

Then note that the claimed result holds as

$$
FMSE_{<x_t>} = MSFE_w+T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}\tilde{\Theta}' x_t x_t'\tilde{\Theta}}_{O_p(T^3)}+O_p(T)
$$

because

$$
\begin{aligned}
\underbrace{A}_{O_p(T^{1/2})} &\equiv T^{-1}\sum\nolimits_{t=1}^{T}\left[\Gamma'\left(\sum\nolimits_{i=1}^{h}u_{t+i}\right)+z_{t+h}-wz_t)(h\Gamma'\mu+\delta-\theta+\tilde{\Theta}' x_t+wz_t\right] \\
&= (h\Gamma'\mu+\delta-\theta)\Gamma'\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}\left(\sum\nolimits_{i=1}^{h}u_{t+i}\right)}_{o_p(1)}+\Gamma' T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{h}u_{t+i}x_t'}_{O_p(T^{3/2})}\tilde{\Theta} \\
&\quad +w\Gamma'\underbrace{T^{-1}\sum\nolimits_{t=1}^{T}\sum\nolimits_{i=1}^{h}u_{t+i}z_t}_{o_p(1)} \\
&\quad +(h\Gamma'\mu+\delta-\theta)T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_{t+h}}_{O_p(T)}+T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_{t+h}x_t'}_{O_p(T^{3/2})}\tilde{\Theta}+wT^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_{t+h}z_t}_{O_p(T)} \\
&\quad -w(h\Gamma'\mu+\delta-\theta)T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t}_{O_p(T)}-w\ T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t x_t'}_{O_p(T^{3/2})}\tilde{\Theta}-w^2T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t^2}_{O_p(T)}
\end{aligned}
$$

and

$$\begin{aligned}
\underbrace{B}_{O_p(T^2)} &\equiv T^{-1}\sum\nolimits_{t=1}^{T}(h\Gamma'\mu+\delta-\theta+\tilde{\Theta}'x_t+wz_t)^2 \\
&= \underbrace{(h\Gamma'\mu+\delta-\theta)^2}_{O(1)}+T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}\tilde{\Theta}'x_t x_t'\tilde{\Theta}}_{O_p(T^3)}+w^2T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t^2}_{O_p(T)} \\
&\quad +2(h\Gamma'\mu+\delta-\theta)T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}x_t'}_{O_p(T^2)}\tilde{\Theta}+2wT^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t x_t'}_{O_p(T^{3/2})}\tilde{\Theta} \\
&\quad +2w(h\Gamma'\mu+\delta-\theta)T^{-1}\underbrace{\sum\nolimits_{t=1}^{T}z_t}_{O_p(T)}
\end{aligned}$$

from (A.10) and Assumption 2.1. Q.E.D.

APPENDIX B: SIMULATION RESULTS OF MSFEs

	$\phi=1,\ \psi=0.9,\ \delta=0,\ \mu=0$					$\phi=1,\ \psi=-0.9,\ \delta=0,\ \mu=0$				
h	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.07	5.89	2.08	2.25	0.83	21.21	6.46	6.46	2.18	−0.88
2	4.13	7.18	4.27	4.74	0.68	4.12	7.42	4.14	4.50	0.77
3	5.94	8.40	6.58	7.25	0.55	21.03	8.26	8.26	6.60	−0.68
4	7.69	9.51	8.70	9.88	0.44	7.70	9.28	7.97	8.98	0.59
5	9.28	10.48	10.29	12.50	0.35	22.02	10.54	10.54	11.41	−0.52
6	10.97	11.55	11.52	15.15	0.27	10.80	11.11	11.03	13.69	0.46
7	12.46	12.54	12.56	17.91	0.20	22.85	12.51	12.51	16.18	−0.41
8	13.97	13.74	13.74	20.80	0.14	14.02	13.19	13.42	18.64	0.36
9	15.48	14.91	14.93	23.56	0.09	24.06	14.58	14.58	21.23	−0.32
10	16.85	15.96	15.98	26.37	0.05	17.18	15.47	15.75	24.19	0.28
11	18.14	16.93	16.95	28.96	0.01	25.65	16.78	16.78	26.68	−0.25
12	19.40	17.98	17.99	31.69	−0.02	19.86	17.47	17.74	29.35	0.22
13	20.85	19.18	19.18	34.70	−0.04	26.97	18.68	18.68	32.12	−0.20
14	22.23	20.32	20.33	37.48	−0.07	22.66	19.57	19.84	34.82	0.17
15	23.45	21.42	21.43	40.25	−0.08	28.36	20.47	20.48	37.13	−0.15
16	24.81	22.54	22.55	43.03	−0.10	24.76	21.28	21.59	39.60	0.14
17	26.05	23.66	23.67	46.01	−0.11	29.92	22.55	22.59	42.65	−0.12
18	27.22	24.68	24.68	48.72	−0.12	27.57	23.44	23.80	45.31	0.11
19	28.17	25.66	25.66	51.39	−0.13	31.54	24.57	24.65	48.33	−0.10
20	29.42	26.79	26.78	26.37	−0.14	29.91	25.48	25.81	24.19	0.08

	$\phi=1,\ \psi=0.9,\ \delta=0.5,\ \mu=0$					$\phi=1,\ \psi=-0.9,\ \delta=0.5,\ \mu=0$				
h	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.28	6.25	2.28	3.29	0.83	21.23	6.78	6.78	3.32	−0.88
2	5.09	8.38	5.24	9.03	0.68	5.00	8.01	5.04	8.69	0.77
3	8.11	10.88	8.87	16.91	0.55	23.73	11.00	11.00	16.91	−0.68
4	11.62	13.88	12.95	27.10	0.44	11.74	13.16	12.01	26.44	0.59
5	15.36	17.12	16.75	39.35	0.34	28.46	16.98	16.98	38.84	−0.52
6	19.49	20.87	20.75	53.76	0.25	20.09	20.20	20.09	53.20	0.46
7	24.51	25.48	25.45	70.74	0.18	35.23	24.92	24.92	70.00	−0.41
8	29.71	30.33	30.32	89.76	0.12	29.90	28.98	29.17	88.62	0.35
9	35.45	35.79	35.78	111.02	0.06	43.72	34.43	34.43	109.67	−0.32
10	41.70	41.75	41.73	134.37	0.02	42.10	40.31	40.62	133.73	0.28
11	48.36	48.09	48.06	159.61	−0.02	55.02	46.55	46.54	158.95	−0.25
12	55.64	55.00	54.99	187.42	−0.06	55.60	53.16	53.43	186.79	0.22
13	62.96	62.08	62.08	216.78	−0.08	68.11	60.25	60.26	216.41	−0.19
14	70.90	69.80	69.80	248.26	−0.11	71.02	68.09	68.39	248.62	0.17
15	79.73	78.46	78.46	282.84	−0.13	83.53	76.13	76.14	282.31	−0.15
16	88.89	87.34	87.34	318.89	−0.15	88.02	84.60	84.88	318.48	0.13
17	98.13	96.54	96.53	357.03	−0.16	101.24	94.20	94.19	357.20	−0.12
18	107.95	106.33	106.32	397.26	−0.17	107.65	103.90	104.23	397.77	0.10
19	118.06	116.31	116.30	438.99	−0.18	121.34	114.63	114.65	440.90	−0.09
20	129.33	127.53	127.53	134.37	−0.19	129.60	125.50	125.79	133.73	0.08

(*Continued*)

(*Continued*)

	$\phi=1,\ \psi=0.9,\ \delta=0.5,\ \mu=0.5$					$\phi=1,\ \psi=-0.9,\ \delta=0.5,\ \mu=0.5$				
h	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.28	6.17	2.28	3.30	0.83	21.61	6.90	6.90	3.30	−0.88
2	5.04	8.36	5.21	9.08	0.68	4.95	8.00	4.96	8.56	0.77
3	8.06	10.83	8.93	17.09	0.55	23.97	10.89	10.89	16.44	−0.67
4	11.55	13.84	12.97	27.45	0.44	11.55	12.98	11.94	26.21	0.59
5	15.39	17.23	16.90	39.83	0.34	28.85	17.02	17.02	38.87	−0.52
6	19.91	21.29	21.22	54.91	0.26	19.88	20.06	20.00	52.98	0.46
7	24.55	25.59	25.55	71.72	0.18	36.09	25.39	25.39	70.84	−0.40
8	29.81	30.54	30.43	90.84	0.12	30.14	29.02	29.36	88.86	0.35
9	35.66	36.12	36.07	112.29	0.07	45.16	35.53	35.53	111.42	−0.31
10	41.85	42.05	42.00	135.89	0.02	42.13	40.12	40.53	133.44	0.28
11	48.36	48.34	48.33	161.28	−0.02	56.37	47.47	47.47	160.35	−0.25
12	55.34	55.20	55.19	188.96	−0.06	56.57	53.77	54.11	188.33	0.22
13	63.02	62.58	62.59	218.85	−0.09	70.07	61.65	61.67	218.69	−0.19
14	70.75	70.13	70.13	250.40	−0.11	72.65	69.27	69.62	251.49	0.17
15	79.41	78.63	78.63	284.87	−0.13	85.85	77.84	77.86	285.79	−0.15
16	88.47	87.68	87.68	321.31	−0.15	89.84	85.96	86.22	321.61	0.13
17	98.06	97.17	97.19	359.83	−0.17	103.95	96.38	96.41	362.06	−0.12
18	108.29	107.35	107.36	400.64	−0.18	109.37	105.07	105.38	400.66	0.10
19	118.55	117.44	117.45	442.43	−0.19	123.71	116.36	116.40	445.36	−0.09
20	129.31	128.22	128.22	135.89	−0.19	130.63	125.97	126.30	133.44	0.08

	$\phi=0.95,\ \psi=0.9,\ \delta=0,\ \mu=0$					$\phi=0.95,\ \psi=-0.9,\ \delta=0,\ \mu=0$				
h	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h\geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.07	5.96	2.07	2.14	0.83	21.05	6.36	6.36	2.08	−0.88
2	4.03	7.01	4.19	4.19	0.68	4.04	7.25	4.06	3.99	0.77
3	5.81	8.08	6.50	6.09	0.56	20.98	8.09	8.09	5.55	−0.67
4	7.46	9.02	8.23	7.80	0.45	7.45	8.97	7.76	7.18	0.59
5	8.95	9.97	9.62	9.40	0.36	21.19	9.75	9.75	8.50	−0.52
6	10.40	10.99	10.79	10.91	0.28	10.32	10.56	10.49	9.86	0.46
7	11.65	11.87	11.79	12.34	0.22	21.70	11.39	11.39	10.80	−0.40
8	12.81	12.57	12.53	13.54	0.16	12.68	11.96	12.20	11.96	0.35
9	13.91	13.39	13.36	14.69	0.12	22.14	12.65	12.65	12.61	−0.31
10	15.18	14.28	14.30	15.66	0.08	14.95	13.25	13.50	13.80	0.28
11	16.17	15.00	15.03	16.53	0.04	22.64	13.97	13.96	14.27	−0.24
12	17.01	15.56	15.58	17.22	0.01	16.82	14.39	14.62	15.38	0.21
13	17.89	16.22	16.24	17.97	−0.01	23.43	15.37	15.39	15.96	−0.19
14	18.84	16.90	16.91	18.53	−0.03	18.43	15.32	15.58	16.28	0.17
15	19.38	17.28	17.29	19.08	−0.05	24.09	16.42	16.44	16.99	−0.15
16	20.02	17.80	17.81	19.46	−0.07	20.24	16.67	16.94	17.60	0.13
17	20.53	18.26	18.28	19.82	−0.08	24.48	17.24	17.26	17.89	−0.12
18	21.09	18.64	18.65	20.28	−0.09	21.43	17.47	17.74	18.47	0.10
19	21.80	19.25	19.26	20.68	−0.10	25.23	18.29	18.33	18.85	−0.09
20	22.16	19.49	19.50	15.66	−0.11	22.43	18.15	18.41	13.80	0.08

(*Continued*)

	$\phi = 0.95,\ \psi = 0.9,\ \delta = 0.5,\ \mu = 0$					$\phi = 0.95,\ \psi = -0.9,\ \delta = 0.5,\ \mu = 0$				
h	F_1	F_0	$F_{[\hat{\rho}_h \geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h \geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.02	5.99	2.02	2.06	0.83	20.63	6.25	6.25	2.09	−0.88
2	3.92	7.16	4.06	4.08	0.69	3.99	7.09	4.01	3.92	0.77
3	5.77	8.25	6.49	6.07	0.57	20.59	8.12	8.12	5.68	−0.67
4	7.48	9.25	8.39	7.87	0.46	7.50	8.97	7.83	7.30	0.59
5	9.01	10.29	9.90	9.62	0.37	21.13	9.93	9.93	8.63	−0.52
6	10.50	11.31	11.15	11.31	0.29	10.44	10.54	10.42	10.02	0.46
7	11.57	11.89	11.81	12.51	0.22	21.75	11.60	11.60	11.29	−0.41
8	12.88	12.71	12.66	13.82	0.16	13.18	12.19	12.35	12.60	0.36
9	14.02	13.49	13.50	15.12	0.11	22.11	12.85	12.85	13.51	−0.32
10	14.87	14.08	14.11	16.18	0.07	15.23	13.40	13.68	14.65	0.28
11	15.67	14.62	14.67	17.25	0.03	22.42	13.94	13.95	15.36	−0.25
12	16.54	15.24	15.26	18.10	0.00	16.94	14.52	14.78	16.43	0.22
13	17.35	15.87	15.88	19.08	−0.02	23.15	15.19	15.19	17.20	−0.19
14	18.28	16.55	16.56	19.90	−0.05	18.42	15.63	15.83	18.00	0.17
15	19.26	17.25	17.27	20.80	−0.07	23.46	16.10	16.13	18.59	−0.15
16	19.87	17.60	17.61	21.37	−0.08	19.84	16.57	16.81	19.46	0.13
17	20.54	18.03	18.04	21.98	−0.10	23.82	16.86	16.89	19.73	−0.12
18	21.18	18.47	18.48	22.67	−0.11	21.14	17.37	17.60	20.86	0.11
19	21.98	19.23	19.23	23.50	−0.12	24.49	17.74	17.78	21.22	−0.09
20	22.59	19.63	19.63	16.18	−0.13	22.19	18.06	18.28	14.65	0.08

	$\phi = 0.95,\ \psi = 0.9,\ \delta = 0.5,\ \mu = 0.5$					$\phi = 0.95,\ \psi = -0.9,\ \delta = 0.5,\ \mu = 0.5$				
h	F_1	F_0	$F_{[\hat{\rho}_h \geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$	F_1	F_0	$F_{[\hat{\rho}_h \geq 1/2]}$	$F_{<x_t>}$	$\hat{\rho}_h$
1	2.08	6.05	2.08	2.14	0.83	20.96	6.40	6.40	2.08	−0.88
2	3.92	7.10	4.08	4.11	0.69	3.91	7.20	3.93	3.84	0.77
3	5.67	8.06	6.33	5.99	0.56	21.06	8.32	8.32	5.70	−0.67
4	7.33	9.17	8.31	7.92	0.46	7.31	9.05	7.75	7.07	0.59
5	8.89	10.15	9.85	9.63	0.36	21.12	9.86	9.86	8.56	−0.52
6	10.28	10.97	10.84	11.12	0.29	10.32	10.80	10.63	9.93	0.46
7	11.53	11.74	11.69	12.61	0.22	21.69	11.56	11.56	11.25	−0.41
8	12.83	12.65	12.62	14.00	0.16	12.73	12.16	12.44	12.30	0.36
9	13.92	13.39	13.38	15.29	0.11	22.09	13.00	13.00	13.58	−0.32
10	14.91	14.03	14.02	16.48	0.07	14.94	13.54	13.84	14.53	0.28
11	15.67	14.48	14.49	17.34	0.03	22.39	14.14	14.14	15.43	−0.25
12	16.61	15.13	15.14	18.29	0.00	16.99	14.86	15.13	16.53	0.22
13	17.57	15.92	15.92	19.19	−0.03	22.90	15.33	15.33	17.25	−0.19
14	18.30	16.44	16.44	19.86	−0.05	18.44	15.66	16.00	17.87	0.17
15	19.11	16.98	16.97	20.64	−0.07	23.20	16.19	16.22	18.66	−0.15
16	19.80	17.59	17.59	21.58	−0.08	19.84	16.54	16.83	19.21	0.13
17	20.43	17.93	17.93	21.96	−0.10	23.61	17.05	17.09	20.02	−0.12
18	21.11	18.43	18.43	22.59	−0.11	20.96	17.49	17.73	20.28	0.10
19	21.75	18.84	18.85	23.17	−0.12	23.88	17.77	17.81	21.33	−0.09
20	22.49	19.42	19.42	16.48	−0.13	21.89	18.16	18.37	14.53	0.08