Essays in Time Series Econometrics

Dissertation

in partial fulfillment of the requirements for the degree of
doctor rerum naturalium (Dr. rer. nat.)

presented to the Department of Statistics
TU Dortmund University

by

Karsten Reichold

Dortmund, May 2023

Dissertation
presented to the Department of Statistics
TU Dortmund University

Dortmund, May 2023

Essays in Time Series Econometrics by Karsten Reichold
Dortmund, May 2023

Supervisors:
Prof. Dr. Martin Wagner (Department of Economics, University of Klagenfurt)

Prof. Dr. Carsten Jentsch (Department of Statistics, TU Dortmund University)

Assessment Committee:
Prof. Dr. Martin Wagner (Department of Economics, University of Klagenfurt)

Prof. Dr. Carsten Jentsch (Department of Statistics, TU Dortmund University)

Prof. Dr. Matei Demetrescu (Department of Statistics, TU Dortmund University)

Jun.-Prof. Dr. Antonia Arsova (Department of Statistics, TU Dortmund University)

Day of Oral Examination:
July 05, 2023

Location:
Dortmund, Germany

To Esther

Preface

This dissertation completes my doctoral studies at the Department of Statistics, TU Dortmund University. The journey started in January 2019 and now, almost four and a half years later, I write these final lines. The period has been shaped by many ups and a few downs, and I am very grateful for all the experiences I have made during my doctoral studies. No doubt, this journey would not have been possible without the support of several people I am lucky enough to have in my life.

First of all, I thank my supervisor Martin Wagner for his tireless support, helpful advice, and constant encouragement since my Bachelor's studies. I cannot emphasize enough that I am beyond thankful to you for challenging me whenever necessary and believing in me throughout these years. Through numerous discussions you have introduced me to Time Series Econometrics in general and to the cointegration literature in particular. Moreover, you have taught me the importance of being precise when formulating ideas, methods and results and I will bear this in mind when working on future projects. In addition, I express my deepest gratitude to my second supervisor Carsten Jentsch for his support, for introducing me to his research community and for his time to discuss research questions in detail. It is a great pleasure working with you. I also thank Matei Demetrescu for his support during the last months of my doctoral studies.

I had many colleagues throughout these years and some of them have even become good friends of mine. I want to take the chance to express my gratitude to Bernd Funovits, Fabian Knorre, Oliver Stypka, Rafael Kawka, Peter Grabarczyk, Lukas Matuschek, Patrick de Matos Ribeiro, and Martin Schumann for fruitful discussions and exciting activities outside university.

For sorting my thoughts, reflecting, or simply relaxing, I go for a run. Therefore, I extend my gratitude to Thomas for introducing me to the wonderful world of long-distance running.

I also thank my long-time friends for forming an oasis where journal rankings or conference deadlines are of secondary importance only. I really enjoy our regular reunions at different places in Germany and I am looking forward to our future events. In particular, I thank Hape for always visiting me wherever I stay for a longer period of time – this means a lot to me.

Dieses Kapitel wäre nicht vollständig, wenn ich mich nicht bei meinen Eltern Silvia und Uwe für alles bedanken würde, was sie je für mich gemacht haben. Ohne eure Unterstützung, euer Vertrauen und eure Fürsorge in allen Lebenslagen würde ich diese Zeilen vermutlich nicht schreiben. Außerdem danke ich meinen Brüdern Stefan und Martin dafür, dass sie immer an meiner Seite stehen und jedes Treffen zu einem unterhaltsamen Erlebnis machen. Einen großen indirekten Anteil an meiner Dissertation haben ohne Zweifel auch meine Großeltern Irmgard und Silver, die vorgelebt haben, dass man mit Mut und Weitsicht vieles bewirken kann.

Zu guter Letzt bedanke ich mich bei Esther für ihre Liebe, ihr Verständnis und ihre Unterstützung. Du hast diese Reise erst so richtig mit Leben gefüllt.

Karsten Reichold

May 2023

Contents

Introduction

This cumulative dissertation consists of three self-contained papers all contributing to the cointegrating regression literature. Cointegrating regression analyses seek to analyze the long-run relationship between integrated (i. e., stochastically trending) processes. They play an important role in, e. g., macroeconomics, environmental economics, and finance, see, e. g., Benati *et al.* (2021), Wagner (2015), and Rad *et al.* (2016) for recent examples. More recently, cointegration-based methods have been proven to be useful in physics (Dahlhaus *et al.*, 2018) and in the analysis of climate change (Phillips *et al.*, 2020).

It is well known in the literature that existing approaches to conduct inference in cointegrating regressions can lead to severe size distortions in finite samples, especially when the data are characterized by large levels of error serial correlation and regressor endogeneity. This makes the tests unreliable in many empirical applications involving stochastically trending variables. To address this issue, this dissertation suggests procedures to reduce these size distortions at the cost of only small power losses under the alternatives.

Each chapter of this dissertation focuses on a particular subfield of the cointegrating regression literature. Chapter 1 is devoted to classical linear cointegrating regressions, i. e., regressions that contain integrated processes as regressors. It combines traditional and self-normalized Wald-type test statistics with a vector autoregressive sieve bootstrap to reduce size distortions of hypothesis tests on the cointegrating vector. To asymptotically justify this method, the chapter proves bootstrap consistency for the traditional and self-normalized test statistics under mild conditions. Monte Carlo simulations complementing the asymptotic results show tremendous reductions in size distortions when bootstrap critical values replace asymptotic critical values. Finally, the empirical illustration indicates that the bootstrap makes a difference when analyzing the validity of the Fisher effect in OECD countries.

Chapter 2, on the other hand, focuses on panels of cointegrating polynomial regressions, i. e., panels of regressions that include an integrated process and its powers as regressors. It derives the asymptotic properties of a group-mean fully modified OLS estimator and of t-type and Wald-type tests based upon it in a fixed cross-section and large time series dimension. Treating the cross-section dimension as fixed allows us to derive cross-section dependence robust test statistics for very general dependence structures. Moreover, the proposed group-mean fully modified OLS estimator and the tests based upon it are invariant to potential non-zero deterministic drifts in the integrated regressors. Both cross-section dependence and non-zero drifts are often observed in empirical applications and the simulation results show that the proposed methods perform very well in these situations. Finally, the empirical illustration shows that accounting for cross-section dependence and deterministic trends makes a difference when analyzing environmental Kuznets curves for carbon dioxide emissions.

Chapters 1 and 2 derive the results under the premise that a (linear or polynomial) cointegrating relationship between the left-hand side variable and the regressors exists, i. e., the regression errors are assumed to be stationary. In contrast, Chapter 3 is devoted to testing for such a (linear) cointegrating relationship between a fixed number of integrated processes. In particular, it derives asymptotic theory for an existing nonparametric variance ratio unit root test (originally proposed to test for an unit root in an observed univariate time series) when applied to regression residuals. A simulation study complements the theoretical analyzes and indicates some performance advantages of the variance ratio test compared to established no-cointegration tests. Finally, an empirical illustration to cryptocurrencies shows the usefulness of the variance ratio test in practice.

The bibliographic details of the three chapters are as follows:

1. Reichold, K., Jentsch, C. (2023). Bootstrap Inference in Cointegrating Regressions: Traditional and Self-Normalized Test Statistics. Revised and Resubmitted to *Journal of Business & Economic Statistics*.

    (a) An earlier working paper version (with a different title) is available on arXiv: arXiv e-print 2204.01373.

    (b) The first working paper version (with a different title) is available as SFB 823 Discussion Paper, TU Dortmund: http://dx.doi.org/10.17877/DE290R-21854.

2. Wagner, M., Reichold, K. (2023). Panel Cointegrating Polynomial Regressions: Group-Mean Fully Modified OLS Estimation and Inference.

    (a) A slightly modified version is published in *Econometric Reviews*, Volume 42, Issue 4, pp. 358–392, https://doi.org/10.1080/07474938.2023.2178141.

    (b) An earlier working paper version is available as SFB 823 Discussion Paper, TU Dortmund: http://dx.doi.org/10.17877/DE290R-19664.

3. Reichold, K. (2023). A Residuals-Based Nonparametric Variance Ratio Test for Cointegration. Reject and Resubmit, *Journal of Time Series Analysis*.

    (a) Earlier working paper versions are available on arXiv: arXiv e-print 2211.06288.

Although all chapters contribute to the cointegrating regression literature, overlap is marginal, as each chapter focuses on a different subfield within the cointegrating regression literature (compare the discussion above). However, it is unavoidable that some basic arguments used to derive asymptotic theory are similar across chapters. A complete list of references, provided at the end of this dissertation, replaces the list of references of each individual chapter. All simulations have been performed in `MATLAB`. Corresponding code and supplementary material are available upon request.

Chapter 1

Bootstrap Inference in Cointegrating Regressions: Traditional and Self-Normalized Test Statistics

Abstract

Traditional tests of hypotheses on the cointegrating vector are well known to suffer from severe size distortions in finite samples, especially when the data are characterized by large levels of endogeneity or error serial correlation. To address this issue, we combine a vector autoregressive (VAR) sieve bootstrap to construct critical values with a self-normalization approach that avoids direct estimation of long-run variance parameters when computing test statistics. To asymptotically justify this method, we prove bootstrap consistency for the self-normalized test statistics under mild conditions. In addition, the underlying bootstrap invariance principle allows us to prove bootstrap consistency also for traditional test statistics based on popular modified OLS estimators. Simulation results show that using bootstrap critical values instead of asymptotic critical values reduces size distortions associated with traditional test statistics considerably, but combining the VAR sieve bootstrap with self-normalization can lead to even less size distorted tests at the cost of only small power losses. We illustrate the usefulness of the VAR sieve bootstrap in empirical applications by analyzing the validity of the Fisher effect in 19 OECD countries.

1.1 Introduction

Cointegration methods have been and are widely used to analyze long-run relationships between stochastically trending variables in many areas such as macroeconomics, environmental economics, and finance, see, e. g., Benati *et al.* (2021), Wagner (2015), and Rad *et al.* (2016) for recent examples. In addition to these classical fields of application, cointegration methods have recently proven to be useful to describe phenomena in physics (Dahlhaus *et al.*, 2018) and climate change (Phillips *et al.*, 2020).

It is standard practice in empirical applications to test linear restrictions on the cointegrating vector by means of *traditional* Wald-type test statistics based on a suitable consistent estimator of the cointegrating vector and on a nonparametric kernel estimator of a long-run variance parameter for standardization. In the presence of endogeneity, the limiting distribution of the OLS estimator is contaminated by second order bias terms making the estimator unsuitable for standard asymptotic inference. The literature provides several endogeneity corrected estimators with a zero-mean Gaussian mixture limiting distribution allowing for asymptotically valid inference based on chi-squared critical values. The most popular estimators are the dynamic OLS (D-OLS) estimator (Phillips and Loretan, 1991; Saikkonen, 1991; Stock and Watson, 1993), the fully modified OLS (FM-OLS) estimator (Phillips and Hansen, 1990), and the integrated modified OLS (IM-OLS) estimator (Vogelsang and Wagner, 2014). Tests based upon these estimators are implemented in several software packages and are thus easy to apply in applications. However, they are all well known to be severely size distorted in finite samples, especially when the data are characterized by large levels of endogeneity or error serial correlation. Similar problems are also observed for tests based on alternative estimators proposed in, e. g., Phillips (2014) and Hwang and Sun (2018).

To address these size distortions, this paper combines a vector autoregressive (VAR) sieve bootstrap to construct critical values with a self-normalization approach that avoids direct estimation of the long-run variance parameter when computing test statistics. In particular, we discuss three *self-normalized* Wald-type test statistics based on the tuning parameter free IM-OLS estimator, which do not rely on a consistent tuning parameter dependent long-run variance estimator but still possess a pivotal limiting distribution under the null hypothesis. The concept of self-normalization has been proven to be useful in the analysis of stationary time series (see, e. g., Kiefer *et al.*, 2000; Shao, 2010a; Shao, 2015) but has not received much attention in the cointegrating regression literature as an alternative to traditional test statistics. As we will see in Section 1.3.2, self-normalization is closely related to, but does not need to be a special case of, the fixed-b approach of Vogelsang and Wagner (2014).

In contrast, the nowadays classical VAR sieve bootstrap (Kreiss, 1992; Bühlmann, 1997; Paparoditis, 1996) has already been applied to cointegrating regressions in various setups. Psaradakis (2001), inspired by the seminal work of Li and Maddala (1997), shows the superior performance when VAR sieve bootstrap critical values replace chi-squared critical values for the traditional Wald-type test based on the FM-OLS estimator (without asymptotically justifying the approach). Park (2002) proves an invariance principle for the bootstrap process under the assumption that the errors form a linear process with i.i.d. increments. The invariance principle allows Chang *et al.* (2006) to prove consistency of the VAR sieve bootstrap for the traditional Wald-type test statistic based on the D-OLS estimator. Although the bootstrap leads to considerable performance advantages over the tests based on asymptotic critical values, it is rarely used in empirical applications.

Alternative bootstrap approaches studied in the cointegrating regression literature are the stationary bootstrap (SB) of Politis and Romano (1994) and the residual-based block bootstrap (RBB) of Paparoditis and Politis (2003), which have been proven to be consistent for the limiting distribution of the OLS estimator of the cointegrating vector in Shin and Hwang (2013) and Jentsch *et al.* (2015), respectively. Moreover, the dependent wild bootstrap (DWB) of Shao (2010b) has been proven to be useful when testing for unit roots (Rho and Shao, 2019), but has not been asymptotically justified in cointegrating regressions yet. In contrast to selecting the order of the

VAR sieve for the VAR sieve bootstrap, however, choosing a suitable geometric distribution for the SB, a suitable block size for the RBB, or a suitable combination of kernel and bandwidth for the DWB seem to be rather challenging tasks in empirical applications.

To prove consistency of the VAR sieve bootstrap for the self-normalized test statistics, we derive a bootstrap invariance principle under mild conditions. In particular, in contrast to Park (2002) and Chang *et al.* (2006), we allow for uncorrelated but not necessarily independent white noise increments. The bootstrap invariance principle might thus be of independent interest and allows us to prove bootstrap consistency also for the traditional Wald-type test statistics based on the D-OLS, FM-OLS, and IM-OLS estimators. With respect to the traditional Wald-type test statistic based on the D-OLS estimator, this paper thus extends the results in Chang *et al.* (2006). Finally, we should emphasize that one of the self-normalized test statistics proposed in this paper has a pivotal limiting null distribution only in case the number of linearly independent restrictions on the cointegrating vector is equal to the dimension of the cointegrating vector. Thus, for this particular test statistic, the VAR sieve bootstrap is key to allow for asymptotically valid inference also for a smaller number of linearly independent restrictions on the cointegrating vector.

The theoretical analysis is complemented by a detailed simulation study assessing the finite sample performance of the traditional and self-normalized tests based on VAR sieve bootstrap critical values. The results reveal tremendous performance advantages of traditional tests based on bootstrap critical values over those based on chi-squared critical values even for small sample sizes. In addition, we find that the self-normalized tests based on asymptotic critical values are considerably less size distorted than the traditional tests based on asymptotic critical values for small to medium levels of endogeneity and error serial correlation at the cost of only small power losses under the alternative. For large levels of endogeneity and error serial correlation, however, self-normalization alone is less advantageous. In these cases, the VAR sieve bootstrap improves the performance of the self-normalized tests considerably, with two of the self-normalized tests often outperforming the traditional tests based on bootstrap critical values.

Finally, we demonstrate the usefulness of the VAR sieve bootstrap in applications by analyzing the validity of the Fisher effect in 19 OECD countries in the three decades prior to the Covid-19 crisis. The Fisher effect states that inflation and the short-term nominal interest rate are in a one-for-one relationship. It is backed by many theoretical models but often rejected in empirical studies, possibly because of the poor performance of estimators and tests in the presence of highly persistent errors typically observed in Fisher equations (Caporale and Pittis, 2004; Westerlund, 2008). Indeed, we find that the traditional and self-normalized tests based on asymptotic critical values tend to reject the Fisher effect for several countries, whereas the tests based on bootstrap critical values indicate the validity of the Fisher effect for almost all countries under consideration.

The paper proceeds as follows: Section 1.2 introduces the model and its underlying assumptions. Section 1.3 reviews the construction of traditional test statistics and discusses three self-normalized test statistics. Section 1.4 presents the VAR sieve bootstrap procedure to construct critical values for the traditional and self-normalized test statistics and proves its asymptotic validity in each case. Section 1.5 assesses the finite sample performance of the proposed methods and Section 1.6 contains the empirical illustration. Section 1.7 summarizes and concludes. All proofs are relegated to the Appendix.

Notation: $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$, $|A|_F = (\mathrm{tr}(A'A))^{1/2}$ denotes the Frobenius norm of a real matrix A, and $\mathrm{diag}(\cdot)$ denotes a (block) diagonal matrix with diagonal elements specified throughout. With $\xrightarrow{w}$ and $\xrightarrow{p}$ we denote weak convergence and convergence in probability, respectively, as $T \to \infty$, with $\mathbb{P}$ denoting the underlying probability measure. Convergence in the bootstrap probability space is denoted by $\xrightarrow{w^*}$ and $\xrightarrow{p^*}$, with $\mathbb{P}^*$ denoting the corresponding probability measure and $\mathbb{E}^*(\cdot)$ denoting the expectation with respect to $\mathbb{P}^*$ (conditional on the data). Throughout, random variables in the bootstrap probability space are indicated by the superscript "$*$".

1.2 The Model and Assumptions

We consider the cointegrating regression model

$$y_t = x_t'\beta + u_t, \tag{1.1}$$

$$x_t = x_{t-1} + v_t, \tag{1.2}$$

for observations $t = 1, \ldots, T$, where y_t is a scalar time series and x_t is an $m \times 1$ vector of time series. For notational brevity, we set $x_0 = 0$ and exclude deterministic components from (1.1). Nevertheless, incorporating deterministic components (fulfilling the condition in equation (14) in Vogelsang and Wagner, 2014) is straightforward and the accompanying software code allows to handle the more general case. To derive asymptotic theory, we have to impose assumptions on the process $\{w_t\}_{t\in\mathbb{Z}} := \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$.

Assumption 1.1. *Let $\{w_t\}_{t\in\mathbb{Z}}$ be an $\mathbb{R}^{1+m}$-valued, strictly stationary, and purely nondeterministic stochastic process of full rank with $\mathbb{E}(w_t) = 0$ and $\mathbb{E}(|w_t|_F^a) < \infty$ for some $a > 2$. The autocovariance matrix function $\Gamma(\cdot)$ of $\{w_t\}_{t\in\mathbb{Z}}$ fulfills $\sum_{h=-\infty}^{\infty}(1+|h|)^k|\Gamma(h)|_F < \infty$ for some $k \geq 3/2$. For the spectral density matrix $f(\cdot)$ of $\{w_t\}_{t\in\mathbb{Z}}$ there exists a constant $c > 0$ such that $\min \sigma(f(\lambda)) \geq c$ for all frequencies $\lambda \in (-\pi, \pi]$, where $\sigma(f(\lambda))$ denotes the spectrum of $f(\cdot)$ at frequency λ.*

Assumption 1.1 is similar to Assumption A in Meyer and Kreiss (2015). The short memory condition implies a continuously differentiable spectral density f, which is particularly bounded from below and from above, uniformly for all frequencies $\lambda \in (-\pi, \pi]$. As shown in Meyer and Kreiss (2015), a process fulfilling Assumption 1.1 does always possess the one-sided representations

$$\Phi(L)w_t = \varepsilon_t \quad \text{and} \quad w_t = \Psi(L)\varepsilon_t, \tag{1.3}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a strictly stationary uncorrelated but not necessarily independent white noise process with positive definite covariance matrix Σ and L denotes the backward shift operator. For $\Phi(z) := I_{m+1} - \sum_{j=1}^{\infty}\Phi_j z^j$ and $\Psi(z) := I_{m+1} + \sum_{j=1}^{\infty}\Psi_j z^j$ it holds that $\det(\Phi(z)) \neq 0$ and $\det(\Psi(z)) \neq 0$ for all $|z| \leq 1$ and $\sum_{j=1}^{\infty}(1+j)^k|\Phi_j|_F < \infty$ and $\sum_{j=1}^{\infty}(1+j)^k|\Psi_j|_F < \infty$ for the k from Assumption 1.1.

Assumption 1.2. *The process* $\{w_t\}_{t\in\mathbb{Z}}$ *has absolutely summable cumulants up to order four. More precisely, it holds for all* $j = 2, \ldots, 4$ *and* $a = [a_1, \ldots, a_j]' \in \{1, \ldots, m+1\}^j$, *that* $\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |cum_a(0, h_2, \ldots, h_j)| < \infty$, *where* $cum_a(0, h_2, \ldots, h_j)$ *denotes the* j*-th joint cumulant of* $w_{a_1,0}, w_{a_2,h_2}, \ldots, w_{a_j,h_j}$ *and* $w_{i,t}$ *denotes the* i*-th element of* w_t.

Let Ω denote the long-run covariance matrix of $\{w_t\}_{t\in\mathbb{Z}}$, i. e.,

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix} = 2\pi f(0) = \sum_{h=-\infty}^{\infty} \mathbb{E}(w_0 w_h') = \Psi(1)\Sigma\Psi(1)'.$$

From $\Sigma > 0$ and $\det(\Psi(1)) \neq 0$ it follows that $\Omega > 0$. In particular, positive definiteness of Ω_{vv} rules out cointegration among the elements of x_t. For later usage, we also define the one-sided long-run covariance matrix $\Delta := \sum_{h=0}^{\infty} \mathbb{E}(w_0 w_h')$ and partition it analogously to Ω. Finally, we assume an invariance principle to hold for $\{w_t\}_{t\in\mathbb{Z}}$.

Assumption 1.3. *Let* $\{w_t\}_{t\in\mathbb{Z}}$ *fulfill*

$$B_T(r) := T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} w_t \xrightarrow{w} B(r) = \begin{bmatrix} B_u(r) \\ B_v(r) \end{bmatrix} = \Omega^{1/2} W(r), \quad 0 \leq r \leq 1, \tag{1.4}$$

where $W(r) = [W_{u\cdot v}(r), W_v(r)']'$ *is an* $(1+m)$*-dimensional vector of independent standard Brownian motions.*

In the following, it is convenient to work with $\Omega^{1/2}$ of the form

$$\Omega^{1/2} = \begin{bmatrix} \Omega_{u\cdot v}^{1/2} & \Omega_{uv}(\Omega_{vv}^{-1/2})' \\ 0 & \Omega_{vv}^{1/2} \end{bmatrix},$$

such that $\Omega^{1/2}(\Omega^{1/2})' = \Omega$, where $\Omega_{u\cdot v} := \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$ is the variance of the scalar Brownian motion $B_{u\cdot v}(r) := B_u(r) - B_v(r)'\Omega_{vv}^{-1}\Omega_{vu}$.

In contrast to the assumptions in Park (2002) and Chang *et al.* (2006), Assumption 1.1 does explicitly not ask for invertibility or causality of the process $\{w_t\}_{t\in\mathbb{Z}}$ with respect to an independent white noise process. Instead, in this paper, the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is an uncorrelated but not necessarily independent white noise process. Assumption 1.2 is of technical nature and satisfied if, e. g., $\{w_t\}_{t\in\mathbb{Z}}$ is α-mixing with strong-mixing coefficients $\alpha(j)$ such that $E(|w_t|_F^{4+\delta}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\delta/(4+\delta)} < \infty$ for some $\delta > 0$, see, e. g., Shao (2010b, p. 221). In particular, Assumption 1.2 requires the existence of fourth moments of $\{w_t\}_{t\in\mathbb{Z}}$. To establish meaningful asymptotic theory, Assumptions 1.1 and 1.2 have to be complemented by an invariance principle in Assumption 1.3. This general formulation of an invariance principle allows for various concepts of choice to quantify weak forms of dependence of the process $\{w_t\}_{t\in\mathbb{Z}}$. These include classical approaches as, e. g., several variants of mixing properties, mixingale-type sequences, and linear processes (see, e. g., Merlevède *et al.*, 2006, for an overview). In addition, the general formulation also allows for more modern approaches that cover, e. g., the general notion of weakly dependent stationary time series discussed in Doukhan and Wintenberger (2007) or physical dependence (Wu, 2007).

1.3 Testing General Linear Hypotheses

1.3.1 Modified OLS Estimation

The OLS estimator $\widehat{\beta}_{\mathrm{OLS}} := \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t$ of β in (1.1) is (rate-T) consistent, but in the presence of endogeneity its limiting distribution is contaminated by second order bias terms. This makes the OLS estimator infeasible for conducting inference based on (simulated) quantile tables of (non)standard distributions. As already mentioned in the introduction, the literature provides several modified estimators which allow for standard asymptotic inference. In the following, we focus on the popular D-OLS, FM-OLS, and IM-OLS estimators. To fix notation, let us briefly review the construction of the estimators.

In comparison with the OLS estimator, the FM-OLS approach rests upon two transformations. First, y_t is replaced by $y_t^+ := y_t - v_t'\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$, where

$$\widehat{\Omega} = \begin{bmatrix} \widehat{\Omega}_{uu} & \widehat{\Omega}_{uv} \\ \widehat{\Omega}_{vu} & \widehat{\Omega}_{vv} \end{bmatrix} := T^{-1} \sum_{i=1}^{T} \sum_{j=1}^{T} \mathcal{K}\left(\frac{|i-j|}{b_T}\right) \begin{bmatrix} \widehat{u}_{\mathrm{OLS},i} \\ v_i \end{bmatrix} \begin{bmatrix} \widehat{u}_{\mathrm{OLS},j} \\ v_j \end{bmatrix}' \tag{1.5}$$

denotes a nonparametric kernel estimator of Ω based on the OLS residuals $\widehat{u}_{\mathrm{OLS},t}$ in (1.1) and the first differences of x_t. Here and in the following, $\mathcal{K}(\cdot)$ denotes a kernel weighting function and b_T a bandwidth parameter. The second transformation requires additive correction factors, given by $T\widehat{\Delta}_{vu}^+$, with $\widehat{\Delta}_{vu}^+ := \widehat{\Delta}_{vu} - \widehat{\Delta}_{vv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$, where

$$\widehat{\Delta} = \begin{bmatrix} \widehat{\Delta}_{uu} & \widehat{\Delta}_{uv} \\ \widehat{\Delta}_{vu} & \widehat{\Delta}_{vv} \end{bmatrix} := T^{-1} \sum_{i=1}^{T} \sum_{j=i}^{T} \mathcal{K}\left(\frac{|i-j|}{b_T}\right) \begin{bmatrix} \widehat{u}_{\mathrm{OLS},i} \\ v_i \end{bmatrix} \begin{bmatrix} \widehat{u}_{\mathrm{OLS},j} \\ v_j \end{bmatrix}'. \tag{1.6}$$

With these definitions in place, the FM-OLS estimator of β in (1.1) is defined as $\widehat{\beta}_{\mathrm{FM}} := \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t y_t^+ - T\widehat{\Delta}_{vu}^+\right)$. Under Assumption 1.3 and under some technical assumptions on the kernel function $\mathcal{K}(\cdot)$ and on the bandwidth parameter b_T ensuring consistency of $\widehat{\Omega}$ and $\widehat{\Delta}$ (see, e. g., Jansson, 2002), Phillips and Hansen (1990) show that

$$T\left(\widehat{\beta}_{\mathrm{FM}} - \beta\right) \xrightarrow{w} \Omega_{u\cdot v}^{1/2} \left(\Omega_{vv}^{-1/2}\right)' \left(\int_0^1 W_v(r) W_v(r)' dr\right)^{-1} \int_0^1 W_v(r) dW_{u\cdot v}(r). \tag{1.7}$$

Conditional upon $W_v(r)$, the limiting distribution of the FM-OLS estimator is Gaussian with zero-mean and covariance matrix $\Omega_{u\cdot v}\left(\Omega_{vv}^{-1/2}\right)' \left(\int_0^1 W_v(r) W_v(r)' dr\right)^{-1} \Omega_{vv}^{-1/2}$.

Let us now turn to the D-OLS estimator. Assumption 1.1 allows to write

$$u_t = \sum_{j=-\infty}^{\infty} \pi_j' v_{t-j} + e_t, \tag{1.8}$$

where e_t is a zero-mean stationary process with $\mathbb{E}\left(e_t v_{t-j}'\right) = 0$ for all $j \in \mathbb{Z}$ and $\sum_{j=-\infty}^{\infty} |\pi_j|_F < \infty$ (cf. Saikkonen, 1991). Replacing u_t in (1.1) with the term on the right-hand side of (1.8) and truncating the infinite sum at both ends yields

$$y_t = x_t'\beta + \sum_{j=-L_1}^{L_2} \pi_j' v_{t-j} + \widetilde{e}_t, \tag{1.9}$$

$t = L_2+1, \ldots, T-L_1$, where $\widetilde{e}_t := e_t + \sum_{j<-L_1, j>L_2} \pi_j' v_{t-j}$ and $L_1, L_2 \geq 0$. To eliminate the error introduced by truncating the infinite sum in (1.8) asymptotically, the numbers of leads and lags have to go to infinity with sample size such that $\lim_{T\to\infty} T^{1/2} \sum_{j<-L_1,j>L_2} |\pi_j|_F = 0$ and $\lim_{T\to\infty} T^{-1}(L_1^3 + L_2^3) = 0$. In practice, the numbers of leads and lags are typically chosen by minimizing an information criterion (see, e. g., Choi and Kurozumi, 2012). The D-OLS estimator $\widehat{\beta}_{\mathrm{D}}$ of β in (1.1) is then defined as the OLS estimator of β in (1.9). Under Assumptions 1.1 – 1.3 it follows that $T\left(\widehat{\beta}_{\mathrm{D}} - \beta\right)$ converges to the limiting distribution given in (1.7), i. e., the limiting distributions of the FM-OLS estimator and the D-OLS estimator coincide (Saikkonen, 1991).

Finally, the IM-OLS estimator $\widehat{\beta}_{\mathrm{IM}}$ of β in (1.1) proposed in Vogelsang and Wagner (2014) is defined as the OLS estimator of β in the augmented partial sums regression

$$S_t^y = S_t^{x\prime}\beta + x_t'\gamma + S_t^u = Z_t'\theta + S_t^u, \tag{1.10}$$

where $S_t^y := \sum_{s=1}^t y_s$, $S_t^x := \sum_{s=1}^t x_s$, $S_t^u := \sum_{s=1}^t u_s$, $Z_t := [S_t^{x\prime}, x_t']'$, and $\theta := [\beta', \gamma']'$. Adding x_t to the partial sums regression serves as an endogeneity correction, which is similar in spirit to the leads and lags augmentation in D-OLS estimation but avoids tuning parameter choices. Denoting the OLS estimator of θ in (1.10) with $\widehat{\theta}_{\mathrm{IM}} := [\widehat{\beta}_{\mathrm{IM}}', \widehat{\gamma}_{\mathrm{IM}}']'$, Vogelsang and Wagner (2014, Theorem 2) show that it holds under Assumption 1.3 that the limiting distribution of $\widehat{\theta}_{\mathrm{IM}}$ is given by

$$\begin{bmatrix} T\left(\widehat{\beta}_{\mathrm{IM}} - \beta\right) \\ \widehat{\gamma}_{\mathrm{IM}} - \Omega_{vv}^{-1}\Omega_{vu} \end{bmatrix} \xrightarrow{w} \Omega_{u\cdot v}^{1/2} \left(\Pi'\right)^{-1} \mathcal{Z}, \tag{1.11}$$

where $\Pi := \operatorname{diag}\left(\Omega_{vv}^{1/2}, \Omega_{vv}^{1/2}\right)$, $\mathcal{Z} := \left(\int_0^1 g(r)g(r)'dr\right)^{-1} \int_0^1 [G(1) - G(r)]\, dW_{u\cdot v}(r)$, $g(r) := [\int_0^r W_v(s)'ds, W_v(r)']'$, and $G(r) := \int_0^r g(s)ds$. As both x_t and S_t^u are I(1) processes, the correlation between $B_v(r)$ and $B_u(r)$ is soaked up in the long-run population regression vector $\Omega_{vv}^{-1}\Omega_{vu}$. Therefore, the correct centering parameter for $\widehat{\gamma}_{\mathrm{IM}}$ in the presence of endogeneity is $\Omega_{vv}^{-1}\Omega_{vu}$ rather than the population value $\gamma = 0$. Conditional upon $W_v(r)$, the limiting distribution in (1.11) is Gaussian with zero-mean and covariance matrix $\Omega_{u\cdot v} V_{\mathrm{IM}}$, where $V_{\mathrm{IM}} = \left(\Pi'\right)^{-1} \widetilde{V}_{\mathrm{IM}} \Pi^{-1}$ and

$$\widetilde{V}_{\mathrm{IM}} := \left(\int_0^1 g(r)g(r)'dr\right)^{-1} \left(\int_0^1 [G(1) - G(r)]\,[G(1) - G(r)]'\,dr\right) \left(\int_0^1 g(r)g(r)'dr\right)^{-1}.$$

1.3.2 Traditional and Self-Normalized Test Statistics

The zero mean Gaussian mixture limiting distributions of the modified estimators allow for standard asymptotic inference based on chi-squared critical values. In the following, we focus on testing $s \leq m$ linearly independent restrictions on $\beta \in \mathbb{R}^m$ in (1.1) summarized as $\mathrm{H}_0\colon R_1\beta = r_0$ versus $\mathrm{H}_1\colon R_1\beta \neq r_0$, where $R_1 \in \mathbb{R}^{s\times m}$ has full row rank s and $r_0 \in \mathbb{R}^s$. It is standard practice in applications to consider *traditional* Wald-type tests of the form, using generic notation,

$$\tau(\widehat{\Omega}_{u\cdot v}) := \left(R_1\widehat{\beta} - r_0\right)' \left[R_1\widehat{\Omega}_{u\cdot v}\widehat{V}R_1'\right]^{-1} \left(R_1\widehat{\beta} - r_0\right), \tag{1.12}$$

where $\widehat{\beta} \in \{\widehat{\beta}_{\mathrm{D}}, \widehat{\beta}_{\mathrm{FM}}, \widehat{\beta}_{\mathrm{IM}}\}$, $\widehat{V}$ is the sample covariance matrix of $\widehat{\beta}$, and $\widehat{\Omega}_{u\cdot v} := \widehat{\Omega}_{uu} - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\widehat{\Omega}_{vu}$ is a consistent kernel estimator of $\Omega_{u\cdot v}$. Whenever necessary, the choice of the estimator is made clear by adding the corresponding subscript to $\tau(\widehat{\Omega}_{u\cdot v})$. For the FM-OLS estimator $\widehat{V}$ is given by the inverse of $\sum_{t=1}^{T} x_t x_t'$ and for the D-OLS estimator $\widehat{V}$ is defined similarly taking into account the augmentation in (1.9). For the IM-OLS estimator $\widehat{V}$ is given by the upper left $(m \times m)$-dimensional block element of the $(2m \times 2m)$-dimensional matrix

$$\widehat{V}_{\mathrm{IM}} := \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} \left(\sum_{t=1}^{T} c_t c_t'\right) \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1}, \tag{1.13}$$

where $c_1 := S_T^Z$ and $c_t := S_T^Z - S_{t-1}^Z$, with $S_t^Z := \sum_{j=1}^{t} Z_j$, for $t = 2, \ldots, T$. In each case, it follows that $\tau(\widehat{\Omega}_{u\cdot v}) \xrightarrow{w} \chi_s^2$, where χ_s^2 denotes a chi-square distribution with s degrees of freedom (see Phillips and Hansen, 1990; Saikkonen, 1991; Vogelsang and Wagner, 2014).

It is well known in the literature that these traditional tests suffer from severe size distortions in finite samples, especially when the data are characterized by large levels of endogeneity or error serial correlation. Clearly, these size distortions are explained by the poor approximation quality of the chi-squared distribution to the finite sample distributions of the traditional test statistics in these cases.

One way to address this issue is the fixed-b approach of Vogelsang and Wagner (2014) based on the tuning parameter free IM-OLS estimator. It allows to tabulate critical values corresponding to the kernel and bandwidth choices made when estimating $\Omega_{u\cdot v}$. However, their simulation results reveal that the performance of the test is still sensitive to the choice of $b = b_T/T \in (0, 1]$. A promising alternative approach is the concept of self-normalization, which has not received much attention in the cointegrating regression literature as an alternative to traditional test statistics. The main idea of self-normalization is to replace a tuning parameter dependent estimator of a long-run variance parameter in the construction of a test statistic with a quantity that is asymptotically proportional to this particular long-run variance parameter and can be directly computed from the data without requiring tuning parameter choices.

Because its construction is completely tuning parameter free, the IM-OLS estimator serves as a natural starting point for developing *self-normalized* Wald-type tests in cointegrating regressions. To simplify the expression of asymptotic results, we rewrite the null hypothesis in terms of the correct centering parameter for $\widehat{\theta}_{\mathrm{IM}}$, given by $[\beta', (\Omega_{vv}^{-1}\Omega_{vu})']'$. To this end, define $R_2 := [R_1, 0_{s\times m}] \in \mathbb{R}^{s\times 2m}$ such that the null hypothesis reads as $R_1\beta = R_2[\beta', (\Omega_{vv}^{-1}\Omega_{vu})']' = r_0$. Clearly, the auxiliary coefficient vector γ is not restricted under the null hypothesis and, in particular, $\Omega_{vv}^{-1}\Omega_{vu}$ does not have to be estimated. Moreover, define

$$\tau_{\mathrm{IM}}(\kappa) := \left(R_2\widehat{\theta}_{\mathrm{IM}} - r_0\right)' \left[R_2\kappa\widehat{V}_{\mathrm{IM}}R_2'\right]^{-1} \left(R_2\widehat{\theta}_{\mathrm{IM}} - r_0\right), \tag{1.14}$$

which is identical to $\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$ defined in (1.12) in case $\kappa = \widehat{\Omega}_{u\cdot v}$. Inspired by Kiefer *et al.* (2000), a straightforward choice for the self-normalizer is given by $\widehat{\eta} := T^{-2}\sum_{t=2}^{T}\left(\sum_{s=2}^{t}\Delta\widehat{S}_s^u\right)^2$, where $\Delta\widehat{S}_t^u := \widehat{S}_t^u - \widehat{S}_{t-1}^u$, $t = 2,\ldots,T$, are the first differences of the OLS residuals $\widehat{S}_t^u := S_t^y - Z_t'\widehat{\theta}_{\mathrm{IM}}$ in the augmented partial sums regression (1.10). The proof of Proposition 1.1 below reveals that the self-normalizer converges weakly to $\Omega_{u\cdot v}\int_0^1 (W_{u\cdot v}(s) - g(s)'\mathcal{Z})^2\,dr$, i. e., its limiting distribution is scale dependent on $\Omega_{u\cdot v}$. Choosing $\kappa = \widehat{\eta}$ in (1.14) thus removes the nuisance parameter $\Omega_{u\cdot v}$ asymptotically, without estimating it directly. The resulting test statistic is closely related to, but not a special case of, the $\widetilde{W}$ statistic considered in Vogelsang and Wagner (2014), compare the discussion in Remark 1.2 below.

Proposition 1.1. *Let the data be generated by* (1.1) *and* (1.2) *and let* $\{w_t\}_{t\in\mathbb{Z}}$ *satisfy Assumption 1.3. Then, under the null hypothesis, its holds that*

$$\tau_{IM}(\widehat{\eta}) \xrightarrow{w} \frac{\left(R_2\,(\Pi')^{-1}\,\mathcal{Z}\right)'(R_2V_{IM}R_2')^{-1}\left(R_2\,(\Pi')^{-1}\,\mathcal{Z}\right)}{\int_0^1 (W_{u\cdot v}(r) - g(r)'\mathcal{Z})^2\,dr} =: \mathcal{G}_{SN}. \tag{1.15}$$

The limiting null distribution in Proposition 1.1 is a ratio of two random variables. It is straightforward to verify that the marginal distribution of the numerator is χ_s^2, whereas the marginal distribution of the denominator is nonstandard but free of any nuisance parameters. However, it follows from Vogelsang and Wagner (2014, Lemma 2) that numerator and denominator are correlated with the correlation depending on nuisance parameters through Π. Hence, tabulating asymptotic critical values for $\tau_{\mathrm{IM}}(\widehat{\eta})$ is not possible in general.

Remark 1.1. *In the special case where the number of linearly independent restrictions on β equals the dimension of β (i. e., in case $s = m$), it follows from the definition of R_2, invertibility of R_1, and simple algebra that*

$$\left(R_2\,(\Pi')^{-1}\,\mathcal{Z}\right)'(R_2V_{IM}R_2')^{-1}\left(R_2\,(\Pi')^{-1}\,\mathcal{Z}\right) = \mathcal{Z}(1)'\widetilde{V}_{IM}(1,1)^{-1}\mathcal{Z}(1),$$

where $\widetilde{V}_{IM}(1,1)$ denotes the upper left $(m\times m)$-dimensional block element of the $(2m\times 2m)$-dimensional matrix $\widetilde{V}_{IM}$ and $\mathcal{Z}(1)$ denotes the vector of the first m elements of $\mathcal{Z}$. Thus, R_2 and Π cancel out algebraically and the correlation between numerator and denominator of the limiting distribution in (1.15) *becomes nuisance parameter free. Table 1.4 in Appendix 1.8.1 provides asymptotically valid critical values for $\tau_{IM}(\widehat{\eta})$ in case $s = m$ for $m = 1,\ldots,4$.*

The use of $\tau_{\mathrm{IM}}(\widehat{\eta})$ in applications is restricted to the case $s = m$. We offer two solutions to overcome this limitation. One solution adjusts the residuals $\widehat{S}_t^u$ such that numerator and denominator of the limiting distribution of the self-normalized test statistic become independent of each other, which results in a pivotal limiting null distribution. The adjustment coincides with the adjustment proposed in Vogelsang and Wagner (2014, p. 746) to allow for fixed-b asymptotics. The second solution, which leaves the test statistic unchanged, is the VAR sieve bootstrap for constructing critical values proposed in Section 1.4.

For the adjustment of the IM-OLS residuals, first define $\widetilde{Z}_t := t\sum_{s=1}^{T} Z_s - \sum_{j=1}^{t-1}\sum_{s=1}^{j} Z_s$, $t = 1,\ldots,T$, and let $\widetilde{Z}_t^{\perp}$ denote the residuals from the regression of $\widetilde{Z}_t$ on Z_t. The adjusted residuals

$\widehat{S}_t^{u\perp}$ are then obtained by regressing $\widehat{S}_t^{u}$ on $\widetilde{Z}_t^{\perp}$. Based on the adjusted residuals, we define the alternative self-normalizer as $\widehat{\eta}^{\perp} := T^{-2}\sum_{t=2}^{T}\left(\sum_{s=2}^{t}\Delta\widehat{S}_s^{u\perp}\right)^2$. This self-normalizer is closely related to, but not a special case of, the kernel estimator of $\Omega_{u\cdot v}$ proposed in Vogelsang and Wagner (2014, $\widetilde{\sigma}^{2*}_{u\cdot v}$ in their notation), which allows for fixed-b inference and is defined as (in our notation)

$$\widetilde{\Omega}_{u\cdot v}^{\perp} := T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathcal{K}\left(\frac{|i-j|}{b_T}\right)\Delta\widehat{S}_i^{u\perp}\Delta\widehat{S}_j^{u\perp}. \tag{1.16}$$

In case $\mathcal{K}(\cdot)$ is the Bartlett kernel and $b_T = T$, similar algebraic arguments as used in Cai and Shintani (2006, Proof of Lemma 1) show that $\widetilde{\Omega}_{u\cdot v}^{\perp}$ is equal to

$$\widetilde{\eta}^{\perp} := \widehat{\eta}^{\perp} + T^{-2}\sum_{t=2}^{T}\left(\sum_{s=2}^{t}\Delta\widehat{S}_s^{u\perp} - \sum_{s=2}^{T}\Delta\widehat{S}_s^{u\perp}\right)^2, \tag{1.17}$$

which also serves as a suitable self-normalizer. In fact, the self-normalized test statistic $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$ coincides with the fixed-b Wald-type test statistic of Vogelsang and Wagner (2014, $\widetilde{W}^{*}$ in their notation) based on the Bartlett kernel and bandwidth equal to sample size.

Remark 1.2. *Similar algebraic arguments as used above reveal that $\widehat{\eta}$ is closely related to, but not a special case of, the kernel estimator of $\Omega_{u\cdot v}$ based on $\Delta\widehat{S}_t^{u}$ rather than $\Delta\widehat{S}_t^{u\perp}$ (denoted as $\widetilde{\sigma}^{2}_{u\cdot v}$ in the notation of Vogelsang and Wagner, 2014, and leading to their $\widetilde{W}$ statistic) because $\sum_{s=2}^{T}\Delta\widehat{S}_t^{u} \neq 0$ in general. Further note that neither $\widetilde{\sigma}^{2}_{u\cdot v}$ nor $\widetilde{\sigma}^{2*}_{u\cdot v}$ (in the notation of Vogelsang and Wagner, 2014) are consistent estimators of $\Omega_{u\cdot v}$ under standard kernel and bandwidth assumptions. In this respect, self-normalization in cointegrating regressions is different from self-normalization in regressions with stationary time series, where prominent self-normalizers are related to consistent (under standard kernel and bandwidth assumptions) kernel estimators of long-run variance parameters (cf., e. g., Kiefer and Vogelsang, 2002; Shao, 2015). Finally, note that OLS residuals in* (1.1) *are not suitable for self-normalization because limiting distributions of functions of OLS residuals will typically be contaminated by second order bias terms.*

The following proposition derives the limiting null distribution of the self-normalized test statistic $\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$. For completeness, it also presents the limiting null distribution of $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$, which coincides with the limiting null distribution of the $\widetilde{W}^{*}$ test statistic of Vogelsang and Wagner (2014, Theorem 3) based on the Bartlett kernel and bandwidth equal to sample size.

Proposition 1.2. *Let the data be generated by* (1.1) *and* (1.2) *and let $\{w_t\}_{t\in\mathbb{Z}}$ satisfy Assumption 1.3. Then, under the null hypothesis, it holds that*

$$\tau_{IM}(\widehat{\eta}^{\perp}) \xrightarrow{w} \frac{\chi_s^2}{\int_0^1 \left(W_{u\cdot v}(r) - h(r)'Q\right)^2 dr} =: \mathcal{G}_{SN}^{\perp}, \tag{1.18}$$

$$\tau_{IM}(\widetilde{\eta}^{\perp}) \xrightarrow{w} \frac{\chi_s^2}{\int_0^1 \left(W_{u\cdot v}(r) - h(r)'Q\right)^2 dr + \int_0^1 \left[\overline{W}_{u\cdot v}(r) - \overline{h}(r)'Q\right]^2 dr} =: \widetilde{\mathcal{G}}_{SN}^{\perp}, \tag{1.19}$$

where $Q := \left(\int_0^1 h(r)h(r)'dr\right)^{-1}\int_0^1 [H(1)-H(r)]\,dW_{u\cdot v}(r)$, $H(r) := \int_0^r h(s)ds$, $h(r) := \left[g(r)', \int_0^r [G(1)-G(s)]'\,ds\right]'$, $\overline{W}_{u\cdot v}(s) := W_{u\cdot v}(s) - W_{u\cdot v}(1)$, *and* $\overline{h}(s) := h(s) - h(1)$.

In (1.18) *as well as in* (1.19) *it holds that the* χ^2_s*-distributed random variable in the numerator is independent of the denominator.*

The limiting null distributions of $\tau_{\text{IM}}(\widehat{\eta}^{\perp})$ and $\tau_{\text{IM}}(\widetilde{\eta}^{\perp})$ are nonstandard but free of any nuisance parameters and only depend on the number of restrictions under the null hypothesis and the number of integrated regressors. Hence, simulating asymptotic critical values is straightforward. Tables 1.5 and 1.6 in Appendix 1.8.1 provide asymptotic critical values for $s \leq m$ and $m = 1, \ldots, 4$.

Remark 1.3. *Jin* et al. *(2006) propose FM-OLS based test statistics with a kernel estimator of* $\Omega_{u\cdot v}$ *based on the FM-OLS residuals and bandwidth equal to sample size. This can be labeled "partial" self-normalization, as the corresponding limiting null distribution accounts for kernel and bandwidth choices to estimate* $\Omega_{u\cdot v}$ *but not for tuning parameter choices related to the FM-OLS estimator. Choosing the IM-OLS estimator instead overcomes this limitation and thus leads to "full" self-normalization (compare also the discussion in Vogelsang and Wagner, 2014, on "partial" fixed-b vs. "full" fixed-b theory).*

1.3.3 Local Asymptotic Power

This section analyzes the asymptotic power properties of the traditional and self-normalized Wald-type tests under local alternatives. To ease exposition of the main arguments, we restrict our attention to the single regressor case ($m = 1$) and consider the null hypothesis $\text{H}_0\colon \beta = \beta_0$. Under local alternatives $\text{H}_1\colon \beta = \beta_0 + cT^{-1}$, $c \in \mathbb{R}$, the limiting distribution of the traditional IM-OLS based Wald-type test statistic is given by

$$\tau_{\text{IM}}(\widehat{\Omega}_{u\cdot v}) \xrightarrow{w} \widetilde{V}_{\text{IM}}(1,1)^{-1} \left(c\,\Omega_{u\cdot v}^{-1/2}\Omega_{vv}^{1/2} + \mathcal{Z}(1) \right)^2, \tag{1.20}$$

with $\widetilde{V}_{\text{IM}}(1,1)$ and $\mathcal{Z}(1)$ as defined in Remark 1.1. The limiting distributions of the traditional FM-OLS and D-OLS based Wald-type test statistics under local alternatives coincide and follow analogously. The limiting distribution of the self-normalized test statistic $\tau_{\text{IM}}(\widehat{\eta}^{\perp})$ under local alternatives is given by

$$\tau_{\text{IM}}(\widehat{\eta}^{\perp}) \xrightarrow{w} \left(\int_0^1 \left(W_{u\cdot v}(r) - h(r)'Q \right)^2 dr \widetilde{V}_{\text{IM}}(1,1) \right)^{-1} \left(c\,\Omega_{u\cdot v}^{-1/2}\Omega_{vv}^{1/2} + \mathcal{Z}(1) \right)^2, \tag{1.21}$$

with analogous results for $\tau_{\text{IM}}(\widehat{\eta})$ and $\tau_{\text{IM}}(\widetilde{\eta}^{\perp})$. For $c = 0$, the limiting distributions coincide with those derived under the null hypothesis in Section 1.3.2 and local asymptotic power of the tests at the nominal α level is equal to α. For $c \neq 0$, local asymptotic power of the tests depends on $c\,\Omega_{u\cdot v}^{-1/2}\Omega_{vv}^{1/2}$. In particular, it follows from the definition of $\Omega_{u\cdot v}$ that local asymptotic power of the tests decreases as the variability in the regression errors increases. To assess the effect of the location parameter c, we plot the simulated power curves as a function of c in Figure 1.1 for $\Omega_{u\cdot v}^{-1/2}\Omega_{vv}^{1/2} = 1$.

For all tests, power increases symmetrically as c moves away from zero. The traditional FM-OLS and D-OLS based tests have identical local asymptotic power, which is somewhat larger than local asymptotic power of the traditional IM-OLS based test. This is not surprising because

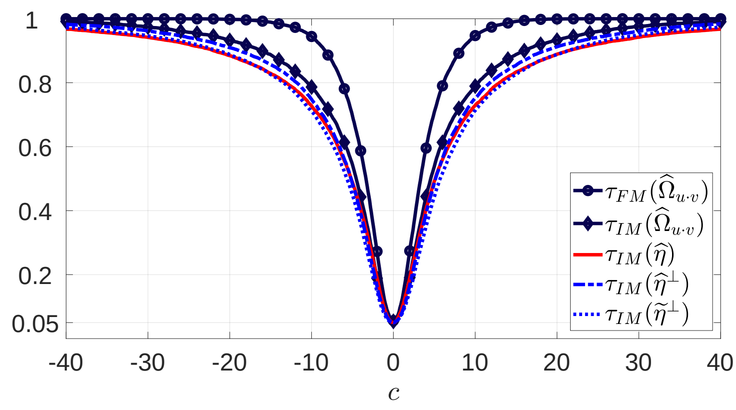


Figure 1.1: Asymptotic power of the traditional and self-normalized Wald-type tests for $\mathrm{H}_0\colon \beta = \beta_0$ at the nominal 5% level under local alternatives $\beta = \beta_0 + cT^{-1}$.
Note: The power curves for $\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$ and $\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$ coincide.

the IM-OLS estimator seems to be asymptotically less efficient than the FM-OLS and D-OLS estimators (cf. Vogelsang and Wagner, 2014, Proposition 2). In general, local asymptotic power of the self-normalized tests is similar to but slightly below local asymptotic power of the traditional IM-OLS based test. This is consistent with the findings in the stationary time series literature, where self-normalization is well known to lead to minor power losses (see, e. g., Kiefer *et al.*, 2000; Shao, 2015). Among the self-normalized tests, $\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$ performs best, while $\tau_{\mathrm{IM}}(\widehat{\eta})$ has some performance advantages over $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$ for small to medium deviations from the null hypothesis, but $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$ catches up as c becomes larger.

1.4 Bootstrap Inference

This section proposes a VAR sieve bootstrap procedure to construct empirical critical values for the traditional and self-normalized Wald-type test statistics. Under the null hypothesis, the bootstrap distribution is expected to serve as a better approximation to the finite sample distributions of these test statistics than the corresponding limiting distributions, especially when the data are characterized by large levels of endogeneity or error serial correlation.

The representation in (1.3) suggests to approximate $\{w_t\}_{t\in\mathbb{Z}} = \{[u_t, v_t']'\}_{t\in\mathbb{Z}}$ by a sequence of VAR processes with increasing order $q \to \infty$ as $T \to \infty$. These VAR approximations can be bootstrapped using the VAR sieve bootstrap. Since u_t is unknown, we fit a finite order VAR to $\widehat{w}_t \coloneqq [\widehat{u}_t, v_t']'$, $t = 1, \dots, T$, where $\widehat{u}_t \coloneqq y_t - x_t'\widehat{\beta}$ denote the residuals in (1.1) based on $\widehat{\beta} \in \{\widehat{\beta}_{\mathrm{D}}, \widehat{\beta}_{\mathrm{FM}}, \widehat{\beta}_{\mathrm{IM}}\}$. Noticeably, it suffices that $T\left(\widehat{\beta} - \beta\right) = O_{\mathbb{P}}(1)$, such that using, e. g., $\widehat{\beta} = \widehat{\beta}_{\mathrm{OLS}}$ is also possible. In the following, let $\widehat{\Phi}_1(q), \dots, \widehat{\Phi}_q(q)$ denote the solution of the sample Yule-Walker equations in the regression of $\widehat{w}_t$ on $\widehat{w}_{t-1}, \dots, \widehat{w}_{t-q}$, $t = q+1, \dots, T$, and denote the corresponding residuals by $\widehat{\varepsilon}_t(q) \coloneqq \widehat{w}_t - \sum_{j=1}^{q} \widehat{\Phi}_j(q)\widehat{w}_{t-j}$. The Yule-Walker estimator is a natural choice, as any finite order VAR estimated by the Yule-Walker estimator is causal and invertible in finite samples, which will be particularly important in the proof of Theorem 1.1 below. The bootstrap scheme to construct critical values consists of four steps and is defined as follows.

Step 1: Obtain the bootstrap sample $(\varepsilon_t^*)_{t=1}^{T}$ by randomly drawing T times with replacement from the centered residuals $\left(\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)\right)_{t=q+1}^{T}$, where $\overline{\widehat{\varepsilon}}_T(q) := (T-q)^{-1}\sum_{t=q+1}^{T}\widehat{\varepsilon}_t(q)$, and construct $(w_t^*)_{t=1}^T$ recursively as $w_t^* = \widehat{\Phi}_1(q)w_{t-1}^* + \ldots + \widehat{\Phi}_q(q)w_{t-q}^* + \varepsilon_t^*$, given initial values $w_{1-q}^*, \ldots, w_0^*$. Partition $w_t^* = [u_t^*, v_t^{*\prime}]'$ analogously to w_t and define $x_t^* := \sum_{s=1}^{t} v_s^*$.

Step 2: Generate data under the null hypothesis H_0: $R_1\beta = r_0$ by defining $y_t^* := x_t^{*\prime}\widehat{\beta}^r + u_t^*$, where $\widehat{\beta}^r$ denotes the restricted version of $\widehat{\beta}$ fulfilling $R_1\widehat{\beta}^r = r_0$.

Step 3: Compute $\widehat{\beta}$ in $y_t^* := x_t^{*\prime}\beta + u_t^*$ and construct the corresponding test statistic for H_0: $R_1\beta = r_0$.

Step 4: Let α denote the desired nominal size of the test. Repeat the previous steps B times, such that $(B+1)(1-\alpha)$ is an integer. Reject the null hypothesis if the test statistic based on the original observations is greater than the $(B+1)(1-\alpha)$-th largest realization of the test statistics based on bootstrap observations.

For later usage, let $\tau^*_{\mathrm{FM}}(\widehat{\Omega}^*_{u\cdot v})$, $\tau^*_{\mathrm{D}}(\widehat{\Omega}^*_{u\cdot v})$, and $\tau^*_{\mathrm{IM}}(\widehat{\Omega}^*_{u\cdot v})$ denote the traditional Wald-type test statistics corresponding to the three modified estimators based on bootstrap observations constructed in Step 3. In particular, note that $\widehat{\Omega}^*_{u\cdot v}$ denotes the estimator of $\Omega_{u\cdot v}$ based on bootstrap observations. Plugging in the long-run variance estimator based on the original observations, $\widehat{\Omega}_{u\cdot v}$, is also possible, but unreported simulation results reveal that this is slightly disadvantageous for the performance of the bootstrap tests in finite samples. Analogously, let $\tau^*_{\mathrm{IM}}(\widehat{\eta}^*)$, $\tau^*_{\mathrm{IM}}(\widehat{\eta}^{\perp *})$, and $\tau^*_{\mathrm{IM}}(\widetilde{\eta}^{\perp *})$ denote the self-normalized Wald-type test statistics based on bootstrap observations.

Remark 1.4. *To eliminate the dependence of the results on the initial values of w_s^*, $1-q \leq s \leq 0$ in applications, we suggest to generate a sufficiently large number of w_t^*'s and keep the last T of them only.*

Remark 1.5. *The restricted IM-OLS estimator of β is given by*

$$\widehat{\beta}^r_{IM} := \begin{bmatrix} I_m \\ 0_{m\times m}\end{bmatrix}' \left(\widehat{\theta}_{IM} - \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} R_2' \left[R_2 \left(\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} R_2'\right]^{-1} \left(R_2\widehat{\theta}_{IM} - r_0\right)\right).$$

The restricted FM-OLS and D-OLS estimators are obtained analogously.

Remark 1.6. *Please note that using the restricted estimator employed in Step 2 also in the construction of $\widehat{w}_t$ has adverse effects under the alternative (cf., e.g., van Giersbergen and Kiviet, 2002; Paparoditis and Politis, 2005).*

To derive asymptotic theory, we have to posit the following technical assumption on the order of the VAR sieve (cf. Assumption 4′ and Remark 7 in Palm *et al.*, 2010).

Assumption 1.4. *Let $q \to \infty$ such that $q = o((T/\ln(T))^{1/3})$ as $T \to \infty$.*

We are now in the position to prove the following bootstrap invariance principle, which is the key ingredient to prove bootstrap consistency for the limiting null distributions of the traditional and self-normalized test statistics.

Theorem 1.1. *Let the data be generated by* (1.1) *and* (1.2), $\{w_t\}_{t\in\mathbb{Z}}$ *satisfy Assumptions 1.1–1.3, and q fulfill Assumption 1.4. Then it holds that*

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} w_t^* \xrightarrow{w^*} \Psi(1)\Sigma^{1/2}W(r), \quad 0 \le r \le 1, \quad in\ \mathbb{P},$$

where $\Psi(1)\Sigma^{1/2}W(r)$ *has covariance matrix* Ω.

Theorem 1.1 extends the bootstrap invariance principle of Park (2002, Theorem 3.3) in the sense that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ has to be an uncorrelated but not necessarily independent white noise process and may thus be of independent interest. Nevertheless, generating the bootstrap quantities $(\varepsilon_t^*)_{t=1}^T$ by drawing independently with replacement from the centered residuals $\left(\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)\right)_{t=q+1}^T$ still allows to capture the entire second order dependence structure of $\{w_t\}_{t\in\mathbb{Z}}$, which is in our context both necessary *and* sufficient for the bootstrap to be consistent. This stems from the fact that the dependence structures in the limiting null distributions of the traditional and self-normalized Wald-type test statistics depend only on the second moments of $\{w_t\}_{t\in\mathbb{Z}}$ and, with respect to second moments, independence and uncorrelatedness are indistinguishable.

The following theorem shows that the VAR sieve bootstrap is consistent for the limiting null distributions of the traditional and self-normalized Wald-type test statistics.

Theorem 1.2. *Let the data be generated by* (1.1) *and* (1.2), $\{w_t\}_{t\in\mathbb{Z}}$ *satisfy Assumptions 1.1–1.3, and q fulfill Assumption 1.4. Then, under both the null hypothesis and the alternative, it holds that* $\tau_{FM}^*(\widehat{\Omega}_{u\cdot v}^*) \xrightarrow{w^*} \chi_s^2$, $\tau_D^*(\widehat{\Omega}_{u\cdot v}^*) \xrightarrow{w^*} \chi_s^2$, $\tau_{IM}^*(\widehat{\Omega}_{u\cdot v}^*) \xrightarrow{w^*} \chi_s^2$, $\tau_{IM}^*(\widehat{\eta}^*) \xrightarrow{w^*} \mathcal{G}_{SN}$, $\tau_{IM}^*(\widehat{\eta}^{\perp *}) \xrightarrow{w^*} \mathcal{G}_{SN}^{\perp}$, *and* $\tau_{IM}^*(\widetilde{\eta}^{\perp *}) \xrightarrow{w^*} \widetilde{\mathcal{G}}_{SN}^{\perp}$ *in* $\mathbb{P}$.

Theorem 1.2 asymptotically justifies the use of VAR sieve bootstrap critical values to conduct inference in cointegrating regresssions based on traditional and self-normalized Wald-type test statistics. In particular, the bootstrap allows to use the self-normalized test statistic $\tau_{\mathrm{IM}}(\widehat{\eta})$ for all $s \le m$, as it accounts for the nuisance parameter dependent correlation structure between numerator and denominator of its limiting null distribution. With respect to the traditional Wald-type test statistic based on the D-OLS estimator, Theorem 1.2 extends the results in Chang *et al.* (2006) in the sense that the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is allowed to be an uncorrelated but not necessarily independent white noise process.

Remark 1.7. *The VAR sieve bootstrap also allows to employ the "textbook" OLS test statistic* $\tau_{OLS}(\widehat{\sigma}_u^2) := \left(R_1\widehat{\beta}_{OLS} - r_0\right)' \left[R_1\widehat{\sigma}_u^2\left(\sum_{t=1}^T x_t x_t'\right)^{-1} R_1'\right]^{-1}\left(R_1\widehat{\beta}_{OLS} - r_0\right)$, *where* $\widehat{\sigma}_u^2 := T^{-1}\sum_{t=1}^T \widehat{u}_{OLS,t}^2$, *which leads to asymptotically valid inference based on chi-squared critical values only in the absence of endogeneity and error serial correlation. However, the finite sample performance of the test turns out to be unsatisfactory (cf. the discussion in Section 1.5).*

1.5 Finite Sample Performance

We generate data according to (1.1) and (1.2) with $m = 2$ regressors, i. e., $y_t = x_{1t}\beta_1 + x_{2t}\beta_2 + u_t$, $x_{it} = x_{i,t-1} + v_{it}$, $i = 1, 2$, for $t = 1, \ldots, T$, where $\beta_1 = \beta_2 = 1$ and $x_{i0} = 0$. The regression errors

u_t and the first differences of the stochastic regressors v_{it} are generated as

$$u_t = \rho_1 u_{t-1} + e_t + \phi e_{t-1} + \rho_2(\nu_{1t} + \nu_{2t}), \quad u_{-100} = 0,$$
$$v_{it} = \nu_{it} + 0.5\nu_{i,t-1}, \quad \nu_{i,-100} = 0, \quad i = 1, 2,$$

for $t = -99, \ldots, 0, 1, \ldots, T$, where $t = -99, \ldots, 0$ serves as a burn-in period to ensure stationarity. The parameters ρ_1 and ρ_2 control the level of error serial correlation and the extent of endogeneity, respectively. For $\phi \neq 0$ the error process contains a first order moving average component. To construct e_t, ν_{1t}, and ν_{2t} we first generate three independent univariate stationary GARCH(1,1) processes $\xi_{jt} = \sigma_{jt}\varepsilon_{jt}$, $j = 1, 2, 3$, where $\sigma_{jt}^2 = a_0 + a_1\xi_{j,t-1}^2 + b_1\sigma_{j,t-1}^2$, with $\xi_{j,-100}^2 = 1$, $\sigma_{j,-100}^2 = 1$, $[\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}]' \sim \mathcal{N}(0, I_3)$ i.i.d. across t, $a_1, b_1 \geq 0$, $a_1 + b_1 < 1$, and $a_0 := 1 - a_1 - b_1$, such that $\mathbb{E}(\xi_{jt}) = 0$ and $\mathbb{E}(\xi_{jt}^2) = \mathbb{E}(\sigma_{jt}^2) = 1$. We then set $[e_t, \nu_{1t}, \nu_{2t}]' = L[\xi_{1t}, \xi_{2t}, \xi_{3t}]'$, where L is the lower triangular matrix of the Cholesky decomposition of the matrix with all main diagonal elements equal to one and all off-diagonal elements equal to ρ_3. In the following we set $\rho_3 = 0.2$ to impose some weak (cross-sectional) correlation between the three GARCH processes. To mimic typical empirical GARCH patterns, we set $a_1 = 0.05$ and $b_1 = 0.94$ (cf. Brüggemann *et al.*, 2016, p. 77). The order $1 \leq q \leq \lfloor T^{1/3} \rfloor =: q_{\max}$ of the VAR sieve is chosen as the one that minimizes either the AIC or the BIC computed on the evaluation period $t = q_{\max}+1, \ldots, T$ (Kilian and Lütkepohl, 2017, p. 56). We present results for $T \in \{75, 100, 250\}$, $\rho_1 = \rho_2 \in \{0, 0.3, 0.6, 0.9\}$ and $\phi \in \{0, 0.3, 0.9\}$. In all cases, the number of Monte Carlo and bootstrap replications is 3,000 and 499, respectively.

Let us start with analyzing the empirical null rejection probabilities of the traditional and self-normalized Wald-type tests based on asymptotic critical values under the null hypothesis $\mathrm{H}_0\colon \beta_1 = 1,\ \beta_2 = 1$. The results are benchmarked against the textbook OLS test $\tau_{\mathrm{OLS}}(\widehat{\sigma}_u^2)$, which leads to asymptotically valid inference only in case $\rho_1 = \rho_2 = \phi = 0$, and against Johansen's (1995) parametric likelihood ratio (LR) test based on the reduced rank quasi maximum likelihood (QML) estimator in a vector error correction model (VECM) for $X_t := [y_t, x_{1t}, x_{2t}]'$ (see Appendix 1.8.2 for more details). The order of the VECM is selected by either AIC or BIC. To select the numbers of leads and lags for the construction of the D-OLS estimator, we follow Choi and Kurozumi (2012) and use BIC (which appears to be the most successful criterion in reducing its RMSE) and the upper bounds used in their simulation study (results based on AIC are qualitatively similar). With respect to long-run covariance matrix estimation, we present results for the Bartlett kernel and the quadratic spectral (QS) kernel together with the corresponding data-dependent bandwidth selection rules of Andrews (1991).

Table 1.1 shows the well known size distortions of the traditional tests whenever ρ_1, ρ_2 is large, or T is small. These size distortions can be almost as severe as those of the textbook OLS test. In contrast, the self-normalized tests are considerably less size-distorted than the traditional tests for $\rho_1, \rho_2 < 0.9$ and even outperform the LR tests in this case. Because the performance of the IM-OLS estimator is comparable with the performance of the D-OLS and FM-OLS estimators in terms of bias and RMSE (see Table 1.7 in Appendix 1.8.2), IM-OLS estimation combined with self-normalization serves as a serious alternative to traditional inference in cointegrating regressions. For $\rho_1, \rho_2 = 0.9$, however, self-normalization becomes less advantageous, reflecting the poor performance of endogeneity corrections in finite samples when the level of endogeneity and

Table 1.1: Empirical sizes of the tests for $H_0\colon \beta_1 = 1,\ \beta_2 = 1$ at 5% level based on asymptotic critical values

						Traditional tests							
			Self-normalized tests			Bartlett kernel			QS kernel			LR tests	
ϕ	ρ_1, ρ_2	$\tau_{\mathrm{OLS}}(\widehat{\sigma}_u^2)$	$\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	LR(AIC)	LR(BIC)
Panel A: $T = 75$													
0	0	0.06	0.04	0.03	0.03	0.13	0.15	0.11	0.17	0.20	0.14	0.13	0.10
	0.3	0.25	0.08	0.05	0.05	0.15	0.21	0.14	0.15	0.23	0.14	0.14	0.15
	0.6	0.69	0.15	0.13	0.08	0.24	0.40	0.19	0.23	0.43	0.18	0.16	0.18
	0.9	0.97	0.66	0.75	0.36	0.42	0.83	0.69	0.54	0.88	0.77	0.27	0.25
0.3	0	0.16	0.06	0.04	0.04	0.15	0.18	0.13	0.17	0.21	0.15	0.14	0.11
	0.3	0.34	0.09	0.07	0.06	0.16	0.23	0.15	0.16	0.25	0.15	0.15	0.16
	0.6	0.69	0.16	0.15	0.09	0.25	0.41	0.22	0.27	0.44	0.23	0.19	0.18
	0.9	0.96	0.65	0.72	0.35	0.46	0.81	0.69	0.63	0.89	0.80	0.32	0.30
0.9	0	0.25	0.07	0.05	0.05	0.16	0.19	0.15	0.17	0.23	0.16	0.18	0.15
	0.3	0.39	0.10	0.08	0.07	0.20	0.26	0.18	0.22	0.30	0.19	0.20	0.17
	0.6	0.67	0.17	0.16	0.10	0.28	0.41	0.26	0.33	0.49	0.31	0.23	0.20
	0.9	0.94	0.62	0.69	0.33	0.52	0.79	0.70	0.73	0.90	0.83	0.39	0.37
Panel B: $T = 100$													
0	0	0.06	0.04	0.03	0.04	0.11	0.13	0.10	0.14	0.17	0.12	0.10	0.08
	0.3	0.27	0.07	0.05	0.05	0.13	0.18	0.13	0.12	0.20	0.13	0.11	0.12
	0.6	0.69	0.12	0.11	0.07	0.22	0.35	0.16	0.21	0.35	0.16	0.13	0.13
	0.9	0.98	0.54	0.64	0.29	0.44	0.77	0.59	0.53	0.82	0.65	0.20	0.17
0.3	0	0.16	0.06	0.04	0.05	0.12	0.16	0.12	0.13	0.18	0.13	0.11	0.10
	0.3	0.35	0.09	0.07	0.06	0.14	0.20	0.14	0.14	0.21	0.14	0.12	0.12
	0.6	0.69	0.14	0.13	0.08	0.24	0.35	0.19	0.24	0.38	0.19	0.15	0.12
	0.9	0.96	0.52	0.62	0.29	0.48	0.76	0.59	0.60	0.84	0.69	0.25	0.22
0.9	0	0.25	0.07	0.05	0.05	0.14	0.18	0.13	0.14	0.19	0.14	0.14	0.14
	0.3	0.41	0.09	0.08	0.06	0.17	0.23	0.16	0.18	0.25	0.17	0.15	0.14
	0.6	0.67	0.14	0.14	0.08	0.26	0.36	0.22	0.28	0.42	0.25	0.18	0.15
	0.9	0.94	0.50	0.59	0.27	0.53	0.74	0.60	0.70	0.86	0.74	0.31	0.27
Panel C: $T = 250$													
0	0	0.06	0.04	0.03	0.04	0.08	0.09	0.07	0.09	0.10	0.08	0.07	0.06
	0.3	0.26	0.06	0.05	0.05	0.10	0.12	0.09	0.09	0.11	0.08	0.08	0.05
	0.6	0.70	0.07	0.06	0.06	0.12	0.21	0.10	0.11	0.19	0.10	0.09	0.05
	0.9	0.98	0.19	0.26	0.12	0.23	0.58	0.26	0.26	0.60	0.29	0.11	0.09
0.3	0	0.16	0.05	0.04	0.05	0.09	0.10	0.08	0.09	0.10	0.08	0.08	0.07
	0.3	0.35	0.06	0.05	0.06	0.11	0.13	0.10	0.10	0.12	0.09	0.09	0.06
	0.6	0.69	0.08	0.07	0.06	0.14	0.21	0.11	0.13	0.21	0.11	0.09	0.06
	0.9	0.98	0.20	0.27	0.13	0.29	0.56	0.29	0.33	0.60	0.34	0.12	0.12
0.9	0	0.25	0.06	0.05	0.05	0.10	0.12	0.09	0.09	0.11	0.09	0.09	0.10
	0.3	0.40	0.07	0.05	0.06	0.12	0.15	0.11	0.11	0.14	0.10	0.10	0.10
	0.6	0.67	0.08	0.08	0.07	0.15	0.22	0.13	0.15	0.23	0.13	0.10	0.12
	0.9	0.96	0.21	0.28	0.14	0.35	0.54	0.32	0.43	0.62	0.41	0.14	0.16

error serial correlation is large. Among the self-normalized tests, the test based on $\widehat{\eta}$ performs best. In particular, adjusting the IM-OLS residuals to remove the correlation between numerator and denominator in the limiting null distribution of $\tau_{\mathrm{IM}}(\widehat{\eta})$ has adverse effects on the performance of the test, especially for $\rho_1, \rho_2 = 0.9$.

Let us now turn to the performance of the tests based on bootstrap critical values. Table 1.2 shows that replacing asymptotic critical values with VAR sieve bootstrap critical values (based on AIC) improves the performance of the traditional tests considerably throughout and also reduces the size distortions of the self-normalized tests, especially in case $\rho_1, \rho_2 = 0.9$. The bootstrap is thus able to account for finite sample effects of both endogeneity corrections and tuning parameter choices. Results based on BIC are similar, compare Table 1.8 in Appendix 1.8.2. Performance differences between traditional and self-normalized tests are negligible for small to medium values of ρ_1, ρ_2. For $\rho_1, \rho_2 = 0.9$, however, the test based on the D-OLS estimator in combination with the Bartlett kernel often performs best among the traditional tests, while the tests based on $\widehat{\eta}$ and

$\widetilde{\eta}^{\perp}$ perform best among the self-normalized tests and often also perform slightly better than the D-OLS test for sample sizes larger than $T = 75$ irrespective of the kernel choice. In case $T = 75$, the D-OLS test based on the Bartlett kernel performs slightly better than the self-normalized test based on $\widehat{\eta}$, which in turn performs slightly better than the D-OLS test based on the QS kernel. Compared to the LR test based on bootstrap critical values (as proposed in Cavaliere *et al.*, 2015) the traditional and self-normalized tests perform very well as long as $\rho_1, \rho_2 < 0.9$, but lead to larger size distortions when $\rho_1, \rho_2 = 0.9$ and $T \in \{75, 100\}$. However, the reduced rank QML estimator is well known to occasionally produce estimates that are far away from the true parameter values (see, e. g., Brüggemann and Lütkepohl, 2005). Thus, it is not surprising that the reduced rank QML estimator has a very large bias and an extremely large RMSE in case $\rho_1, \rho_2 = 0.9$ and $T \in \{75, 100\}$ compared to the (modified) OLS estimators, see Table 1.7 in Appendix 1.8.2. This susceptibility to producing outliers can lead to misleading test decisions based on the LR test in applications.

Table 1.2 also contains the empirical null rejection probabilities of the test statistic $\tau_{\mathrm{IM}}(1)$, whose limiting null distribution is given by $\Omega_{u\cdot v}\chi^2_2$, in conjunction with VAR sieve bootstrap critical values. The test generally performs worse than the traditional and self-normalized tests. This indicates that removing the long-run variance parameter $\Omega_{u\cdot v}$ asymptotically – either by direct estimation or by self-normalization – is beneficial in finite samples. Similarly, the traditional and self-normalized tests also outperform the textbook OLS test, with the differences being most pronounced for $\rho_1, \rho_2 = 0.9$. Thus, it is advantageous to first to account for endogeneity and error serial correlation in the construction of the test statistic before using the VAR sieve bootstrap to construct critical values. We observe similar results also for other choices of the GARCH parameters a_1, b_1, and ρ_3. In particular, the tests based on bootstrap critical values are considerably less size distorted than those based on asymptotic critical values even in case e_t, v_{1t}, and v_{2t} are i.i.d. standard normal and independent of each other ($a_1 = b_1 = \rho_3 = 0$), compare Tables 1.9 and 1.10 in Appendix 1.8.2.

To analyze the properties of the tests under deviations from the null hypothesis, we generate data for $\beta_1 = \beta_2 \in [1, 1.2]$ using 21 values on a grid with mesh size 0.01. To account for the large performance differences under the null hypothesis, we follow Cavaliere *et al.* (2015, p. 826) and present results based on asymptotic and bootstrap critical values corresponding to the nominal size $\widetilde{\alpha}$ that yields empirical null rejection probabilities equal to 5% under the null hypothesis. All size-corrected power curves thus start at 0.05 for $\beta_1 = \beta_2 = 1$. Figure 1.2 displays illustrative size-corrected power curves of the self-normalized test statistics, the traditional test statistics based on the Bartlett kernel and the LR test statistic based on AIC in combination with asymptotic critical values (top row) and bootstrap critical values (bottom row) for $\rho_1, \rho_2 \in \{0.3, 0.6, 0.9\}$ in case $T = 100$ and $\phi = 0.3$. For $\rho_1, \rho_2 = 0.3$ the results are in line with the local asymptotic power results analyzed in Section 1.3.3. The traditional tests based on the FM-OLS and D-OLS estimators are slightly more powerful than the traditional test based on the IM-OLS estimator, which in turn is slightly more powerful than the self-normalized tests. Moreover, the self-normalized tests are as powerful as the LR test. The power loss when replacing asymptotic critical values with bootstrap critical values is negligible for all tests. Increasing ρ_1, ρ_2 reduces power of the tests without altering the relative performance differences with one exception. For $\rho_1, \rho_2 \in \{0.6, 0.9\}$ the LR test is considerably less powerful than the traditional and self-normalized tests, irrespective of whether

Table 1.2: Empirical sizes of the tests for $H_0\colon \beta_1 = 1,\ \beta_2 = 1$ at 5% level based on bootstrap critical values

							Traditional tests							
				Self-normalized tests			Bartlett kernel			QS kernel			LR tests	
ϕ	ρ_1, ρ_2	$\tau^*_{\mathrm{OLS}}(\widehat{\sigma}^2_u)$	$\tau^*_{\mathrm{IM}}(1)$	$\tau^*_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	LR*(AIC)	LR*(BIC)
Panel A: $T = 75$														
0	0	0.07	0.10	0.07	0.07	0.07	0.06	0.06	0.07	0.05	0.05	0.07	0.06	0.07
	0.3	0.09	0.12	0.08	0.07	0.08	0.06	0.06	0.08	0.06	0.05	0.08	0.07	0.11
	0.6	0.12	0.16	0.09	0.09	0.08	0.09	0.08	0.09	0.09	0.07	0.09	0.07	0.13
	0.9	0.33	0.53	0.24	0.28	0.19	0.14	0.22	0.25	0.18	0.24	0.28	0.08	0.09
0.3	0	0.06	0.10	0.07	0.06	0.07	0.06	0.06	0.07	0.05	0.05	0.06	0.07	0.07
	0.3	0.08	0.13	0.07	0.07	0.08	0.06	0.06	0.08	0.06	0.05	0.08	0.06	0.10
	0.6	0.11	0.18	0.09	0.09	0.09	0.09	0.07	0.10	0.08	0.07	0.09	0.07	0.12
	0.9	0.31	0.53	0.24	0.27	0.19	0.16	0.22	0.26	0.21	0.24	0.29	0.09	0.11
0.9	0	0.09	0.16	0.09	0.10	0.09	0.07	0.08	0.10	0.05	0.06	0.09	0.09	0.09
	0.3	0.10	0.19	0.10	0.11	0.10	0.08	0.08	0.11	0.07	0.06	0.09	0.08	0.09
	0.6	0.14	0.24	0.12	0.13	0.11	0.10	0.08	0.13	0.09	0.07	0.12	0.08	0.11
	0.9	0.31	0.55	0.26	0.29	0.21	0.18	0.23	0.27	0.24	0.24	0.30	0.09	0.11
Panel B: $T = 100$														
0	0	0.07	0.08	0.07	0.07	0.07	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.06
	0.3	0.08	0.10	0.07	0.07	0.07	0.06	0.06	0.08	0.06	0.05	0.07	0.06	0.09
	0.6	0.11	0.13	0.08	0.08	0.07	0.09	0.06	0.09	0.08	0.06	0.09	0.06	0.09
	0.9	0.29	0.44	0.17	0.22	0.15	0.15	0.18	0.19	0.17	0.18	0.21	0.07	0.06
0.3	0	0.07	0.09	0.07	0.06	0.07	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.06
	0.3	0.07	0.10	0.07	0.07	0.07	0.06	0.06	0.08	0.06	0.05	0.07	0.07	0.08
	0.6	0.10	0.15	0.08	0.09	0.08	0.09	0.07	0.10	0.09	0.06	0.09	0.08	0.07
	0.9	0.28	0.44	0.18	0.22	0.16	0.17	0.18	0.20	0.20	0.20	0.22	0.09	0.08
0.9	0	0.09	0.13	0.09	0.09	0.08	0.07	0.07	0.09	0.05	0.06	0.08	0.08	0.10
	0.3	0.10	0.16	0.10	0.10	0.09	0.08	0.08	0.10	0.07	0.06	0.09	0.08	0.09
	0.6	0.13	0.21	0.10	0.11	0.10	0.10	0.09	0.11	0.10	0.07	0.10	0.07	0.07
	0.9	0.28	0.46	0.19	0.23	0.17	0.19	0.19	0.22	0.23	0.19	0.24	0.09	0.09
Panel C: $T = 250$														
0	0	0.06	0.06	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.05	0.05
	0.3	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.06	0.06	0.04
	0.6	0.08	0.09	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.06	0.03
	0.9	0.13	0.19	0.08	0.10	0.08	0.09	0.09	0.09	0.10	0.10	0.10	0.06	0.06
0.3	0	0.07	0.07	0.06	0.06	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.05	0.06
	0.3	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.06	0.07	0.04
	0.6	0.08	0.09	0.06	0.06	0.06	0.07	0.06	0.07	0.06	0.06	0.06	0.07	0.04
	0.9	0.14	0.20	0.08	0.10	0.08	0.10	0.10	0.10	0.11	0.10	0.11	0.07	0.08
0.9	0	0.08	0.08	0.06	0.07	0.06	0.05	0.07	0.07	0.05	0.06	0.07	0.06	0.09
	0.3	0.08	0.10	0.07	0.07	0.07	0.06	0.07	0.07	0.05	0.06	0.07	0.06	0.08
	0.6	0.10	0.12	0.07	0.08	0.07	0.07	0.07	0.08	0.07	0.06	0.07	0.06	0.09
	0.9	0.16	0.24	0.09	0.12	0.10	0.12	0.11	0.11	0.13	0.09	0.12	0.06	0.10

Notes: Superscript "$*$" signifies the use of bootstrap critical values. The VAR sieve bootstrap is based on AIC.

it is based on asymptotic or bootstrap critical values. The difference is even more pronounced for $T = 75$ but becomes smaller as the sample size increases. The small size distortions of the LR test based on bootstrap critical values under the null hypothesis are thus accompanied by relatively low power under the alternative. With respect to the moving average component in the regression errors, we find that power of the tests is generally largest for $\phi = 0$ and becomes smaller as ϕ increases. The effect is most pronounced for the LR test, especially for $T \in \{75, 100\}$. Increasing the sample size is clearly beneficial for all tests, especially in case $\rho_1, \rho_2 = 0.9$, compare Figure 1.3 in Appendix 1.8.2, which shows the results for $T = 250$. Finally, note that size-corrected power of the self-normalized tests is not necessarily lower than size-corrected power of the traditional tests, see, e. g., Figure 1.4 in Appendix 1.8.2 for $\rho_1, \rho_2 = 0.9$, which shows the size-corrected power curves of the traditional tests based on the QS kernel (and of the LR test based on BIC).

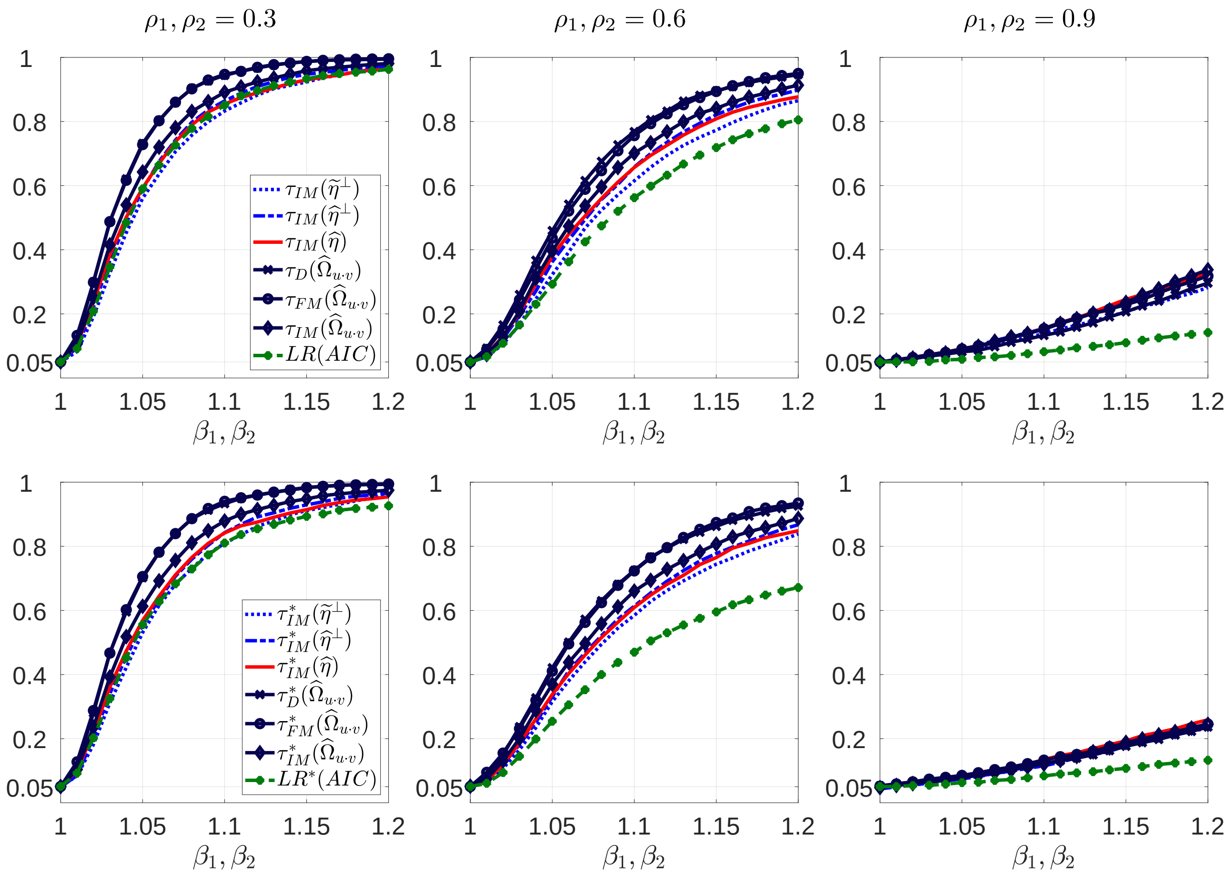


Figure 1.2: Size-corrected power of the tests for H_0: $\beta_1 = 1$, $\beta_2 = 1$ at 5% level based on asymptotic critical values (top row) and bootstrap critical values (bottom row) for $T = 100$ and $\phi = 0.3$.
Note: Long-run variance parameters are estimated using the Bartlett kernel and the VAR sieve bootstrap is based on AIC.

1.6 Empirical Illustration: The Fisher Effect

Many empirical studies suggest that inflation π_t and the short-term nominal interest rate i_t do not cointegrate with the slope of inflation being equal to one. This finding is at odds with the Fisher effect, which is backed by many theoretical models (see, e. g., Westerlund, 2008, for a brief description of underlying economic theory). The errors u_t in the Fisher equation $i_t = \alpha + \beta\pi_t + u_t$, $t = 1, \ldots, T$, are likely to be highly persistent even in case cointegration between π_t and i_t prevails. Consequently, the Fisher effect might be rejected even if it exists because traditional tests are prone to severe size distortions in this case (cf., e. g., Caporale and Pittis, 2004; Westerlund, 2008). Since self-normalization is less advantageous when the errors are highly persistent, we expect similar results also for the self-normalized tests.

This section illustrates the usefulness of the VAR sieve bootstrap when analyzing the validity of the Fisher effect. We consider the relationship between inflation and the short-term nominal interest rate for 19 OECD countries between 1990Q1 and 2019Q4 ($T = 120$) using quarterly data (measured at annual rates in percentages) obtained from the OECD databases Economic Outlook and Main Economic Indicators (see Table 1.3 for details). As a simple persistence measure of the regression errors in the Fisher equation, we regress the OLS residuals on their first lag. The empirical first order autocorrelations lie between 0.88 and 0.95 – indeed indicating highly persistent errors.

To shed some light on the integratedness of the variables, we employ the augmented Dickey-Fuller (ADF) test based on generalized least squares (GLS) demeaning (Elliott *et al.*, 1996) and the KPSS test (Kwiatkowski *et al.*, 1992). All tests in this section are carried out at the nominal 10% level and the bandwidth for long-run variance estimation is always selected with the data-dependent rule of Andrews (1991). Results for the short-term interest rates are unambiguous. The ADF-GLS test (based on AIC with a maximum of five lags) does not reject the null hypothesis of a unit root for any country, whereas the KPSS test (based on the Bartlett kernel) rejects the null hypothesis of stationarity for all countries except Iceland, see Table 1.11 in Appendix 1.8.3. The table also presents evidence for a unit root in inflation, but the results are less persuasive. The ADF-GLS test rejects the null hypothesis of a unit root for five countries and the KPSS test decides in favor of stationarity for eleven countries. However, only for Austria, Belgium, and Ireland do both tests decide in favor of stationarity. The results are in line with some parts of the literature questioning an exact unit root in inflation (see, e. g., Jensen, 2009). Nevertheless, it is common practice in applications to treat both the interest rate and the inflation rate as integrated processes of order one (cf., e. g., Caporale and Pittis, 2004, p. 35). To test for cointegration, we employ the group-mean and pooled panel no-cointegration tests developed in Westerlund (2008). The tests are more powerful than single equation no-cointegration tests, especially in the presence of highly persistent errors. Using again the Bartlett kernel for long-run variance estimation, both tests reject the null hypothesis of no-cointegration. The results are robust to different choices for the maximal number of common factors. Moreover, we obtain similar results when the test statistics are constructed under the restriction that $\beta = 1$, which already provides some evidence for the validity of the Fisher effect in the individual countries.

We now test the null hypothesis $\beta = 1$ for all countries separately using the same test statistics as already analyzed in Section 1.5. Table 1.3 summarizes the results. In all but two countries, at least one of the nine tests rejects the validity of the Fisher effect when using asymptotic critical values. Moreover, for six countries at least five tests decide against the Fisher effect. Using VAR sieve bootstrap critical values based on AIC (with 499 replications) instead yields different results, with those based on BIC – reported in Table 1.12 in Appendix 1.8.3 – being similar. For 14 countries none of the tests rejects the validity of the Fisher effect and for additional four countries at most two tests reject the Fisher effect. Hence, after accounting for highly persistent errors, the empirical results support the Fisher effect in OECD countries, which is consistent with many theoretical models. Finally, note that for Italy all bootstrap tests reject the null hypothesis, which serves as strong evidence against the validity of the Fisher effect in this particular country.

1.7 Summary and Conclusions

To address the severe size distortions of hypotheses tests in cointegrating regressions, this paper combines a VAR sieve bootstrap to construct critical values with a self-normalization approach that avoids direct estimation of long-run variance parameters when computing test statistics. To prove bootstrap consistency, we derive a bootstrap invariance principle which allows for uncorrelated

Table 1.3: Realizations of test statistics for $\mathrm{H}_0\colon \beta = 1$

				Traditional tests					
	Self-normalized tests			Bartlett kernel			QS kernel		
Country	$\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$
Australia	**75.03**	**553.76**	5.22	0.04	1.00	**4.66**	0.04	0.81	**4.10**
Austria	<u>**1387.15**</u>	**2779.14**	52.04	0.98	**2.76**	**4.20**	0.78	2.30	**3.31**
Belgium	**945.79**	**3215.69**	31.43	0.34	1.22	**4.34**	0.26	1.26	**3.32**
Canada	<u>**780.87**</u>	**1506.10**	49.42	0.03	1.21	**6.86**	0.03	0.84	**6.55**
Germany	<u>**1543.31**</u>	<u>**3428.59**</u>	**87.20**	**3.03**	**6.77**	**7.02**	2.50	**5.56**	**5.78**
Denmark	**57.10**	**188.85**	1.68	0.68	**4.47**	1.38	0.59	**4.21**	1.20
Spain	**221.66**	**271.51**	**135.61**	**4.64**	**18.63**	<u>**16.51**</u>	**4.31**	**16.65**	<u>**15.33**</u>
Finland	**396.46**	**796.37**	**73.81**	0.35	**4.08**	**8.61**	0.30	**3.40**	**7.30**
France	**214.28**	**1532.73**	**97.37**	2.13	**6.30**	**9.87**	1.77	**4.97**	**8.23**
United Kingdom	**429.67**	**1324.88**	18.83	0.31	1.41	2.03	0.29	1.73	1.90
Ireland	23.67	88.01	2.00	0.04	0.07	0.75	0.03	0.13	0.60
Iceland	15.17	17.59	1.90	1.47	1.63	2.47	1.21	1.33	2.04
Italy	<u>**481.56**</u>	<u>**736.29**</u>	<u>**254.43**</u>	<u>**14.64**</u>	<u>**36.67**</u>	<u>**33.04**</u>	<u>**14.84**</u>	<u>**37.03**</u>	<u>**33.50**</u>
Netherlands	**193.87**	**671.68**	57.02	1.29	0.90	**3.75**	1.16	0.45	**3.37**
Norway	**370.22**	**813.18**	44.70	0.02	0.48	**4.42**	0.02	0.25	**4.31**
New Zealand	**159.91**	**432.83**	2.00	0.13	0.67	1.62	0.10	0.57	1.27
Sweden	**368.28**	**1542.48**	24.60	0.24	0.93	**5.94**	0.25	0.64	**6.29**
United States	**206.56**	**405.61**	53.45	1.12	1.10	**4.62**	0.95	0.92	**3.92**
South Africa	16.43	30.25	3.35	0.54	0.38	0.60	0.50	0.24	0.56

Notes: **Bold** numbers indicate significance at the nominal 10% level based on asymptotic critical values, whereas <u>underlined</u> numbers indicate significance at the nominal 10% level based on VAR sieve bootstrap critical values using AIC. The inflation rates and short-term interest rates are available on https://data.oecd.org/price/inflation-cpi.htm and https://data.oecd.org/interest/short-term-interest-rates.htm, respectively (Accessed: March 24, 2022).

but not necessarily independent white noise increments and might be of independent interest. In particular, it allows us to prove bootstrap consistency also for traditional test statistics based on popular modified OLS estimators. Simulation results show that the VAR sieve bootstrap reduces size distortions of hypotheses tests in cointegrating regressions considerably, with two self-normalized test statistics often outperforming the traditional test statistics at the cost of only small power losses. Among the traditional test statistics, the one based on the D-OLS estimator in combination with the Bartlett kernel performs best and turns out to be the closest competitor to the self-normalized test statistics. For the QS kernel, however, the D-OLS based test statistic becomes less competitive. Finally, the empirical illustration demonstrates that replacing asymptotic critical values with VAR sieve bootstrap critical values when analyzing the validity of the Fisher effect in OECD countries leads to alternative conclusions, which are more in line with economic theory.

Possible extensions of the proposed methods to, e. g., panels of cointegrating regressions are currently under investigation. Finally, note that a crucial assumption of the paper is that the regressors possess an exact unit root, while parts of the literature already allow for nearly integrated regressors (see, e. g., Phillips and Magdalinos, 2009; Müller and Watson, 2013; Hwang and Valdes, 2023). However, extending the VAR sieve bootstrap to this setting is not straightforward at all, as constructing bootstrap regressors requires some knowledge of the true local to unity parameters, which are not consistently estimable. Whether the approaches in Phillips *et al.* (2001), Phillips (2022), or Hwang and Valdes (2023) help to overcome this limitation will be examined in future research.

Acknowledgements

We are grateful to the editor Atsushi Inoue, an associate editor, and four referees for several insightful and constructive comments that have led to significant changes and improvements of the paper. We further thank Katharina Hees, Fabian Knorre, and participants at the Econometrics Colloquium at the University of Konstanz, the IAAE 2021 Annual Conference, the 2021 Asian and North American Summer Meetings of the Econometric Society, and the XII$_t$ Workshop in Time Series Econometrics in Zaragoza for helpful comments. Parts of this research were conducted while Karsten Reichold held a position at the University of Klagenfurt. The authors have no competing interests to declare.

1.8 Appendix

1.8.1 Asymptotic Critical Values

To simulate asymptotic critical values for the self-normalized test statistics, we approximate standard Brownian motions with normalized sums of 10,000 i.i.d. standard normal random variables and approximate the corresponding integrals accordingly. We tabulate critical values based on 10,000 Monte Carlo replications for various choices of m and s and different deterministic regressors in the model $y_t = d_t'\delta + x_t'\beta + u_t$. Table 1.4 displays critical values for $\tau_{\mathrm{IM}}(\widehat{\eta})$ in case $s = m$, whereas Tables 1.5 and 1.6 display critical values for $\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$ and $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$, respectively, for $s \leq m$.

Table 1.4: Asymptotic critical values for $\tau_{\mathrm{IM}}(\widehat{\eta})$

%	$s = m = 1$	$s = m = 2$	$s = m = 3$	$s = m = 4$
Panel A: No deterministic regressors				
90.0	36.52	122.05	239.61	399.56
95.0	56.59	166.72	311.99	505.48
97.5	78.72	216.55	385.87	628.92
99.0	120.18	286.41	490.05	759.33
Panel B: $d_t = 1$				
90.0	63.80	168.27	304.10	476.69
95.0	95.47	232.12	392.99	593.92
97.5	134.95	291.93	487.56	712.47
99.0	186.28	379.48	597.20	870.72
Panel C: $d_t = [1, t]'$				
90.0	90.33	207.46	361.72	541.86
95.0	133.13	281.36	457.89	682.79
97.5	183.47	355.65	562.45	804.92
99.0	243.48	460.43	708.85	967.07
Panel D: $d_t = [1, t, t^2]'$				
90.0	115.03	244.49	416.04	602.70
95.0	165.89	329.89	526.60	756.21
97.5	216.76	398.40	633.99	892.73
99.0	289.76	510.98	799.74	1060.42
Panel E: $d_t = [1, t, t^2, t^3]'$				
90.0	136.71	290.12	462.65	673.86
95.0	197.68	375.53	583.50	849.22
97.5	263.32	465.10	713.27	992.71
99.0	351.88	581.93	891.45	1206.54

Notes: Critical values depend on the deterministic regressors in the model $y_t = d_t'\delta + x_t'\beta + u_t$. In the absence of deterministic regressors, the model reduces to (1.1). s denotes the number of linearly independent restrictions under the null hypothesis on the coefficients corresponding to the m integrated regressors x_t.

Table 1.5: Asymptotic critical values for $\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$

	$m=1$	$m=2$		$m=3$			$m=4$			
%	$s=1$	$s=1$	$s=2$	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$	$s=4$
Panel A: No deterministic regressors										
90.0	96.54	166.42	293.57	236.48	419.51	578.76	314.30	556.36	756.59	943.89
95.0	142.94	244.30	401.14	350.09	567.82	744.79	471.73	741.97	964.69	1175.09
97.5	193.92	336.23	519.00	488.99	704.00	899.24	634.50	907.73	1174.09	1411.81
99.0	277.97	453.55	666.54	654.89	880.21	1137.23	849.80	1188.62	1462.50	1738.58
Panel B: $d_t = 1$										
90.0	134.97	208.54	363.71	276.67	496.62	666.96	364.40	637.85	862.48	1063.50
95.0	201.02	305.08	501.62	410.92	661.96	863.81	535.71	818.80	1084.90	1312.78
97.5	275.28	406.10	620.11	561.97	819.51	1050.81	707.94	1018.90	1318.49	1579.26
99.0	393.49	558.42	807.29	749.96	1054.24	1301.78	942.26	1342.73	1612.33	1889.18
Panel C: $d_t = [1, t]'$										
90.0	190.66	257.02	446.54	329.87	574.89	782.45	403.32	704.21	959.97	1203.93
95.0	279.53	375.36	606.18	475.42	764.00	1012.18	582.17	926.78	1211.47	1485.34
97.5	377.44	496.59	757.11	655.43	980.47	1234.03	774.90	1148.83	1474.29	1780.90
99.0	533.19	666.36	950.33	906.94	1227.60	1507.77	1038.12	1492.33	1790.50	2138.93
Panel D: $d_t = [1, t, t^2]'$										
90.0	210.50	272.01	481.06	348.76	606.70	831.37	425.58	742.19	1010.63	1268.76
95.0	305.52	395.48	649.51	519.21	815.02	1056.50	634.84	976.08	1287.99	1567.06
97.5	413.25	535.90	797.93	684.90	1037.40	1311.09	817.63	1230.69	1535.12	1842.51
99.0	563.99	725.63	990.81	962.61	1313.25	1633.04	1056.64	1562.85	1890.96	2224.02
Panel E: $d_t = [1, t, t^2, t^3]'$										
90.0	232.81	301.71	534.19	376.05	636.88	884.17	457.80	774.69	1057.38	1341.91
95.0	337.10	443.42	699.16	539.19	850.73	1128.92	659.33	1022.97	1370.91	1668.09
97.5	463.23	577.35	872.26	723.96	1082.00	1377.80	865.86	1295.33	1651.39	1973.51
99.0	630.98	793.15	1088.12	1009.81	1407.99	1656.51	1162.26	1628.12	2003.57	2447.55

Note: See notes to Table 1.4.

Table 1.6: Asymptotic critical values for $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$

	$m=1$	$m=2$		$m=3$			$m=4$			
%	$s=1$	$s=1$	$s=2$	$s=1$	$s=2$	$s=3$	$s=1$	$s=2$	$s=3$	$s=4$
Panel A: No deterministic regressors										
90.0	31.17	55.66	101.62	77.15	143.61	201.83	104.41	190.12	265.10	341.34
95.0	50.05	86.11	140.29	119.69	200.58	271.42	156.84	267.78	360.27	441.01
97.5	70.10	119.77	186.17	166.98	262.92	342.76	224.47	346.51	448.52	541.91
99.0	105.10	170.54	246.93	242.51	349.84	448.97	326.05	450.10	571.11	678.05
Panel B: $d_t = 1$										
90.0	45.01	71.31	125.68	92.46	169.99	236.01	122.22	217.80	309.88	387.55
95.0	68.35	104.56	175.27	139.80	238.75	317.78	188.00	304.88	400.09	495.37
97.5	98.76	146.22	232.64	202.26	315.59	406.82	266.13	378.18	491.60	590.77
99.0	146.11	213.20	299.64	299.57	405.54	523.10	346.66	490.62	625.07	739.01
Panel C: $d_t = [1, t]'$										
90.0	62.20	87.00	153.96	111.39	200.06	276.88	134.81	244.37	335.37	432.73
95.0	95.29	130.06	217.87	168.89	275.45	366.57	208.49	331.73	444.18	552.70
97.5	135.63	178.04	287.62	232.76	355.40	457.69	282.60	426.11	557.58	673.97
99.0	191.58	248.13	359.29	335.84	469.64	611.68	379.30	554.57	711.90	851.77
Panel D: $d_t = [1, t, t^2]'$										
90.0	68.57	91.02	164.53	117.45	212.78	295.20	144.32	260.53	358.65	455.70
95.0	104.75	137.19	227.99	177.22	292.15	391.81	216.69	356.30	479.79	585.59
97.5	146.06	190.77	293.31	251.60	382.53	501.02	294.84	463.69	597.52	720.36
99.0	207.70	271.29	383.54	369.29	517.33	649.04	407.87	598.79	747.23	879.40
Panel E: $d_t = [1, t, t^2, t^3]'$										
90.0	76.23	100.00	181.90	126.90	224.69	313.29	154.82	271.37	373.04	482.67
95.0	115.24	153.14	255.47	192.13	309.00	413.24	230.12	365.11	496.19	619.74
97.5	158.92	212.59	331.46	267.33	397.29	519.51	317.70	470.32	638.75	786.57
99.0	232.98	291.22	428.16	376.51	531.63	673.24	433.59	630.16	823.91	959.16

Note: See notes to Table 1.4.

1.8.2 Additional Finite Sample Results

Johansen's Likelihood Ratio Test

The data generating process (DGP) in Section 1.5 can be expressed as $X_t = \Pi_1 X_{t-1} + \epsilon_t$, where $X_t = [y_t, x_{1t}, x_{2t}]'$, $\epsilon_t = A_0^{-1}[u_t, v_{1t}, v_{2t}]'$, and $\Pi_1 = A_0^{-1} A_1$, with

$$A_0 \coloneqq \begin{bmatrix} 1 & -\beta_1 & -\beta_2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad A_1 \coloneqq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The corresponding error correction model of order $k \geq 1$ with one cointegrated relation is given by

$$\Delta X_t = ab' X_{t-1} + \sum_{l=1}^{k-1} \Gamma_l \Delta X_{t-l} + \epsilon_t, \tag{1.22}$$

where $\Delta X_s \coloneqq X_s - X_{s-1}$, $a = [-1, 0, 0]'$ and $b = [1, -\beta_1, -\beta_2]'$. The order $1 \leq k \leq \lfloor T^{1/3} \rfloor =: k_{\max}$ is determined by either AIC or BIC computed on the evaluation period $t = k_{\max} + 1, \ldots, T$ (cf. Kilian and Lütkepohl, 2017, p. 56). Testing the null hypothesis $\mathrm{H}_0\colon \beta_1 = 1,\ \beta_2 = 1$ in the DGP in Section 1.5 corresponds to testing the null hypothesis $\mathrm{H}_0\colon b = [1, -1, -1]'$ in (1.22). The reduced rank QML estimator of Johansen (1995) is used to estimate b and the null hypothesis is tested by means of the LR test statistic of Johansen (1995). If the assumptions in Johansen (1995) are fulfilled, the LR test statistic follows a chi-squared distribution with two degrees of freedom asymptotically under the null hypothesis.

Estimator and Test Performance

Table 1.7: Bias and RMSE of the estimators of β_1

		Bias × 100							RMSE × 100						
					FM-OLS		QML					FM-OLS		QML	
ϕ	ρ_1, ρ_2	OLS	IM-OLS	D-OLS	Bartlett	QS	AIC	BIC	OLS	IM-OLS	D-OLS	Bartlett	QS	AIC	BIC
Panel A: $T = 75$															
0	0	0.38	-0.25	-0.27	0.04	0.11	-0.05	-0.17	3.12	5.15	3.52	3.28	3.42	15.31	3.61
	0.3	2.05	-0.25	-0.08	0.66	0.79	-0.45	-0.21	5.64	7.29	4.91	5.02	5.38	19.79	5.92
	0.6	6.91	0.88	1.65	3.64	3.70	-6.82	-0.67	13.24	12.72	11.68	11.12	11.97	374.48	39.17
	0.9	30.40	19.63	12.69	24.93	25.00	-606.33	10.47	46.55	58.08	46.22	48.09	55.54	31796.03	950.30
0.3	0	0.62	-0.28	-0.18	0.13	0.19	0.39	-0.23	4.05	6.67	4.46	4.24	4.42	20.65	4.93
	0.3	2.39	-0.28	0.10	0.87	1.00	0.42	-0.51	6.91	9.46	7.43	6.50	6.93	23.56	16.41
	0.6	7.45	0.91	1.52	4.16	4.30	1.25	-2.15	15.05	16.45	16.59	13.63	14.75	77.59	161.73
	0.9	31.84	20.44	12.89	26.44	26.37	75.50	-43.35	50.91	68.99	58.15	54.10	88.03	6141.08	1687.86
0.9	0	1.11	-0.35	0.09	0.34	0.41	-0.66	-0.28	5.98	9.74	8.06	6.27	6.56	20.00	8.05
	0.3	3.06	-0.33	0.35	1.37	1.51	-0.95	0.61	9.54	13.83	12.87	9.56	10.16	76.13	16.51
	0.6	8.52	0.96	1.60	5.14	5.59	-3.55	11.10	18.96	23.95	26.29	18.48	19.82	354.12	485.48
	0.9	34.72	22.08	13.54	29.43	29.69	-31.68	-41.48	60.55	92.30	83.42	66.97	122.62	4171.58	4348.58
Panel B: $T = 100$															
0	0	0.33	-0.07	-0.15	0.05	0.09	0.13	0.01	2.36	3.91	2.37	2.38	2.46	6.22	2.71
	0.3	1.60	-0.03	0.03	0.45	0.51	-0.14	0.03	4.29	5.55	3.25	3.55	3.72	9.96	4.20
	0.6	5.36	0.74	1.53	2.48	2.44	-2.05	-2.25	10.25	9.71	6.63	7.76	8.25	115.53	115.31
	0.9	25.50	14.76	11.97	20.06	20.35	-54.65	-26.69	39.15	45.30	33.92	39.03	42.27	3497.87	1741.91
0.3	0	0.54	-0.07	-0.07	0.12	0.15	0.02	0.00	3.08	5.07	2.97	3.09	3.18	3.97	3.65
	0.3	1.90	-0.01	0.19	0.60	0.65	0.09	0.05	5.30	7.20	4.21	4.62	4.84	11.40	6.16
	0.6	5.85	0.83	1.71	2.92	2.94	-5.09	-6.13	11.72	12.56	8.85	9.67	10.28	289.60	290.91
	0.9	26.89	15.60	12.88	21.49	22.36	12.16	27.51	42.85	53.98	42.04	44.37	50.19	903.93	1264.81
0.9	0	0.97	-0.06	0.12	0.30	0.31	0.24	0.03	4.57	7.39	4.42	4.56	4.71	9.54	5.14
	0.3	2.50	0.03	0.49	1.00	1.05	0.06	0.60	7.36	10.51	6.40	6.84	7.18	21.90	8.02
	0.6	6.83	1.01	2.04	3.72	3.86	-1.24	-15.58	14.86	18.27	13.33	13.42	14.37	168.09	920.77
	0.9	29.67	17.27	14.19	24.28	25.81	123.50	-44.37	50.98	72.40	58.58	55.89	66.60	4255.25	1543.85
Panel C: $T = 250$															
0	0	0.14	0.01	-0.06	0.02	0.02	0.01	0.02	0.91	1.59	0.91	0.91	0.92	0.95	0.93
	0.3	0.68	0.03	0.02	0.12	0.12	0.01	-0.01	1.69	2.27	1.25	1.33	1.35	1.38	1.37
	0.6	2.39	0.21	0.46	0.79	0.69	0.04	-0.23	4.33	3.96	2.38	2.80	2.84	2.50	2.53
	0.9	13.50	5.01	4.25	8.83	8.67	-0.17	-1.89	20.99	17.71	11.78	17.39	18.36	15.07	18.51
0.3	0	0.23	0.02	-0.03	0.04	0.04	0.01	0.02	1.19	2.07	1.14	1.18	1.19	1.25	1.21
	0.3	0.81	0.05	0.10	0.17	0.16	-0.00	0.05	2.08	2.95	1.61	1.74	1.77	1.83	1.80
	0.6	2.61	0.24	0.55	0.93	0.84	0.01	-0.12	4.93	5.14	3.06	3.54	3.63	3.33	3.31
	0.9	14.26	5.32	4.98	9.64	9.82	-1.37	-2.71	22.81	21.88	14.73	19.96	21.40	36.14	105.98
0.9	0	0.41	0.04	0.08	0.09	0.09	0.02	0.03	1.77	3.02	1.67	1.74	1.77	1.91	1.79
	0.3	1.07	0.07	0.25	0.30	0.29	0.07	0.04	2.90	4.30	2.41	2.58	2.65	2.87	2.64
	0.6	3.04	0.31	0.75	1.21	1.15	0.03	-0.01	6.21	7.50	4.44	5.01	5.19	6.09	4.76
	0.9	15.77	5.95	6.43	11.20	11.85	5.25	-0.65	26.78	30.54	20.66	25.18	27.23	290.19	54.54

Note: Results for β_2 are similar and therefore not reported. QML denotes the reduced rank quasi maximum likelihood estimator of Johansen (1995).

Table 1.8: Empirical sizes of the tests for $\mathrm{H}_0\colon \beta_1 = 1,\ \beta_2 = 1$ at 5% level based on bootstrap critical values

							Traditional tests							
				Self-normalized tests			Bartlett kernel			QS kernel			LR tests	
ϕ	ρ_1, ρ_2	$\tau^*_{\mathrm{OLS}}(\widehat{\sigma}^2_u)$	$\tau^*_{\mathrm{IM}}(1)$	$\tau^*_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	LR*(AIC)	LR*(BIC)
Panel A: $T = 75$														
0	0	0.06	0.08	0.06	0.06	0.06	0.05	0.06	0.07	0.05	0.05	0.06	0.06	0.07
	0.3	0.09	0.10	0.07	0.06	0.07	0.06	0.06	0.07	0.06	0.05	0.06	0.07	0.11
	0.6	0.13	0.12	0.07	0.07	0.06	0.08	0.08	0.07	0.08	0.07	0.07	0.07	0.13
	0.9	0.34	0.48	0.20	0.24	0.16	0.12	0.22	0.22	0.16	0.25	0.26	0.08	0.09
0.3	0	0.05	0.07	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.07	0.07
	0.3	0.06	0.08	0.05	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.06	0.06	0.10
	0.6	0.08	0.10	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.12
	0.9	0.26	0.45	0.15	0.17	0.14	0.11	0.19	0.19	0.18	0.24	0.26	0.09	0.11
0.9	0	0.03	0.06	0.04	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.06	0.09	0.09
	0.3	0.05	0.09	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.08	0.09
	0.6	0.07	0.14	0.06	0.07	0.07	0.07	0.06	0.08	0.07	0.07	0.08	0.08	0.11
	0.9	0.24	0.47	0.18	0.20	0.16	0.15	0.20	0.21	0.22	0.24	0.29	0.09	0.11
Panel B: $T = 100$														
0	0	0.06	0.07	0.06	0.06	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.06	0.06
	0.3	0.09	0.08	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.05	0.07	0.06	0.09
	0.6	0.12	0.09	0.06	0.06	0.06	0.08	0.07	0.06	0.08	0.07	0.07	0.06	0.09
	0.9	0.29	0.36	0.14	0.18	0.13	0.12	0.17	0.16	0.15	0.20	0.19	0.07	0.06
0.3	0	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.06	0.06	0.06
	0.3	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.07	0.08
	0.6	0.07	0.07	0.04	0.05	0.05	0.06	0.05	0.06	0.07	0.06	0.06	0.08	0.07
	0.9	0.21	0.34	0.12	0.13	0.11	0.12	0.14	0.15	0.17	0.19	0.19	0.09	0.08
0.9	0	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.08	0.10
	0.3	0.06	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.06	0.08	0.09
	0.6	0.09	0.13	0.07	0.07	0.07	0.08	0.07	0.08	0.08	0.06	0.08	0.07	0.07
	0.9	0.25	0.40	0.16	0.19	0.15	0.17	0.17	0.19	0.22	0.19	0.23	0.09	0.09
Panel C: $T = 250$														
0	0	0.06	0.06	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	0.3	0.08	0.06	0.05	0.05	0.06	0.05	0.06	0.05	0.05	0.06	0.05	0.06	0.04
	0.6	0.09	0.06	0.05	0.05	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.06	0.03
	0.9	0.12	0.17	0.07	0.09	0.08	0.08	0.10	0.09	0.08	0.09	0.09	0.06	0.06
0.3	0	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
	0.3	0.06	0.04	0.05	0.04	0.05	0.04	0.05	0.04	0.04	0.05	0.05	0.07	0.04
	0.6	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.07	0.04
	0.9	0.12	0.20	0.08	0.10	0.08	0.10	0.10	0.10	0.11	0.09	0.10	0.07	0.08
0.9	0	0.09	0.09	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.09
	0.3	0.10	0.12	0.07	0.07	0.07	0.07	0.07	0.08	0.06	0.06	0.07	0.06	0.08
	0.6	0.12	0.13	0.07	0.08	0.07	0.08	0.07	0.08	0.07	0.06	0.07	0.06	0.09
	0.9	0.18	0.23	0.10	0.12	0.10	0.12	0.11	0.11	0.13	0.10	0.11	0.06	0.10

Notes: Superscript "$*$" signifies the use of bootstrap critical values. The VAR sieve bootstrap is based on BIC.

Table 1.9: Empirical sizes of the tests for $H_0\colon \beta_1 = 1,\ \beta_2 = 1$ at 5% level based on asymptotic critical values in the i.i.d. case ($a_1 = b_1 = \rho_3 = 0$)

						Traditional tests							
			Self-normalized tests			Bartlett kernel			QS kernel			LR tests	
ϕ	ρ_1, ρ_2	$\tau_{\text{OLS}}(\widehat{\sigma}_u^2)$	$\tau_{\text{IM}}(\widetilde{\eta}^{\perp})$	$\tau_{\text{IM}}(\widehat{\eta}^{\perp})$	$\tau_{\text{IM}}(\widehat{\eta})$	$\tau_{\text{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\text{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\text{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\text{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\text{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\text{IM}}(\widehat{\Omega}_{u\cdot v})$	LR(AIC)	LR(BIC)
Panel A: $T = 75$													
0	0	0.06	0.04	0.02	0.03	0.13	0.16	0.11	0.17	0.20	0.14	0.12	0.09
	0.3	0.21	0.07	0.05	0.05	0.16	0.20	0.14	0.17	0.23	0.15	0.14	0.13
	0.6	0.60	0.16	0.14	0.09	0.25	0.35	0.23	0.25	0.37	0.22	0.17	0.19
	0.9	0.95	0.63	0.70	0.35	0.49	0.79	0.66	0.60	0.84	0.73	0.30	0.32
0.3	0	0.15	0.06	0.04	0.04	0.15	0.18	0.13	0.17	0.21	0.15	0.13	0.09
	0.3	0.30	0.09	0.07	0.06	0.18	0.22	0.17	0.18	0.24	0.17	0.16	0.14
	0.6	0.60	0.18	0.16	0.09	0.27	0.36	0.25	0.29	0.40	0.27	0.20	0.19
	0.9	0.93	0.61	0.67	0.33	0.52	0.77	0.66	0.67	0.85	0.76	0.36	0.36
0.9	0	0.23	0.07	0.05	0.05	0.17	0.20	0.15	0.18	0.22	0.16	0.18	0.14
	0.3	0.34	0.10	0.08	0.06	0.21	0.26	0.19	0.23	0.29	0.21	0.21	0.18
	0.6	0.58	0.19	0.17	0.10	0.29	0.38	0.29	0.35	0.45	0.34	0.25	0.22
	0.9	0.90	0.59	0.64	0.30	0.55	0.74	0.68	0.76	0.87	0.82	0.42	0.41
Panel B: $T = 100$													
0	0	0.06	0.04	0.02	0.03	0.11	0.14	0.10	0.14	0.17	0.12	0.10	0.08
	0.3	0.23	0.07	0.05	0.05	0.14	0.18	0.14	0.14	0.18	0.13	0.11	0.11
	0.6	0.62	0.12	0.11	0.08	0.24	0.31	0.19	0.23	0.31	0.18	0.13	0.14
	0.9	0.96	0.50	0.60	0.27	0.51	0.73	0.57	0.58	0.79	0.64	0.23	0.21
0.3	0	0.16	0.05	0.04	0.05	0.13	0.15	0.12	0.14	0.17	0.13	0.11	0.09
	0.3	0.32	0.08	0.06	0.06	0.16	0.19	0.15	0.15	0.20	0.14	0.12	0.12
	0.6	0.61	0.14	0.13	0.09	0.25	0.32	0.22	0.25	0.33	0.22	0.15	0.14
	0.9	0.94	0.50	0.58	0.27	0.53	0.71	0.58	0.66	0.81	0.69	0.28	0.26
0.9	0	0.24	0.07	0.05	0.05	0.15	0.17	0.13	0.15	0.18	0.14	0.14	0.14
	0.3	0.36	0.09	0.07	0.06	0.19	0.22	0.17	0.19	0.24	0.17	0.16	0.16
	0.6	0.59	0.14	0.14	0.09	0.27	0.33	0.25	0.31	0.37	0.28	0.19	0.16
	0.9	0.90	0.49	0.56	0.25	0.56	0.69	0.59	0.73	0.84	0.74	0.34	0.32
Panel C: $T = 250$													
0	0	0.05	0.04	0.03	0.04	0.08	0.09	0.07	0.08	0.10	0.08	0.07	0.06
	0.3	0.23	0.05	0.05	0.05	0.10	0.12	0.09	0.09	0.11	0.09	0.08	0.06
	0.6	0.61	0.08	0.07	0.06	0.16	0.19	0.11	0.14	0.17	0.10	0.09	0.05
	0.9	0.96	0.20	0.26	0.13	0.31	0.52	0.28	0.33	0.55	0.31	0.12	0.07
0.3	0	0.15	0.05	0.04	0.05	0.09	0.11	0.09	0.08	0.10	0.08	0.07	0.07
	0.3	0.31	0.06	0.05	0.05	0.10	0.13	0.10	0.09	0.12	0.09	0.08	0.06
	0.6	0.60	0.08	0.08	0.06	0.16	0.20	0.13	0.15	0.19	0.12	0.10	0.06
	0.9	0.94	0.21	0.27	0.13	0.35	0.50	0.31	0.40	0.54	0.36	0.13	0.11
0.9	0	0.23	0.05	0.05	0.05	0.10	0.12	0.09	0.09	0.11	0.09	0.09	0.11
	0.3	0.36	0.06	0.05	0.06	0.11	0.15	0.11	0.11	0.14	0.10	0.10	0.10
	0.6	0.59	0.09	0.09	0.07	0.17	0.21	0.15	0.16	0.21	0.14	0.10	0.11
	0.9	0.91	0.22	0.27	0.14	0.40	0.48	0.35	0.47	0.57	0.42	0.17	0.18

Table 1.10: Empirical sizes of the tests for $H_0\colon \beta_1 = 1$, $\beta_2 = 1$ at 5% level based on bootstrap critical values in the i.i.d. case ($a_1 = b_1 = \rho_3 = 0$)

							Traditional tests							
				Self-normalized tests			Bartlett kernel			QS kernel			LR tests	
ϕ	ρ_1, ρ_2	$\tau^*_{\mathrm{OLS}}(\widehat{\sigma}^2_u)$	$\tau^*_{\mathrm{IM}}(1)$	$\tau^*_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau^*_{\mathrm{IM}}(\widehat{\eta})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau^*_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	LR*(AIC)	LR*(BIC)
Panel A: $T = 75$														
0	0	0.06	0.09	0.06	0.06	0.07	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.06
	0.3	0.08	0.11	0.07	0.07	0.07	0.06	0.06	0.07	0.05	0.05	0.06	0.07	0.09
	0.6	0.12	0.16	0.08	0.09	0.08	0.08	0.08	0.09	0.08	0.07	0.08	0.06	0.13
	0.9	0.33	0.48	0.23	0.26	0.20	0.16	0.23	0.24	0.20	0.23	0.26	0.08	0.14
0.3	0	0.06	0.09	0.07	0.06	0.07	0.05	0.06	0.07	0.04	0.05	0.06	0.06	0.06
	0.3	0.07	0.12	0.07	0.07	0.07	0.06	0.06	0.07	0.05	0.05	0.07	0.07	0.08
	0.6	0.11	0.17	0.09	0.09	0.09	0.08	0.07	0.09	0.07	0.06	0.08	0.07	0.11
	0.9	0.30	0.48	0.22	0.25	0.19	0.16	0.21	0.24	0.22	0.22	0.28	0.09	0.12
0.9	0	0.09	0.16	0.09	0.09	0.10	0.07	0.08	0.09	0.05	0.06	0.08	0.09	0.09
	0.3	0.10	0.19	0.10	0.11	0.10	0.08	0.08	0.11	0.06	0.06	0.09	0.08	0.09
	0.6	0.13	0.25	0.13	0.14	0.12	0.09	0.09	0.13	0.08	0.07	0.11	0.09	0.10
	0.9	0.28	0.51	0.24	0.28	0.19	0.19	0.21	0.25	0.24	0.22	0.28	0.10	0.11
Panel B: $T = 100$														
0	0	0.06	0.08	0.06	0.06	0.06	0.05	0.06	0.06	0.04	0.05	0.06	0.06	0.06
	0.3	0.08	0.09	0.06	0.06	0.07	0.06	0.06	0.07	0.05	0.05	0.06	0.06	0.07
	0.6	0.11	0.13	0.07	0.07	0.07	0.08	0.07	0.08	0.07	0.06	0.07	0.07	0.10
	0.9	0.28	0.39	0.17	0.20	0.14	0.15	0.18	0.18	0.18	0.18	0.20	0.07	0.08
0.3	0	0.06	0.08	0.06	0.06	0.07	0.05	0.05	0.06	0.04	0.05	0.06	0.06	0.06
	0.3	0.07	0.10	0.07	0.06	0.07	0.06	0.06	0.07	0.05	0.05	0.06	0.06	0.07
	0.6	0.10	0.14	0.08	0.08	0.08	0.08	0.07	0.09	0.07	0.06	0.08	0.07	0.08
	0.9	0.27	0.39	0.17	0.20	0.15	0.17	0.18	0.19	0.20	0.18	0.21	0.09	0.08
0.9	0	0.09	0.13	0.08	0.08	0.08	0.06	0.07	0.08	0.05	0.06	0.07	0.08	0.10
	0.3	0.10	0.15	0.09	0.09	0.09	0.07	0.08	0.09	0.06	0.06	0.08	0.07	0.10
	0.6	0.12	0.21	0.10	0.10	0.10	0.09	0.08	0.11	0.08	0.07	0.09	0.07	0.08
	0.9	0.25	0.42	0.18	0.21	0.16	0.18	0.17	0.20	0.22	0.17	0.23	0.09	0.10
Panel C: $T = 250$														
0	0	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.04	0.05	0.06	0.05	0.05
	0.3	0.06	0.07	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.05	0.06	0.05	0.04
	0.6	0.07	0.07	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.05	0.06	0.06	0.03
	0.9	0.14	0.18	0.07	0.10	0.07	0.09	0.10	0.09	0.09	0.09	0.09	0.06	0.04
0.3	0	0.06	0.07	0.05	0.06	0.06	0.05	0.06	0.06	0.04	0.05	0.06	0.05	0.05
	0.3	0.06	0.07	0.05	0.06	0.06	0.05	0.06	0.06	0.05	0.05	0.06	0.06	0.05
	0.6	0.08	0.08	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.07	0.06	0.04
	0.9	0.15	0.19	0.07	0.09	0.08	0.10	0.10	0.09	0.11	0.09	0.10	0.07	0.06
0.9	0	0.07	0.08	0.06	0.07	0.06	0.05	0.07	0.07	0.05	0.06	0.06	0.07	0.09
	0.3	0.08	0.09	0.06	0.07	0.07	0.05	0.07	0.07	0.05	0.06	0.07	0.06	0.08
	0.6	0.09	0.11	0.07	0.07	0.07	0.07	0.07	0.08	0.06	0.06	0.07	0.06	0.08
	0.9	0.16	0.23	0.09	0.11	0.09	0.12	0.10	0.11	0.12	0.09	0.11	0.06	0.09

Note: See notes to Table 1.2.

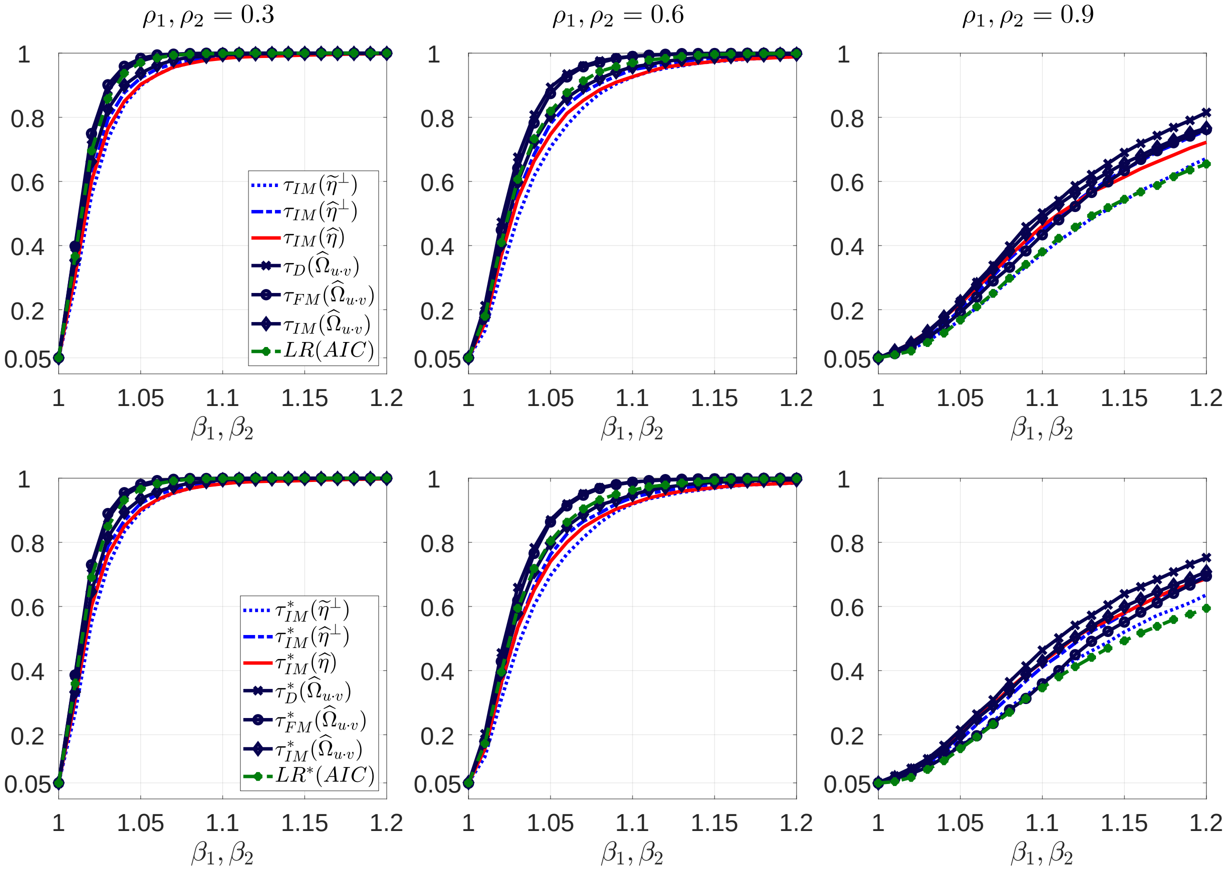


Figure 1.3: Size-corrected power of the tests for H_0: $\beta_1 = 1$, $\beta_2 = 1$ at 5% level based on asymptotic critical values (top row) and bootstrap critical values (bottom row) for $T = 250$ and $\phi = 0.3$.

Note: See note to Figure 1.2.

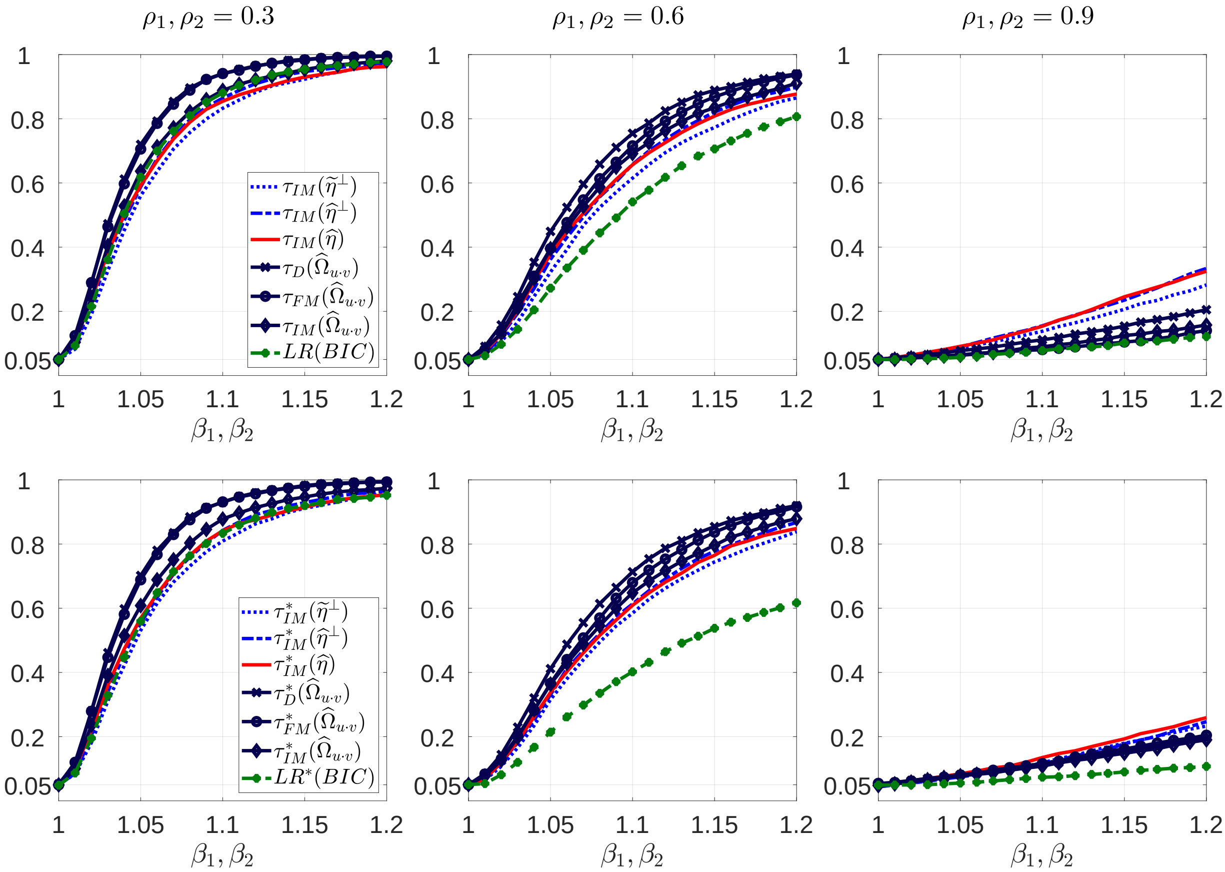


Figure 1.4: Size-corrected power of the tests for H_0: $\beta_1 = 1$, $\beta_2 = 1$ at 5% level based on asymptotic critical values (top row) and bootstrap critical values (bottom row) for $T = 100$ and $\phi = 0.3$.

Note: Long-run variance parameters are estimated using the QS kernel and the VAR sieve bootstrap is based on AIC.

1.8.3 Additional Empirical Results

Table 1.11: Realizations of test statistics

	i_t		π_t	
Country	ADF-GLS	KPSS	ADF-GLS	KPSS
Australia	0.17	**0.53**	-0.68	0.24
Austria	-0.12	**0.40**	**-1.99**	0.30
Belgium	0.28	**0.41**	**-1.91**	0.24
Canada	-0.41	**0.48**	-0.83	0.26
Germany	-0.25	**0.40**	**-2.08**	**0.35**
Denmark	0.06	**0.44**	-0.69	**0.39**
Spain	1.11	**0.39**	-0.19	**0.48**
Finland	0.19	**0.46**	-0.40	0.32
France	0.01	**0.41**	-0.92	**0.38**
United Kingdom	-0.02	**0.43**	-1.34	0.32
Ireland	-0.38	**0.54**	**-2.13**	0.31
Iceland	-0.46	0.31	-0.28	0.11
Italy	0.18	**0.38**	-0.26	**0.43**
Netherlands	-0.13	**0.40**	**-2.54**	**0.44**
Norway	-0.37	**0.45**	-1.18	0.17
New Zealand	0.18	**0.44**	-0.61	0.21
Sweden	0.48	**0.42**	-0.63	0.34
United States	-0.80	**0.36**	-1.11	**0.55**
South Africa	-0.67	**0.38**	-0.65	**0.40**

Note: **Bold** numbers indicate significance at the nominal 10% level.

Table 1.12: Realizations of test statistics for $\mathrm{H}_0\colon \beta = 1$

				Traditional tests					
	Self-normalized tests			Bartlett kernel			QS kernel		
Country	$\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$	$\tau_{\mathrm{IM}}(\widehat{\eta})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{D}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{FM}}(\widehat{\Omega}_{u\cdot v})$	$\tau_{\mathrm{IM}}(\widehat{\Omega}_{u\cdot v})$
Australia	**75.03**	**553.76**	5.22	0.04	1.00	**4.66**	0.04	0.81	**4.10**
Austria	**<u>1387.15</u>**	**2779.14**	52.04	0.98	**2.76**	**4.20**	0.78	2.30	**3.31**
Belgium	**945.79**	**3215.69**	31.43	0.34	1.22	**4.34**	0.26	1.26	**3.32**
Canada	**780.87**	**1506.10**	49.42	0.03	1.21	**6.86**	0.03	0.84	**6.55**
Germany	**<u>1543.31</u>**	**<u>3428.59</u>**	**87.20**	**3.03**	**6.77**	**7.02**	2.50	**5.56**	**5.78**
Denmark	**57.10**	**188.85**	1.68	0.68	**4.47**	1.38	0.59	**4.21**	1.20
Spain	**221.66**	**271.51**	**135.61**	**4.64**	**18.63**	<u>16.51</u>	**4.31**	**16.65**	<u>15.33</u>
Finland	**396.46**	**796.37**	**73.81**	0.35	**4.08**	**8.61**	0.30	**3.40**	**7.30**
France	**214.28**	**1532.73**	**97.37**	2.13	**6.30**	**9.87**	1.77	**4.97**	**8.23**
United Kingdom	**429.67**	**1324.88**	18.83	0.31	1.41	2.03	0.29	1.73	1.90
Ireland	23.67	88.01	2.00	0.04	0.07	0.75	0.03	0.13	0.60
Iceland	15.17	17.59	1.90	1.47	1.63	2.47	1.21	1.33	2.04
Italy	**<u>481.56</u>**	**<u>736.29</u>**	**<u>254.43</u>**	**14.64**	**<u>36.67</u>**	**<u>33.04</u>**	**14.84**	**<u>37.03</u>**	**<u>33.50</u>**
Netherlands	**193.87**	**671.68**	57.02	1.29	0.90	**3.75**	1.16	0.45	**3.37**
Norway	**370.22**	**813.18**	44.70	0.02	0.48	**4.42**	0.02	0.25	**4.31**
New Zealand	**159.91**	**432.83**	2.00	0.13	0.67	1.62	0.10	0.57	1.27
Sweden	**368.28**	**<u>1542.48</u>**	24.60	0.24	0.93	**5.94**	0.25	0.64	**6.29**
United States	**<u>206.56</u>**	**405.61**	53.45	1.12	1.10	**4.62**	0.95	0.92	**3.92**
South Africa	16.43	30.25	3.35	0.54	0.38	0.60	0.50	0.24	0.56

Notes: **Bold** numbers indicate significance at the nominal 10% level based on asymptotic critical values, whereas <u>underlined</u> numbers indicate significance at the nominal 10% level based on VAR sieve bootstrap critical values using BIC.

1.8.4 Proofs of Main Results

Proof of Proposition 1.1. Vogelsang and Wagner (2014, Lemma 2) show that

$$T^{-1/2}\sum_{t=2}^{\lfloor rT \rfloor}\Delta\widehat{S}_t^u \xrightarrow{w} \Omega_{u\cdot v}^{1/2}\left(W_{u\cdot v}(r)-g(r)'\mathcal{Z}\right), \quad 0\leq r\leq 1.$$

The continuous mapping theorem thus yields

$$\widehat{\eta}=T^{-1}\sum_{t=2}^{T}\left(T^{-1/2}\sum_{s=2}^{t}\Delta\widehat{S}_s^u\right)^2 \xrightarrow{w} \Omega_{u\cdot v}\int_0^1\left(W_{u\cdot v}(r)-g(r)'\mathcal{Z}\right)^2 dr$$

and the final result follows from standard arguments similar to those used in Vogelsang and Wagner (2014, Proof of Theorem 3). ☐

Proof of Proposition 1.2. The result for $\tau_{\mathrm{IM}}(\widehat{\eta}^{\perp})$ follows with the same arguments as used in the proof of Proposition 1.1 by noting that

$$T^{-1/2}\sum_{t=2}^{\lfloor rT \rfloor}\Delta\widehat{S}_s^{u\perp} \xrightarrow{w} \Omega_{u\cdot v}^{1/2}\left(W_{u\cdot v}(r)-h(r)'Q\right), \quad 0\leq r\leq 1,$$

as shown in Lemma 2 in Vogelsang and Wagner (2014). Moreover, $\tau_{\mathrm{IM}}(\widetilde{\eta}^{\perp})$ coincides with the $\widetilde{W}^*$ test statistic in Vogelsang and Wagner (2014) based on the Bartlett kernel and $b=1$. The result thus follows from Theorem 3 in Vogelsang and Wagner (2014). ☐

The proof of the remaining results relies on the following lemma.

Lemma 1.1. *With the definitions in Section 1.4 it holds under Assumptions 1.1–1.4 that*

$$\max_{q+1\leq t\leq T}|\widehat{w}_t-w_t|_F=O_{\mathbb{P}}(T^{-1/2}) \tag{1.23}$$

and

$$q^{1/2}\sum_{j=1}^{q}|\widehat{\Phi}_j(q)-\widetilde{\Phi}_j(q)|_F=O_{\mathbb{P}}(q^3/T)=o_{\mathbb{P}}(1), \tag{1.24}$$

where $\widetilde{\Phi}_1(q),\ldots,\widetilde{\Phi}_q(q)$ *denote the solution of the sample Yule-Walker equations in the regression of* w_t *on* $w_{t-1},\ldots,w_{t-q}$, $t=q+1,\ldots,T$.

Proof. See Appendix 1.8.5. ☐

The following two key ingredients in the proof of Lemma 1.1 are also useful hereafter. First, it holds under Assumptions 1.1 and 1.4 that

$$q^{3/2}\sup_{1\leq j\leq q}|\widetilde{\Phi}_j(q)-\Phi_j(q)|_F=q^{3/2}O_{\mathbb{P}}((\ln(T)/T)^{1/2})=O_{\mathbb{P}}(1), \tag{1.25}$$

compare Meyer and Kreiss (2015, Remark 3.3.), where $\Phi_1(q),\ldots,\Phi_q(q)$ denote the solution of the population Yule-Walker equations based on the true moments. Second, under Assumption 1.1 with

$k \geq 3/2$, there exist constants $q_0 \in \mathbb{N}$ and $c < \infty$ such that

$$\sum_{j=1}^{q}(1+j)^k|\Phi_j(q) - \Phi_j|_F \leq c \sum_{j=q+1}^{\infty}(1+j)^k|\Phi_j|_F, \tag{1.26}$$

for all $q \geq q_0$ and the right-hand side converges to zero as $q \to \infty$, see Meyer and Kreiss (2015, Lemma 3.1).[1]

We use Lemma 1.1 to prove the following two lemmas, which are then used to prove Lemma 1.4.

Lemma 1.2. *It holds under Assumptions 1.1–1.4 that*

$$\mathbb{E}^*\left(|\varepsilon_t^*|_F^a\right) = (T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)|_F^a = O_{\mathbb{P}}(1) \ \textit{in}\ \mathbb{P},$$

for the $a > 2$ from Assumption 1.1.

Proof. See Appendix 1.8.5. □

Lemma 1.3. *It holds under Assumptions 1.1–1.4 that*

$$\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right) = (T-q)^{-1}\sum_{t=q+1}^{T}\left(\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)\right)\left(\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)\right)' = \Sigma + o_{\mathbb{P}}(1) \ \textit{in}\ \mathbb{P}.$$

Proof. See Appendix 1.8.5. □

Lemma 1.4. *It holds under Assumptions 1.1–1.4 that*

$$W_T^*(r) \coloneqq T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor}\varepsilon_t^* \xrightarrow{w^*} \Sigma^{1/2}W(r), \quad 0 \leq r \leq 1, \quad \textit{in}\ \mathbb{P},$$

with $\Sigma^{1/2}(\Sigma^{1/2})' = \Sigma$.

Proof. See Appendix 1.8.5. □

Proof of Theorem 1.1. Using similar arguments as Palm *et al.* (2010, p. 670), it follows that

$$T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor} w_t^* = \left(I - \sum_{j=1}^{q}\widehat{\Phi}_j(q)\right)^{-1} W_T^*(r) + T^{-1/2}(\overline{w}_0^* - \overline{w}_{\lfloor rT \rfloor}^*),$$

where $\overline{w}_{t-1}^* \coloneqq \left(I - \sum_{j=1}^{q}\widehat{\Phi}_j(q)\right)^{-1}\sum_{i=1}^{q}\left(\sum_{j=i}^{q}\widehat{\Phi}_j(q)\right)w_{t-i}^*$. Given Lemma 1.4, it remains to show that

$$I - \sum_{j=1}^{q}\widehat{\Phi}_j(q) \xrightarrow{p} \Phi(1) \tag{1.27}$$

[1] This result is known as the generalized Baxter's inequality, see Baxter (1962) and Hannan and Deistler (1988, p. 269).

and

$$\mathbb{P}^*\left(\max_{0\le t\le T}|T^{-1/2}\overline{w}_t^*|_F > \delta\right) = o_{\mathbb{P}}(1). \tag{1.28}$$

We first show (1.27). From Lemma 1.1, (1.25), and (1.26) we obtain

$$\begin{aligned}
|I-\sum_{j=1}^{q}\widehat{\Phi}_j(q)-\Phi(1)|_F &\le \sum_{j=1}^{q}|\widehat{\Phi}_j(q)-\widetilde{\Phi}_j(q)|_F + \sum_{j=1}^{q}|\widetilde{\Phi}_j(q)-\Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q}|\Phi_j(q)-\Phi_j|_F + \sum_{j=q+1}^{\infty}|\Phi_j|_F \\
&\le \sum_{j=1}^{q}|\widehat{\Phi}_j(q)-\widetilde{\Phi}_j(q)|_F + q\sup_{1\le j\le q}|\widetilde{\Phi}_j(q)-\Phi_j(q)|_F \\
&\quad + c\sum_{j=q+1}^{\infty}|\Phi_j|_F + \sum_{j=q+1}^{\infty}|\Phi_j|_F \\
&= o_{\mathbb{P}}(1)+o_{\mathbb{P}}(1)+o(1)+o(1) = o_{\mathbb{P}}(1).
\end{aligned}$$

To prove (1.28), note that it follows from strict stationarity of $\{\overline{w}_t^*\}_{t\in\mathbb{Z}}$ and Markov's inequality, that

$$\begin{aligned}
\mathbb{P}^*\left(\max_{0\le t\le T}|T^{-1/2}\overline{w}_t^*|_F > \delta\right) &\le \sum_{t=0}^{T}\mathbb{P}^*\left(|T^{-1/2}\overline{w}_t^*|_F > \delta\right) \\
&\le (T+1)\mathbb{P}^*\left(|T^{-1/2}\overline{w}_t^*|_F > \delta\right) \\
&\le \delta^{-a}(T^{1-a/2}+T^{-a/2})\mathbb{E}^*\left(|\overline{w}_t^*|_F^a\right),
\end{aligned}$$

with the $a > 2$ from Assumption 1.1, compare Park (2002, p. 486). Similarly as in Palm *et al.* (2010, p. 671), we obtain

$$\mathbb{E}^*\left(|\overline{w}_t^*|_F^a\right) \le c\,(m+1)^{a/2-1}\left(\sum_{j=0}^{\infty}|\overline{\widehat{\Psi}}_j(q)|_F^2\right)^{a/2}\mathbb{E}^*\left(|\varepsilon_t^*|_F^a\right),$$

for some constant c and $\overline{\widehat{\Psi}}_j(q) \coloneqq \sum_{i=j+1}^{\infty}\widehat{\Psi}_i(q)$, where the matrices $\widehat{\Psi}_j(q)$ are determined by the power series expansion of the inverse of $I-\sum_{j=1}^{q}\widehat{\Phi}_j(q)z^j$. As discussed in Palm *et al.* (2010, p. 671), it follows that $\sum_{j=0}^{\infty}|\overline{\widehat{\Psi}}_j(q)|_F^2 = O_{\mathbb{P}}(1)$ if we can show that $\sum_{j=1}^{q}j^{1/2}|\widehat{\Psi}_j(q)|_F = O_{\mathbb{P}}(1)$, which in turn holds if $\sum_{j=1}^{q} j^{1/2}|\widehat{\Phi}_j(q)|_F = O_{\mathbb{P}}(1)$. Using again Lemma 1.1, (1.25), and (1.26), we obtain

$$\begin{aligned}
\sum_{j=1}^{q} j^{1/2}|\widehat{\Phi}_j(q)|_F &\leq \sum_{j=1}^{q} j^{1/2}|\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F + \sum_{j=1}^{q} j^{1/2}|\widetilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q} j^{1/2}|\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2}|\Phi_j|_F \\
&\leq q^{1/2}\sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F + q^{1/2}\sum_{j=1}^{q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q}(1+j)|\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2}|\Phi_j|_F \\
&\leq q^{1/2}\sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F + q^{3/2}\sup_{1\leq j\leq q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F \\
&\quad + \sum_{j=1}^{q}(1+j)|\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} j^{1/2}|\Phi_j|_F \\
&= o_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) + o(1) + O(1) = O_{\mathbb{P}}(1).
\end{aligned}$$

This completes the proof, since $\mathbb{E}^*\left(|\varepsilon_t^*|_F^a\right) = O_{\mathbb{P}}(1)$ by Lemma 1.2 for the $a > 2$ from Assumption 1.1. $\square$

Proof of Theorem 1.2. Given the invariance principle for $\{w_t^*\}_{t\in\mathbb{Z}}$ in Theorem 1.1, it follows with similar arguments as used in the proofs of Propositions 1.1 and 1.2 that $\tau_{\mathrm{IM}}^*(\widehat{\eta}^*) \xrightarrow{w^*} \mathcal{G}_{\mathrm{SN}}$, $\tau_{\mathrm{IM}}^*(\widehat{\eta}^{\perp *}) \xrightarrow{w^*} \mathcal{G}_{\mathrm{SN}}^{\perp}$, and $\tau_{\mathrm{IM}}^*(\widetilde{\eta}^{\perp *}) \xrightarrow{w^*} \widetilde{\mathcal{G}}_{\mathrm{SN}}^{\perp}$ in $\mathbb{P}$.

Let us now turn to the traditional Wald-type test statistics. We first prove that the kernel estimator of Ω based on the bootstrap sample, defined as

$$\widehat{\Omega}^* = \begin{bmatrix} \widehat{\Omega}_{uu}^* & \widehat{\Omega}_{uv}^* \\ \widehat{\Omega}_{vu}^* & \widehat{\Omega}_{vv}^* \end{bmatrix} := T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{T}\mathcal{K}\left(\frac{|i-j|}{b_T}\right)\widehat{w}_{\mathrm{OLS},i}^*\widehat{w}_{\mathrm{OLS},j}^{*\prime}, \tag{1.29}$$

where $\widehat{w}_{\mathrm{OLS},t}^* := [\widehat{u}_{\mathrm{OLS},t}^*, v_t^{*\prime}]'$, with $\widehat{u}_{\mathrm{OLS},t}^*$ denoting the OLS residuals in the bootstrap regression $y_t^* = x_t^{*\prime}\beta + u_t^*$, converges in $\xrightarrow{p^*}$ to Ω in $\mathbb{P}$. To this end, let $f^*(\cdot)$ denote the spectral density matrix of $\{w_t^*\}_{t\in\mathbb{Z}}$, i. e.,

$$\begin{aligned}
f^*(\lambda) &:= \left(\sum_{j=-\infty}^{\infty}\widehat{\Psi}_j(q)e^{ij\lambda}\right) f_\varepsilon^*(\lambda)\left(\sum_{j=-\infty}^{\infty}\widehat{\Psi}_j(q)e^{-ij\lambda}\right) \\
&= \left(I - \sum_{j=1}^{q}\widehat{\Phi}_j(q)e^{ij\lambda}\right)^{-1} f_\varepsilon^*(\lambda)\left(I - \sum_{j=1}^{q}\widehat{\Phi}_j(q)e^{-ij\lambda}\right)^{-1},
\end{aligned} \tag{1.30}$$

with $\widehat{\Psi}_j(q)$ as defined in the proof of Theorem 1.1 and $f_\varepsilon^*(\lambda) := (2\pi)^{-1}\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right)$ denoting the spectral density matrix of $\{\varepsilon_t^*\}_{t\in\mathbb{Z}}$. It follows from Chang *et al.* (2006, Lemma A.3) that $\sup_\lambda|f^*(\lambda) - f(\lambda)|_F \xrightarrow{p^*} 0$ in $\mathbb{P}$. In particular, for $\lambda = 0$, we obtain $|\Omega^* - \Omega|_F \xrightarrow{p^*} 0$ in $\mathbb{P}$, where $\Omega^* = 2\pi f^*(0)$

and $\Omega = 2\pi f(0)$. Hence,

$$|\widehat{\Omega}^* - \Omega|_F \leq |\widehat{\Omega}^* - \Omega^*|_F + |\Omega^* - \Omega|_F = |\widehat{\Omega}^* - \Omega^*|_F + o_{\mathbb{P}}(1). \tag{1.31}$$

It thus remains to show that $|\widehat{\Omega}^* - \Omega^*|_F \xrightarrow{p^*} 0$ in $\mathbb{P}$, but this follows under standard kernel and bandwidth assumptions similarly as in Jansson (2002). Consistency of $\widehat{\Omega}^*$ implies $\widehat{\Omega}^*_{u\cdot v} = \widehat{\Omega}^*_{uu} - \widehat{\Omega}^*_{uv}\left(\widehat{\Omega}^*_{vv}\right)^{-1}\widehat{\Omega}^*_{vu} \xrightarrow{p^*} \Omega_{u\cdot v}$ in $\mathbb{P}$. Together with the invariance principle for $\{w^*_t\}_{t\in\mathbb{Z}}$ in Theorem 1.1 it thus follows from similar arguments as used in Vogelsang and Wagner (2014, Proof of Theorem 3) that $\tau^*_{\mathrm{IM}}(\widehat{\Omega}^*_{u\cdot v}) \xrightarrow{w^*} \chi^2_s$ in $\mathbb{P}$. Now, let $z^*_t := \sum_{s=1}^t w^*_s$. Analogously to the proof of Lemma 3.4 in Chang *et al.* (2006), we obtain

$$T^{-2}\sum_{t=1}^T z^*_t z^{*\prime}_t \xrightarrow{w^*} \int_0^1 B(r)B(r)'dr \text{ in } \mathbb{P},$$
$$T^{-1}\sum_{t=1}^T z^*_{t-1} w^{*\prime}_t \xrightarrow{w^*} \int_0^1 B(r)dB(r)' + \Delta \text{ in } \mathbb{P}.$$

It is thus straightforward to show that $\tau^*_{\mathrm{FM}}(\widehat{\Omega}^*_{u\cdot v}) \xrightarrow{w^*} \chi^2_s$ in $\mathbb{P}$. Finally, to prove the result for the traditional Wald-type test statistic based on the D-OLS estimator, note that Chang *et al.* (2006) verify that a representation as in (1.8) also holds for the bootstrap errors, i. e., we have

$$u^*_t = \sum_{j=-\infty}^{\infty} \pi^{*\prime}_j v^*_{t-j} + e^*_t \text{ in } \mathbb{P},$$

where $\mathbb{E}^*\left(e^*_t v^{*\prime}_{t-j}\right) = 0$ for all $j \in \mathbb{Z}$ and $\sum_{j=-\infty}^{\infty} |\pi^*_j|_F < \infty$ in $\mathbb{P}$. It thus follows from Chang *et al.* (2006, Theorem 3.7) that $\tau^*_{\mathrm{D}}(\widehat{\Omega}^*_{u\cdot v}) \xrightarrow{w^*} \chi^2_s$ in $\mathbb{P}$. $\square$

1.8.5 Proofs of Auxiliary Results

Proof of Lemma 1.1. The solution of the sample Yule-Walker equations in the regression of w_t on $w_{t-1}, \ldots, w_{t-q}$, $t = q+1, \ldots, T$, can be written in compact form as a $((m+1) \times q(m+1))$-dimensional matrix

$$\widetilde{\mathbf{\Phi}}(q) := [\widetilde{\Phi}_1(q), \ldots, \widetilde{\Phi}_q(q)] = \widetilde{\Gamma}\widetilde{G}^{-1},$$

with the $((m+1) \times q(m+1))$-dimensional matrix $\widetilde{\Gamma} := \left[\widetilde{\Gamma}(1), \ldots, \widetilde{\Gamma}(q)\right]$, the $(q(m+1) \times q(m+1))$-dimensional matrix $\widetilde{G} := \left(\widetilde{\Gamma}(s-r)\right)_{r,s=1,\ldots,q}$ and the $((m+1) \times (m+1))$-dimensional empirical autocovariance matrix of $w_1, \ldots, w_T$ at lag $-q+1 \leq h \leq q$, given by

$$\widetilde{\Gamma}(h) := T^{-1}\sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T)(w_t - \overline{w}_T)',$$

where $\overline{w}_T := T^{-1}\sum_{t=1}^T w_t$. Analogously, the solution of the sample Yule-Walker equations in the regression of $\widehat{w}_t$ on $\widehat{w}_{t-1}, \dots, \widehat{w}_{t-q}$, $t = q+1, \dots, T$, can be written in compact form as

$$\widehat{\mathbf{\Phi}}(q) \coloneqq [\widehat{\Phi}_1(q), \dots, \widehat{\Phi}_q(q)] = \widehat{\Gamma}\widehat{G}^{-1},$$

with $\widehat{\Gamma} \coloneqq \left[\widehat{\Gamma}(1), \dots, \widehat{\Gamma}(q)\right]$, $\widehat{G} \coloneqq \left(\widehat{\Gamma}(s-r)\right)_{r,s=1,\dots,q}$ and $\widehat{\Gamma}(h)$ the empirical autocovariance matrix of $\widehat{w}_1, \dots, \widehat{w}_T$ at lag $-q+1 \leq h \leq q$. Taking the difference of $\widetilde{\mathbf{\Phi}}(q)$ and $\widehat{\mathbf{\Phi}}(q)$, adding and subtracting $\widehat{\Gamma}\widetilde{G}^{-1}$ and using

$$\widetilde{G}^{-1} - \widehat{G}^{-1} = \widetilde{G}^{-1}(\widehat{G} - \widetilde{G})\widehat{G}^{-1}$$

leads to

$$\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q) = \widehat{\Gamma}\widetilde{G}^{-1}(\widetilde{G} - \widehat{G})\widehat{G}^{-1} - (\widetilde{\Gamma} - \widehat{\Gamma})\widetilde{G}^{-1}$$

Hence, we have to consider $\widetilde{G} - \widehat{G}$ in more detail ($\widetilde{\Gamma} - \widehat{\Gamma}$ works similarly). A typical block element of $\widetilde{G} - \widehat{G}$ is

$$\begin{aligned}
\widetilde{\Gamma}(h) - \widehat{\Gamma}(h) = T^{-1} & \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T)(w_t - \overline{w}_T)' \\
& - T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (\widehat{w}_{t+h} - \overline{\widehat{w}}_T)(\widehat{w}_t - \overline{\widehat{w}}_T)' \\
& = T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T)(w_t - \widehat{w}_t - (\overline{w}_T - \overline{\widehat{w}}_T))' \\
& - T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (\widehat{w}_{t+h} - w_{t+h} - (\overline{\widehat{w}}_T - \overline{w}_T))(\widehat{w}_t - \overline{\widehat{w}}_T)' \\
& = A_1(h) - A_2(h),
\end{aligned}$$

with an obvious definition for $A_1(h)$ and $A_2(h)$. Let us consider $A_1(h)$ in more detail ($A_2(h)$ works similarly). Using $\widehat{w}_t = [\widehat{u}_t, v_t']'$ and $\widehat{u}_t = y_t - x_t'\widehat{\beta}$ together with the model equations (1.1) and (1.2), we get

$$w_t - \widehat{w}_t = \begin{bmatrix} u_t - \widehat{u}_t \\ 0_{m\times 1} \end{bmatrix} = \begin{bmatrix} y_t - x_t'\beta - (y_t - x_t'\widehat{\beta}) \\ 0_{m\times 1} \end{bmatrix} = \begin{bmatrix} x_t'(\widehat{\beta} - \beta) \\ 0_{m\times 1} \end{bmatrix}. \tag{1.32}$$

Before we continue, note that the last equality implies

$$\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \leq T^{-1/2} \max_{1\leq t\leq T} |T^{-1/2}x_t|_F |T\left(\widehat{\beta} - \beta\right)|_F = O_{\mathbb{P}}(T^{-1/2}),$$

since $T\left(\widehat{\beta} - \beta\right) = O_{\mathbb{P}}(1)$ and $\max_{1\leq t\leq T} |T^{-1/2}x_t|_F = \sup_{0\leq r\leq 1} |T^{-1/2}x_{\lfloor rT\rfloor}|_F$ converges by Assumption 1.3 and the continuous mapping theorem to $\sup_{0\leq r\leq 1} |B_v(r)|_F = O_{\mathbb{P}}(1)$. This proves (1.23).

We now proceed with the proof of (1.24). From (1.32) we obtain

$$
\begin{aligned}
A_1(h) &= T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T) \begin{bmatrix} (x_t - \overline{x}_T)'(\widehat{\beta} - \beta) \\ 0_{m\times 1} \end{bmatrix}' \\
&= T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T)[(x_t - \overline{x}_T)'(\widehat{\beta} - \beta), 0_{1\times m}] \\
&= \widetilde{\Gamma}_{w,x}(h)(\widehat{\beta} - \beta)e_1',
\end{aligned}
$$

where $\widetilde{\Gamma}_{w,x}(h) := T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (w_{t+h} - \overline{w}_T)(x_t - \overline{x}_T)'$ is $((m+1) \times m)$-dimensional and $e_1 := [1, 0_{1\times m}]'$ is the first $(m+1)$-dimensional unit vector. Denoting the part of $\widetilde{G} - \widehat{G}$ that consists of block-entries $A_1(h)$ by $\widetilde{G}_1 - \widehat{G}_1$, we get

$$
\widetilde{G}_1 - \widehat{G}_1 = \left[\widetilde{\Gamma}_{w,x}(s-r)(\widehat{\beta} - \beta)e_1'\right]_{r,s=1,\ldots,q} = \widetilde{\Gamma}_{w,x}(I_q \otimes ((\widehat{\beta} - \beta)e_1')),
$$

where $\widetilde{\Gamma}_{w,x} := (\widetilde{\Gamma}_{w,x}(s-r))_{r,s=1,\ldots,q}$ is $(q(m+1) \times qm)$-dimensional. For the second factor we have

$$
\begin{aligned}
|I_q \otimes ((\widehat{\beta} - \beta)e_1')|_F &= \sqrt{\operatorname{tr}\left((I_q \otimes ((\widehat{\beta} - \beta)e_1'))' I_q \otimes ((\widehat{\beta} - \beta)e_1')\right)} \\
&= \sqrt{\operatorname{tr}\left(I_q \otimes (e_1(\widehat{\beta} - \beta)'(\widehat{\beta} - \beta)e_1')\right)} \\
&= \sqrt{\operatorname{tr}\left(I_q \otimes \operatorname{diag}\left(\sum_{i=1}^{m}(\widehat{\beta}_i - \beta_i)^2, 0, \ldots, 0\right)\right)} \\
&= \sqrt{q\sum_{i=1}^{m}(\widehat{\beta}_i - \beta_i)^2} \\
&= q^{1/2}T^{-1}|T(\widehat{\beta} - \beta)|_F \\
&= O_P(q^{1/2}T^{-1}).
\end{aligned}
$$

Next, let us consider $\widetilde{\Gamma}_{w,x}$ in more detail. Recall that $w_t = [u_t, v_t']'$ and $x_t = \sum_{k=1}^{t} v_k$. To avoid lengthy index notation, w.l.o.g. we can assume that $m = 1$ and consider the second element of w_t only (the first element works similarly). We thus consider the scalar quantity

$$
\widetilde{\Gamma}_{w,x}(h) = T^{-1} \sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}} (v_{t+h} - \overline{v}_T) \left(\sum_{k=1}^{t} v_k - T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{i} v_j\right).
$$

Taking the expectation of the squared Frobenius norm of the corresponding $(q \times q)$-dimensional matrix $\widetilde{\Gamma}_{w,x}$ and combining the sums over r and s, leads to

$$\begin{aligned}
&\mathbb{E}(|\widetilde{\Gamma}_{w,x}|_F^2) \\
&= \sum_{r,s=1}^{q} T^{-2} \sum_{t_1,t_2=\max\{1,1-(s-r)\}}^{\min\{T,T-(s-r)\}} \mathbb{E}\Bigg[(v_{t_1+s-r} - \overline{v}_T)\left(\sum_{k_1=1}^{t_1} v_{k_1} - T^{-1}\sum_{i_1=1}^{T}\sum_{j_1=1}^{i_1} v_{j_1}\right) \\
&\qquad\qquad \times (v_{t_2+s-r} - \overline{v}_T)\left(\sum_{k_2=1}^{t_2} v_{k_2} - T^{-1}\sum_{i_2=1}^{T}\sum_{j_2=1}^{i_2} v_{j_2}\right)\Bigg] \\
&= T^{-2}\sum_{h=-q+1}^{q-1}(q-|h|)\sum_{t_1,t_2=\max\{1,1-h\}}^{\min\{T,T-h\}} \mathbb{E}\Bigg[(v_{t_1+h} - \overline{v}_T)\left(\sum_{k_1=1}^{t_1} v_{k_1} - T^{-1}\sum_{i_1=1}^{T}\sum_{j_1=1}^{i_1} v_{j_1}\right) \\
&\qquad\qquad \times (v_{t_2+h} - \overline{v}_T)\left(\sum_{k_2=1}^{t_2} v_{k_2} - T^{-1}\sum_{i_2=1}^{T}\sum_{j_2=1}^{i_2} v_{j_2}\right)\Bigg].
\end{aligned}$$

Note that the last expectation is of the form $\mathbb{E}(ABCD)$ with $\mathbb{E}(A) = \mathbb{E}(B) = \mathbb{E}(C) = \mathbb{E}(D) = 0$. Hence, by using common rules for joint cumulants of centered random variables (see, e. g., Brillinger, 1981), we get

$$\begin{aligned}
&\mathbb{E}(ABCD) \\
&= \mathrm{cum}(A,B,C,D) + \mathbb{E}(AB)\mathbb{E}(CD) + \mathbb{E}(AC)\mathbb{E}(BD) + \mathbb{E}(AD)\mathbb{E}(BC) \\
&= \mathrm{cum}(A,B,C,D) + \mathrm{Cov}(A,B)\mathrm{Cov}(C,D) + \mathrm{Cov}(A,C)\mathrm{Cov}(B,D) \\
&\quad + \mathrm{Cov}(A,D)\mathrm{Cov}(B,C),
\end{aligned}$$

where $\mathrm{Cov}(\cdot,\cdot)$ denotes the covariance of two random variables. Hence, the first term corresponding to the fourth-order cumulant becomes

$$\begin{aligned}
T^{-2}\sum_{h=-q+1}^{q-1}(q-|h|)\sum_{t_1,t_2=\max\{1,1-h\}}^{\min\{T,T-h\}} \mathrm{cum}\Bigg(& v_{t_1+h} - \overline{v}_T, \sum_{k_1=1}^{t_1} v_{k_1} - T^{-1}\sum_{i_1=1}^{T}\sum_{j_1=1}^{i_1} v_{j_1}, \\
& v_{t_2+h} - \overline{v}_T, \sum_{k_2=1}^{t_2} v_{k_2} - T^{-1}\sum_{i_2=1}^{T}\sum_{j_2=1}^{i_2} v_{j_2}\Bigg),
\end{aligned}$$

leading to $2^4 = 16$ terms when exapnding the cumulant. Exemplarily, for the absolute value of the first one (the others work similarly), we get from the common calculation rules for cumulants

$$\begin{aligned}
&|T^{-2}\sum_{h=-q+1}^{q-1}(q-|h|)\sum_{t_1,t_2=\max\{1,1-h\}}^{\min\{T,T-h\}} \mathrm{cum}\left(v_{t_1+h}, \sum_{k_1=1}^{t_1} v_{k_1}, v_{t_2+h}, \sum_{k_2=1}^{t_2} v_{k_2}\right)| \\
&\leq T^{-2}\sum_{h=-q+1}^{q-1}|q-|h||\sum_{t_1,t_2=\max\{1,1-h\}}^{\min\{T,T-h\}}\sum_{k_1=1}^{t_1}\sum_{k_2=1}^{t_2}|\mathrm{cum}\left(v_{t_1+h}, v_{k_1}, v_{t_2+h}, v_{k_2}\right)| \\
&\leq \frac{q}{T^2}\sum_{h=-q+1}^{q-1}\sum_{t_1,t_2=1}^{T}\sum_{k_1,k_2=1}^{T}|\mathrm{cum}\left(v_{t_1+h}, v_{k_1}, v_{t_2+h}, v_{k_2}\right)|.
\end{aligned}$$

By combining the sums over t_1 and t_2 and those over k_1 and k_2, respectively, the above term becomes

$$
\begin{aligned}
&\frac{q}{T^2}\sum_{h=-q+1}^{q-1}\sum_{l=-(T-1)}^{T-1}\sum_{s=\max\{1,1-l\}}^{\min\{T,T-l\}}\sum_{i=-(T-1)}^{T-1}\sum_{j=\max\{1,1-i\}}^{\min\{T,T-i\}}\left|\operatorname{cum}\left(v_{s+l+h},v_{j+i},v_{s+h},v_j\right)\right|\\
&\leq \frac{q}{T^2}\sum_{h=-q+1}^{q-1}\sum_{i,l=-(T-1)}^{T-1}\sum_{s,j=1}^{T}\left|\operatorname{cum}\left(v_{s+l+h},v_{j+i},v_{s+h},v_j\right)\right|\\
&\leq \frac{q}{T^2}\sum_{h=-q+1}^{q-1}\sum_{i,l=-(T-1)}^{T-1}\sum_{k=-(T-1)}^{T-1}\sum_{r=\max\{1,1-k\}}^{\min\{T,T-k\}}\left|\operatorname{cum}\left(v_{r+k+l+h},v_{r+i},v_{r+k+h},v_r\right)\right|\\
&\leq \frac{q}{T}\sum_{h=-q+1}^{q-1}\sum_{i,l=-(T-1)}^{T-1}\sum_{k=-(T-1)}^{T-1}\left|\operatorname{cum}\left(v_{k+l+h},v_i,v_{k+h},v_0\right)\right|,
\end{aligned}
$$

where we also combined the sums over s and j and made use of the (strict) stationarity of $\{v_t\}_{t\in\mathbb{Z}}$. Finally, combining the sums over h and k, we obtain the bound

$$
\begin{aligned}
&\frac{q}{T}\sum_{r=-(T+q-2)}^{T+q-2}(2q-1)\sum_{i,l=-(T-1)}^{T-1}\left|\operatorname{cum}\left(v_{r+l},v_i,v_r,v_0\right)\right|\\
&\leq 2\frac{q^2}{T}\sum_{j=-(2T+q-3)}^{2T+q-3}\sum_{r=-(T+q-2)}^{T+q-2}\sum_{i=-(T-1)}^{T-1}\left|\operatorname{cum}\left(v_j,v_i,v_r,v_0\right)\right|\\
&\leq 2\frac{q^2}{T}\sum_{j,i,r=-\infty}^{\infty}\left|\operatorname{cum}\left(v_j,v_i,v_r,v_0\right)\right| = O(q^2T^{-1}),
\end{aligned}
$$

due to the summability condition imposed on the fourth order cumulants in Assumption 1.2. Hence, this term vanishes asymptotically. However, the leading term is the term corresponding to $\operatorname{Cov}(A,B)\operatorname{Cov}(C,D)$. That is, we have to consider

$$
\begin{aligned}
&T^{-2}\sum_{h=-q+1}^{q-1}(q-|h|)\sum_{t_1,t_2=\max\{1,1-h\}}^{\min\{T,T-h\}}\operatorname{Cov}(v_{t_1+h}-\overline{v}_T,\sum_{k_1=1}^{t_1}v_{k_1}-T^{-1}\sum_{i_1=1}^{T}\sum_{j_1=1}^{i_1}v_{j_1})\\
&\qquad\qquad\times\operatorname{Cov}(v_{t_2+h}-\overline{v}_T,\sum_{k_2=1}^{t_2}v_{k_2}-T^{-1}\sum_{i_2=1}^{T}\sum_{j_2=1}^{i_2}v_{j_2})\\
&=\sum_{h=-q+1}^{q-1}(q-|h|)\left(T^{-1}\sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}}\operatorname{Cov}(v_{t+h}-\overline{v}_T,\sum_{k=1}^{t}v_k-T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{i}v_j)\right)^2.
\end{aligned}
$$

Hence, we have to compute

$$
T^{-1}\sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}}\operatorname{Cov}(v_{t+h}-\overline{v}_T,\sum_{k=1}^{t}v_k-T^{-1}\sum_{i=1}^{T}\sum_{j=1}^{i}v_j).
$$

This leads to four terms to consider, which can be treated with similar arguments. For the first term we get

$$T^{-1}\sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}}\mathrm{Cov}(v_{t+h},\sum_{k=1}^{t}v_k)=T^{-1}\sum_{t=\max\{1,1-h\}}^{\min\{T,T-h\}}\sum_{k=1}^{t}\gamma_v(t+h-k),$$

where $\gamma_v(h)$ is the covariance of the one-dimensional process (still assumed for notational brevity) $\{v_t\}_{t\in\mathbb{Z}}$ at lag h. E. g., for $h\geq 0$ it holds that

$$T^{-1}\sum_{t=1}^{T-h}\sum_{k=1}^{t}\gamma_v(t+h-k)=T^{-1}\sum_{j=h}^{T-1}(T-j)\gamma_v(j)\leq\sum_{j=h}^{T-1}\gamma_v(j)$$

and its absolute value can be bounded by $\sum_{j=-\infty}^{\infty}|\gamma_v(j)|<\infty$ due to the second-order cumulant condition imposed in Assumption 1.2. Similar arguments yield the same bound for $h<0$ and for the other three terms. Hence, the term of $\mathbb{E}(|\widetilde{\Gamma}_{w,x}|_F^2)$ that corresponds to $\mathrm{Cov}(A,B)\mathrm{Cov}(C,D)$ is of order $O(q^2)$.

In total we thus have $\mathbb{E}(|\widetilde{\Gamma}_{w,x}|_F^2)=O(q^2)$. Note that we have proven this result for $m=1$. However, as m is fixed, the result also holds for $m>1$. Therefore, we obtain for the $(q(m+1)\times qm)$-dimensional matrix $\widetilde{\Gamma}_{w,x}$ that $|\widetilde{\Gamma}_{w,x}|_F=O_{\mathbb{P}}(q)$. It follows that $|\widetilde{G}_1-\widehat{G}_1|_F=O_{\mathbb{P}}(q)O_P(q^{1/2}T^{-1})=O_{\mathbb{P}}(q^{3/2}T^{-1})$ and similarly also $|\widetilde{G}-\widehat{G}|_F=O_{\mathbb{P}}(q)O_{\mathbb{P}}(q^{1/2}T^{-1})=O_{\mathbb{P}}(q^{3/2}T^{-1})$.

Further, we have to consider $\widetilde{G}^{-1}$. In the following, let $\mu_{\min}(A)$ and $\mu_{\max}(A)$ denote the smallest and largest eigenvalue of a matrix A, respectively, and define the $(q(m+1)\times q(m+1))$-dimensional matrix $G:=(\Gamma(s-r))_{r,s=1,\dots,q}$. Similar to the above, using the fourth-order cumulant condition from Assumption 1.2, we can show that $|\widetilde{G}-G|_F=O_{\mathbb{P}}(qT^{-1/2})=o_{\mathbb{P}}(1)$. Then, to show boundedness in probability of $\widetilde{G}^{-1}$ (similar for $\widehat{G}^{-1}$), for all $\epsilon>0$, we have to find a $K<\infty$ and a $T_0<\infty$ both large enough such that it holds for all $T>T_0$ that

$$\mathbb{P}(|\widetilde{G}^{-1}|_2>K)<\epsilon,$$

where $|A|_2$ denotes the spectral norm of a matrix A. Let $\epsilon>0$. Then, due to positive semi-definiteness of $\widetilde{G}$ by construction and invertibility, see Meyer and Kreiss (2015, Lemma 3.4 and Remark 3.2), we have positive definiteness of $\widetilde{G}$ and, consequently, of $\widetilde{G}^{-1}$. Hence,

$$\begin{aligned}\mathbb{P}(|\widetilde{G}^{-1}|_2>K)=&\mathbb{P}(\mu_{\max}(\widetilde{G}^{-1})>K)=\mathbb{P}(\mu_{\min}^{-1}(\widetilde{G})>K)\\=&\mathbb{P}(\mu_{\min}(\widetilde{G})<\frac{1}{K},|\widetilde{G}-G|_2\geq\delta)\\&+\mathbb{P}(\mu_{\min}(\widetilde{G})<\frac{1}{K},|\widetilde{G}-G|_2<\delta).\end{aligned}$$

Further, as $|\widetilde{G}-G|_2\leq|\widetilde{G}-G|_F=o_{\mathbb{P}}(1)$, for any $\delta>0$, we can choose T_0 large enough to have $\mathbb{P}(|\widetilde{G}-G|_2\geq\delta)\leq\epsilon$. Thus, the first term on the right-hand side can be bounded by $\mathbb{P}(|\widetilde{G}-G|_2\geq\delta)\leq\epsilon$. Now consider the second term. As $\widetilde{G}$, G, and hence also $\widetilde{G}-G$ are symmetric with real-valued entries, these matrices are Hermitian such that Weyl's theorem (see, e. g., Theorem 4.3.1 in Horn and Johnson, 2012) applies, leading to the inequality $\mu_{\min}(G)+\mu_{\min}(\widetilde{G}-G)\leq\mu_{\min}(\widetilde{G})$. It follows that

$$\begin{aligned}
&\mathbb{P}(\mu_{\min}(\widetilde{G}) < \frac{1}{K}, |\widetilde{G} - G|_2 < \delta) \\
&\leq \mathbb{P}(\mu_{\min}(G) + \mu_{\min}(\widetilde{G} - G) < \frac{1}{K}, |\widetilde{G} - G|_2 < \delta) \\
&= \mathbb{P}(\mu_{\min}(G) < \frac{1}{K} - \mu_{\min}(\widetilde{G} - G), |\widetilde{G} - G|_2 < \delta).
\end{aligned}$$

From symmetry of $\widetilde{G} - G$ we get that the eigenvalues of $(\widetilde{G} - G)'(\widetilde{G} - G)$ are exactly the squared eigenvalues of $\widetilde{G} - G$. Hence, the bound

$$|\widetilde{G} - G|_2 = \sqrt{\mu_{\max}((\widetilde{G} - G)'(\widetilde{G} - G))} < \delta$$

implies also $\mu_{\min}(\widetilde{G} - G) \geq -\delta$ such that the last right-hand side can be bounded by

$$\mathbb{P}\left(\mu_{\min}(G) < \frac{1}{K} + \delta, |\widetilde{G} - G|_2 < \delta\right) \leq \mathbb{P}\left(\mu_{\min}(G) < \frac{1}{K} + \delta\right). \tag{1.33}$$

Next, note that $\mu_{\min}(G) \geq \widetilde{c}$ for some constant $\widetilde{c} > 0$ by Assumption 1.1. Therefore, the right-hand side in (1.33) becomes zero if we choose $\delta < \widetilde{c}/2$ small enough and $K > 2/\widetilde{c}$ large enough, such that $\frac{1}{K} + \delta < \widetilde{c}$. This completes the proof of $|\widetilde{G}^{-1}|_2 = O_{\mathbb{P}}(1)$. Furthermore, from Assumption 1.1 we also get $|\widetilde{G}|_2 = O_{\mathbb{P}}(1)$ and similarly $|\widehat{G}|_2 = O_{\mathbb{P}}(1)$ and $|\widehat{G}|_2 = O_{\mathbb{P}}(1)$. Altogether, it holds that

$$\begin{aligned}
|\widetilde{\boldsymbol{\Phi}}(q) - \widehat{\boldsymbol{\Phi}}(q)|_2 &\leq |\widetilde{\Gamma} - \widehat{\Gamma}|_2 |\widetilde{G}^{-1}|_2 + |\widehat{\Gamma}|_2 |\widetilde{G}^{-1}|_2 |\widetilde{G} - \widehat{G}|_2 |\widehat{G}^{-1}|_2 \\
&\leq |\widetilde{G} - \widehat{G}|_2 \left(|\widetilde{G}^{-1}|_2 + |\widehat{G}|_2 |\widetilde{G}^{-1}|_2 |\widehat{G}^{-1}|_2\right) \\
&\leq |\widetilde{G} - \widehat{G}|_F \left(|\widetilde{G}^{-1}|_2 + |\widehat{G}|_2 |\widetilde{G}^{-1}|_2 |\widehat{G}^{-1}|_2\right) \\
&= O_{\mathbb{P}}(q^{3/2} T^{-1}) \left(O_{\mathbb{P}}(1) + O_{\mathbb{P}}(1) O_{\mathbb{P}}(1) O_{\mathbb{P}}(1)\right) \\
&= O_{\mathbb{P}}(q^{3/2} T^{-1}).
\end{aligned}$$

Since $\widetilde{\boldsymbol{\Phi}}(q) - \widehat{\boldsymbol{\Phi}}(q)$ is $((m+1) \times q(m+1))$-dimensional and m is fixed, it follows that (see, e. g. Gentle, 2007)

$$|\widetilde{\boldsymbol{\Phi}}(q) - \widehat{\boldsymbol{\Phi}}(q)|_F \leq \sqrt{m+1}\ |\widetilde{\boldsymbol{\Phi}}(q) - \widehat{\boldsymbol{\Phi}}(q)|_2 = O_{\mathbb{P}}(q^{3/2} T^{-1}).$$

This implies

$$q^{1/2} \sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F \leq q^{3/2} |\widetilde{\boldsymbol{\Phi}}(q) - \widehat{\boldsymbol{\Phi}}(q)|_F = O_{\mathbb{P}}(q^3 T^{-1}),$$

which is $o_{\mathbb{P}}(1)$ since $q^3 T^{-1} = o(1)$ by Assumption 1.4. □

In the proofs of Lemma 1.2 and 1.3 we repeatedly use the fact that by convexity it holds that $|\sum_{i=1}^k z_i|^a \leq k^{a-1} \sum_{i=1}^k |z_i|^a$, for all $a, k \geq 1$.

Proof of Lemma 1.2. Let $\widetilde{\varepsilon}_t(q) := w_t - \sum_{j=1}^q \widetilde{\Phi}_j(q) w_{t-j}$, $t = q+1, \ldots, T$, denote the Yule-Walker residuals in the regression of w_t on $w_{t-1}, \ldots, w_{t-q}$, $t = q+1, \ldots, T$ and define $\overline{\widetilde{\varepsilon}}_T(q) :=$

$(T-q)^{-1}\sum_{t=q+1}^{T}\widetilde{\varepsilon}_t(q)$.[2] For $q+1 \le t \le T$ we have

$$\begin{aligned}
&|\widehat{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a \\
&\le \left(|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F + |\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F + |\overline{\widehat{\varepsilon}}_T(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F\right)^a \\
&\le \left(|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F + |\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F + (T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F\right)^a \\
&\le 3^{a-1}\left(|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F^a + |\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a + \left((T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F\right)^a\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&(T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a \\
&\le 3^{a-1}\Bigg((T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F^a + (T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a \\
&\qquad + \left((T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F\right)^a\Bigg) \\
&= 3^{a-1}\left(F_{T,a} + (T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a + (F_{T,1})^a\right),
\end{aligned}$$

where $F_{T,a} := (T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F^a$. Let us consider $F_{T,a}$ in more detail. For $q+1 \le$

[2] In the following, a denotes the fixed $a > 2$ from Assumption 1.1. However, the results also hold for $\widetilde{a}$, with $1 \le \widetilde{a} < a$.

$t \leq T$ we have

$$
\begin{aligned}
&|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F \\
&= |\widehat{w}_t - \sum_{j=1}^{q} \widehat{\Phi}_j(q)\widehat{w}_{t-j} - (w_t - \sum_{j=1}^{q} \widetilde{\Phi}_j(q) w_{t-j})|_F \\
&= |\widehat{w}_t - w_t - \sum_{j=1}^{q} \left(\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q) + \widetilde{\Phi}_j(q)\right)(\widehat{w}_{t-j} - w_{t-j} + w_{t-j}) + \sum_{j=1}^{q} \widetilde{\Phi}_j(q) w_{t-j}|_F \\
&= |\widehat{w}_t - w_t - \sum_{j=1}^{q} \left(\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)\right)(\widehat{w}_{t-j} - w_{t-j} + w_{t-j}) - \sum_{j=1}^{q} \widetilde{\Phi}_j(q)\left(\widehat{w}_{t-j} - w_{t-j}\right)|_F \\
&\leq \max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F + \sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F |\widehat{w}_{t-j} - w_{t-j}|_F \\
&\quad + \sum_{j=1}^{q} |\widetilde{\Phi}_j(q)|_F |\widehat{w}_{t-j} - w_{t-j}|_F + \sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F |w_{t-j}|_F \\
&\leq \max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F + \max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F \\
&\quad + \max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \sum_{j=1}^{q} |\widetilde{\Phi}_j(q)|_F + \sqrt{m+1}\, q |\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F \; q^{-1} \sum_{j=1}^{q} |w_{t-j}|_F,
\end{aligned}
$$

where we have used that

$$
|\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F \leq |\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F |[0, I_{m+1}, 0]'|_F = \sqrt{m+1} |\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F .
$$

Therefore,

$$
\begin{aligned}
&|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F^a \\
&\leq 4^{a-1} \left(\left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a + \left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a \left(\sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F \right)^a \right. \\
&\quad + \left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a \left(\sum_{j=1}^{q} |\widetilde{\Phi}_j(q)|_F \right)^a \\
&\quad \left. + \left(\sqrt{m+1}\, q |\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F \right)^a \left(q^{-1} \sum_{j=1}^{q} |w_{t-j}|_F \right)^a \right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
F_{T,a} \leq 4^{a-1} &\left(\left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a + \left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a \left(\sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F \right)^a \right. \\
&\quad + \left(\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F \right)^a \left(\sum_{j=1}^{q} |\widetilde{\Phi}_j(q)|_F \right)^a \\
&\quad \left. + \left(\sqrt{m+1}\, q |\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F \right)^a (T-q)^{-1} \sum_{t=q+1}^{T} \left(q^{-1} \sum_{j=1}^{q} |w_{t-j}|_F \right)^a \right).
\end{aligned}
$$

From Lemma 1.1 we have $\max_{1\leq t\leq T} |\widehat{w}_t - w_t|_F = O_{\mathbb{P}}(T^{-1/2})$ and $\sum_{j=1}^{q} |\widehat{\Phi}_j(q) - \widetilde{\Phi}_j(q)|_F = O_{\mathbb{P}}(q^{5/2}T^{-1})$. Moreover, the proof of Lemma 1.1 shows that $q|\widehat{\mathbf{\Phi}}(q) - \widetilde{\mathbf{\Phi}}(q)|_F = O_{\mathbb{P}}(q^{5/2}T^{-1})$. Further, by (1.25), (1.26) and Assumption 1.1 we have

$$\begin{aligned}\sum_{j=1}^{q} |\widetilde{\Phi}_j(q)|_F &\leq \sum_{j=1}^{q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F + \sum_{j=1}^{q} |\Phi_j(q) - \Phi_j|_F + \sum_{j=1}^{q} |\Phi_j|_F \\ &\leq q \sup_{1\leq j\leq q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F + c\sum_{j=q+1}^{\infty} |\Phi_j|_F + \sum_{j=1}^{\infty} |\Phi_j|_F \\ &= O_{\mathbb{P}}(1).\end{aligned}$$

Finally, note that

$$\begin{aligned}(T-q)^{-1}\sum_{t=q+1}^{T}\left(q^{-1}\sum_{j=1}^{q}|w_{t-j}|_F\right)^a &\leq (T-q)^{-1}\sum_{t=q+1}^{T} q^{-a}q^{a-1}\sum_{j=1}^{q}|w_{t-j}|_F^a \\ &= (T-q)^{-1}q^{-1}\sum_{t=q+1}^{T}\sum_{j=1}^{q}|w_{t-j}|_F^a \\ &\leq (T-q)^{-1}\sum_{t=1}^{T-1}|w_t|_F^a,\end{aligned}$$

where the last inequality follows from the fact that each element in the double sum occurs at most q times, i. e., $\sum_{t=q+1}^{T}\sum_{j=1}^{q}|w_{t-j}|_F^a \leq q\sum_{t=1}^{T-1}|w_t|_F^a$. From $\sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) < \infty$ (by stationarity of $\{w_t\}_{t\in\mathbb{Z}}$) and Markov's inequality, it follows that $(T-q)^{-1}\sum_{t=1}^{T-1}|w_t|_F^a = O_{\mathbb{P}}(1)$. In total, we thus have $F_{T,a} = O_{\mathbb{P}}((q^{5/2}T^{-1})^a) = o_{\mathbb{P}}(1)$. Therefore,

$$(T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)|_F^a \leq 3^{a-1}\left((T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a + o_{\mathbb{P}}(1)\right).$$

It thus remains to show that $(T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a = O_{\mathbb{P}}(1)$. We now follow Park (2002, Proof of Lemma 3.2) and Palm *et al.* (2010, Proof of Lemma 2). Define $\varepsilon_t(q) := \varepsilon_t + \sum_{j=q+1}^{\infty}\Phi_j w_{t-j}$ and note that

$$\begin{aligned}&(T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \overline{\widetilde{\varepsilon}}_T(q)|_F^a \\ &= (T-q)^{-1}\sum_{t=q+1}^{T}|\widetilde{\varepsilon}_t(q) - \varepsilon_t(q) + \varepsilon_t(q) - \varepsilon_t + \varepsilon_t - \overline{\widetilde{\varepsilon}}_T(q)|_F^a \\ &\leq 4^{a-1}\left(A_{T,a} + B_{T,a} + C_{T,a} + D_{T,a}\right),\end{aligned}$$

where

$$
\begin{aligned}
A_{T,a} &\coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\varepsilon_t|_F^a, \\
B_{T,a} &\coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\varepsilon_t(q) - \varepsilon_t|_F^a = (T-q)^{-1} \sum_{t=q+1}^{T} |\sum_{j=q+1}^{\infty} \Phi_j w_{t-j}|_F^a, \\
C_{T,a} &\coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\widetilde{\varepsilon}_t(q) - \varepsilon_t(q)|_F^a, \\
D_{T,a} &\coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\overline{\widetilde{\varepsilon}}_T(q)|_F^a = |\overline{\widetilde{\varepsilon}}_T(q)|_F^a = |(T-q)^{-1} \sum_{t=q+1}^{T} \widetilde{\varepsilon}_t(q)|_F^a.
\end{aligned}
$$

We first consider $B_{T,a}$. Note that $\mathbb{E}\left(|B_{T,a}|_F\right) \leq \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|\varepsilon_t(q) - \varepsilon_t|_F^a\right)$. It follows from Minkowski's inequality that

$$
\begin{aligned}
\mathbb{E}\left(|\varepsilon_t(q) - \varepsilon_t|_F^a\right) &= \mathbb{E}\left(|\sum_{j=q+1}^{\infty} \Phi_j w_{t-j}|_F^a\right) = \left(\left[\mathbb{E}\left(|\sum_{j=q+1}^{\infty} \Phi_j w_{t-j}|_F^a\right)\right]^{1/a}\right)^a \\
&\leq \left(\sum_{j=q+1}^{\infty} \left[\mathbb{E}\left(|\Phi_j w_{t-j}|_F^a\right)\right]^{1/a}\right)^a \leq \left(\sum_{j=q+1}^{\infty} \left[\mathbb{E}\left(|\Phi_j|_F^a |w_{t-j}|_F^a\right)\right]^{1/a}\right)^a \\
&\leq \left(\sum_{j=q+1}^{\infty} |\Phi_j|_F \left[\mathbb{E}\left(|w_{t-j}|_F^a\right)\right]^{1/a}\right)^a \leq \left(\sum_{j=q+1}^{\infty} |\Phi_j|_F \left[\sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right)\right]^{1/a}\right)^a \\
&\leq \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) \left(\sum_{j=q+1}^{\infty} |\Phi_j|_F\right)^a.
\end{aligned}
$$

From Assumption 1.1 we have $\sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) < \infty$ and $\sum_{j=q+1}^{\infty} |\Phi_j|_F = o(1)$. Markov's inequality thus yields $B_{T,a} = o_{\mathbb{P}}(1)$. Analogously, $\mathbb{E}\left(|A_{T,a}|_F\right) \leq \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|\varepsilon_t|_F^a\right)$. Using Minkowski's inequality, we have as above

$$
\begin{aligned}
\mathbb{E}\left(|\varepsilon_t|_F^a\right) &= \mathbb{E}\left(|w_t - \sum_{j=1}^{\infty} \Phi_j w_{t-j}|_F^a\right) \leq 2^{a-1}\left(\mathbb{E}\left(|w_t|_F^a\right) + \mathbb{E}\left(|\sum_{j=1}^{\infty} \Phi_j w_{t-j}|_F^a\right)\right) \\
&\leq 2^{a-1}\left(\sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) + \mathbb{E}\left(|\sum_{j=1}^{\infty} \Phi_j w_{t-j}|_F^a\right)\right) \\
&\leq 2^{a-1} \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right)\left(1 + \left(\sum_{j=1}^{\infty} |\Phi_j|_F\right)^a\right) < \infty.
\end{aligned}
$$

Markov's inequality finally yields $A_{T,a} = O_{\mathbb{P}}(1)$. We now turn to $C_{T,a}$. By definition,

$$
\begin{aligned}
\widetilde{\varepsilon}_t(q) &= w_t - \sum_{j=1}^{q} \widetilde{\Phi}_j(q) w_{t-j} = \varepsilon_t(q) - \sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j\right) w_{t-j} \\
&= \varepsilon_t(q) - \sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j(q)\right) w_{t-j} - \sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j\right) w_{t-j}.
\end{aligned}
$$

Hence,

$$|\widetilde{\varepsilon}_t(q) - \varepsilon_t(q)|_F^a \leq 2^{a-1} \left(|\sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j(q)\right) w_{t-j}|_F^a + |\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j\right) w_{t-j}|_F^a \right).$$

It follows that $C_{T,a} = 2^{a-1}\left(C_{1T,a} + C_{2T,a}\right)$, where

$$C_{1T,a} \coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j(q)\right) w_{t-j}|_F^a,$$
$$C_{2T,a} \coloneqq (T-q)^{-1} \sum_{t=q+1}^{T} |\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j\right) w_{t-j}|_F^a.$$

We consider both terms separately. First note that

$$\begin{aligned} C_{1T,a} &\leq q^{a-1}(T-q)^{-1} \sum_{t=q+1}^{T} \sum_{j=1}^{q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F^a |w_{t-j}|_F^a \\ &\leq q^{a-1} \left(\sup_{1\leq j\leq q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F \right)^a (T-q)^{-1} \sum_{t=q+1}^{T} \sum_{j=1}^{q} |w_{t-j}|_F^a \\ &\leq \left(q \sup_{1\leq j\leq q} |\widetilde{\Phi}_j(q) - \Phi_j(q)|_F \right)^a (T-q)^{-1} \sum_{t=1}^{T-1} |w_t|_F^a, \end{aligned}$$

where the third inequality follows again from the fact that $\sum_{t=q+1}^{T}\sum_{j=1}^{q} |w_{t-j}|_F^a \leq q \sum_{t=1}^{T-1} |w_t|_F^a$. As $(T-q)^{-1}\sum_{t=1}^{T-1} |w_t|_F^a = O_{\mathbb{P}}(1)$ it follows from (1.25) that $C_{1T,a} = o_{\mathbb{P}}(1)$. Moreover, Minkowski's inequality yields

$$\begin{aligned} \mathbb{E}\left(|C_{2T,a}|_F\right) &= (T-q)^{-1} \sum_{t=q+1}^{T} \mathbb{E}\left(|\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j\right) w_{t-j}|_F^a \right) \\ &= (T-q)^{-1} \sum_{t=q+1}^{T} \left(\left[\mathbb{E}\left(|\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j\right) w_{t-j}|_F^a \right) \right]^{1/a} \right)^a \\ &\leq (T-q)^{-1} \sum_{t=q+1}^{T} \left(\sum_{j=1}^{q} \left[\mathbb{E}\left(| \left(\Phi_j(q) - \Phi_j\right) w_{t-j}|_F^a\right)\right]^{1/a} \right)^a \\ &\leq (T-q)^{-1} \sum_{t=q+1}^{T} \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) \left(\sum_{j=1}^{q} |\Phi_j(q) - \Phi_j|_F \right)^a \\ &= \sup_{t\in\mathbb{Z}} \mathbb{E}\left(|w_t|_F^a\right) \left(\sum_{j=1}^{q} |\Phi_j(q) - \Phi_j|_F \right)^a. \end{aligned}$$

From (1.26) and Markov's inequality it follows that $C_{2T,a} = o_{\mathbb{P}}(1)$. In total we thus have $C_{T,a} =$

$o_{\mathbb{P}}(1)$. Finally, we consider $D_{T,a}$. It holds that $(T-q)^{-1}\sum_{t=q+1}^{T}\widetilde{\varepsilon}_t(q) = D_{1T}+D_{2T}+D_{3T}$, where

$$\begin{aligned}
D_{1T} &\coloneqq (T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t,\\
D_{2T} &\coloneqq (T-q)^{-1}\sum_{t=q+1}^{T}(\varepsilon_t(q)-\varepsilon_t),\\
D_{3T} &\coloneqq (T-q)^{-1}\sum_{t=q+1}^{T}(\widetilde{\varepsilon}_t(q)-\varepsilon_t(q)).
\end{aligned}$$

By Chebyshev's weak law of large numbers (White, 2001, p. 25) it holds that $D_{1T}\xrightarrow{p}\mathbb{E}(\varepsilon_t)=0$, i. e., $D_{1T}=o_{\mathbb{P}}(1)$. Moreover, $|D_{2T}|_F\leq B_{T,1}=o_{\mathbb{P}}(1)$ and $|D_{3T}|_F\leq C_{T,1}=o_{\mathbb{P}}(1)$. By the continuous mapping theorem we thus have $D_T=o_{\mathbb{P}}(1)$. This completes the proof. $\square$

Proof of Lemma 1.3. It follows from Assumption 1.2 that

$$|(T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t\varepsilon_t'-\Sigma|_F=o_{\mathbb{P}}(1).$$

Therefore,

$$|\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right)-\Sigma|_F\leq|\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right)-(T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t\varepsilon_t'|_F+o_{\mathbb{P}}(1).$$

Moreover,

$$\begin{aligned}
&|\mathbb{E}^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right)-(T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t\varepsilon_t'|_F\\
&=|(T-q)^{-1}\sum_{t=q+1}^{T}\left(\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)\right)\left(\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)\right)'-\varepsilon_t\varepsilon_t'|_F\\
&=|(T-q)^{-1}\sum_{t=q+1}^{T}\left(\left[\left(\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)\right)-\varepsilon_t\right]\left(\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)\right)'\right.\\
&\qquad\qquad\left.+\varepsilon_t\left[\left(\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)\right)-\varepsilon_t\right]'\right)|_F\\
&\leq E_{1T}+E_{2T},
\end{aligned}$$

where

$$\begin{aligned}
E_{1T}&=(T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)-\varepsilon_t|_F|\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)|_F,\\
E_{2T}&=(T-q)^{-1}\sum_{t=q+1}^{T}|\widehat{\varepsilon}_t(q)-\overline{\widehat{\varepsilon}}_T(q)-\varepsilon_t|_F|\varepsilon_t|_F.
\end{aligned}$$

The Cauchy-Schwarz inequality yields

$$E_{1T} \leq \left((T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F^2 (T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)|_F^2 \right)^{1/2},$$

$$E_{2T} \leq \left((T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F^2 (T-q)^{-1} \sum_{t=q+1}^{T} |\varepsilon_t|_F^2 \right)^{1/2}.$$

From the proof of Lemma 1.2 we have $(T-q)^{-1} \sum_{t=q+1}^{T} |\varepsilon_t|_F^2 = A_{T,2} = O_{\mathbb{P}}(1)$ and $(T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q)|_F^2 = O_{\mathbb{P}}(1)$. It thus remains to show that $(T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F^2 = o_{\mathbb{P}}(1)$. To this end note that

$$\begin{aligned}
&|\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F \\
&\leq |\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F + |\widetilde{\varepsilon}_t(q) - (w_t - \sum_{j=1}^{q} \Phi_j(q) w_{t-j})|_F \\
&\quad + |w_t - \sum_{j=1}^{q} \Phi_j(q) w_{t-j} - \varepsilon_t|_F + |\overline{\widehat{\varepsilon}}_T(q)|_F \\
&= |\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F + |\sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j(q) \right) w_{t-j}|_F \\
&\quad + |\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j \right) w_{t-j}|_F + |\sum_{j=q+1}^{\infty} \Phi_j w_{t-j}|_F + |\overline{\widehat{\varepsilon}}_T(q)|_F.
\end{aligned}$$

Hence,

$$\begin{aligned}
&|\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F^2 \\
&\leq 5 \Bigg(|\widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q)|_F^2 + |\sum_{j=1}^{q} \left(\widetilde{\Phi}_j(q) - \Phi_j(q) \right) w_{t-j}|_F^2 \\
&\quad + |\sum_{j=1}^{q} \left(\Phi_j(q) - \Phi_j \right) w_{t-j}|_F^2 + |\sum_{j=q+1}^{\infty} \Phi_j w_{t-j}|_F^2 + |\overline{\widehat{\varepsilon}}_T(q)|_F^2 \Bigg).
\end{aligned}$$

In the notation of the proof of Lemma 1.2, we obtain

$$\begin{aligned}
(T-q)^{-1} \sum_{t=q+1}^{T} |\widehat{\varepsilon}_t(q) - \overline{\widehat{\varepsilon}}_T(q) - \varepsilon_t|_F^2 &\leq 5 \left(F_{T,2} + C_{1T,2} + C_{2T,2} + B_{T,2} + |\overline{\widehat{\varepsilon}}_T(q)|_F^2 \right) \\
&= 5 |\overline{\widehat{\varepsilon}}_T(q)|_F^2 + o_{\mathbb{P}}(1).
\end{aligned}$$

From $\widehat{\varepsilon}_t(q) = \widehat{\varepsilon}_t(q) - \widetilde{\varepsilon}_t(q) + \widetilde{\varepsilon}_t(q) - \varepsilon_t(q) + \varepsilon_t(q) - \varepsilon_t + \varepsilon_t$, with $\varepsilon_t(q)$ as defined in the proof of

Lemma 1.2, it follows that

$$
\begin{aligned}
|\bar{\bar{\varepsilon}}_T(q)|_F &\leq |(T-q)^{-1}\sum_{t=q+1}^{T}\widehat{\varepsilon}_t(q)-\widetilde{\varepsilon}_t(q)|_F + |(T-q)^{-1}\sum_{t=q+1}^{T}\widetilde{\varepsilon}_t(q)-\varepsilon_t(q)|_F \\
&\quad + |(T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t(q)-\varepsilon_t|_F + |(T-q)^{-1}\sum_{t=q+1}^{T}\varepsilon_t|_F \\
&\leq F_{T,1}+C_{T,1}+B_{T,1}+|D_{1T}|_F = o_{\mathbb{P}}(1).
\end{aligned}
$$

This completes the proof. □

Proof of Lemma 1.4. Given Lemma 1.2 and Lemma 1.3, the result follows immediately from Einmahl (1987), compare Chang *et al.* (2006, p. 714).[3] □

[3] For more details we refer to the pre-print of Palm *et al.* (2010), which is available on https://www.stephansmeekes.nl/research (Accessed: January 11, 2023).

Chapter 2

Panel Cointegrating Polynomial Regressions: Group-Mean Fully Modified OLS Estimation and Inference

Abstract

We develop group-mean fully modified OLS estimation and inference for panels of cointegrating polynomial regressions, i. e., regressions that include an integrated process and its powers as explanatory variables. The stationary errors are allowed to be serially correlated, the integrated regressors – allowed to contain drifts – to be endogenous and, as usual in the panel literature, we include individual specific fixed effects and also allow for individual specific time trends. We consider a fixed cross-section dimension and asymptotics in the time dimension only. Within this setting we develop cross-section dependence robust inference for the group-mean estimator. In both the simulations as well as an illustrative application estimating environmental Kuznets curves for carbon dioxide emissions we compare our group-mean fully modified OLS approach with a recently proposed pooled FM-OLS approach.

2.1 Introduction

We develop *group-mean* fully modified OLS (FM-OLS) estimation and inference for panels of cointegrating polynomial regressions (CPRs) in a large time ($T \to \infty$) and finite cross-section (N fixed) setting. Cointegrating polynomial regressions, a term coined by Wagner and Hong (2016), include deterministic variables as well as integrated processes, potentially with drifts, and their powers as regressors. The stochastic regressors are allowed to be endogenous and the stationary errors are allowed to be serially correlated. For notational brevity we only discuss a simple specification, the cubic CPR with only one integrated regressor, see (2.1) and (2.2) below. The cubic and – probably even more so – the quadratic single regressor CPR are the most widely-used specifications for the analysis of, e. g., environmental Kuznets curves (EKCs). Thus, considering the cubic case simplifies notation considerably whilst containing all elements required for a typical EKC analysis. All results extend, at the price of increased notational rather than mathematical complexity, straightforwardly to higher order powers and multiple integrated regressors, compare for a pure time series setting in Wagner and Hong (2016). With respect to deterministic regressors we consider individual specific intercepts only or individual specific intercepts and individual specific linear trends; also this can be generalized without additional mathematical complexities to more general deterministic regressors.

The paper is closely related to de Jong and Wagner (2022), who consider *pooled* FM-OLS-type estimators for CPRs in a setting with both a large cross-section and time dimension and with a cross-sectional i.i.d. structure. Considering a finite cross-section dimension and asymptotic analysis only for a large time series dimension renders it, of course, impossible to develop a joint or sequential asymptotic normality result for the group-mean FM-OLS estimator.[1] However, the finite cross-section dimension offers some room to consider a more general setting than de Jong and Wagner (2022) in two important ways: First, we allow for the presence of *drifts*, i. e., linear time trends, in the integrated regressors, which are a prominent feature in many macroeconomic and financial time series. The presence of drifts, as is known from standard unit root and cointegration analysis, see, e. g., West (1988), can lead to asymptotic normality of estimated coefficients even in the time series unit root case. We show that similar results holds also in the CPR case, in which higher order polynomial trends are the dominant features of the powers of the integrated regressors with drifts. It turns out that whether and if so for which slope coefficients asymptotic normality prevails, depends in addition to the presence of drifts also upon the presence or absence of individual specific linear trends in the regression model. In this respect it is important to note that for applying the developed estimators and tests no knowledge concerning the presence or absence of drifts is required. Second, we allow for very general forms of cross-section dependence by providing *robust* test statistics that lead to asymptotically valid inference despite cross-section dependence. As is well-known, for macro-panels, which is an important difference to classical micro-panels, the assumption of cross-sectional independence is very likely unrealistic. Consequently, being able to perform cross-section dependence robust inference in conjunction with our group-mean estimator increases applicability substantially, *nota bene* without the need to posit a specific model for cross-section dependence like, e. g., a factor structure.

In a simulation study we compare the group-mean estimators, both OLS and FM-OLS, with the pooled FM-OLS estimator of de Jong and Wagner (2022). In addition to assessing estimator performance we also compare the performance, i. e., null rejection probabilities and "size-corrected" power, of a variety of tests based upon these estimators. Many of the results are as expected and in line with asymptotic theory, e. g., the strong negative effects of error serial correlation, endogeneity and cross-section dependence on the performance of the estimators, where – as expected

[1] Given that many macro panel data sets have a small cross-section dimension, e. g., also two of the data sets used in our illustration with six and 19 countries, it is not ex ante clear that it is always necessary or beneficial to consider large cross-section dimensions. Of course, in situations with N large compared to T, asymptotics in N in addition to T is important and useful. One main value added that large N asymptotics provides – at the standard $\sqrt{N}$-rate – in addition to large T asymptotics, is unconditional asymptotic normality of estimators (under appropriate assumptions). Of course, in case of large N, especially large with respect to T, asymptotics in N is an important aspect. However, unconditional asymptotic normality is not necessary for asymptotic standard inference, which can be based on a conditional asymptotic normality results when only $T \to \infty$.

– the group-mean OLS estimator is most strongly affected. By construction, the pooled FM-OLS estimator leads in most cases to the smallest bias and RMSE. The overall conclusion for hypothesis testing is to use the cross-section robust version of tests based on the group-mean FM-OLS estimator. These tests are, by construction, least affected by cross-section dependence and are much less affected than, e. g., the test based on the pooled FM-OLS estimator by large values of error serial correlation and regressor endogeneity (and are the only ones asymptotically valid in the presence of cross-section dependence). Furthermore, even in the absence of cross-section dependence, the cross-section robust test statistics perform at least at par with the non-robust counterparts. Altogether, this makes the cross-section robust tests based on the group-mean FM-OLS estimator the preferred choice.

We briefly illustrate the developed methodology by estimating environmental Kuznets curves for carbon dioxide emissions using the same data sets as de Jong and Wagner (2022), i. e., two long data sets with $N = 6$ and $N = 19$ countries and about $T = 130$ observations over time and one wide data set with $N = 89$ countries and $T = 54$ observations over time. The EKC hypothesis postulates an inverse U-shaped relationship between measures of economic development, typically GDP per capita, and measures of pollution or emissions. The term EKC refers by analogy to the inverted U-shaped relationship between the level of economic development and income inequality postulated by Simon Kuznets (1955) in his 1954 presidential address to the American Economic Association.[2] A key finding in our illustrative application is that cross-section robust inference makes a difference. The coefficient to the third order power of the logarithm of GDP per capita is significantly different from zero only for the wide data set, both with or without individual specific linear trends included. Relying only upon standard inference indicates the necessity for a cubic specification also for the two long data sets for one of the specifications. The group-mean FM-OLS-based turning points for the long data sets are larger than those found in de Jong and Wagner (2022) for the $N = 6$ data set and very similar for the $N = 19$ data set. For the wide data set group-mean FM-OLS estimation leads to very small or no turning points. For this data set pooled estimation leads to more *plausible* results.

The paper is organized as follows: Section 2.2 presents the setting, the assumptions and the theoretical results, separated – for didactic reasons – in three subsections according to different settings concerning the absence or presence of drifts. Section 2.3 contains some illustrative simulation results. Section 2.4 briefly illustrates the method by estimating EKCs for carbon dioxide emissions using, as mentioned above, the same data sets as de Jong and Wagner (2022) and Section 2.5 briefly summarizes and concludes. The proofs are relegated to Appendix A and Appendix B provides the country list for the wide data set. Supplementary Material available upon request contains additional simulation results.

[2]The empirical EKC literature started about 30 years ago, with early important contributions including Grossman and Krueger (1993) or Holtz-Eakin and Selden (1995). Early survey papers like Stern (2004) or Yandle *et al.* (2004) already count more than 100 refereed publications, with the number growing steadily since then. For more discussion on the empirical literature and theoretical underpinnings of the EKC see, e. g., Wagner (2015). Inverted U-shaped relationships also feature prominently in modelling the relationship between energy or material intensity and GDP per capita (see, e. g., Labson and Crompton, 1993; Malenbaum, 1978). In the exchange rate target zone literature predictive regressions involving an exchange rate and its powers as explanatory variables are widely used (see, e. g., Darvas, 2008; Svensson, 1992). In either of these literatures typically only quadratic or cubic polynomials are considered. Thus, also from this perspective it suffices to describe the estimator in this paper for the cubic specification.

We use the following notation: $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$ and $\operatorname{diag}(\cdot)$ denotes a diagonal matrix. With $\Rightarrow$, $\xrightarrow{p}$ and $\xrightarrow{d}$ we denote weak convergence, convergence in probability and convergence in distribution, respectively, as $T \to \infty$. Brownian motion with variance specified in the context is denoted by $B(r)$ and $W(r)$ denotes a standard Wiener process. $\operatorname{Var}(z)$ denotes the covariance matrix of a vector z and $\operatorname{Cov}(z_1, z_2)$ denotes the cross-covariance matrix of two vectors z_1 and z_2.

2.2 Theory

As mentioned in the introduction, in this paper we discuss the cubic specification with a single unit root regressor only. With respect to deterministic regressors, we allow for either individual specific intercepts (i. e., fixed effects) only or individual specific intercepts and individual specific linear time trends. The integrated regressors x_{it} *potentially* contain individual specific drifts μ_i, i. e.,:

$$y_{it} = \alpha_i + \delta_i t + x_{it}\beta_1 + x_{it}^2\beta_2 + x_{it}^3\beta_3 + u_{it}, \tag{2.1}$$

$$x_{it} = \mu_i + x_{i,t-1} + v_{it}. \tag{2.2}$$

Mainly to relate the paper to de Jong and Wagner (2022), see the discussion below Assumption 2.3, we use the same assumptions as in that (companion) paper. Thus, we assume that the cross-sectionally independent error processes $\{\eta_{it}\}_{t\in\mathbb{Z}} := \{(u_{it}, v_{it})'\}_{t\in\mathbb{Z}}$ are random linear processes fulfilling a functional central limit theorem similar to Phillips and Moon (1999, Lemma 3), i. e.,:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \eta_{it} \Rightarrow B_i(r) = \Omega_i^{1/2} W_i(r), \quad 0 \le r \le 1, \tag{2.3}$$

where $W_i(r) := (W_{u_i}(r), W_{v_i}(r))'$, with $B_i(r)$ partitioned analogously, is a bivariate standard Wiener process. The random long-run covariance matrices are partitioned as:

$$\Omega_i := \begin{pmatrix} \Omega_{u_i u_i} & \Omega_{u_i v_i} \\ \Omega_{v_i u_i} & \Omega_{v_i v_i} \end{pmatrix}. \tag{2.4}$$

For later usage we also define the half long-run covariance matrices partitioned analogously, i. e.,:

$$\Delta_i := \begin{pmatrix} \Delta_{u_i u_i} & \Delta_{u_i v_i} \\ \Delta_{v_i u_i} & \Delta_{v_i v_i} \end{pmatrix}, \tag{2.5}$$

with consequently $\Omega_i = \Delta_i + \Delta_i' - \Sigma_i$, where Σ_i is the random contemporaneous covariance matrix. More specifically, this leads to the assumption:

Assumption 2.1. *The random processes $\{\eta_{it}\}_{t\in\mathbb{Z}}$ are independent across $i = 1, \dots, N$, the random matrices (Δ_i, Σ_i) are independent of the Wiener processes $W_i(r)$ for $i = 1, \dots, N$ and Ω_i is positive definite almost surely for $i = 1, \dots, N$.*

Given the primary focus on the slope parameter vector $\beta := (\beta_1, \beta_2, \beta_3)'$, it is convenient to use *uniform notation*, $\tilde{y}_{it}$ and $\tilde{X}_{it}$, for both demeaned and demeaned and linearly detrended variables. In the demeaning only case, we thus have:

$$\tilde{y}_{it} := y_{it} - \bar{y}_{i.} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}, \tag{2.6}$$

with analogously defined quantities for x_{it} (and its powers), u_{it} and v_{it}. Stacking defines:

$$\tilde{X}_{it} := \begin{pmatrix} \tilde{x}_{it} \\ \widetilde{x_{it}^2} \\ \widetilde{x_{it}^3} \end{pmatrix} = \begin{pmatrix} x_{it} - \overline{x}_{i.} \\ x_{it}^2 - \overline{x_{i.}^2} \\ x_{it}^3 - \overline{x_{i.}^3} \end{pmatrix}. \tag{2.7}$$

In case of demeaning and linear detrending, we have, using generic notation z_{it}, for y_{it}, x_{it} and its powers:

$$\tilde{z}_{it} := z_{it} - \frac{4T-6t+2}{T-1}\overline{z_{i.}} - \frac{-6T+12t-6}{(T-1)(T+1)}\sum_{t=1}^{T}\left(\frac{t}{T}\right) z_{it}, \tag{2.8}$$

leading to a correspondingly demeaned and detrended stacked vector $\tilde{X}_{it}$, with $\tilde{u}_{it}$ and $\tilde{v}_{it}$ again defined analogously.[3]

The exact form of the results depends, in addition to the specification of the deterministic components in the regression equation, also on whether the regressors x_{it} include a (non-zero) drift or not. It is therefore convenient to structure the discussion according to the following cases: zero drifts $\mu_i = 0$, $i = 1, \ldots, N$, non-zero drifts $\mu_i \neq 0$, $i = 1, \ldots, N$ and the general case $\mu_i \in \mathbb{R}$, $i = 1, \ldots, N$.

2.2.1 Zero Drifts

To complete the formulation of the assumptions required in this case, define $G_T := \mathrm{diag}(T^{-1}, T^{-3/2}, T^{-2})$ and $A_i := \left(1, 2\int_0^1 B_{v_i}(r)dr, 3\int_0^1 B_{v_i}^2(r)dr\right)'$. To capture the effects of demeaning, demeaning and linear detrending – or of the "removal" of more general trend functions – define for an (integrable stochastic process) $P(r)$ and an asymptotically regular trend function $D(r)$ for $0 \leq r \leq 1$:

$$\tilde{P}(r) := P(r) - D(r)\left(\int_0^1 D(s)D(s)'ds\right)^{-1}\int_0^1 D(s)P(s)ds, \tag{2.9}$$

which for the case of demeaning, of course, simplifies to $\tilde{P}(r) = P(r) - \int_0^1 P(s)ds$.[4] The notation allows to (generically) define $\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) := \left(\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}), \widetilde{\mathbf{B}^{\mathbf{2}}_{\mathbf{v_i}}}(\mathbf{r}), \widetilde{\mathbf{B}^{\mathbf{3}}_{\mathbf{v_i}}}(\mathbf{r})\right)'$, corresponding to the deterministic specification considered. Using this notation we assume:

Assumption 2.2. *For $i = 1, \ldots, N$ and $0 \leq r \leq 1$ it holds that:*

[3] Clearly, more general (asymptotically) regular trend functions can be considered, e. g., higher order polynomial time trends. A trend function $D(r), 0 \leq r \leq 1$ is called asymptotically regular, if $\int_0^1 D(r)D(r)'dr$ is positive definite.

[4] As is well-known, in case of demeaning and linear detrending, $\tilde{P}(r) = P(r) - (4-6r)\int_0^1 P(s)ds - (-6 + 12r)\int_0^1 sP(s)ds$.

(a) $T^{1/2}G_T\tilde{X}_{i\lfloor rT\rfloor}\Rightarrow\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})$,

(b) $G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\xrightarrow{d}\int_0^1\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\mathbf{d}\mathbf{B}_{\mathbf{u_i}}(\mathbf{r})+\mathbf{\Delta}_{\mathbf{v_i u_i}}\mathbf{A_i}$,

(c) $G_T\sum_{t=1}^{T}\tilde{X}_{it}v_{it}\xrightarrow{d}\int_0^1\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\mathbf{d}\mathbf{B}_{\mathbf{v_i}}(\mathbf{r})+\mathbf{\Delta}_{\mathbf{v_i v_i}}\mathbf{A_i}$,

with all quantities converging jointly.

As usual in fully modified OLS type estimation, consistent non-parametric kernel estimators of long-run covariances and half long-run covariances – based on the OLS residuals $\hat{u}_{it}$ from (2.1) and $v_{it}=\Delta x_{it}$ – are required. This in turn requires appropriate kernel and bandwidth choices, compare, e. g., Jansson (2002).[5]

Assumption 2.3. *The cross-sectionally independent estimators $\hat{\Delta}_i$ and $\hat{\Sigma}_i$ satisfy $\hat{\Delta}_i\xrightarrow{p}\Delta_i$ and $\hat{\Sigma}_i\xrightarrow{p}\Sigma_i$ for $i=1,\ldots,N$. By definition, this implies cross-sectional independence and consistency of $\hat{\Omega}_i:=\hat{\Delta}_i+\hat{\Delta}_i'-\hat{\Sigma}_i$ for $i=1,\ldots,N$.*

For brevity we abstain from formulating primitive assumptions that generate our Assumptions 2.2 and 2.3 that are, in fact, convergence results. The literature provides several – by now well-understood – routes to derive these results from primitive assumptions using near epoch dependence concepts, martingale difference sequences or linear processes (see, e. g., de Jong, 2002; Ibragimov and Phillips, 2008; Park and Phillips, 2001). Our formulations and assumptions are similar to de Jong and Wagner (2022) who in turn build upon Phillips and Moon (1999). However, in a finite N setting, as considered in this paper, one can replace the random linear process framework without any (substantial) loss with more classical assumptions as posited, e. g., in Wagner and Hong (2016) in a time series setting. As discussed below in Remark 2.4, the random linear process framework provides *fundamental* value added only in case of $N\to\infty$, see also the discussion in de Jong and Wagner (2022).

We are now ready to define the group-mean fully modified OLS estimator as the cross-sectional average of the individual specific fully modified OLS estimators (as developed in Wagner and Hong, 2016) of the coefficient vector β. More precisely, we define for $i=1,\ldots,N$ the FM-OLS estimator of β from the i-th cross-section member – computed from individual specifically demeaned or individual specifically demeaned and linearly detrended data – as:

$$\hat{\beta}^{+}(i):=\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{y}_{it}^{+}-C_i\right), \tag{2.10}$$

where $\tilde{y}_{it}^{+}:=\tilde{y}_{it}-\Delta x_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}$ and $C_i:=\hat{\Delta}_{v_iu_i}^{+}\left(T,2\sum_{t=1}^{T}x_{it},3\sum_{t=1}^{T}x_{it}^2\right)'$, with $\hat{\Delta}_{v_iu_i}^{+}:=\hat{\Delta}_{v_iu_i}-\hat{\Delta}_{v_iv_i}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}$.[6] The cross-sectional average of $\hat{\beta}^{+}(i)$ defines the *group-mean* fully modified OLS

[5] To maintain cross-sectional independence of the individual specific estimators, the long-run covariance matrix estimators need to be cross-sectionally independent as well. The asymptotic analysis considered in de Jong and Wagner (2022), with also $N\to\infty$ after $T\to\infty$, allows for more flexibility in this respect.

[6] Note that performing FM-OLS calculations for a time series dimension ranging from $t=1,\ldots,T$ *implicitly* assumes that observations are available for $t=0,\ldots,T$ as the construction of $\tilde{y}_{it}^{+}$ implies that one loses the first observation.

estimator:

$$\hat{\beta}^{+} := \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{+}(i). \tag{2.11}$$

The following proposition derives its asymptotic distribution as the time series dimension $T \to \infty$, for fixed cross-section dimension N.

Proposition 2.1. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i = 0,\ i = 1, \ldots, N$ *and let Assumptions 2.1, 2.2 and 2.3 be in place. Then it holds for* $T \to \infty$, conditional upon Δ_i, Σ_i *and* $W_{v_i}(r)$ *for* $i = 1, \ldots, N$, *that:*

$$G_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V^{+}\right), \tag{2.12}$$

where $\mathcal{N}\left(0, V^{+}\right)$ *denotes a normal distribution with expectation zero and conditional covariance matrix:*

$$V^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i\cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})'\mathbf{dr}\right)^{-1} = \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i\cdot v_i}\tilde{M}_{ii}^{-1}, \tag{2.13}$$

with $\Omega_{u_i\cdot v_i} := \Omega_{u_iu_i} - \Omega_{u_iv_i}\Omega_{v_iv_i}^{-1}\Omega_{v_iu_i} > 0$ *given by the conditional variance of* $B_{u_i\cdot v_i}(r) := B_{u_i}(r) - \Omega_{u_iv_i}\Omega_{v_iv_i}^{-1}B_{v_i}(r)$ *and* $\tilde{M}_{ii}$ *defined by the last equality.*

Under our assumptions, the natural consistent estimator of V^{+} is:

$$\hat{V}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1} = G_T^{-1}\hat{S}^{+}G_T^{-1}, \tag{2.14}$$

with $\hat{\Omega}_{u_i\cdot v_i} := \hat{\Omega}_{u_iu_i} - \hat{\Omega}_{u_iv_i}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}$ and $\hat{S}^{+}$ defined by the last equality.

The conditional normal limit in conjunction with the availability of a consistent estimator of the covariance matrix as given in (2.14) leads to standard asymptotic inference. To obtain standard asymptotic behavior of hypothesis tests, we have to take into account that the components of the vector $\hat{\beta}^{+}$ converge at different rates, an issue discussed in detail, e. g., in Sims *et al.* (1990, Section 4) or Wagner and Hong (2016, Section 2.2, p. 1297). It suffices to assume that the constraint matrix fulfills the (asymptotic) restriction posited in the following corollary.

Corollary 2.1. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i = 0,\ i = 1, \ldots, N$, *and let Assumptions 2.1, 2.2 and 2.3 be in place. Consider* s *linearly independent restrictions collected in:*

$$H_0 : R\beta = r, \tag{2.15}$$

with $R \in \mathbb{R}^{s\times 3}$, $r \in \mathbb{R}^{s}$ *and assume that there exists a non-singular matrix* $G_R \in \mathbb{R}^{s\times s}$ *such that* $\lim_{T\to\infty} G_R R G_T = R^{*}$, *with* $R^{*} \in \mathbb{R}^{s\times 3}$ *of rank* s. *Then it holds under the null hypothesis that the Wald-type statistic:*

$$W^{+} := \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.16}$$

is chi-squared distributed with s degrees of freedom as $T \to \infty$. In case $s = 1$, of course, also a t-type test can be considered:

$$t^+ := \frac{R\hat{\beta}^+ - r}{\sqrt{R\hat{S}^+ R'}}, \tag{2.17}$$

which is under the null hypothesis asymptotically standard normally distributed as $T \to \infty$.

Inference on α_i and δ_i is also possible. Similarly, as an observation for later when drifts are considered, also inference on μ_i using, e. g., augmented Dickey-Fuller type regressions can be performed.

Remark 2.1. *The group-mean estimator is robust to many forms of cross-section dependence, i. e., it remains consistent with a zero mean Gaussian mixture limiting distribution despite cross-section dependence. Of course, the covariance matrix of the asymptotic distribution changes – depending upon the form and extent of cross-section dependence. Given that we consider a fixed N setting, it suffices to simply consider a multivariate version of our assumptions ensuring joint convergence of all quantities for $i = 1, \ldots, N$.*

The key quantity required for "robust" inference is (a consistent estimator of) the asymptotic covariance matrix of the group-mean FM-OLS estimator in case of cross-section dependence. To this end denote with $\tilde{M}_{ij} := \int_0^1 \mathbf{\tilde{B}_{v_i}(r)\tilde{B}_{v_j}(r)'dr}$ and with $\Omega_{u_i \cdot v_i; u_j \cdot v_j}$ the "constant" in the quadratic covariation of the processes $B_{u_i \cdot v_i}(r)$ and $B_{u_j \cdot v_j}(r)$.[7] *The asymptotic covariance matrix of the group-mean estimator given in (2.11) changes from the expression given in (2.13) to the "sandwich" form:*

$$V^+_{rob} := \frac{1}{N^2} \sum_{i,j=1}^{N} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \tilde{M}_{ii}^{-1} \tilde{M}_{ij} \tilde{M}_{jj}^{-1}. \tag{2.18}$$

It is important to note that the above result allows for very general forms of cross-section dependencies, as long as V^+_{rob} is invertible. As an (extreme) example, consider the case $x_{it} = x_t$ for $i = 1, \ldots, N$, i. e., the integrated regressor is the same for all cross-section members, which is an extreme form of cross-unit cointegration, compare Wagner and Hlouskova (2010). In this case $\tilde{M}_{ii} = \tilde{M}_{jj} = \tilde{M}_{ij} = \tilde{M}$ for $i, j = 1, \ldots, N$ and $V^+_{rob} = \left(\frac{1}{N^2}\sum_{i,j=1}^{N} \Omega_{u_i \cdot v; u_j \cdot v}\right) \tilde{M}^{-1}$, using simplified notation $\Delta x_t = v_t$ in Ω. The term in brackets simplifies in this case to $\frac{1}{N^2} 1_N' \Omega_{u \cdot v} 1_N$, with $1_N := [1, \ldots, 1]' \in \mathbb{R}^N$ and $\Omega_{u \cdot v} = \Omega_{uu} - \Omega_{uv}\Omega_{vv}^{-1}\Omega_{vu}$. Thus, positive definiteness of $\Omega_{u \cdot v}$ is in this example sufficient for robust inference. This example highlights the wide applicability of robust inference based on the group-mean estimator, without having to posit a model for cross-section dependence, e. g., common stochastic trends or a factor structure.[8]

[7]Given that we consider the quadratic covariation between Brownian motions, this constant is of course simply the covariance between $B_{u_i \cdot v_i}(1)$ and $B_{u_j \cdot v_j}(1)$.

[8]We abstain from positing an explicit set of assumptions for brevity as the discussion in the remark makes clear that any set of sufficient assumptions has to extend the *marginal* assumptions posited so far to hold jointly with cross-section dependence allowed for. Clearly, in the presence of cross-section dependence the estimators of the joint long-run covariance matrix cannot will not feature cross-sectional independence by construction, compare Footnote 5.

A consistent estimator of the asymptotic covariance matrix V^{+}_{rob} is given by:

$$\begin{aligned}\hat{V}^{+}_{rob} &:= \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}G_T\right)^{-1}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{jt}G_T\right)\left(G_T\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}'_{jt}G_T\right)^{-1}\\ &= G_T^{-1}\frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{jt}\right)\left(\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}'_{jt}\right)^{-1}G_T^{-1}\\ &=: G_T^{-1}\hat{S}^{+}_{rob}G_T^{-1}, \end{aligned}\tag{2.19}$$

with $\hat{S}^{+}_{rob}$ defined by the last equality. Since $\Omega_{u_i\cdot v_i;u_j\cdot v_j} = \Omega_{u_iu_j} - \Omega_{u_iv_i}\Omega^{-1}_{v_iv_i}\Omega_{v_iu_j} - \Omega_{u_jv_j}\Omega^{-1}_{v_jv_j}\Omega_{v_ju_i} + \Omega_{u_iv_i}\Omega^{-1}_{v_iv_i}\Omega_{v_iv_j}\Omega^{-1}_{v_jv_j}\Omega_{v_ju_j}$, we obtain the estimator $\hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j}$ by replacing the unknown long-run variances and covariances in the expression just given for $\Omega_{u_i\cdot v_i;u_j\cdot v_j}$ by consistent estimators. "Robust" Wald-type and t-type test statistics can now be defined similarly to the Wald-type and t-type test statistics defined in (2.16) *and* (2.17)*, with $\hat{S}^{+}_{rob}$ as defined in* (2.19) *in place of $\hat{S}^{+}$, i. e.:*

$$W^{+}_{rob} := \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}_{rob}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.20}$$

$$t^{+}_{rob} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\hat{S}^{+}_{rob}R'}}, \tag{2.21}$$

which are under the null hypothesis asymptotically chi-squared distributed with s degrees of freedom and standard normally distributed, respectively, as $T \to \infty$.

Remark 2.2. *In panel data settings often rather than individual specific time trends, time effects are considered – most commonly in conjunction with individual effects – in a two-way effects specification. Time effects also do not invalidate consistency of the group-mean estimator. However, the limiting distribution is in this case contaminated by second order bias terms related to the presence of cross-sectional averages of time series limits. In the two-way case, with individual specific intercepts and time effects, the transformed regressor vector, e. g., is given by $\check{X}_{it} := \tilde{X}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{X}_{jt}$, with $\tilde{X}_{it},\ i = 1,\ldots,N$ as defined in* (2.7)*. This leads to a partial sum limit (compare Assumption 2.2) of the form $T^{1/2}G_T\check{X}_{i\lfloor rT\rfloor} \Rightarrow \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) - \frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r}) =: \check{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})$. Thus, the cross-section dependence induced by two-way demeaning shows up in the limit partial sum processes, which in turn leads to second order bias terms also in the limit of $G_T\sum_{t=1}^{T}\check{X}_{it}\check{u}^{+}_{it}$, with $\check{u}^{+}_{it} := \check{u}_{it} - \Delta x_{it}\hat{\Omega}^{-1}_{v_iv_i}\hat{\Omega}_{v_iu_i}$ and $\check{u}_{it} := \tilde{u}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{u}_{jt}$. Under appropriate assumptions $\frac{1}{N}\sum_{j=1}^{N}\tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r})$ fulfills a law of large numbers for $N \to \infty$. A corresponding result is the basis for the derivation of the large N and large T asymptotic distribution of the pooled estimator in de Jong and Wagner (2022) in the two-way effects case.*

Remark 2.3. *Considering time effects in a* cross-sectionally homogenous *case, with $\Delta_i = \Delta$ a.s. and $\Sigma_i = \Sigma$ a.s for $i = 1,\ldots,N$, allows to alternatively adjust the group-mean estimator to achieve asymptotically valid inference by using $\check{y}^{+}_{it} := \check{y}_{it} - \Delta\check{x}_{it}\hat{\Omega}^{-1}_{vv}\hat{\Omega}_{vu}$, where $\check{y}_{it} := \tilde{y}_{it} - \frac{1}{N}\sum_{j=1}^{N}\tilde{y}_{jt}$, with*

$\tilde{y}_{it}$ *as defined in* (2.6) *for* $i = 1, \ldots, N$, *as transformed dependent variable and:*[9]

$$\check{C}_i := \hat{\Delta}^+_{vu} \left(\left(\frac{N-1}{N} \right)^2 \left(T, 2\sum_{t=1}^{T} x_{it}, 3\sum_{t=1}^{T} x_{it}^2 \right)' + \frac{1}{N^2} \sum_{j \neq i} \left(T, 2\sum_{t=1}^{T} x_{jt}, 3\sum_{t=1}^{T} x_{jt}^2 \right)' \right) \tag{2.22}$$

as additive correction term when estimating the parameters of the i*-th equation with FM-OLS. This leads to the following homogeneous group-mean estimator:*

$$\check{\beta}^+_{HOM} := \frac{1}{N} \sum_{i=1}^{N} \check{\beta}^+(i), \tag{2.23}$$

where:

$$\check{\beta}^+(i) := \left(\sum_{t=1}^{T} \check{X}_{it} \check{X}_{it}' \right)^{-1} \left(\sum_{t=1}^{T} \check{X}_{it} \check{y}_{it}^+ - \check{C}_i \right), \quad i = 1, \ldots, N. \tag{2.24}$$

The asymptotic (conditional) covariance matrix of the homogenous group-mean FM-OLS estimator is given by $V^+_{HOM} := \Omega_{u \cdot v} \frac{1}{N^2} \sum_{i=1}^{N} \tilde{M}_{ii}^{-1}$, *compare* (2.13). *The homogenous versions of the Wald- and t-type statistics follow straightforwardly.*

Remark 2.4. *Note that under (additional) assumptions that ensure the existence of required moments, in particular of* $\mathbb{E}(\Omega_{u_i \cdot v_i} \tilde{M}_{ii}^{-1})$, *it follows in case of cross-sectional independence that:*

$$\sqrt{N} G_T^{-1} \left(\hat{\beta}^+ - \beta \right) \xrightarrow{d} \mathcal{N} \left(0, \mathbb{E}(\Omega_{u_i \cdot v_i} \tilde{M}_{ii}^{-1}) \right), \tag{2.25}$$

as $N \to \infty$ *after* $T \to \infty$. *An estimator of the covariance matrix of this limiting distribution is given by* $N\hat{V}^+$, *with* $\hat{V}^+$ *the "finite* N*" covariance matrix estimator given below Proposition 2.1 in* (2.14).

2.2.2 Non-Zero Drifts

Let us now consider the case with non-zero drifts, i. e., $\mu_i \neq 0,\ i = 1, \ldots, N$. In this case, the integrated regressor:

$$x_{it} = \mu_i + x_{i,t-1} + v_{it} = \mu_i t + \sum_{s=1}^{t} v_{is} + x_{i0} = \mu_i t + x_{it}^o + x_{i0}, \tag{2.26}$$

is asymptotically *dominated* by the deterministic linear trend $\mu_i t$ rather than the stochastic trend $x_{it}^o := \sum_{s=1}^{t} v_{is}$. For later usage define $\tilde{X}_{it}^o$ similarly to $\tilde{X}_{it}$ in (2.7), with x_{it}^o and its powers in place of x_{it} and its powers.

The implications of the dominance of a deterministic trend component on unit root and cointegration analysis have been investigated in the linear time series case already by West (1988), and, in the context of FM-OLS estimation, in Phillips and Hansen (1990, Remark (e), p. 105). For the

[9] In this case, e. g., the homogenous long-run covariance matrix Ω can be estimated by the cross-sectional average of individual specific long-run covariance matrix estimators, i. e., $\hat{\Omega} := \frac{1}{N} \sum_{i=1}^{N} \hat{\Omega}_i$; and similarly for the other required matrices.

second and third powers of x_{it}, the higher order deterministic (monomial) quadratic or cubic time trends are the dominant elements. This, of course, leads to similar asymptotic normality results as that of West (1988) in a *linear cointegration setting.* However, in our context, the deterministic trend will not be dominant in $\tilde{x}_{it}$, when *both* demeaning and linear detrending take place. In this case the deterministic component is exactly annihilated in the demeaned and detrended variable $\tilde{x}_{it}$. Consequently, in this case, the coefficient to the first power of the integrated regressor will have a unit root type asymptotic distribution rather than a normal asymptotic distribution.[10]

It is maybe worth mentioning that the presence of non-zero drifts $\mu_i \neq 0$ does not imply changes in the construction of the transformed dependent variable $\tilde{y}_{it}^+$. Commencing from $\Delta x_{it} = \mu_i + v_{it}$ immediately leads to:

$$\tilde{y}_{it}^+ = \tilde{y}_{it} - \Delta x_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i} = \tilde{y}_{it} - v_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i} - \mu_i\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}. \tag{2.27}$$

This in turn implies that $\tilde{u}_{it}^+ = \tilde{u}_{it} - v_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i} - \mu_i\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}$. Consequently, the scaled partial sum process $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor}\tilde{u}_{it}^+$ diverges, being non-centered. Nevertheless, $\sum_{t=1}^{T}\tilde{X}_{it} = 0$ implies that – after appropriate scaling – the cross product term $\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^+$ converges (conditionally) to a Gaussian mixture limit (with integrator and integrand independent of each other) plus an additive bias term to be subtracted. This is the key result allowing for asymptotically valid inference based upon the group-mean FM-OLS estimator. Thus, the definition and computation of group-mean FM-OLS estimator is unaffected by or invariant to the presence of non-zero drifts.

Remark 2.5. *In relation to the above, a word of caution may be in order concerning long-run covariance estimation, typically based on the OLS residuals $\hat{u}_{it}$ of* (2.1) *in conjunction with the first difference Δx_{it} of x_{it}. If one uses, as is sometimes done, an estimator that does not center the variables prior to autocovariance estimation, the resultant estimator will diverge due to non-zero expectation μ_i of Δx_{it}. By construction $\hat{u}_{it}$ does not have to be centered in any of our specifications as they all include at least an intercept as deterministic variable. If it is known that $\mu_i = 0$, also Δx_{it} need not be centered.*

Extending Proposition 2.1 to the case of non-zero drifts requires the definition of a few additional quantities, including the scaling matrices $H_T := \text{diag}\left(T^{-3/2}, T^{-5/2}, T^{-7/2}\right)$ and $K_T := \text{diag}\left(T^{-1}, T^{-5/2}, T^{-7/2}\right)$. Furthermore, for $i = 1, \ldots, N$ define:

$$J_i(r) := \begin{pmatrix} \mu_i & & \\ & \mu_i^2 & \\ & & \mu_i^3 \end{pmatrix} \begin{pmatrix} r - 1/2 \\ r^2 - 1/3 \\ r^3 - 1/4 \end{pmatrix} =: \mathcal{D}(\mu_i) \begin{pmatrix} r - 1/2 \\ r^2 - 1/3 \\ r^3 - 1/4 \end{pmatrix} \tag{2.28}$$

$$L_i(r) := \begin{pmatrix} 1 & & \\ & \mu_i^2 & \\ & & \mu_i^3 \end{pmatrix} \begin{pmatrix} \tilde{B}_{v_i}(r) \\ r^2 - r + 1/6 \\ r^3 - 9/10r + 1/5 \end{pmatrix} =: \mathcal{E}(\mu_i) \begin{pmatrix} \tilde{B}_{v_i}(r) \\ r^2 - r + 1/6 \\ r^3 - 9/10r + 1/5 \end{pmatrix} \tag{2.29}$$

[10]For a full analysis of the impacts of the presence of deterministic trends in the regression equation and/or the regressors for a *more general* cointegrating polynomial regression specification – in the time series case – see Reichold and Wagner (2022).

Proposition 2.2. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i \neq 0$, $i = 1, \ldots, N$ *and let Assumptions 2.1, 2.2 for* $\tilde{X}_{it}^{o}$ *and 2.3 be in place.*

(i) In case individual specific intercepts but no individual specific linear trends are included in (2.1)*, it holds for* $T \to \infty$*, conditional upon* Δ_i *and* Σ_i *for* $i = 1, \ldots, N$ *that:*

$$H_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\alpha}^{+}\right), \tag{2.30}$$

with $V_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}$ *for* $i = 1, \ldots, N$.

(ii) In case individual specific intercepts and linear trends are included in (2.1)*, it holds for* $T \to \infty$*, conditional upon* Δ_i*,* Σ_i *and* $W_{v_i}(r)$ *for* $i = 1, \ldots, N$ *that:*

$$K_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\alpha,\delta}^{+}\right), \tag{2.31}$$

with $V_{\alpha,\delta}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1}$ *for* $i = 1, \ldots, N$.

Proposition 2.2 shows that the two cases – with or without individual specific trends – lead to different asymptotic distributions of the group-mean FM-OLS estimator. Case (i), without individual specific trends, leads to a *West*-type asymptotic normality result for all elements of β, more clearly (unconditionally) visible in case Δ_i and Σ_i are considered non-random. It is convenient to rewrite V_{α}^{+} as:

$$V_{\alpha}^{+} = \frac{1}{N^2}\sum_{i=1}^{N}\Omega_{u_i \cdot v_i}\left(\mathcal{D}(\mu_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\mu_i)\right)^{-1} \tag{2.32}$$

This leads immediately to *two* estimators of V_{α}^{+}, one similar to the estimator $\hat{V}^{+}$ given in (2.14) and the second commencing from the closed form expression for the limit result, i. e.:

$$\hat{V}_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i \cdot v_i}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1} = H_T^{-1}\hat{S}^{+}H_T^{-1}, \tag{2.33}$$

and:

$$\tilde{V}_{\alpha}^{+} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i \cdot v_i}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\hat{\mu}_i)\right)^{-1}, \tag{2.34}$$

with, e. g., $\hat{\mu}_i := \frac{1}{T}\sum_{t=1}^{T}\Delta x_{it}$.

In case (ii), with individual specific intercepts and linear trends included, the coefficient to $\tilde{x}_{it}$ has, as mentioned above, a *unit root* type limiting distribution and only the coefficients to the higher order powers have a *West*-type asymptotic normal distribution. This implies that a "direct" estimator of $V^{+}_{\alpha,\delta}$, similar in spirit to $\tilde{V}^{+}_{\alpha}$, can only be constructed for the lower 2×2 block, i. e.:

$$\hat{V}^{+}_{\alpha,\delta} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\left(K_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}K_T\right)^{-1} = K_T^{-1}\hat{S}^{+}K_T^{-1}, \tag{2.35}$$

and:

$$\tilde{V}^{+}_{\alpha,\delta} := \frac{1}{N^2}\sum_{i=1}^{N}\hat{\Omega}_{u_i\cdot v_i}\begin{pmatrix} \frac{1}{T^2}\sum_{t=1}^{T}(\tilde{x}_{it})^2 & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}} \\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \hat{\mu}_i^4/180 & \hat{\mu}_i^5/120 \\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}} & \hat{\mu}_i^5/120 & 9\hat{\mu}_i^6/700 \end{pmatrix}^{-1}. \tag{2.36}$$

The above considerations lead to the following corollary summarizing the test options in case of non-zero drifts.

Corollary 2.2. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i \neq 0$, $i = 1, \ldots, N$ *and let Assumptions 2.1, 2.2 for* $\tilde{X}^{o}_{it}$ *and 2.3 be in place. Consider* s *linearly independent restrictions collected in* $H_0 : R\beta = r$, *with* $R \in \mathbb{R}^{s\times 3}$, $r \in \mathbb{R}^s$ *and assume that there exists a non-singular matrix* $G_R \in \mathbb{R}^{s\times s}$ *and a matrix* $R^* \in \mathbb{R}^{s\times 3}$ *of rank* s *such that* $\lim_{T\to\infty} G_R R H_T = R^*$ *(in the individual specific intercepts only case) or* $\lim_{T\to\infty} G_R R K_T = R^*$ *(in the individual specific intercepts and linear trends case).*

In both, the individual specific intercepts only and the individual specific intercepts and linear trends cases, the Wald- and (in case $s = 1$*)* t*-type statistics:*

$$W^{+} = \left(R\hat{\beta}^{+} - r\right)'\left(R\hat{S}^{+}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.37}$$

$$t^{+} = \frac{R\hat{\beta}^{+} - r}{\sqrt{R\hat{S}^{+}R'}}, \tag{2.38}$$

already defined in (2.16) *and* (2.17)*, are under the null hypothesis chi-squared distributed with* s *degrees of freedom and standard normally distributed, respectively, as* $T \to \infty$.

Furthermore, in the individual specific intercepts only case, the test statistics can alternatively (asymptotically equivalently) be defined as:

$$W^{+}_{\alpha} := \left(R\hat{\beta}^{+} - r\right)'\left(R\tilde{S}^{+}_{\alpha}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.39}$$

$$t^{+}_{\alpha} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\tilde{S}^{+}_{\alpha}R'}}, \tag{2.40}$$

with $\tilde{S}^{+}_{\alpha} := H_T\tilde{V}^{+}_{\alpha}H_T$.

In the individual specific intercepts and linear trends case, the test statistics can alternatively *(asymptotically equivalently) be defined as:*

$$W^+_{\alpha,\delta} := \left(R\hat{\beta}^+ - r\right)' \left(R\tilde{S}^+_{\alpha,\delta}R'\right)^{-1} \left(R\hat{\beta}^+ - r\right) \tag{2.41}$$

$$t^+_{\alpha,\delta} := \frac{R\hat{\beta}^+ - r}{\sqrt{R\tilde{S}^+_{\alpha,\delta}R'}}, \tag{2.42}$$

with $\tilde{S}^+_{\alpha,\delta} := K_T\tilde{V}^+_{\alpha,\delta}K_T$. *Under the null hypothesis the four additionally considered test statistics are asymptotically are asymptotically chi-squared or standard normally distributed, respectively, as* $T \to \infty$.

Remark 2.6. *Similar to Remark 2.1 in Subsection 2.2.1, the results can be extended to allow for cross-section dependence; based again upon any suitable modification of the assumptions to ensure the necessary joint convergence results. The precise form of the asymptotic results will depend upon the deterministic components in* (2.1)*. With individual specific intercepts only, the covariance matrix of the asymptotic distribution is in case of cross-section dependence given by:*

$$\begin{aligned} V^+_{\alpha,rob} &:= \frac{1}{N^2}\sum_{i,j=1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j} \left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1} \int_0^1 J_i(r)J_j(r)'dr \left(\int_0^1 J_j(r)J_j(r)'dr\right)^{-1} \\ &= \frac{1}{N^2}\sum_{i,j=1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j} \left(\mathcal{D}(\mu_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\mu_j)\right)^{-1}. \end{aligned} \tag{2.43}$$

In case that both individual specific intercepts and linear trends are included in (2.1)*, the covariance matrix of the asymptotic distribution is given by:*

$$\begin{aligned} V^+_{\alpha,\delta,rob} &:= \frac{1}{N^2}\sum_{i,j=1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j} \left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1} \int_0^1 L_i(r)L_j(r)'dr \left(\int_0^1 L_j(r)L_j(r)'dr\right)^{-1} \\ &=: \frac{1}{N^2}\sum_{i,j=1}^{N} \Omega_{u_i\cdot v_i;u_j\cdot v_j}\mathcal{C}(i,j), \end{aligned} \tag{2.44}$$

with $\mathcal{C}(i,j)$ *defined by the last equality. Considering again:*

$$\hat{S}^+_{rob} = \frac{1}{N^2}\sum_{i,j=1}^{N} \hat{\Omega}_{u_i\cdot v_i;u_j\cdot v_j} \left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{it}\right)^{-1} \left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}'_{jt}\right) \left(\sum_{t=1}^{T}\tilde{X}_{jt}\tilde{X}'_{jt}\right)^{-1}, \tag{2.45}$$

as defined already in (2.19)*, immediately leads to consistent estimators in both cases, given by* $\hat{V}^+_{\alpha,rob} := H_T^{-1}\hat{S}^+_{rob}H_T^{-1}$ *or* $\hat{V}^+_{\alpha,\delta,rob} := K_T^{-1}\hat{S}^+_{rob}K_T^{-1}$*, respectively. Entirely analogously to Remark 2.1, using* $\hat{S}^+_{rob}$ *in the definition of the robust test statistics* W^+_{rob} *and* t^+_{rob} *(in case* $s = 1$*) given in* (2.20) *and* (2.21)*, leads to chi-squared and standard normal inference, respectively, as* $T \to \infty$.

Note for completeness that the test statistics W^+_{α} *and* t^+_{α} *defined in* (2.39) *and* (2.40) *in Corol-*

lary 2.2 can also be "robustified" straightforwardly. Considering:

$$\tilde{V}^{+}_{\alpha,rob} := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i v_i;u_j \cdot v_j}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix}1/12 & 1/12 & 3/40\\ 1/12 & 4/45 & 1/12\\ 3/40 & 1/12 & 9/112\end{pmatrix}\mathcal{D}(\hat{\mu}_j)\right)^{-1} \tag{2.46}$$

and $\tilde{S}^{+}_{\alpha,rob} := H_T \tilde{V}^{+}_{\alpha,rob} H_T$ *allows to define:*

$$W^{+}_{\alpha,rob} := \left(R\hat{\beta}^{+} - r\right)'\left(R\tilde{S}^{+}_{\alpha,rob}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.47}$$

$$t^{+}_{\alpha,rob} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\tilde{S}^{+}_{\alpha,rob}R'}}. \tag{2.48}$$

Analogously, $W^{+}_{\alpha,\delta}$ *and* $t^{+}_{\alpha,\delta}$ *can be "robustified" by constructing a "direct" estimator of* $V^{+}_{\alpha,\delta,rob}$*. To be precise,* $\tilde{V}^{+}_{\alpha,\delta}$ *defined in* (2.36) *has to be replaced by:*

$$\tilde{V}^{+}_{\alpha,\delta,rob} := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i \cdot v_i;u_j \cdot v_j}\tilde{\mathcal{C}}(i,j) := \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i \cdot v_i;u_j \cdot v_j}\tilde{\mathcal{A}}(i)^{-1}\tilde{\mathcal{B}}(i,j)\tilde{\mathcal{A}}(j)^{-1} \tag{2.49}$$

$$\tilde{\mathcal{A}}(i) := \begin{pmatrix}\frac{1}{T^2}\sum_{t=1}^{T}(\tilde{x}_{it})^2 & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}}\\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{it}} & \hat{\mu}_i^4/180 & \hat{\mu}_i^5/120\\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{it}} & \hat{\mu}_i^5/120 & 9\hat{\mu}_i^6/700\end{pmatrix}, \quad i = 1,\dots,N, \tag{2.50}$$

$$\tilde{\mathcal{B}}(i,j) := \begin{pmatrix}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{jt} & \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{jt}} & \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{jt}}\\ \frac{1}{T^{7/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^2_{jt}} & \hat{\mu}_i^2\hat{\mu}_j^2/180 & \hat{\mu}_i^2\hat{\mu}_j^3/120\\ \frac{1}{T^{9/2}}\sum_{t=1}^{T}\tilde{x}_{it}\widetilde{x^3_{jt}} & \hat{\mu}_i^3\hat{\mu}_j^2/120 & 9\hat{\mu}_i^3\hat{\mu}_j^3/700\end{pmatrix}, \quad i,j = 1,\dots,N. \tag{2.51}$$

Based upon this, defining $\tilde{S}^{+}_{\alpha,\delta,rob} := K_T \tilde{V}^{+}_{\alpha,rob} K_T$ *leads to the robust versions of the "direct" test statistics, i. e.:*

$$W^{+}_{\alpha,\delta,rob} := \left(R\hat{\beta}^{+} - r\right)'\left(R\tilde{S}^{+}_{\alpha,\delta,rob}R'\right)^{-1}\left(R\hat{\beta}^{+} - r\right) \tag{2.52}$$

$$t^{+}_{\alpha,\delta,rob} := \frac{R\hat{\beta}^{+} - r}{\sqrt{R\tilde{S}^{+}_{\alpha,\delta,rob}R'}}. \tag{2.53}$$

Under the null hypothesis the test statistics are asymptotically chi-squared or standard normally distributed, respectively, as $T \to \infty$*.*

Remark 2.7. *In case of individual specific intercepts in* (2.1) *only, also the OLS estimator allows for asymptotically valid standard inference, as noted by West (1988) in the context of linear time series cointegrating regressions. Proper scaling by a consistent estimator of the long-run variance of the errors* u_{it} *suffices. Therefore, in this case one can consider a group-mean OLS estimator:*

$$\hat{\beta} := \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}(i), \tag{2.54}$$

with:

$$\hat{\beta}(i) := \left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}'\right)^{-1} \sum_{t=1}^{T} \tilde{X}_{it}\tilde{y}_{it}, \quad i = 1, \ldots, N. \tag{2.55}$$

Under the assumptions of Proposition 2.2 it holds for $T \to \infty$, conditional upon Δ_i and Σ_i for $i = 1, \ldots, N$, that:

$$H_T^{-1}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{N^2}\sum_{i=1}^{N} \Omega_{u_i u_i}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\right) \tag{2.56}$$

Therefore, exactly as discussed in Corollary 2.2, group-mean OLS based Wald- and t-type test statistics can be defined using two different estimators of the covariance matrix, analogous to using either $\hat{S}^+$ or $\tilde{S}_\alpha^+$, where in both matrices $\hat{\Omega}_{u_i \cdot v_i}$ is replaced by $\hat{\Omega}_{u_i u_i}$ for $i = 1, \ldots, N$. More precisely, constructing:

$$\hat{S} := \frac{1}{N^2}\sum_{i=1}^{N} \hat{\Omega}_{u_i u_i}\left(\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}'\right)^{-1} \tag{2.57}$$

$$\tilde{S}_\alpha := \frac{1}{N^2}\sum_{i=1}^{N} \hat{\Omega}_{u_i u_i} H_T \left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix} 1/12 & 1/12 & 3/40 \\ 1/12 & 4/45 & 1/12 \\ 3/40 & 1/12 & 9/112 \end{pmatrix}\mathcal{D}(\hat{\mu}_i)\right)^{-1} H_T, \tag{2.58}$$

with, as before, $H_T = diag(T^{-3/2}, T^{-5/2}, T^{-7/2})$, $\mathcal{D}(\hat{\mu}_i) = diag(\hat{\mu}_i, \hat{\mu}_i^2, \hat{\mu}_i^3)$ and $\hat{\Omega}_{u_i u_i}$ an estimator of the long-run variance of u_{it}, allows to define corresponding Wald- and (in case $s = 1$) t-type statistics:

$$W := \left(R\hat{\beta} - r\right)'\left(R\hat{S}R'\right)^{-1}\left(R\hat{\beta} - r\right) \tag{2.59}$$

$$t := \frac{R\hat{\beta} - r}{\sqrt{R\hat{S}R'}}, \tag{2.60}$$

and

$$W_\alpha := \left(R\hat{\beta} - r\right)'\left(R\tilde{S}_\alpha R'\right)^{-1}\left(R\hat{\beta} - r\right) \tag{2.61}$$

$$t_\alpha := \frac{R\hat{\beta} - r}{\sqrt{R\tilde{S}_\alpha R'}}. \tag{2.62}$$

Furthermore, similar to Remarks 2.1 and 2.6, cross-section dependence can be accommodated, i. e., the group-mean OLS estimator can also be used to perform robust inference, again in two ways. One variant is given by:

$$W_{rob} := \left(R\hat{\beta} - r\right)'\left(R\hat{S}_{rob}R'\right)^{-1}\left(R\hat{\beta} - r\right) \tag{2.63}$$

$$t_{rob} := \frac{R\hat{\beta} - r}{\sqrt{R\hat{S}_{rob}R'}}, \tag{2.64}$$

with $\hat{S}_{rob}$ similar to $\hat{S}^{+}_{rob}$ as defined in (2.19)*, but with $\hat{\Omega}_{u_i u_j}$ in place of $\hat{\Omega}_{u_i \cdot v_i; u_j \cdot v_j}$. The second possibility resembles the result discussed in Remark 2.6. The corresponding test statistics are given by:*

$$W_{\alpha,rob} := \left(R\hat{\beta} - r\right)' \left(R\tilde{S}_{\alpha,rob}R'\right)^{-1} \left(R\hat{\beta} - r\right) \tag{2.65}$$

$$t_{\alpha,rob} := \frac{R\hat{\beta} - r}{\sqrt{R\tilde{S}_{\alpha,rob}R'}}, \tag{2.66}$$

with $\tilde{S}_{\alpha,rob}$ similar to $\tilde{S}^{+}_{\alpha,rob}$, but with $\hat{\Omega}_{u_i u_j}$ in place of $\hat{\Omega}_{u_i \cdot v_i; u_j \cdot v_j}$ in $\tilde{V}^{+}_{\alpha,rob}$ as defined in (2.46)*. Under the null hypothesis all considered test statistics are asymptotically chi-squared or standard normally distributed, respectively, as $T \to \infty$.*

2.2.3 Zero or Non-Zero Drifts

We are now ready to discuss the "general" case concerning drifts, with drifts present or absent in any cross-section member. It is important to stress again that for using the developed estimators and tests based upon them no knowledge concerning the presence or absence of drifts is required. As in the previous subsection it is convenient to first discuss the case with individual specific intercepts only on the one hand and the case with individual specific intercepts and linear trends on the other hand separately.

In the individual specific intercepts only case, it follows from a combination of the results of Propositions 2.1 and 2.2 that the asymptotic behavior of the group-mean estimator only depends on the individual specific estimators $\hat{\beta}^{+}(i)$ calculated from cross-section members with zero drifts, since these converge at a slower rate than the estimators corresponding to cross-section members with non-zero drifts in the integrated regressor. It is clear that this "sorts out itself" in the limiting distributions and there are no implications for either the definition or the useage of the considered test statistics.

In case of individual specific intercepts and linear trends, Proposition 2.2 shows that the coefficient to the first power of the integrated regressor, β_1, is estimated with (the standard unit root) rate T, irrespective of whether a non-zero drift is present or not. Therefore, the limiting distribution of the first component of $\hat{\beta}^{+}$ will depend upon all cross-section member specific estimates of β_1. For β_2 and β_3, the situation is exactly as in the individual specific intercepts only case, with the limiting distribution only depending upon the individual specific estimators corresponding to cross-section members with zero drifts in the integrated regressor.

For notational convenience only, consider the cross-section members ordered in $i = 1, \ldots, N_0$ cross-section members with zero drifts and $i = N_0 + 1, \ldots, N$ cross-section members with non-zero drifts; noting that N_0 can range from zero (non-zero drifts in all cross-section members) to N (all cross-section members with zero drifts in x_{it}). Furthermore, define the following scaling matrices:

$$Q_T := \begin{cases} G_T & \text{if } N_0 > 0 \\ H_T & \text{if } N_0 = 0 \end{cases} \quad \text{and} \quad R_T := \begin{cases} G_T & \text{if } N_0 > 0 \\ K_T & \text{if } N_0 = 0 \end{cases}. \tag{2.67}$$

Proposition 2.3. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i \in \mathbb{R}$, $i = 1, \ldots, N$ *and let Assumptions 2.1, 2.2 for* $\tilde{X}_{it}^{o}$ *and 2.3 be in place.*

(i) In case individual specific intercepts but no individual specific linear trends are included in (2.1)*, it holds for* $T \to \infty$*, conditional upon* Δ_i*,* Σ_i *and* $W_{v_i}(r)$ *for* $i = 1, \ldots, N$ *that:*

$$Q_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{N_0}^{+}\right), \tag{2.68}$$

with:

$$V_{N_0}^{+} := \begin{cases} \frac{1}{N^2}\sum_{i=1}^{N_0} \Omega_{u_i \cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})'\mathbf{dr}\right)^{-1} & \text{if } N_0 > 0 \\ V_{\alpha}^{+} & \text{if } N_0 = 0 \end{cases}. \tag{2.69}$$

(ii) In case individual specific intercepts and linear trends are included in (2.1)*, it holds for* $T \to \infty$*, conditional upon* Δ_i*,* Σ_i *and* $W_{v_i}(r)$ *for* $i = 1, \ldots, N$ *that:*

$$R_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, V_{N_0}^{+}\right), \tag{2.70}$$

with:

$$V_{N_0}^{+} := \begin{cases} \frac{1}{N^2}\sum_{i=1}^{N_0} \Omega_{u_i \cdot v_i}\left(\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})'\mathbf{dr}\right)^{-1} & \\ \quad + \frac{1}{N^2}\sum_{i=N_0+1}^{N} \Omega_{u_i \cdot v_i}\begin{pmatrix} \left(\int_0^1 L_i(r)L_i(r)'dr\right)^{-1}_{[1,1]} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{if } N_0 > 0 \\ V_{\alpha,\delta}^{+} & \text{if } N_0 = 0 \end{cases}, \tag{2.71}$$

with $_{[1,1]}$ *indicating the* $(1,1)$ *element of the (*3×3 *inverted) matrix.*

The second term in the covariance matrix $V_{N_0}^{+}$ in item (ii) in case $N_0 > 0$ reflects the above-mentioned fact that the coefficient to the first power of the integrated regressor is estimated at rate T irrespective of whether the drift is zero or non-zero – as in either case linear detrending removes a potential deterministic linear trend from the corresponding regressor. The asymptotic distribution immediately leads to Wald- and t-type test statistics.

Corollary 2.3. *Let the data be generated by* (2.1) *and* (2.2) *with* $\mu_i \in \mathbb{R}$, $i = 1, \ldots, N$ *and let Assumptions 2.1, 2.2 for* $\tilde{X}_{it}^{o}$ *and 2.3 be in place. Consider* s *linearly independent restrictions collected in* $H_0 : R\beta = r$ *with* $R \in \mathbb{R}^{s \times 3}$*,* $r \in \mathbb{R}^{s}$ *and assume that there exists a non-singular matrix* $G_R \in \mathbb{R}^{s \times s}$ *and a matrix* $R^* \in \mathbb{R}^{s \times 3}$ *of rank* s *such that* $\lim_{T\to\infty} G_R R Q_T = R^*$ *(in the individual specific intercepts only case) or* $\lim_{T\to\infty} G_R R R_T = R^*$ *(in the individual specific intercepts and linear trends case). In both, the individual specific intercepts only and the individual* *specific intercepts and linear trends case, the Wald- and (in case $s = 1$) t-type statistics:*

$$W = \left(R\hat{\beta}^{+} - r\right)' \left(R\hat{S}^{+} R'\right)^{-1} \left(R\hat{\beta}^{+} - r\right), \tag{2.72}$$

$$t = \frac{R\hat{\beta}^{+} - r}{\sqrt{R\hat{S}^{+} R'}} \tag{2.73}$$

already defined in (2.16) *and* (2.17)*, are under the null hypothesis chi-squared distributed with s degrees of freedom and standard normally distributed, respectively, as $T \to \infty$.*

Remark 2.8. *As in the previous subsections, cf. Remarks 2.1 and 2.6, the group-mean FM-OLS estimator remains consistent with a zero mean (conditional) normal limiting distribution in case of cross-section dependencies; with the assumptions correspondingly adjusted. The key input for performing "robust" inference is again a consistent estimator of the covariance matrix of the asymptotic distribution.*

In case individual specific intercepts only are included (2.1)*, the asymptotic covariance matrix is in case of cross-section dependence given by:*

$$V^{+}_{N_0,rob} := \begin{cases} \frac{1}{N^2} \sum_{i,j=1}^{N_0} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \tilde{M}_{ii}^{-1} \tilde{M}_{ij} \tilde{M}_{jj}^{-1} & \text{if } N_0 > 0 \\ V^{+}_{\alpha,rob} & \text{if } N_0 = 0 \end{cases}. \tag{2.74}$$

In case both individual specific intercepts and linear trends are included in (2.1)*, the asymptotic covariance matrix is given by:*

$$V^{+}_{N_0,rob} := \begin{cases} \frac{1}{N^2} \sum_{i,j=1}^{N_0} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \tilde{M}_{ii}^{-1} \tilde{M}_{ij} \tilde{M}_{jj}^{-1} & \\ \quad + \frac{1}{N^2} \sum_{i,j=N_0+1}^{N} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \begin{pmatrix} \mathcal{C}(i,j)_{[1,1]} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \\ \quad + \frac{1}{N^2} \sum_{i=1}^{N_0} \sum_{j=N_0+1}^{N} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \begin{pmatrix} \mathcal{F}(i,j)_{[1,1]} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \text{if } N_0 > 0 \\ \quad + \frac{1}{N^2} \sum_{i=N_0+1}^{N} \sum_{j=1}^{N_0} \Omega_{u_i \cdot v_i; u_j \cdot v_j} \begin{pmatrix} \mathcal{K}(i,j)_{[1,1]} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \\ V^{+}_{\alpha,\delta,rob} & \text{if } N_0 = 0 \end{cases}, \tag{2.75}$$

with:

$$\mathcal{F}(i,j) := \left(\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})' \mathbf{dr}\right)^{-1} \int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \mathbf{L_j}(\mathbf{r})' \mathbf{dr} \left(\int_{\mathbf{0}}^{\mathbf{1}} \mathbf{L_j}(\mathbf{r}) \mathbf{L_j}(\mathbf{r})' \mathbf{dr}\right)^{\mathbf{-1}}, \tag{2.76}$$

$$\mathcal{K}(i,j) := \left(\int_0^1 L_i(r) L_i(r)' dr\right)^{-1} \int_0^1 L_i(r) \tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r})' \mathbf{dr} \left(\int_{\mathbf{0}}^{\mathbf{1}} \tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r}) \tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r})' \mathbf{dr}\right), \tag{2.77}$$

for $i, j = 1, \ldots, N$.

For performing robust inference, however, the fact that the asymptotic covariance matrices are case-dependent with respect to both N_0 and whether or not individual specific linear trends are *included in* (2.1)*, has no consequences. The robust test statistics* W_{rob}^{+} *and* t_{rob}^{+} *defined in* (2.20) *and* (2.21)*, using* $\hat{S}_{rob}^{+}$ *defined in* (2.19)*, lead to chi-squared respectively standard normal inference under the null hypothesis as* $T \to \infty$*. This follows using similar arguments as in Proposition 2.3 and Corollary 2.3.*

We abstain from a detailed discussion of constructing test statistics based on "direct" estimators of the covariance matrix. Doing so would in practice necessitate knowledge concerning the presence or absence of non-zero drifts in the integrated regressors in the individual cross-section members. Whilst this knowledge, as unlikely as this may be, could in some applications indeed be available and one could construct individual specific "direct" estimators, we do not provide – notationally more cumbersome rather than mathematically more complicated – details here. For the same reason we also abstain from considering OLS rather than FM-OLS estimation in the cross-section members with non-zero drifts and do not define a *mixed* OLS-FM-OLS group-mean estimator. The corresponding analysis is again notationally more cumbersome rather than mathematically more complex.

2.3 Finite Sample Performance

We generate, commencing from de Jong and Wagner (2022), data according to (2.1) and (2.2), i. e.:

$$y_{it} = \alpha_i + \delta_i t + x_{it}\beta_1 + x_{it}^2\beta_2 + x_{it}^3\beta_3 + u_{it}, \tag{2.78}$$

$$x_{it} = \mu_i + x_{i,t-1} + v_{it}, \quad x_{i0} = 0, \tag{2.79}$$

with slope parameters $\beta_1 = 5$, $\beta_2 = -3$ and $\beta_3 = 0.3$. The regression errors u_{it} and v_{it} are generated as:

$$u_{it} = \rho_{1i}u_{i,t-1} + \varepsilon_{it} + \rho_{2i}\nu_{it}, \tag{2.80}$$

$$v_{it} = 0.1\left(\nu_{it} + 0.5\nu_{i,t-1}\right), \tag{2.81}$$

with $(\varepsilon_{1t}, \ldots, \varepsilon_{Nt})' \sim \mathcal{N}(0, \Sigma)$ and $(\nu_{1t}, \ldots, \nu_{Nt})' \sim \mathcal{N}(0, \Sigma)$, i.i.d. across $t = 0, 1, \ldots, T$, where:

$$\Sigma = \begin{pmatrix} 1 & \rho_3 & \ldots & \rho_3 \\ \rho_3 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_3 \\ \rho_3 & \ldots & \rho_3 & 1 \end{pmatrix}. \tag{2.82}$$

The parameters ρ_{1i} and ρ_{2i} control the level of serial correlation in the error terms u_{it} and the extent of regressor endogeneity, respectively, whereas the parameter ρ_3 controls the extent of cross-section dependence. The parameters ρ_{i1}, ρ_{i2} are cross-sectionally i.i.d. and independent of $(\varepsilon_{it}, \nu_{it})'$, $t = 1, \ldots, T$. In particular we consider $\rho_{1i} = \rho_1 + \mathcal{U}_{1i}$ and $\rho_{2i} = \rho_2 + \mathcal{U}_{2i}$ with $\mathcal{U}_{1i}, \mathcal{U}_{2i}$ i.i.d. uniform random variables over the interval $[-0.05, 0.05]$, with $\rho_1, \rho_2 \in \{0, 0.3, 0.6, 0.9\}$.[11] Furthermore, we consider also $\rho_3 \in \{0, 0.3, 0.6, 0.9\}$. The individual effects α_i are i.i.d. $\mathcal{N}(-45, 5)$ and independent of all other random quantities. For the individual specific time trends we consider two cases: (i) $\delta_i = 0$ for $i = 1, \ldots, N$ and (ii) δ_i i.i.d. $\mathcal{N}(-0.01, 0.01)$, independent of all other random quantities. In the former case the variables are demeaned and in the second case the variables are demeaned and linearly detrended for the construction of the estimators, compare (2.10) and (2.11).

With respect to drifts, μ_i, we consider three cases: Two boundary cases, one with all drift parameters equal to zero, i. e., $\mu_i = \mu = 0$, and one with all drift parameters equal to $\mu_i = \mu = 0.02$.[12] Furthermore, we consider an "intermediate case", with half of the individual-specific drifts equal to zero and the other half equal to 0.02. The simulation setting covers all combinations of $N \in \{10, 20, 100\}$ and $T \in \{100, 250, 500\}$. For every considered setting, the number of replications is 5,000 and all test decisions are performed at the 5% nominal level. The reported results rely upon long-run covariance estimation using the Bartlett kernel in conjunction with the data-dependent bandwidth rule of Andrews (1991). As indicated already in the introduction, the Supplementary Material contains a number of additional tables and figures.

We start by considering Bias and root mean squared error (RMSE) of three estimators: The group-mean OLS estimator, labelled $\hat{\beta}$, the group-mean FM-OLS estimator $\hat{\beta}^+$ and the pooled FM-OLS estimator of de Jong and Wagner (2022), labelled $\hat{\beta}^+_{\mathrm{P}}$.[13] In general, see as an illustration the results for β_1, with $\mu_i \neq 0$ for $i = 1, \ldots, N$, in Tables 2.1 and 2.2, the presence of individual specific trends adversely affects estimator performance, both in terms of bias and RMSE. This almost necessarily implies, as will be seen also below, a corresponding detrimental impact also on test performance.[14] As expected, increasing the sample size, either the cross-section dimension N or (with a stronger positive effect) the time series dimension T leads to improved performance. As also expected, increasing any of the ρ-parameters that govern error serial correlation, regressor endogeneity or cross-section dependence, respectively, leads to performance deterioration. In this respect it turns out that RMSE is more strongly affected by cross-section dependence than bias, which does not react strongly to cross-section dependence. By construction, as the pooled FM-OLS estimator estimates only one set of slope coefficients, the pooled FM-OLS estimator mostly outperforms the group-mean FM-OLS estimator both in terms of bias and RMSE. Only for β_1 in the individual specific intercepts only case, see Tables 2.1 and 2.2, the group-mean FM-OLS estimator leads in several case to smaller bias than the pooled FM-OLS estimator (more pronounced for large ρ-values and smaller sample sizes), albeit in conjunction with higher RMSE. However, this is not the case

[11] The addition of cross-sectionally i.i.d. random variables to the coefficients ρ_1 and ρ_2 is a simple way of generating data in a random linear process fashion. Considering non-random ρ_{1i} and ρ_{2i} leads, as expected, to very similar results.
Our way of introducing cross-section dependence is inspired by Wagner and Hlouskova (2010) who consider three specifications for modelling cross-section dependence. We consider their constant correlation setting.

[12] Setting all non-zero drift parameters equal to 0.02 is for simplicity only. The results are very similar when the non-zero drifts are independently drawn from the interval $[0.01, 0.03]$. The point value for $\mu = 0.02$ and the interval $[0.01, 0.03]$ are inspired by the arithmetic means of the annual GDP per capita growth rates for 19 countries in the long data set analyzed in Section 2.4. The country-specific arithmetic means range from 0.013 to 0.024, and the arithmetic mean over all countries of the country-specific mean growth rates is equal to 0.018.

[13] de Jong and Wagner (2022) do not consider the case of individual specific linear trends but consider time effects, compare Remark 2.2. It is straightforward to adjust – and implement – the pooled FM-OLS estimator to include individual specific (linear) time trends, using demeaned and linearly detrended observations.

[14] Tables 7 and 8 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$.

Table 2.1: Bias and RMSE of the estimators of β_1 in the individual specific intercepts only case with non-zero drifts.

		$N = 10$			$N = 20$			$N = 100$		
T	ρ_1, ρ_2	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$
					Bias, $\rho_3 = 0$					
100	0	−0.01	−0.01	−0.00	−0.00	0.00	0.00	0.00	0.00	0.00
	0.3	0.23	0.04	0.01	0.24	0.05	0.01	0.24	0.05	0.01
	0.6	0.88	0.21	0.10	0.90	0.22	0.09	0.90	0.22	0.08
	0.9	3.23	0.54	0.94	3.24	0.52	0.95	3.23	0.57	0.89
250	0	0.00	0.00	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00
	0.3	0.08	0.01	0.00	0.08	0.01	0.00	0.08	0.01	0.00
	0.6	0.36	0.06	0.03	0.36	0.06	0.03	0.36	0.06	0.02
	0.9	1.79	0.24	0.33	1.78	0.24	0.31	1.79	0.25	0.30
500	0	0.00	0.00	−0.00	−0.00	−0.00	−0.00	0.00	0.00	−0.00
	0.3	0.03	0.00	0.00	0.03	0.00	0.00	0.03	0.00	0.00
	0.6	0.14	0.02	0.01	0.14	0.02	0.01	0.14	0.02	0.01
	0.9	0.86	0.13	0.11	0.86	0.12	0.11	0.87	0.14	0.10
					Bias, $\rho_3 = 0.9$					
100	0	−0.02	−0.02	−0.01	−0.03	−0.02	−0.02	0.00	0.00	0.01
	0.3	0.21	0.02	0.00	0.20	0.01	−0.01	0.24	0.05	0.03
	0.6	0.86	0.18	0.11	0.86	0.17	0.10	0.90	0.23	0.15
	0.9	3.23	0.39	0.68	3.22	0.35	0.72	3.28	0.47	0.90
250	0	−0.00	−0.00	−0.00	−0.01	−0.01	−0.00	−0.00	−0.00	0.00
	0.3	0.08	0.01	0.00	0.08	−0.00	0.00	0.08	0.01	0.01
	0.6	0.36	0.06	0.03	0.35	0.05	0.03	0.36	0.07	0.04
	0.9	1.78	0.22	0.23	1.77	0.17	0.22	1.81	0.22	0.27
500	0	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	0.00	0.00	−0.00
	0.3	0.03	0.00	0.00	0.03	−0.00	−0.00	0.03	0.00	−0.00
	0.6	0.14	0.02	0.01	0.13	0.01	0.00	0.14	0.02	0.01
	0.9	0.84	0.10	0.09	0.85	0.10	0.07	0.84	0.12	0.09
					RMSE, $\rho_3 = 0$					
100	0	0.47	0.48	0.11	0.33	0.33	0.07	0.15	0.15	0.03
	0.3	0.66	0.63	0.16	0.50	0.44	0.10	0.31	0.20	0.04
	0.6	1.32	0.97	0.29	1.14	0.71	0.19	0.95	0.37	0.10
	0.9	4.03	2.69	1.37	3.66	1.95	1.14	3.32	1.04	0.93
250	0	0.14	0.14	0.05	0.10	0.10	0.03	0.04	0.04	0.01
	0.3	0.21	0.19	0.07	0.16	0.14	0.05	0.10	0.06	0.02
	0.6	0.50	0.33	0.13	0.44	0.24	0.09	0.38	0.12	0.04
	0.9	2.15	1.16	0.65	1.97	0.85	0.49	1.83	0.46	0.34
500	0	0.05	0.05	0.03	0.04	0.04	0.02	0.02	0.02	0.01
	0.3	0.08	0.08	0.04	0.06	0.05	0.02	0.04	0.02	0.01
	0.6	0.20	0.13	0.07	0.17	0.09	0.04	0.15	0.05	0.02
	0.9	1.06	0.54	0.32	0.95	0.39	0.23	0.89	0.22	0.13
					RMSE, $\rho_3 = 0.9$					
100	0	1.10	1.11	0.63	1.05	1.06	0.57	1.03	1.04	0.53
	0.3	1.50	1.46	0.84	1.41	1.40	0.77	1.38	1.36	0.71
	0.6	2.54	2.25	1.34	2.40	2.13	1.23	2.32	2.05	1.14
	0.9	6.46	6.37	4.13	6.14	5.82	3.73	5.95	5.58	3.42
250	0	0.35	0.35	0.24	0.35	0.35	0.22	0.35	0.35	0.22
	0.3	0.50	0.49	0.34	0.50	0.49	0.31	0.50	0.48	0.30
	0.6	0.95	0.82	0.57	0.96	0.82	0.53	0.95	0.81	0.51
	0.9	3.48	2.87	2.04	3.50	2.85	1.93	3.44	2.79	1.84
500	0	0.14	0.14	0.11	0.14	0.14	0.10	0.14	0.14	0.10
	0.3	0.21	0.20	0.15	0.21	0.20	0.15	0.21	0.20	0.15
	0.6	0.42	0.35	0.27	0.41	0.35	0.25	0.43	0.35	0.25
	0.9	1.84	1.41	1.05	1.80	1.38	1.01	1.85	1.40	1.00

Note: The column labels $\hat{\beta}_1$, $\hat{\beta}_1^+$ and $\hat{\beta}_{P,1}^+$ denote the group-mean OLS estimator, the group-mean FM-OLS estimator and the pooled FM-OLS estimator, respectively, of β_1.

for β_2 and β_3, see Tables 9 to 16 in the Supplementary Material, and should thus not be over-interpreted. Increasing values of ρ_1, ρ_2 lead to performance advantages of group-mean FM-OLS over group-mean OLS with – as expected – basically no differences between these two estimators

Table 2.2: Bias and RMSE of the estimators of β_1 in the individual specific intercepts and linear trends case with non-zero drifts.

		$N = 10$			$N = 20$			$N = 100$		
T	ρ_1, ρ_2	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$	$\hat{\beta}_1$	$\hat{\beta}_1^+$	$\hat{\beta}_{P,1}^+$
					Bias, $\rho_3 = 0$					
100	0	−0.01	−0.01	−0.00	−0.00	0.00	−0.00	0.00	0.00	0.00
	0.3	0.31	0.10	0.03	0.32	0.11	0.03	0.32	0.11	0.02
	0.6	1.23	0.67	0.30	1.25	0.69	0.27	1.24	0.69	0.25
	0.9	4.69	3.96	2.91	4.73	3.97	2.82	4.73	3.99	2.74
250	0	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00
	0.3	0.13	0.03	0.01	0.13	0.03	0.01	0.13	0.03	0.01
	0.6	0.58	0.24	0.10	0.57	0.23	0.09	0.57	0.23	0.08
	0.9	3.06	2.22	1.34	3.02	2.19	1.25	3.04	2.21	1.18
500	0	0.00	0.00	0.00	−0.00	−0.00	−0.00	0.00	0.00	0.00
	0.3	0.06	0.01	0.00	0.06	0.01	0.00	0.07	0.01	0.00
	0.6	0.30	0.10	0.04	0.30	0.09	0.03	0.30	0.10	0.03
	0.9	1.92	1.23	0.60	1.92	1.23	0.56	1.94	1.25	0.53
					Bias, $\rho_3 = 0.9$					
100	0	−0.03	−0.03	−0.02	−0.03	−0.03	−0.02	0.01	0.01	0.01
	0.3	0.29	0.08	0.04	0.28	0.07	0.03	0.32	0.12	0.07
	0.6	1.21	0.65	0.46	1.20	0.63	0.44	1.25	0.71	0.49
	0.9	4.75	4.01	3.53	4.73	3.94	3.46	4.75	4.09	3.56
250	0	−0.00	0.00	−0.00	−0.01	−0.01	−0.00	−0.00	0.00	0.00
	0.3	0.13	0.03	0.02	0.12	0.02	0.01	0.13	0.03	0.02
	0.6	0.58	0.24	0.16	0.56	0.22	0.15	0.58	0.24	0.17
	0.9	3.07	2.25	1.84	3.00	2.15	1.76	3.11	2.27	1.82
500	0	−0.00	−0.00	−0.00	−0.01	−0.01	−0.01	0.00	0.00	0.00
	0.3	0.06	0.01	0.01	0.06	0.00	−0.00	0.06	0.01	0.01
	0.6	0.29	0.09	0.07	0.30	0.09	0.06	0.29	0.10	0.07
	0.9	1.89	1.21	0.94	1.94	1.23	0.93	1.91	1.25	0.93
					RMSE, $\rho_3 = 0$					
100	0	0.49	0.50	0.14	0.34	0.35	0.09	0.15	0.15	0.04
	0.3	0.72	0.66	0.20	0.55	0.47	0.13	0.38	0.23	0.06
	0.6	1.59	1.20	0.45	1.44	0.99	0.35	1.28	0.75	0.27
	0.9	5.19	4.54	3.11	4.99	4.29	2.91	4.78	4.06	2.75
250	0	0.15	0.15	0.07	0.10	0.10	0.04	0.05	0.05	0.02
	0.3	0.25	0.21	0.09	0.19	0.15	0.06	0.14	0.07	0.03
	0.6	0.69	0.42	0.19	0.63	0.34	0.14	0.59	0.26	0.09
	0.9	3.31	2.51	1.52	3.14	2.33	1.34	3.06	2.24	1.20
500	0	0.06	0.06	0.03	0.04	0.04	0.02	0.02	0.02	0.01
	0.3	0.11	0.08	0.05	0.09	0.06	0.03	0.07	0.03	0.01
	0.6	0.34	0.17	0.09	0.32	0.14	0.07	0.30	0.11	0.04
	0.9	2.05	1.39	0.72	1.98	1.31	0.63	1.95	1.27	0.55
					RMSE, $\rho_3 = 0.9$					
100	0	1.15	1.16	0.70	1.09	1.11	0.64	1.06	1.07	0.60
	0.3	1.57	1.52	0.95	1.49	1.46	0.88	1.44	1.41	0.82
	0.6	2.77	2.44	1.60	2.64	2.32	1.49	2.54	2.23	1.39
	0.9	7.19	6.73	5.27	6.95	6.39	4.97	6.75	6.27	4.83
250	0	0.37	0.37	0.27	0.36	0.37	0.26	0.37	0.37	0.25
	0.3	0.54	0.52	0.38	0.54	0.51	0.36	0.53	0.51	0.35
	0.6	1.11	0.90	0.67	1.11	0.89	0.64	1.11	0.90	0.62
	0.9	4.39	3.67	2.90	4.34	3.59	2.79	4.37	3.62	2.75
500	0	0.16	0.16	0.13	0.15	0.15	0.12	0.16	0.16	0.12
	0.3	0.24	0.22	0.18	0.23	0.22	0.17	0.24	0.22	0.17
	0.6	0.54	0.40	0.32	0.53	0.40	0.31	0.54	0.41	0.31
	0.9	2.68	2.04	1.59	2.67	2.01	1.54	2.69	2.03	1.54

Note: See note to Table 2.1.

for $\rho_1, \rho_2 = 0$.

The (asymptotic) implications of the absence or presence of drifts manifest themselves also in the finite sample results. In the individual specific intercepts only case, bias and RMSE of all components of the OLS and FM-OLS group-mean estimators of β are smaller in the presence than in the absence of drifts; compare, e. g., for β_1 Table 2.1 with Table 7 in the Supplementary Material. Exactly in line with asymptotic theory (Proposition 2.3), bias and RMSE of the OLS and FM-OLS group-mean estimators of β_1 are not affected by the absence or presence of drifts in the individual specific intercepts and linear trends case, compare Table 2.2 with Table 8 in the Supplementary Material.

To assess test performance we consider in total five different test statistics evaluated under the null hypothesis by means of empirical null rejection probabilities and under a sequence of 20 alternatives by means of "size-corrected" power. We consider two test statistics based on the group-mean OLS estimator: The first is a *textbook* version of a group-mean OLS estimator based test, labelled W_{TB}, using in $\hat{S}$, as defined in (2.57), instead of $\hat{\Omega}_{u_i u_i}$ a textbook variance estimator given by $\hat{\sigma}^2_{u_i} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}^2_{it}$. This test serves as a "textbook" OLS test benchmark and leads to asymptotically valid inference only when $\rho_{i1} = \rho_{i2} = \rho_3 = 0$ for $i = 1, \ldots, N$. The second group-mean OLS based test statistic is W_{rob} as defined in (2.63). As discussed in Remark 2.7, asymptotic validity of this test for all values of the ρ-parameters hinges critically upon drifts being present in all cross-section members, which in practice is almost certainly unknown. We, of course, consider both standard and robust inference based on the group-mean OLS estimator, i. e., W^+ as defined in (2.16) and W^+_{rob} as defined in (2.20). Finally, for comparison, we also include the Wald-type test based on the pooled estimator of de Jong and Wagner (2022), labelled as W^+_{P}.[15] Specifically, we consider the null hypothesis $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$. To assess power, we generate data for a sequence of 20 alternative values for the vector β. Reflecting the different convergence rates of the components of β, we choose (including also the null values) 21 equidistant values for β_1 in the interval $[5, 7]$, for β_2 in the interval $[-3, -2]$ and for β_3 in the interval $[0.3, 0.7]$. The selection of tests does not include the "direct" tests as they do not provide any extra value added. The simulations have shown that for small values of T they are very conservative, with empirical null rejection probabilities often very close to zero, and for large values of T their performance is (as expected) very similar to the performance of their "non-direct" counterparts.

As indicated already above, also the tests – as an immediate consequence of estimator performance – generally perform better in the individual specific intercepts only case than when also linear trends are included. This effect becomes more pronounced for increasing ρ-parameters, see and compare, e. g., Tables 2.3 and 2.4 for the results in case $\mu_i \neq 0$, $i = 1, \ldots, N$.[16] Many of the observed features are in line with expectations: First, size distortions increase with increasing ρ-parameters. This effect occurs most visibly for W_{TB}, which, as mentioned, only leads to asymptotically valid inference in case all ρ-parameters are equal to zero. If N is *large* compared to T, we observe the phenomenon of "size-divergence" (see, e. g., Wagner and Hlouskova, 2010), i. e., increasing size distortions for increasing N and fixed (small) T.[17] The (relative) behavior of W^+ and W^+_{rob} is also as expected: Both tests are, by construction, less adversely affected than, e. g., W when ρ_{i1}, ρ_{i2} increase, at least for small values of ρ_3. Increasing ρ_3 leads to smaller size distortions – partly

[15] For the pooled estimator the so-called "standard" covariance estimator is used, see de Jong and Wagner (2022) for details.

[16] Tables 17 and 18 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$. The absence or presence of drifts exhibits very limited impact on the null rejection probabilities.

[17] The test based on the pooled estimator of de Jong and Wagner (2022) is particularly strongly affected by size-divergence.

substantially smaller size distortions – of W^+_{rob} than of W^+. This indicates that "robust" inference indeed works. The group-mean OLS-based robust test W_{rob} is much less affected by increasing ρ_3 than one would expect, with this being driven by our DGP that generates strong contemporaneous cross-section dependence for large values of ρ_3. The test based on the pooled FM-OLS estimator of de Jong and Wagner (2022) is very strongly adversely affected by cross-section dependence, visible already for $\rho_3 = 0.3$. W^+_{P} is strongly outperformed by W^+_{rob} and even by W_{rob} in case of cross-section dependence. Altogether, in case of unknown forms of error serial correlation, regressor endogeneity and cross-section dependence, W^+_{rob} is the overall best performing test with the smallest size distortions under the null hypothesis. W^+_{rob} performs similarly to W^+ even when all ρ-parameters are equal to zero and is thus, from the null rejection probabilities perspective, the best choice.

We close the simulation section by looking at "size-corrected" power. Figures 2.1 and 2.2 display results for $T = 100$, $\rho_1, \rho_2 = 0.6$ and $\mu_i \neq 0$, $i = 1, \ldots, N$ for the individual specific intercepts only and the individual specific intercepts and linear trends cases, respectively.[18] Some observations emerge: First, whilst the empirical null rejection probabilities are hardly affected by the absence or presence of drifts, size corrected power is higher when all drifts are non-zero. Second, larger values of ρ_{i1}, ρ_{i2} lead to smaller size-corrected power. Third, size-corrected power increases unequivocally with an increasing time dimension T, whereas increasing N has only minor impact on size-corrected power in case of cross-section dependence. Fourth, effectively by construction, the test based on the pooled estimator of de Jong and Wagner (2022) exhibits the highest size-corrected power (which, however, has to be seen in conjunction with the very large size distortions in case of cross-section dependence). Fifth, size-corrected power is often the second highest for W^+_{rob} and is for large values of ρ_3 closely followed by size-corrected power of W_{rob}. These findings, in conjunction with the behavior under the null hypothesis, lead to the conclusion that in applications, where one typically does not know the dependence structure, it is the best choice to use W^+_{rob}, i. e., the robust version of the group-mean FM-OLS based test statistic.[19]

2.4 An Illustration: The Environmental Kuznets Curve for Carbon Dioxide Emissions

In this section we briefly illustrate the group-mean FM-OLS estimator as well as inference based upon it by estimating environmental Kuznets curves (EKCs) for carbon dioxide (CO_2) emissions. The dependent variable is the logarithm of per capita CO_2 emissions and the explanatory variables are log per capita GDP and its powers. We consider both the quadratic and the cubic specification as well as the inclusion of individual specific intercepts only and of both individual specific intercepts and linear trends. Long-run covariance estimation uses the Bartlett kernel and the Andrews (1991) bandwidth selection rule.

We use exactly the same data as de Jong and Wagner (2022). These are the *long data set* with

[18] Figures 4 and 5 in the Supplementary Material provide the corresponding results for $\mu_i = 0$ for $i = 1, \ldots, N$.

[19] Note that even in the absence of cross-section dependence, i. e., $\rho_3 = 0$, the "non-robust" version of the Wald-type test, W^+, does not have larger size-corrected power than W^+_{rob}.

Table 2.3: Empirical null rejection probabilities of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ in the individual specific intercepts only case with non-zero drifts.

		$N = 10$					$N = 20$					$N = 100$				
T	ρ_1, ρ_2	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P
									$\rho_3 = 0$							
100	0	0.06	0.07	0.08	0.08	0.06	0.07	0.08	0.09	0.09	0.07	0.06	0.07	0.07	0.07	0.06
	0.3	0.22	0.11	0.12	0.12	0.11	0.25	0.13	0.13	0.13	0.11	0.40	0.25	0.11	0.11	0.11
	0.6	0.58	0.20	0.15	0.15	0.18	0.67	0.27	0.15	0.14	0.20	0.95	0.68	0.16	0.16	0.35
	0.9	0.89	0.29	0.26	0.25	0.55	0.94	0.32	0.24	0.23	0.71	1.00	0.54	0.34	0.23	0.99
250	0	0.05	0.06	0.07	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.06	0.06	0.05
	0.3	0.26	0.11	0.10	0.11	0.09	0.27	0.12	0.10	0.10	0.08	0.47	0.25	0.09	0.09	0.08
	0.6	0.69	0.22	0.13	0.13	0.12	0.77	0.32	0.12	0.12	0.12	0.99	0.81	0.14	0.12	0.16
	0.9	0.97	0.39	0.23	0.22	0.30	0.99	0.47	0.21	0.18	0.36	1.00	0.86	0.29	0.17	0.76
500	0	0.05	0.05	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05
	0.3	0.25	0.09	0.09	0.09	0.08	0.28	0.10	0.09	0.09	0.08	0.45	0.22	0.08	0.08	0.07
	0.6	0.69	0.19	0.11	0.11	0.10	0.77	0.27	0.10	0.10	0.10	0.99	0.77	0.12	0.10	0.10
	0.9	0.98	0.40	0.20	0.20	0.19	0.99	0.51	0.20	0.18	0.22	1.00	0.92	0.28	0.13	0.41
									$\rho_3 = 0.3$							
100	0	0.13	0.07	0.15	0.08	0.21	0.18	0.08	0.21	0.08	0.33	0.41	0.07	0.43	0.08	0.70
	0.3	0.33	0.12	0.22	0.13	0.28	0.40	0.14	0.27	0.13	0.41	0.67	0.17	0.49	0.13	0.75
	0.6	0.66	0.21	0.26	0.17	0.36	0.74	0.24	0.31	0.17	0.49	0.94	0.41	0.52	0.18	0.81
	0.9	0.91	0.33	0.37	0.31	0.63	0.94	0.35	0.41	0.29	0.75	0.99	0.44	0.62	0.32	0.94
250	0	0.18	0.06	0.20	0.06	0.26	0.28	0.06	0.29	0.06	0.41	0.61	0.06	0.62	0.06	0.79
	0.3	0.44	0.11	0.27	0.11	0.32	0.55	0.12	0.36	0.11	0.47	0.81	0.13	0.67	0.10	0.82
	0.6	0.78	0.21	0.31	0.15	0.36	0.86	0.25	0.40	0.15	0.50	0.97	0.36	0.69	0.14	0.84
	0.9	0.97	0.38	0.42	0.27	0.51	0.98	0.40	0.47	0.24	0.64	1.00	0.50	0.72	0.25	0.91
500	0	0.25	0.06	0.25	0.06	0.32	0.40	0.06	0.41	0.06	0.51	0.76	0.06	0.76	0.06	0.87
	0.3	0.52	0.09	0.31	0.10	0.37	0.65	0.10	0.46	0.10	0.55	0.88	0.11	0.79	0.10	0.89
	0.6	0.82	0.17	0.35	0.13	0.39	0.89	0.19	0.50	0.12	0.57	0.97	0.25	0.81	0.13	0.89
	0.9	0.98	0.35	0.46	0.23	0.46	0.99	0.39	0.58	0.24	0.63	1.00	0.45	0.83	0.22	0.91
									$\rho_3 = 0.6$							
100	0	0.30	0.08	0.33	0.09	0.40	0.43	0.07	0.45	0.08	0.58	0.76	0.07	0.78	0.08	0.89
	0.3	0.52	0.14	0.41	0.15	0.48	0.62	0.13	0.52	0.13	0.64	0.88	0.14	0.80	0.14	0.91
	0.6	0.78	0.22	0.45	0.21	0.54	0.85	0.23	0.56	0.19	0.69	0.97	0.28	0.82	0.20	0.93
	0.9	0.95	0.37	0.56	0.38	0.72	0.96	0.38	0.63	0.38	0.84	0.99	0.43	0.87	0.42	0.97
250	0	0.41	0.06	0.43	0.07	0.48	0.56	0.06	0.58	0.06	0.66	0.86	0.06	0.86	0.07	0.93
	0.3	0.64	0.11	0.49	0.12	0.53	0.77	0.10	0.64	0.11	0.71	0.94	0.12	0.88	0.11	0.94
	0.6	0.87	0.20	0.54	0.17	0.57	0.93	0.21	0.67	0.16	0.74	0.99	0.25	0.89	0.17	0.95
	0.9	0.98	0.38	0.63	0.32	0.68	0.99	0.38	0.73	0.32	0.81	1.00	0.42	0.91	0.33	0.96
500	0	0.50	0.06	0.51	0.06	0.55	0.68	0.06	0.68	0.06	0.73	0.92	0.06	0.93	0.06	0.96
	0.3	0.73	0.10	0.56	0.10	0.60	0.84	0.10	0.72	0.10	0.77	0.97	0.10	0.94	0.10	0.96
	0.6	0.91	0.16	0.59	0.14	0.62	0.95	0.16	0.75	0.14	0.78	0.99	0.17	0.95	0.14	0.97
	0.9	0.99	0.34	0.69	0.26	0.67	1.00	0.35	0.80	0.28	0.82	1.00	0.35	0.96	0.26	0.98
									$\rho_3 = 0.9$							
100	0	0.66	0.09	0.68	0.10	0.70	0.81	0.08	0.83	0.09	0.84	0.97	0.09	0.98	0.10	0.98
	0.3	0.80	0.15	0.74	0.17	0.75	0.90	0.13	0.86	0.16	0.87	0.99	0.15	0.98	0.16	0.99
	0.6	0.92	0.22	0.78	0.25	0.79	0.96	0.21	0.88	0.25	0.89	1.00	0.23	0.99	0.25	0.99
	0.9	0.98	0.44	0.84	0.50	0.87	0.99	0.44	0.91	0.49	0.94	1.00	0.45	0.98	0.50	0.99
250	0	0.71	0.06	0.73	0.06	0.73	0.86	0.06	0.87	0.06	0.88	0.98	0.06	0.98	0.07	0.99
	0.3	0.86	0.10	0.77	0.12	0.78	0.94	0.10	0.89	0.11	0.91	0.99	0.11	0.99	0.12	0.99
	0.6	0.95	0.18	0.81	0.17	0.80	0.98	0.17	0.91	0.17	0.91	1.00	0.18	0.99	0.18	0.99
	0.9	1.00	0.38	0.85	0.38	0.85	1.00	0.39	0.93	0.38	0.93	1.00	0.40	0.99	0.38	0.99
500	0	0.76	0.05	0.76	0.06	0.77	0.89	0.06	0.89	0.06	0.90	0.99	0.06	0.99	0.06	0.99
	0.3	0.89	0.09	0.80	0.09	0.80	0.95	0.10	0.91	0.10	0.91	0.99	0.09	0.99	0.10	0.99
	0.6	0.96	0.14	0.82	0.13	0.82	0.99	0.15	0.92	0.14	0.92	1.00	0.14	0.99	0.13	0.99
	0.9	1.00	0.31	0.86	0.28	0.85	1.00	0.32	0.94	0.29	0.94	1.00	0.32	0.99	0.28	0.99

Note: The column labels are as defined in the main text of Section 2.3.

Table 2.4: Empirical null rejection probabilities of Wald-type tests for $H_0 : \beta_1 = 5, \beta_2 = -3, \beta_3 = 0.3$ in the individual specific intercepts and linear trends case with non-zero drifts.

		$N = 10$					$N = 20$					$N = 100$				
T	ρ_1, ρ_2	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P	W_{TB}	W_{rob}	W^+	W^+_{rob}	W^+_P
									$\rho_3 = 0$							
100	0	0.06	0.08	0.09	0.09	0.07	0.07	0.08	0.09	0.09	0.07	0.06	0.07	0.08	0.08	0.07
	0.3	0.26	0.16	0.14	0.15	0.13	0.30	0.19	0.15	0.15	0.14	0.55	0.39	0.15	0.15	0.15
	0.6	0.74	0.40	0.30	0.30	0.35	0.83	0.53	0.36	0.35	0.47	0.99	0.94	0.71	0.65	0.94
	0.9	0.98	0.81	0.83	0.82	0.99	1.00	0.90	0.92	0.91	1.00	1.00	1.00	1.00	1.00	1.00
250	0	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06
	0.3	0.35	0.18	0.12	0.12	0.10	0.42	0.23	0.11	0.11	0.09	0.77	0.58	0.12	0.12	0.10
	0.6	0.91	0.62	0.29	0.29	0.21	0.96	0.78	0.36	0.35	0.26	1.00	1.00	0.76	0.70	0.65
	0.9	1.00	0.94	0.92	0.91	0.89	1.00	0.98	0.97	0.97	0.98	1.00	1.00	1.00	1.00	1.00
500	0	0.05	0.05	0.06	0.06	0.06	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.06	0.05
	0.3	0.42	0.20	0.09	0.09	0.09	0.55	0.32	0.10	0.10	0.08	0.95	0.85	0.12	0.11	0.08
	0.6	0.98	0.78	0.24	0.24	0.14	1.00	0.95	0.35	0.33	0.16	1.00	1.00	0.82	0.76	0.40
	0.9	1.00	0.99	0.94	0.93	0.70	1.00	1.00	1.00	0.99	0.89	1.00	1.00	1.00	1.00	1.00
									$\rho_3 = 0.3$							
100	0	0.11	0.08	0.15	0.09	0.17	0.17	0.08	0.20	0.09	0.28	0.39	0.08	0.42	0.09	0.63
	0.3	0.35	0.16	0.22	0.15	0.24	0.44	0.18	0.28	0.15	0.37	0.72	0.25	0.50	0.15	0.71
	0.6	0.79	0.42	0.40	0.32	0.45	0.86	0.49	0.49	0.37	0.59	0.98	0.71	0.77	0.47	0.89
	0.9	0.99	0.81	0.86	0.83	0.97	0.99	0.87	0.92	0.89	0.99	1.00	0.97	0.99	0.98	1.00
250	0	0.16	0.06	0.17	0.07	0.20	0.24	0.06	0.26	0.07	0.33	0.58	0.06	0.59	0.06	0.75
	0.3	0.50	0.17	0.25	0.12	0.26	0.61	0.19	0.34	0.12	0.39	0.87	0.26	0.66	0.12	0.78
	0.6	0.92	0.56	0.42	0.28	0.37	0.96	0.65	0.56	0.32	0.52	1.00	0.81	0.83	0.39	0.87
	0.9	1.00	0.90	0.91	0.87	0.89	1.00	0.93	0.95	0.91	0.97	1.00	0.98	1.00	0.97	1.00
500	0	0.19	0.06	0.20	0.06	0.25	0.33	0.06	0.34	0.06	0.41	0.72	0.06	0.73	0.06	0.82
	0.3	0.59	0.17	0.27	0.10	0.29	0.72	0.21	0.41	0.11	0.47	0.93	0.29	0.77	0.10	0.85
	0.6	0.97	0.65	0.41	0.22	0.36	0.99	0.76	0.57	0.25	0.53	1.00	0.88	0.88	0.32	0.89
	0.9	1.00	0.96	0.93	0.86	0.78	1.00	0.98	0.97	0.93	0.92	1.00	0.99	1.00	0.97	1.00
									$\rho_3 = 0.6$							
100	0	0.30	0.08	0.34	0.10	0.39	0.44	0.08	0.48	0.09	0.59	0.78	0.08	0.80	0.09	0.89
	0.3	0.54	0.15	0.43	0.17	0.48	0.66	0.16	0.55	0.16	0.66	0.90	0.19	0.83	0.16	0.91
	0.6	0.85	0.38	0.57	0.33	0.61	0.90	0.41	0.67	0.33	0.75	0.99	0.51	0.89	0.38	0.94
	0.9	0.99	0.78	0.88	0.79	0.95	1.00	0.81	0.93	0.83	0.98	1.00	0.89	0.99	0.90	1.00
250	0	0.40	0.07	0.42	0.07	0.46	0.56	0.06	0.58	0.07	0.66	0.88	0.06	0.88	0.06	0.93
	0.3	0.68	0.15	0.50	0.13	0.53	0.80	0.14	0.64	0.12	0.71	0.95	0.16	0.90	0.12	0.94
	0.6	0.94	0.42	0.61	0.25	0.61	0.97	0.45	0.73	0.25	0.76	1.00	0.51	0.93	0.27	0.95
	0.9	1.00	0.83	0.91	0.79	0.90	1.00	0.85	0.95	0.81	0.95	1.00	0.90	0.99	0.84	1.00
500	0	0.47	0.06	0.48	0.06	0.52	0.67	0.06	0.67	0.06	0.73	0.93	0.06	0.93	0.06	0.96
	0.3	0.74	0.13	0.55	0.10	0.57	0.85	0.14	0.72	0.10	0.76	0.98	0.15	0.95	0.12	0.97
	0.6	0.97	0.42	0.64	0.19	0.62	0.98	0.45	0.78	0.19	0.79	1.00	0.50	0.96	0.21	0.97
	0.9	1.00	0.85	0.91	0.71	0.84	1.00	0.87	0.95	0.73	0.94	1.00	0.91	0.99	0.78	0.99
									$\rho_3 = 0.9$							
100	0	0.67	0.10	0.70	0.11	0.71	0.83	0.09	0.86	0.10	0.87	0.98	0.10	0.98	0.10	0.98
	0.3	0.83	0.16	0.77	0.18	0.77	0.92	0.16	0.89	0.17	0.89	0.99	0.17	0.98	0.18	0.99
	0.6	0.94	0.31	0.83	0.32	0.83	0.97	0.32	0.91	0.32	0.91	1.00	0.34	0.99	0.33	0.99
	0.9	1.00	0.72	0.94	0.75	0.95	1.00	0.73	0.97	0.76	0.98	1.00	0.75	0.99	0.77	1.00
250	0	0.72	0.06	0.73	0.07	0.74	0.87	0.07	0.88	0.07	0.89	0.98	0.07	0.98	0.07	0.99
	0.3	0.87	0.12	0.78	0.12	0.79	0.94	0.12	0.91	0.13	0.91	0.99	0.13	0.99	0.13	0.99
	0.6	0.97	0.27	0.83	0.21	0.83	0.99	0.27	0.92	0.22	0.93	1.00	0.28	0.99	0.23	0.99
	0.9	1.00	0.69	0.94	0.66	0.94	1.00	0.69	0.97	0.66	0.97	1.00	0.70	1.00	0.68	1.00
500	0	0.77	0.06	0.77	0.06	0.78	0.89	0.06	0.89	0.06	0.90	0.99	0.06	0.99	0.06	0.99
	0.3	0.90	0.10	0.81	0.10	0.81	0.96	0.11	0.91	0.10	0.92	1.00	0.10	0.99	0.11	0.99
	0.6	0.98	0.23	0.84	0.16	0.84	0.99	0.24	0.93	0.16	0.93	1.00	0.26	0.99	0.16	0.99
	0.9	1.00	0.64	0.94	0.54	0.93	1.00	0.65	0.97	0.54	0.97	1.00	0.67	1.00	0.55	1.00

Note: See note to Table 2.3.

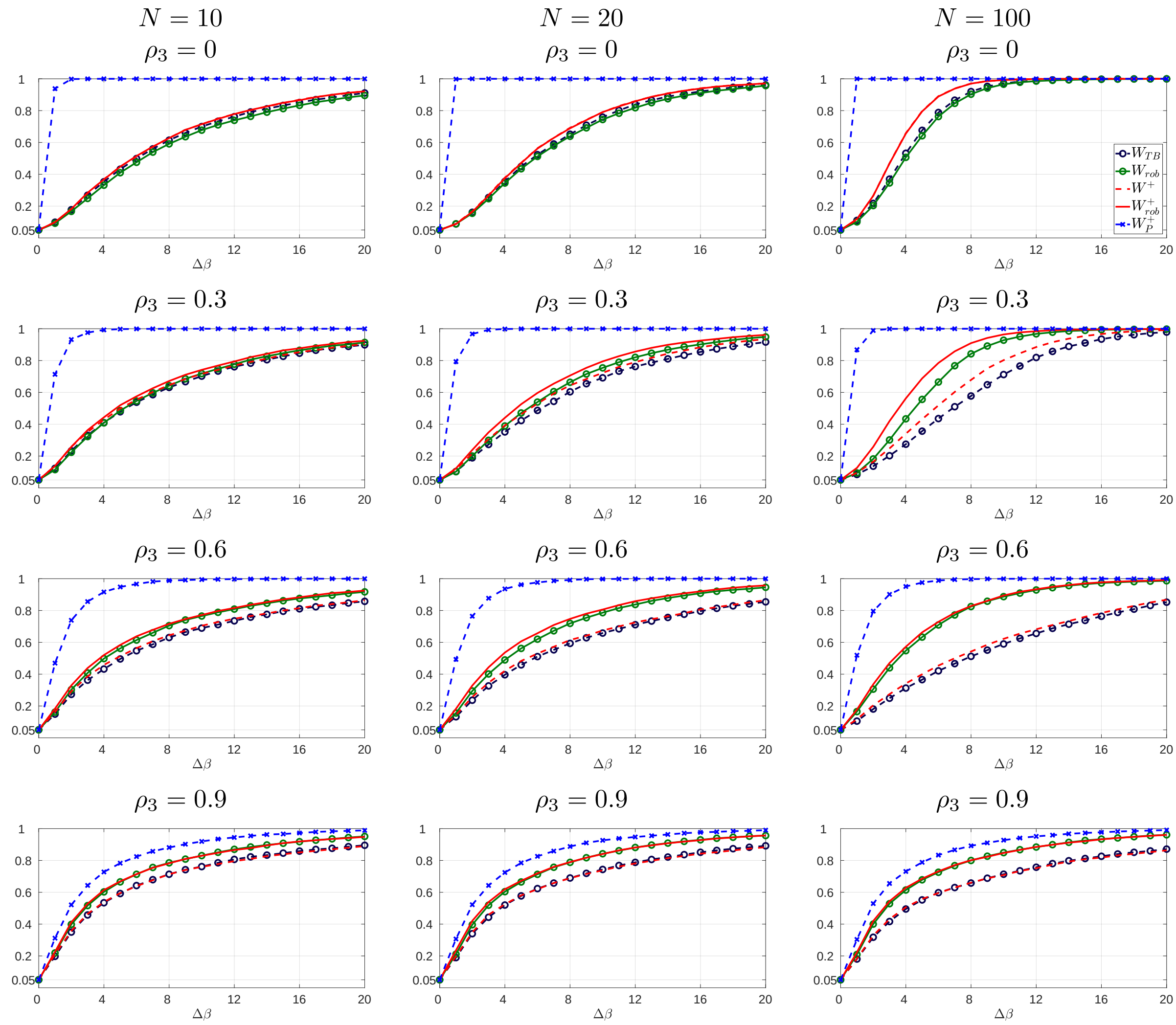


Figure 2.1: Size corrected power of the tests for $T = 100$ and $\rho_1, \rho_2 = 0.6$ in the individual specific intercepts only case with non-zero drifts.

Note: The axis label $\Delta\beta$ indicates, see also the description in the main text, the difference between the parameter vector under the null hypothesis and for the considered alternatives, i. e., $\beta_{H_1} = \beta + j \times \Delta\beta$, with $\beta = (5, -3, 0.3)'$, $\Delta\beta = (0.1, 0.05, 0.02)'$ and $j = 0, 1, \ldots, 20$ (displayed on the horizontal axis).

$N = 19$ countries for $T = 136$ years and the *wide data set* with $N = 89$ countries and $T = 54$ years. The long data set has originally been used in Wagner *et al.* (2020) and ranges from 1878 – 2013 for 19 early industrialized countries.[20] We also consider a subset comprising six of these 19 countries analyzed in more detail in a seemingly unrelated regressions setting in Wagner *et al.* (2020). These six countries are Austria (AT), Belgium (BE), Finland (FI), the Netherlands (NL), Switzerland (CH) and the United Kingdom (UK), with data for these countries available from 1870 – 2013, leading to a sample size of $T = 144$. The country list for the wide data set, with time span 1960 –

[20]The 19 countries are given by Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and USA. Note that the data are in fact available from 1870 onwards, with the exception of CO_2 emissions for New Zealand. Considering all 19 countries with 1878 as starting point is merely done to use exactly the same balanced panel data set as de Jong and Wagner (2022). Of course, whether the panel is balanced or not is irrelevant even from a computational perspective for group-mean estimation. A detailed description of the data including the sources is contained in Wagner *et al.* (2020).

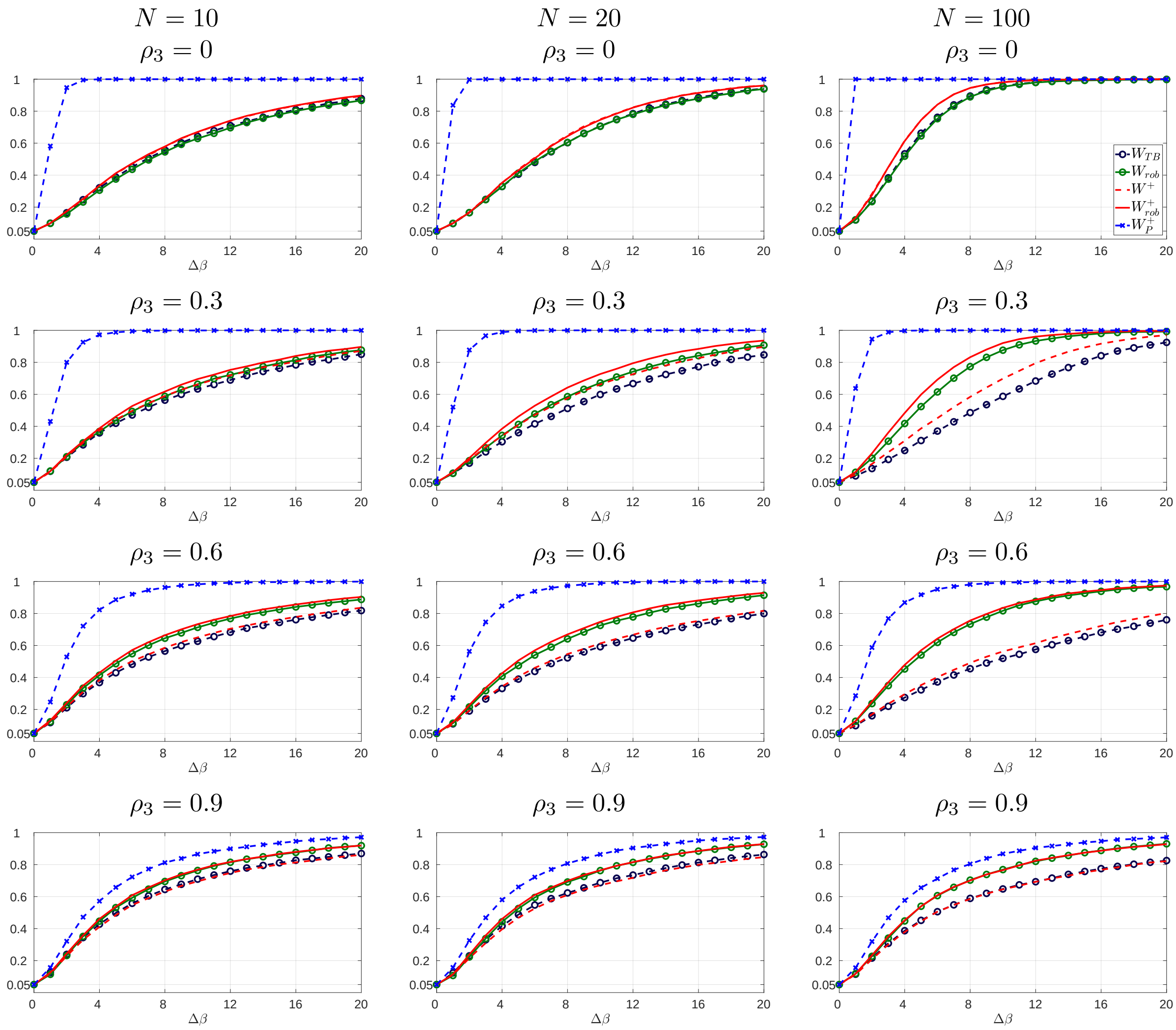


Figure 2.2: Size corrected power of the tests for $T = 100$ and $\rho_1, \rho_2 = 0.6$ in the individual specific intercepts and linear trends only case with non-zero drifts.
Note: See note to Figure 2.1.

2013, is available in Table 2.6 in Appendix 2.6.2.

Table 2.5 shows all estimation results – including “standard” and “robust” t-statistics – as well as the implied turning points (TPs). To facilitate comparison with de Jong and Wagner (2022) also the TPs obtained in that paper are included in the rows labeled “TP de J&W”. The upper panel considers individual specific intercepts only and the lower panel considers individual specific intercepts and linear trends. The left block-column shows the results for the quadratic specification and the right block-column shows the results for the cubic specification. The first question to be addressed concerns the polynomial degree of the EKC, i. e., whether a cubic specification has to be considered or the quadratic specification suffices. With respect to this question it turns out that robust inference leads to different conclusions than standard inference. For both $N = 6$ and $N = 19$ the use of robust inference leads to insignificant coefficients to the third power of the logarithm of per capita GDP; for both the intercept only and the intercept and trend case. For the wide data set with $N = 89$ the cubic specification is required, in the sense that both standard and – more importantly – robust t-statistics indicate significance of the third order coefficient for

Table 2.5: Group-mean fully modified OLS EKC estimation results

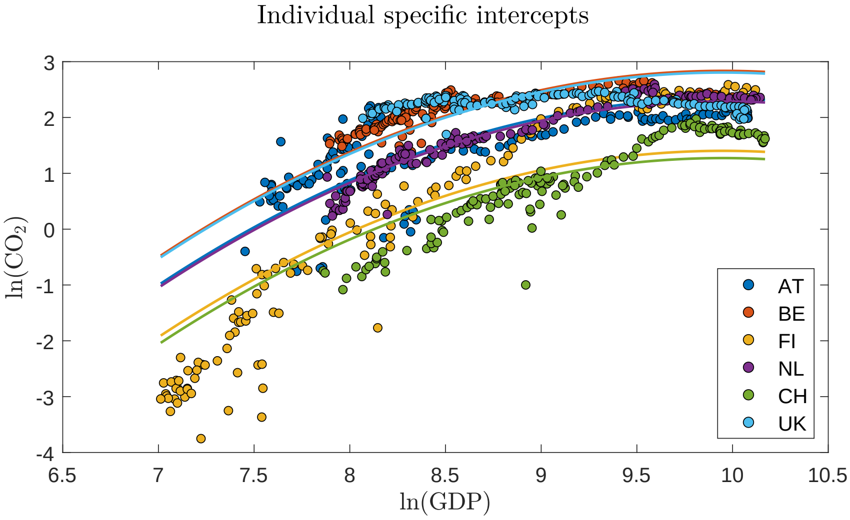
	quadratic specification			cubic specification		
	$N=6$	$N=19$	$N=89$	$N=6$	$N=19$	$N=89$
	Individual specific intercepts only					
β_1	7.63	8.24	9.16	−26.22	0.46	1061.85
	(**9.66**)	(**15.38**)	(**3.26**)	(*−1.65*)	(0.04)	(**2.53**)
	[**6.06**]	[**6.46**]	[**3.06**]	[−1.03]	[0.02]	[**2.46**]
β_2	−0.38	−0.42	−0.44	3.43	0.43	−148.79
	(**−8.65**)	(**−13.98**)	(**−2.44**)	(*1.92*)	(0.33)	(**−2.72**)
	[**−5.43**]	[**−5.83**]	[**−2.30**]	[1.20]	[0.14]	[**−2.65**]
β_3				−0.14	−0.03	6.80
				(**−2.13**)	(−0.64)	(**2.84**)
				[−1.33]	[−0.27]	[**2.78**]
TP GM	20,951	19,470	35,596	16,854	19,587	4,211
				548	1	510
TP de J&W	14,051	20,054	531,260	–	–	43,231
				–	–	443
	Individual specific intercepts and linear trends					
β_1	9.92	8.74	11.54	15.83	26.69	−952.79
	(**15.58**)	(**18.58**)	(**4.72**)	(1.57)	(**2.87**)	(**−2.75**)
	[**12.22**]	[**8.16**]	[**4.11**]	[1.18]	[1.45]	[**−2.77**]
β_2	−0.48	−0.43	−0.59	−1.18	−2.51	114.93
	(**−14.06**)	(**−17.38**)	(**−3.86**)	(−1.04)	(**−2.43**)	(**2.55**)
	[**−10.95**]	[**−7.41**]	[**−3.40**]	[−0.78]	[−1.21]	[**2.56**]
β_3				0.03	0.08	−4.71
				(0.67)	(**2.11**)	(**−2.40**)
				[0.50]	[1.03]	[**−2.42**]
TP GM	33,743	25,889	17,027	1.2×10^7	–	–
				94,276	–	–
TP de J&W	23,967	26,284	72,329	–	–	29,519
				–	–	578

Notes: "Standard" t-statistics, defined (2.17), in parentheses and "robust" t-statistics, defined in (2.21), in square brackets. *Italic* numbers indicate significance at the 10% nominal level and **bold** numbers indicate significance at the 5% significance level. The turning points based on the group-mean estimator (TP GM) are computed as $\exp\left(-\frac{\hat\beta_1}{2\hat\beta_2}\right)$ in the quadratic case and as $\exp\left(-\frac{\hat\beta_2}{3\hat\beta_3}(\pm1)(-\frac{\hat\beta_1}{3\hat\beta_3}+(\frac{\hat\beta_2}{3\hat\beta_3})^2)^{1/2}\right)$ in the cubic case. The symbol "–" indicates the absence of turning points for the estimated polynomial. The row labeled "TP de J&W" contains the turning points given in de Jong and Wagner (2022, Tables 7–8) using the pooled FM-OLS estimator in a slightly different specification with, in the lower panel, (common) time effects instead of individual specific linear time trends.

both specifications of the deterministic component.

Based on the above, we focus on the findings with the quadratic specification for the $N = 6$ and $N = 19$ data sets. For both specifications of the deterministic components, the coefficient to the squared logarithm of per capita GDP is (significantly) negative, with both standard and robust t-statistics. For $N = 6$, the turning points differ substantially between the group-mean estimator and the pooled estimator of de Jong and Wagner (2022) and are substantially larger for the group-mean estimator. For $N = 19$ the differences in the turning points between the group-mean and pooled estimators are negligible.[21]

[21] The sample range for the $N = 6$ data set is from 1,725 to 26,102 and for the $N = 19$ data set the sample range is from 794 to 31,933 (measured in 1990 Geary-Khamis dollars). Therefore the group-mean turning point when including individual specific intercepts and linear trends is out of sample for the $N = 6$ data set.



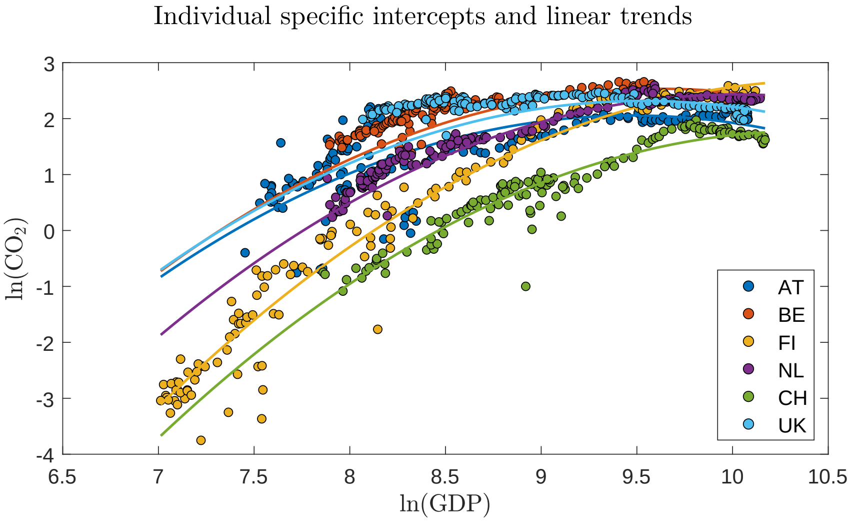


Figure 2.3: Scatter plot and estimated EKC relationship for CO_2 emissions over the period 1870–2013 for the $N = 6$ data set.

Notes: The curves display the results of inserting 144 equidistant points from the sample range of ln(GDP) in the quadratic relationship estimated with group-mean fully modified OLS estimator and adding the individual specific intercepts (top panel) or the individual specific intercepts and linear time trends (bottom panel), with corresponding values of the time trend given by $t = 1, \ldots, 144$.

Figure 2.3 shows the impact of including individual specific linear trends (in the lower graph) in addition to individual specific intercepts only (in the upper graph) on the estimated EKCs for $N = 6$. Including individual specific linear time trends (obviously) leads to a better fit, in particular for Finland and Switzerland, both for the low GDP values, i. e., for the beginning of the sample period, and the high GDP values, i. e., for the end of the sample period. Thus, the different "average levels" of log per capita emissions are well captured by the individual specific intercepts, the individual specific trends allow in addition to account to some extent for "curvature differences" across countries. On the question of poolability of the EKC across these countries see also Wagner *et al.* (2020), who in fact only find evidence for – in the words of that paper – *partial poolability* of the slope coefficients for Belgium, the Netherlands and the UK. Against this background this empirical section is to be interpreted merely as an illustration. For larger values of N, of course, the seemingly unrelated regressions based analysis of Wagner *et al.* (2020) is not feasible and one needs to resort to panel-type methods of one kind or another with the corresponding cross-sectional pooling imposed.

Let us close this illustration section with a brief look at the cubic specification results for the $N = 89$ wide data set. One striking feature for this data set is that the signs of the coefficient to the third power differ between the two specifications of the deterministic component, with $\beta_3 > 0$ in the intercept only case and $\beta_3 < 0$ in the intercept and linear trend case.[22] The group-mean estimator leads to two turning points at small values in the intercept only case, with the larger turning point corresponding to U-type behavior, and to a monotonic relationship in the intercept and linear trend specification. For the wide data set it thus appears that the pooled estimator leads to – notwithstanding all issues concerning poolability – more "useful" turning points.

2.5 Summary and Conclusions

This paper extends the toolkit for parameter estimation and inference in panels of cointegrating polynomial regressions with a group-mean fully modified OLS approach, which complements the pooled FM-OLS approach of de Jong and Wagner (2022). The consideration of a group-mean rather than a pooled estimation approach is not the only difference between the two papers. The present paper gains a lot of mileage from considering a fixed cross-section setting, which allows to include two features not considered in de Jong and Wagner (2022). First, we allow for the (potential) presence of drifts in the integrated regressors, which increases applicability substantially. Second, we provide cross-section "robust" inference for the group-mean OLS and FM-OLS estimators. Asymptotically valid inference is, as discussed, possible under minimal restrictions on the form and extent of cross-section dependence. No specific model of cross-section dependence, e. g., a factor structure, has to be posited. It is important to stress again that computation of the developed estimators and tests does not require any knowledge concerning the presence or absence of drifts and/or cross-section dependence.

The simulation results are, by and large, as expected, with one important exception regarding hypothesis testing: Using the cross-section robust version of the group-mean FM-OLS estimator based tests is unequivocally the best choice, as the robust version of the tests performs at least as good as the non-robust version of the tests even in the absence of cross-section dependence. The test based on the pooled estimator of de Jong and Wagner (2022) is very strongly adversely affected by cross-section dependence.

The illustrative application conveys two messages: First, cross-section robust inference makes a difference. In our illustration it indicates, unlike standard inference, that the quadratic specification is sufficient for the long data sets and that a cubic formulation is only required for the wide data set. The wide data set with $N = 89$ (larger than $T = 54$) indicates potential advantages of the pooled estimator in case of large cross-section dimension compared to the time series dimension,

[22] $\beta_3 > 0$ implies that the fitted polynomial diverges to plus infinity for log per capita GDP tending to infinity. Consequently, in case of turning points being present, the larger turning point corresponds to U-type rather than an inverted U-type behavior. Note for completeness, see de Jong and Wagner (2022, Table 8), that the pooled estimator leads to negative third order coefficients for both specifications for this data set.

i. e., benefits of resorting to an asymptotic approximation also in the cross-section dimension.

Two (related) issues remain open for future research: First, an analysis of the asymptotic behavior of the group-mean estimator in the two-way fixed effects case, i. e., with both individual and time specific fixed effects. This will require, second, asymptotic analysis in a large time and large cross-section setting, which is in any case important for panels with N large compared to T. Letting $N \to \infty$ requires that potential cross-section dependence will have to be considered more restrictively than in our fixed N setting; not only with respect to robust inference but also for obtaining, e. g., a sequential (unconditional) asymptotic normality result for the estimated coefficients.

Acknowledgements

The authors gratefully acknowledge partial financial support from Deutsche Forschungsgemeinschaft via the Collaborative Research Center SFB823 *Statistical Modelling of Nonlinear Dynamic Processes* (Projects A3 and A4). Furthermore, we thank the editor Esfandiar Maasoumi and two anonymous reviewers for a number of insightful comments that have led to significant changes and improvements of the paper. In addition we thank Fabian Knorre for his suggestion to present the empirical results as in Figure 2.3 and conference participants at the 2021 Annual Conference of the Verein für Socialpolitik in Regensburg. The views expressed in this paper are solely those of the authors and not necessarily those of the Bank of Slovenia or the European System of Central Banks. On top of this the usual disclaimer applies.

2.6 Appendix

2.6.1 Proofs

Proof of Proposition 2.1. The starting point is:

$$\begin{aligned} G_T^{-1}\left(\hat{\beta}^+ - \beta\right) &= \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^+(i) - \beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T\right)^{-1}\left(G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^+ - G_T C_i\right), \end{aligned} \tag{2.83}$$

with:

$$\tilde{u}_{it}^+ := \tilde{u}_{it} - \Delta x_{it}\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i} = \tilde{u}_{it} - (\mu_i + v_{it})\hat{\Omega}_{v_i v_i}^{-1}\hat{\Omega}_{v_i u_i}. \tag{2.84}$$

Since $\mu_i = 0$ for all $i = 1, \dots, N$, it follows directly from Assumptions 2.2 and 2.3 that:

$$G_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'G_T \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr, \tag{2.85}$$

$$G_T \sum_{t=1}^{T} \tilde{X}_{it} \tilde{u}_{it}^{+} \xrightarrow{d} \int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \mathbf{d}\mathbf{B}_{\mathbf{u_i}\cdot\mathbf{v_i}}(\mathbf{r}) + \boldsymbol{\Delta}_{\mathbf{v_i u_i}}^{+} \mathbf{A_i}, \tag{2.86}$$

and:

$$G_T C_i \xrightarrow{d} \Delta_{v_i u_i}^{+} A_i, \tag{2.87}$$

where $\Delta_{v_i u_i}^{+} := \Delta_{v_i u_i} - \Delta_{v_i v_i} \Omega_{v_i v_i}^{-1} \Omega_{v_i u_i}$ and A_i as given in the main text, with all quantities converging jointly. This immediately implies – for the parameter estimator corresponding to the i-th equation – that:

$$G_T^{-1} \left(\hat{\beta}^{+}(i) - \beta \right) \xrightarrow{d} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r) \tilde{\mathbf{B}}_{v_i}(r)' dr \right)^{-1} \int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \mathbf{d}\mathbf{B}_{\mathbf{u_i}\cdot\mathbf{v_i}}(\mathbf{r}). \tag{2.88}$$

Conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ the limiting distribution given in (2.88) is normal with expectation zero and covariance matrix $\Omega_{u_i \cdot v_i} \left(\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})' \mathbf{dr} \right)^{-1}$. Cross-sectional independence (Assumption 2.1) thus implies the – conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \ldots, N$ – asymptotic normality result for the group-mean FM-OLS estimator given in the main text in (2.12) and (2.13). □

Proof of Corollary 2.1. Under the null hypothesis the Wald-type statistic given in (2.16) is equal to:

$$\begin{aligned} W^{+} = & \left((G_R R G_T) G_T^{-1} (\hat{\beta}^{+} - \beta) \right)' \left((G_R R G_T) \hat{V}^{+} (G_R R G_T)' \right)^{-1} \\ & \times \left((G_R R G_T) G_T^{-1} (\hat{\beta}^{+} - \beta) \right). \end{aligned} \tag{2.89}$$

With the (asymptotic) restriction on the constraint matrix R posited in the main text in place and with $\hat{V}^{+} = G_T^{-1} \hat{S}^{+} G_T^{-1}$ converging in distribution to V^{+}, it follows from Proposition 2.1 that:

$$W^{+} \xrightarrow{d} (R^* \mathcal{Z})' \left(R^* V^{+} R^{*\prime} \right)^{-1} (R^* \mathcal{Z}), \tag{2.90}$$

with $\mathcal{Z}$ conditionally $\mathcal{N}(0, V^{+})$ distributed. This shows the conditional – and hence unconditional – asymptotic chi-squared null distribution of the Wald-type statistic. In case $s = 1$ analogous arguments lead to the result for the t-type test. □

Proof of Remark 2.1. Similar arguments as used in the proof of Proposition 2.1 show that:

$$G_T^{-1} \left(\hat{\beta}^{+} - \beta \right) \xrightarrow{d} \frac{1}{N} \sum_{i=1}^{N} \left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r) \tilde{\mathbf{B}}_{v_i}(r)' dr \right)^{-1} \int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r}) \mathbf{d}\mathbf{B}_{\mathbf{u_i}\cdot\mathbf{v_i}}(\mathbf{r}). \tag{2.91}$$

Conditional upon Δ, Σ and $W_v(r)$ the limiting distribution given in (2.91) is normal with expectation zero and covariance matrix:

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\mathbf{dB}_{\mathbf{u_i}\cdot\mathbf{v_i}}(\mathbf{r})\right\} \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\mathrm{Cov}\left\{\left(\int_0^1 \tilde{\mathbf{B}}_{v_i}(r)\tilde{\mathbf{B}}_{v_i}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\mathbf{dB}_{\mathbf{u_i}\cdot\mathbf{v_i}}(\mathbf{r}),\right. \\
&\qquad\qquad \left.\left(\int_0^1 \tilde{\mathbf{B}}_{v_j}(r)\tilde{\mathbf{B}}_{v_j}(r)'dr\right)^{-1}\int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r})\mathbf{dB}_{\mathbf{u_j}\cdot\mathbf{v_j}}(\mathbf{r})\right\} \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\tilde{M}_{ii}^{-1}\tilde{M}_{ij}\tilde{M}_{jj}^{-1} = V_{\mathrm{rob}}^{+}, \qquad (2.92)
\end{aligned}$$

where $\tilde{M}_{ij} = \int_0^1 \tilde{\mathbf{B}}_{\mathbf{v_i}}(\mathbf{r})\tilde{\mathbf{B}}_{\mathbf{v_j}}(\mathbf{r})'\mathbf{dr}$ and $\Omega_{u_i\cdot v_i;u_j\cdot v_j}$ is the constant in the quadratic covariation of the processes $B_{u_i\cdot v_i}(r)$ and $B_{u_j\cdot v_j}(r)$ and is defined in the main text.

It is straightforward to verify that $\hat{V}_{\mathrm{rob}}^{+} = G_T^{-1}\hat{S}_{\mathrm{rob}}^{+}G_T^{-1}$ converges in distribution to V_{rob}^{+}. Therefore, the null limiting distributions of W_{rob}^{+} and t_{rob}^{+} can be derived with exactly the same arguments as used in the proof of Corollary 2.1. □

Proof of Proposition 2.2. We first consider the case with individual specific intercepts but no individual specific linear trends included in (2.1). The proof for the case with both individual specific intercepts and individual specific linear trends included in (2.1) is considered afterwards and is based upon similar arguments.

(i) Similar to the proof of Proposition 2.1 the starting point is given by:

$$\begin{aligned}
H_T^{-1}\left(\hat{\beta}^{+}-\beta\right) &= \frac{1}{N}\sum_{i=1}^{N}H_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+}-H_TC_i\right). \qquad (2.93)
\end{aligned}$$

By definition of $\tilde{u}_{it}^{+}$ (see, e. g., the proof of Proposition 2.1) it follows that:

$$\begin{aligned}
\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+} &= \sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it}-v_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}) - \mu_i\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}\sum_{t=1}^{T}\tilde{X}_{it} \\
&= \sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it}-v_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}), \qquad (2.94)
\end{aligned}$$

where the last equality follows from the fact that by construction $\sum_{t=1}^{T} \tilde{X}_{it} = 0$. This implies:

$$
\begin{aligned}
&H_T^{-1}\left(\hat{\beta}^{+} - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N} H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'H_T\right)^{-1}\left(H_T\sum_{t=1}^{T}\tilde{X}_{it}(\tilde{u}_{it} - v_{it}\hat{\Omega}_{v_iv_i}^{-1}\hat{\Omega}_{v_iu_i}) - H_TC_i\right). \qquad (2.95)
\end{aligned}
$$

As the deterministic trends (asymptotically) dominate the elements of $\tilde{X}_{it}$, it follows that $T^{1/2}H_T\tilde{X}_{i\lfloor rT\rfloor} \Rightarrow J_i(r)$, $H_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} \xrightarrow{d} \int_0^1 J_i(r)dB_{u_i}(r)$, $H_T\sum_{t=1}^{T}\tilde{X}_{it}v_{it} \xrightarrow{d} \int_0^1 J_i(r)dB_{v_i}(r)$ and $H_TC_i = o_{\mathbb{P}}(1)$, with all quantities converging jointly, with $J_i(r)$ as defined in the main text in (2.28).[23] This immediately implies – for the parameter estimator from the i-th equation – that:

$$
H_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \xrightarrow{d} \left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r). \qquad (2.96)
$$

Conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ the limiting distribution given in (2.96) is normal with expectation zero and covariance matrix $\Omega_{u_i\cdot v_i}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}$. Cross-sectional independence (Assumption 2.1) thus implies the – conditional upon Δ_i, Σ_i and $W_{v_i}(r)$ for $i = 1, \ldots, N$ – asymptotic normality result for the group-mean estimator given in the main text in (2.30).

(ii) Analogously, the starting point for showing (2.31) is given by:

$$
\begin{aligned}
K_T^{-1}\left(\hat{\beta}^{+} - \beta\right) &= \frac{1}{N}\sum_{i=1}^{N} K_T^{-1}\left(\hat{\beta}^{+}(i) - \beta\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(K_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'K_T\right)^{-1}\left(K_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}^{+} - K_TC_i\right). \qquad (2.97)
\end{aligned}
$$

As described in the main text, as a result of demeaning and linear detrending, the linear trend that asymptotically dominates x_{it} is exactly annihilated in $\tilde{x}_{it}$. This is reflected in the following joint convergence results that can be derived with similar calculations as in Reichold and Wagner (2022, Proof of Lemma 2). First, $T^{1/2}K_T\tilde{X}_{i\lfloor rT\rfloor} \Rightarrow L_i(r)$, with $L_i(r)$ as defined in the main text in (2.29). Moreover:

$$
K_T\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} \xrightarrow{d} \int_0^1 L_i(r)dB_{u_i}(r) + (\Delta_{v_iu_i}, 0, 0)', \qquad (2.98)
$$

$$
K_T\sum_{t=1}^{T}\tilde{X}_{it}v_{it} \xrightarrow{d} \int_0^1 L_i(r)dB_{v_i}(r) + (\Delta_{v_iv_i}, 0, 0)', \qquad (2.99)
$$

and $K_TC_i \xrightarrow{d} (\Delta_{v_iu_i}^{+}, 0, 0)'$. The remaining parts of the proof are similar to the corresponding parts of the proof of (i) and are therefore omitted. □

[23] For more details we refer to Reichold and Wagner (2022, Lemma 2).

Proof of Corollary 2.2. The proof is based on similar arguments as the proof of Corollary 2.1 and therefore omitted. □

Proof of Remark 2.6. For sake of brevity we only consider the individual specific intercepts only case here in detail. The proof is entirely analogous for the individual specific intercepts and linear trends case.

It follows from the proof of Proposition 2.2 that:

$$H_T^{-1}\left(\hat{\beta}^+ - \beta\right) \xrightarrow{d} \frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r). \tag{2.100}$$

Conditional upon Δ, Σ and $W_v(r)$ the limiting distribution given in (2.100) is normal with expectation zero and covariance matrix:

$$\begin{aligned}
&\mathrm{Var}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r)\right\}\\
&=\frac{1}{N^2}\sum_{i,j=1}^{N}\mathrm{Cov}\left\{\left(\int_0^1 J_i(r)J_i(r)'dr\right)^{-1}\int_0^1 J_i(r)dB_{u_i\cdot v_i}(r),\right.\\
&\qquad\qquad\left.\left(\int_0^1 J_j(r)J_j(r)'dr\right)^{-1}\int_0^1 J_j(r)dB_{u_j\cdot v_j}(r)\right\}\\
&=\frac{1}{N^2}\sum_{i,j=1}^{N}\Omega_{u_i\cdot v_i;u_j\cdot v_j}\left(\mathcal{D}(\mu_i)\begin{pmatrix}1/12 & 1/12 & 3/40\\ 1/12 & 4/45 & 1/12\\ 3/40 & 1/12 & 9/112\end{pmatrix}\mathcal{D}(\mu_j)\right)^{-1} = V^+_{\alpha,\mathrm{rob}},
\end{aligned}$$

where $\Omega_{u_i\cdot v_i;u_j\cdot v_j}$ is the constant in the quadratic covariation of the processes $B_{u_i\cdot v_i}(r)$ and $B_{u_j\cdot v_j}(r)$ and is defined in the main text.

It is straightforward to verify that both $\hat{V}^+_{\alpha,\mathrm{rob}} = H_T^{-1}\hat{S}^+_{\mathrm{rob}}H_T^{-1}$ and:

$$\tilde{V}^+_{\alpha,\mathrm{rob}} = \frac{1}{N^2}\sum_{i,j=1}^{N}\hat{\Omega}_{u_i\dot{v}_i;u_j\cdot v_j}\left(\mathcal{D}(\hat{\mu}_i)\begin{pmatrix}1/12 & 1/12 & 3/40\\ 1/12 & 4/45 & 1/12\\ 3/40 & 1/12 & 9/112\end{pmatrix}\mathcal{D}(\hat{\mu}_j)\right)^{-1} \tag{2.101}$$

converge in distribution to $V^+_{\alpha,\mathrm{rob}}$. Therefore, the limiting distributions of W^+_{rob} and $W^+_{\alpha,\mathrm{rob}}$ can be shown to be chi-squared with s degrees of freedom under the null hypothesis using exactly the same arguments as in the proof of Corollary 2.1. Similarly, in case $s = 1$, t^+_{rob} and $t^+_{\alpha,\mathrm{rob}}$ can be shown to be asymptotically standard normally distributed under the null hypothesis. □

Proof of Proposition 2.3. The case $N_0 = 0$ is contained in the (proof of) Proposition 2.2. The results for $N_0 > 0$ follow from combining the results of Propositions 2.1 and 2.2. As in the proof of Proposition 2.2, we commence with the individual specific intercepts only case before turning to the individual specific intercepts and linear trends case.

(i) First note that the appropriate scaling matrix for the individual specific estimators $\hat{\beta}^+(i)$ calculated from cross-section members with zero drifts in the integrated regressor is G_T, whereas the appropriate scaling matrix for the individual specific estimators $\hat{\beta}^+(i)$ calculated from cross-section members with non-zero drifts in the integrated regressor is H_T. This implies:

$$\begin{aligned} G_T^{-1}\left(\hat{\beta}^{+}-\beta\right) &= \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N} G_T^{-1}H_T H_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) + o_{\mathbb{P}}(1), \end{aligned} \tag{2.102}$$

where the last equality follows from $G_T^{-1}H_T = \operatorname{diag}(T^{-1/2}, T^{-1}, T^{-3/2})$ and $H_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) = O_{\mathbb{P}}(1)$ for $i = N_0+1,\ldots,N$. Hence, the asymptotic behavior of the group-mean estimator only depends on the individual specific estimators $\hat{\beta}^{+}(i)$ calculated from cross-section members with zero drifts in the integrated regressor, since these converge at a slower rate than the estimators corresponding to cross-section members with non-zero drifts. The rest of the proof is analogous to the proof of Proposition 2.1 and therefore omitted.

(ii) As in (i), the appropriate scaling matrix depends upon the absence or presence of a non-zero drift in the integrated regressor. In the former case the appropriate scaling matrix is again given by G_T, whereas it is given by K_T in the presence of a non-zero drift. Therefore:

$$\begin{aligned} & G_T^{-1}\left(\hat{\beta}^{+}-\beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N} G_T^{-1}K_T K_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) \\ &= \frac{1}{N}\sum_{i=1}^{N_0} G_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) + \frac{1}{N}\sum_{i=N_0+1}^{N} \begin{bmatrix} T\left(\hat{\beta}_1^{+}(i)-\beta_1\right) & & \\ & o_{\mathbb{P}}(1) & \\ & & o_{\mathbb{P}}(1) \end{bmatrix}, \end{aligned} \tag{2.103}$$

where $\hat{\beta}_1^{+}(i)$ denotes the first element of $\hat{\beta}^{+}(i)$. The last equality follows from $G_T^{-1}K_T = \operatorname{diag}(1, T^{-1}, T^{-3/2})$ and $K_T^{-1}\left(\hat{\beta}^{+}(i)-\beta\right) = O_{\mathbb{P}}(1)$, for $i = N_0+1,\ldots,N$. In contrast to (i), the limiting distribution of the first component of $\hat{\beta}^{+}$ depends upon all cross-section member specific estimates of β_1, reflecting the fact that the coefficient to the first power of the integrated regressor is estimated at rate T irrespective of whether the drift is zero or non-zero – as in any case linear detrending removes a potentially present linear trend from the corresponding regressor. For the coefficients β_2 and β_3, the situation is exactly as in (i), with the limiting distribution only depending upon the individual specific estimators corresponding to cross-section members with zero drifts in the integrated regressor, since these are converging at a slower rate than the estimators corresponding to cross-section members with non-zero drifts. The rest of the proof is similar to the proof of Proposition 2.1 and therefore omitted. □

Proof of Corollary 2.3. The proof is based on similar arguments as the proofs of Corollary 2.1 and 2.2 and therefore omitted. □

Proof of Remark 2.8. The case $N_0 = 0$ has already been considered in the proof of Remark 2.6. The results for $N_0 > 0$ follow from combining the results of Proposition 2.3 and Remark 2.1, compare also the proofs of Remarks 2.1 and 2.6. The proof is therefore omitted. □

2.6.2 Country List for the Wide Data Set

Table 2.6: Country list for the wide data set.

Albania	Algeria	Angola	Argentina	Australia
Austria	Bahrain	Barbados	Belgium	Bolivia
Brazil	Bulgaria	Cambodia	Cameroon	Canada
Chile	China	Colombia	Costa Rica	Cote d'Ivoire
Cyprus	Denmark	Dominican Republic	DR Congo	Ecuador
Egypt	Ethiopia	Finland	France	Germany
Ghana	Greece	Guatemala	Hong Kong	Hungary
Iceland	India	Indonesia	Iran	Iraq
Ireland	Israel	Italy	Jamaica	Japan
Jordan	Kenya	Luxembourg	Madagascar	Mali
Malta	Mexico	Morocco	Mozambique	Myanmar
Netherlands	New Zealand	Niger	Nigeria	Norway
Pakistan	Peru	Philippines	Poland	Portugal
Romania	Saudi Arabia	Senegal	Singapore	South Africa
South Korea	Spain	Sri Lanka	Saint Lucia	Sudan
Sweden	Switzerland	Syria	Tanzania	Trinidad and Tobago
Tunisia	Turkey	Uganda	United Kingdom	United States
Uruguay	Venezuela	Vietnam	Yemen	

Chapter 3

A Residuals-Based Nonparametric Variance Ratio Test for Cointegration

Abstract

This paper derives asymptotic theory for Breitung's (2002, Journal of Econometrics 108, 343–363) nonparameteric variance ratio unit root test when applied to regression residuals. The test requires neither the specification of the correlation structure in the data nor the choice of tuning parameters. Compared with popular residuals-based no-cointegration tests, the variance ratio test is less prone to size distortions but has smaller local asymptotic power. However, this paper shows that local asymptotic power properties do not serve as a useful indicator for the power of residuals-based no-cointegration tests in finite samples. In terms of size-corrected power, the variance ratio test performs relatively well and, in particular, does not suffer from power reversal problems detected for, e. g., the frequently used augmented Dickey-Fuller type no-cointegration test. An application to daily prices of cryptocurrencies illustrates the usefulness of the variance ratio test in practice.

3.1 Introduction

Analyzing the relationship between stochastically trending (economic) time series in a single-equation regression framework entails the risk of obtaining misleading spurious regression results (Granger and Newbold, 1974; Phillips, 1986). Practitioners thus rely on statistical tests to assess whether the time series are cointegrated. In this context, it is popular in applications to employ so-called *no-cointegration* tests for the null hypothesis of no cointegration against the alternative of cointegration based on regression residuals estimated by ordinary least squares (OLS).

Alternative approaches include, e. g., single-equation cointegration tests based on conditional error correction models (e. g., Kremers *et al.*, 1992; Zivot, 2000) and system-based tests (e. g., Phillips and Ouliaris, 1990; Johansen, 1991; Shintani, 2001; Breitung, 2002; Harris and Poskitt, 2004; Cai and Shintani, 2006). System-based approaches have the advantage that they do not require the specification of a left-hand side variable and may also allow to test for the number of (linearly independent) cointegrating relations in the system. If, however, there exist reasons for a specific choice of the left-hand side variable, it is convenient and intuitively appealing to analyze the relationship between the variables in a single-equation framework.

Among the most popular residuals-based no-cointegration tests are the parametric augmented Dickey-Fuller (ADF, Dickey and Fuller, 1979; Said and Dickey, 1984) type test, proposed in Engle and Granger (1987) and asymptotically justified in Phillips and Ouliaris (1990), the semiparametric $\widehat{Z}_\alpha$ test (Phillips, 1987; Phillips and Ouliaris, 1990), and the parametric MSB test (Perron and Ng, 1996; Pesavento, 2007).[1] The three tests share the common feature that they require the choice of tuning parameters (e. g., the number of lags in an auxiliary regression and/or kernel and bandwidth choices to estimate a long-run variance parameter) to accommodate the correlation structure in the data. Although these tuning parameter choices allow for asymptotically valid inference, they are likely to have adverse effects on the performance of the tests in finite samples.

In contrast, this paper proposes a nonparametric no-cointegration test, which requires neither the specification of the correlation structure in the data nor the choice of tuning parameters. The test is an extension of Breitung's (2002) nonparametric variance ratio unit root test (originally applied to observed univariate time series) to regression residuals.[2] The test statistic is easy to compute as it is defined as a (re-scaled) ratio between the sample variances of the regression residuals and their partial sums. Under the null hypothesis, the sample variances converge to random variables whose distributions are scale dependent on the same long-run variance parameter. This makes the limiting null distribution of the test statistic nuisance parameter free without estimating the long-run variance parameter directly. Under the alternative of cointegration, the test statistic converges to zero at rate equal to sample size, which makes the test consistent. In the following, we refer to the test as the (nonparametric) variance ratio (no-cointegration) test.[3]

The paper derives asymptotic theory for the variance ratio no-cointegration test in a setting that allows for the presence of deterministic time trends both in the regression equation and in the regressors. In the presence of deterministic components, we derive the asymptotic properties of the variance ratio no-cointegration test under both OLS detrending and general least squares (GLS) detrending. Moreover, the paper compares the variance ratio test in terms of local asymptotic power, empirical size and size-corrected power with the ADF, $\widehat{Z}_\alpha$ and MSB tests in a detailed simulation study.

We follow, e. g., Pesavento (2007) and Perron and Rodríguez (2016) and impose a directional restriction on the model, which excludes cointegration between the right-hand side variables. In this case, local asymptotic power of the variance ratio test and its competitors is a function of a single nuisance parameter, R^2, which measures the long-run correlation between the regression errors and the regressors (cf. Pesavento, 2004; 2007). In addition, we construct a simulation setting that allows to analyze the effects of different short-run dynamics in the data generating process

[1] The MSB test and the $\widehat{Z}_\alpha$ test are popular representatives of the class of M tests (Stock, 1999; Perron and Ng, 1996; Ng and Perron, 2001) and the class of Z tests (Phillips and Perron, 1988), respectively.

[2] Breitung's (2002) test, in turn, is a generalization of the unit root test of Shin and Schmidt (1992).

[3] Please note that Westerlund (2005) has extended Breitung's (2002) unit root test to test for cointegration in panel data. However, little is known about the asymptotic and finite sample properties of Breitung's (2002) unit root test when applied to regression residuals in a pure time series setting. In particular, the nonparametric variance ratio no-cointegration test is not considered in the two insightful contributions of Pesavento (2007) and Perron and Rodríguez (2016).

(DGP) on the performance of the tests in finite samples while controlling for effects of R^2. This justifies a comparison between local asymptotic power of the tests and their power in finite samples. The results reveal that local asymptotic power properties do not serve as a useful indicator for the performance of residuals-based no-cointegration tests in finite samples and we explain why Pesavento (2007, p. 127) and Perron and Rodríguez (2016, p. 99) come to opposing conclusions.

Finally, an empirical illustration applies the variance ratio test and its competitors to daily prices of the four cryptocurrencies with highest market capitalization. Test decisions are heterogeneous across tests, but the variance ratio test, the ADF test, and the ADF test based on a modified information criterion provide reliable evidence for the presence of cointegration between the four cryptocurrencies with the highest market capitalization. The results are in line with those in related literature pointing towards the presence of cointegrating relationships in the cryptocurrency market (cf., e. g. Keilbar and Zhang, 2021; Bykhovskaya and Gorin, 2022b).

The paper proceeds as follows: Section 3.2 introduces the model and its underlying assumptions. Section 3.3 defines the variance ratio no-cointegration test under OLS- and GLS detrending and derives its asymptotic properties. Section 3.4 analyzes the performance of the variance ratio test in finite samples and Section 3.5 contains the empirical illustration. Section 3.6 summarizes and concludes. All proofs are relegated to the Appendix, which also contains additional asymptotic and finite sample results. Supplementary Material, available on the author's homepage, provides further finite sample results.

Throughout, $\lfloor x \rfloor$ denotes the integer part of a real number x, 1_m denotes the m-dimensional vector of ones and I_m denotes the $(m \times m)$-dimensional identity matrix. The symbols $\xrightarrow{p}$ and $\xrightarrow{w}$ signify convergence in probability and weak convergence, respectively, as the sample size $T \to \infty$.

3.2 The Model and Assumptions

We consider the model

$$x_t = \mu + x_{t-1} + v_t = x_0 + \mu t + \sum_{s=1}^{t} v_s \tag{3.1}$$

$$y_t = d_t' \tau + x_t' \beta + u_t \tag{3.2}$$

$$u_t = \rho u_{t-1} + \xi_t, \tag{3.3}$$

for $t = 1, \dots, T$, where $\{y_t\}_{t \in \mathbb{Z}}$ is a scalar integrated process, $\{x_t\}_{t \in \mathbb{Z}}$ is a m-dimensional vector of integrated processes with potentially non-zero deterministic drift $\mu \in \mathbb{R}^m$, and $u_0 = O_{\mathbb{P}}(1)$.[4] The p-dimensional vector d_t contains the deterministic components included in the model. Under the null hypothesis $\rho = 1$ the error $\{u_t\}_{t \in \mathbb{Z}}$ is an integrated process, i. e., there exists no cointegrating relation between $\{y_t\}_{t \in \mathbb{Z}}$ and $\{x_t\}_{t \in \mathbb{Z}}$. Under the alternative $|\rho| < 1$, however, the error process $\{u_t\}_{t \in \mathbb{Z}}$ is stationary and $\{y_t\}_{t \in \mathbb{Z}}$ and $\{x_t\}_{t \in \mathbb{Z}}$ are cointegrated with (normalized) cointegrating vector $[1, -\beta']'$, $\beta \neq 0$.

[4]For the asymptotic results in this paper it in fact suffices to assume $T^{-1/2} u_0 = o_{\mathbb{P}}(1)$. Section 3.4.3 analyzes the impact of a large initial value u_0 of order $T^{1/2}$ on the finite sample performance of the variance ratio test.

Assumption 3.1. *Let* $\{w_t\}_{t\in\mathbb{Z}} := \{[\xi_t, v_t']'\}_{t\in\mathbb{Z}}$ *be a strictly stationary and ergodic process with* $\mathbb{E}(w_t) = 0$*, finite covariance matrix* $\mathbb{E}(w_t w_t') > 0$ *and continuous spectral density matrix* $f_{ww}(\lambda)$ *on* $(-\pi, \pi]$*. Moreover,* $\{w_t\}_{t\in\mathbb{Z}}$ *fulfills a functional central limit theorem of the form*

$$T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor} w_t \xrightarrow{w} B(r) = \Omega^{1/2}W(r), \quad 0 \leq r \leq 1, \tag{3.4}$$

where $W(r) = [W_{\xi\cdot v}(r), W_v(r)']'$ *is an* $(1+m)$*-dimensional vector of independent standard Brownian motions and*

$$\Omega = \begin{bmatrix}\Omega_{\xi\xi} & \Omega_{\xi v}\\ \Omega_{v\xi} & \Omega_{vv}\end{bmatrix} := 2\pi f_{ww}(0) > 0 \tag{3.5}$$

denotes the long-run covariance matrix of $\{w_t\}_{t\in\mathbb{Z}}$*.*

In particular, positive definiteness of Ω_{vv} rules out cointegration among the elements of $\{x_t\}_{t\in\mathbb{Z}}$. To express asymptotic results, it is convenient to work with

$$\Omega^{1/2} = \begin{bmatrix}\Omega_{\xi\cdot v}^{1/2} & \Omega_{\xi v}(\Omega_{vv}^{-1/2})'\\ 0 & \Omega_{vv}^{1/2}\end{bmatrix}, \tag{3.6}$$

where $\Omega_{\xi\cdot v} := \Omega_{\xi\xi} - \Omega_{\xi v}\Omega_{vv}^{-1}\Omega_{v\xi}$, such that $\Omega^{1/2}(\Omega^{1/2})' = \Omega$. For later usage, we partition $B(r) = [B_\xi(r), B_v(r)']'$ analogously to the partitioning of $W(r)$.

The asymptotic results depend on the specification of the deterministic components in (3.1) and (3.2). This paper considers three different cases: no deterministics (D0), intercept only (D1) and intercept and linear trend (D2). In case D0, we set $x_0 = \mu = 0$ and remove $d_t\tau$ from (3.2). In case D1, we allow for $x_0 = O_{\mathbb{P}}(1)$ and set $\mu = 0$ and $d_t = 1$. Finally, in case D2, we allow for $x_0 = O_{\mathbb{P}}(1)$ and a deterministic trend in x_t and set $d_t = [1, t]'$.[5]

3.3 Asymptotic Theory

3.3.1 Preliminary Data Detrending

If deterministic components are included in (3.2), it is standard practice in applications to first detrend $z_t := [y_t, x_t']'$ according to the choice of d_t using ordinary least squares (OLS). In this case, the OLS detrended time series are defined as

$$\widetilde{z}_t' = [\widetilde{y}_t, \widetilde{x}_t'] := z_t' - d_t'\left(\sum_{s=1}^{T} d_s d_s'\right)^{-1}\sum_{s=1}^{T} d_s z_s'. \tag{3.7}$$

[5]Perron and Rodríguez (2016) also consider the case with a deterministic trend in the regressors but only a constant (rather than a constant and a linear trend) in the regression. Following Pesavento (2007), we abstain from considering this case as it implies that the limiting null distributions of the test statistics depend on the drift parameter μ being exactly zero or not – and thus leaves the opportunity of choosing wrong critical values in applications. Considering more general deterministic components, e. g., polynomial time trends, in (3.2) and/or (3.1) is also possible but beyond the scope of this paper.

In case D1 and D2 the definition reduces to $\widetilde{z}_t = z_t - T^{-1}\sum_{s=1}^T z_s$ and

$$\widetilde{z}_t = z_t - \frac{4T-6t+2}{T-1}\frac{1}{T}\sum_{s=1}^T z_s - \frac{-6T+12t-6}{(T-1)(T+1)}\sum_{s=1}^T \left(\frac{s}{T}\right) z_s, \tag{3.8}$$

respectively. For notational brevity, we define $\widetilde{z}_t = [\widetilde{y}_t, \widetilde{x}_t']' := z_t$ in case D0. With these definitions in place, it follows from (3.2) that

$$\widetilde{y}_t = \widetilde{x}_t'\beta + \widetilde{u}_t, \tag{3.9}$$

where $\widetilde{u}_t$ is defined analogously to $\widetilde{z}_t$. The OLS residuals in (3.9) are given by

$$\widehat{u}_t := \widetilde{y}_t - \widetilde{x}_t'\widehat{\beta} = \widetilde{u}_t - \widetilde{x}_t'\left(\widehat{\beta}-\beta\right) = \widetilde{u}_t - \widetilde{x}_t'\left(\sum_{s=1}^T \widetilde{x}_s\widetilde{x}_s'\right)^{-1}\sum_{s=1}^T \widetilde{x}_s\widetilde{u}_s, \tag{3.10}$$

where $\widehat{\beta}$ denotes the OLS estimator of β in (3.9).

To capture the asymptotic effects of detrending, define for a potentially multivariate integrable stochastic process $P(r)$, $0 \le r \le 1$, the detrended process $\widetilde{P}(r) = P(r) - \int_0^1 P(s)ds$ in case D1 and $\widetilde{P}(r) = P(r) - (4-6r)\int_0^1 P(s)ds - (12r-6)\int_0^1 sP(s)ds$ in case D2. In case D0, we set $\widetilde{P}(r) := P(r)$.

Remark 3.1. *Perron and Rodríguez (2016) suggest to use GLS detrended data rather than OLS detrended data in cases D1 and D2 to increase local asymptotic power of residuals-based no-cointegration tests. For* $\bar{\rho} := 1 + \bar{c}/T$, *with some constant* $\bar{c} \le 0$, *define* $z_t^{\bar{\rho}} := z_t - \bar{\rho} z_{t-1}$ *and* $d_t^{\bar{\rho}} := d_t - \bar{\rho} d_{t-1}$, $t = 2, \dots, T$. *The GLS detrended variables are constructed as* $\widetilde{y}_t^{(GLS)} := y_t - d_t'\widehat{\tau}^*$ *and* $\widetilde{x}_t^{(GLS)} := x_t - \widehat{\Psi}_x^{*\prime} d_t$, *where* $[\widehat{\tau}^*, \widehat{\Psi}_x^*] := (D^{\bar{\rho}\prime}D^{\bar{\rho}})^{-1}D^{\bar{\rho}\prime}Z^{\bar{\rho}}$, *with* $D^{\bar{\rho}} := [d_1, d_2^{\bar{\rho}}, \dots, d_t^{\bar{\rho}}]'$ *and* $Z^{\bar{\rho}} := [z_1, z_2^{\bar{\rho}}, \dots, z_T^{\bar{\rho}}]'$.[6] *The corresponding test statistics are then based on the OLS residuals* $\widehat{u}_t^{(GLS)}$ *in the regression* $\widetilde{y}_t^{(GLS)} = \widetilde{x}_t^{(GLS)\prime}\beta + \widetilde{u}_t^{(GLS)}$, *where* $\widetilde{u}_t^{(GLS)} := u_t - d_t'(\widehat{\tau}^* - \tau) + \widehat{\beta}'\widehat{\Psi}_x^{*\prime} d_t$.

In what follows, the main text focuses in cases D1 and D2 on OLS detrended data, whereas a series of remarks is dedicated to the use of GLS detrended data.

3.3.2 The Variance Ratio Test

Typically, residuals-based no-cointegration tests require tuning parameter choices (e. g., the number of lags in an auxiliary regression and/or kernel and bandwidth choices to estimate a long-run variance parameter) to accommodate the correlation structure in the data. Various simulation results indicate that these tuning parameter choices can affect the finite sample performance of no-cointegration tests considerably. This is particularly unfavorable in situations where different tuning parameter choices lead to different test decisions.

To obtain a tuning parameter free no-cointegration test, we apply the nonparametric variance ratio test of Breitung (2002), originally proposed to test for a unit root in an observed univariate time series, to the OLS residuals $\widehat{u}_t$ defined in (3.10), i. e., based on OLS detrended data. The variance

[6] Including d_1 and z_1 in the definitions of $D^{\bar{\rho}}$ and $Z^{\bar{\rho}}$, respectively, is important, see, e. g., Breitung and Taylor (2003).

ratio test statistic is thus defined as

$$\mathrm{VR} := \frac{T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\widehat{u}_s\right)^2}{\sum_{t=1}^{T}\widehat{u}_t^2} = \frac{\widehat{\eta}_T}{T^{-2}\sum_{t=1}^{T}\widehat{u}_t^2}, \tag{3.11}$$

where $\widehat{\eta}_T := T^{-4}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\widehat{u}_s\right)^2$.

In contrast to, e. g., the MSB and $\widehat{Z}_\alpha$ test statistics, the variance ratio test statistic does not depend on a consistent estimator of the long-run variance parameter $\Omega_{\xi\cdot v}$. Instead, it depends on a random variable $\widehat{\eta}_T$, whose limiting null distribution is, as we shall see in the proof of Proposition 3.1, scale dependent on $\Omega_{\xi\cdot v}$. In this sense, the variance ratio test statistic may be interpreted as a self-normalized version of the MSB test statistic.[7] Scale dependence of the limiting null distribution of η_T on $\Omega_{\xi\cdot v}$ is key to obtain a nuisance parameter free limiting distribution of the variance ratio test statistic under the null hypothesis of no cointegration.

Proposition 3.1. *Let $\{x_t\}_{t\in\mathbb{Z}}$, $\{y_t\}_{t\in\mathbb{Z}}$, and $\{u_t\}_{t\in\mathbb{Z}}$ be generated by (3.1), (3.2), and (3.3), respectively, and let $\{w_t\}_{t\in\mathbb{Z}}$ satisfy Assumption 3.1. Then it holds under the null hypothesis of no cointegration ($\rho = 1$) in cases D0, D1, and D2 that*

$$VR \xrightarrow{w} \frac{\int_0^1\left(\int_0^r \widetilde{W}^+_{\xi\cdot v}(s)ds\right)^2 dr}{\int_0^1\left(\widetilde{W}^+_{\xi\cdot v}(r)\right)^2 dr}, \tag{3.12}$$

where

$$\widetilde{W}^+_{\xi\cdot v}(r) := \widetilde{W}_{\xi\cdot v}(r) - \widetilde{W}_v(r)'\left(\int_0^1 \widetilde{W}_v(s)\widetilde{W}_v(s)'ds\right)^{-1}\int_0^1 \widetilde{W}_v(s)\widetilde{W}_{\xi\cdot v}(s)ds. \tag{3.13}$$

The limiting null distribution of the variance ratio test statistic is nonstandard but free of any nuisance parameters. It only depends on the dimension m of x_t, through the dimension of $W_v(r)$, and on the deterministic component in (3.2), as the detrended processes appear in the limit.[8]

We proceed with analyzing the behavior of the variance ratio test statistic under the alternative of cointegration.

Proposition 3.2. *Let $\{x_t\}_{t\in\mathbb{Z}}$, $\{y_t\}_{t\in\mathbb{Z}}$, and $\{u_t\}_{t\in\mathbb{Z}}$ be generated by (3.1), (3.2), and (3.3), respectively, and let $\{w_t\}_{t\in\mathbb{Z}}$ satisfy Assumption 3.1. Then it holds under the alternative of cointegration ($|\rho| < 1$) in cases D0, D1, and D2 that $VR = O_{\mathbb{P}}(T^{-1})$.*

Proposition 3.2 shows that in the presence of cointegration, the variance ratio test statistic converges to zero at rate equal to sample size. Hence, the variance ratio test is a left-tailed test, rejecting the null hypothesis of no cointegration in favor of the alternative of cointegration for small (i. e., close to zero) realizations of VR. Table 3.4 in Appendix 3.7.1 provides corresponding asymptotic critical values in cases D0, D1, and D2 for $m = 1, \dots, 5$.

[7] This self-normalizing feature of the variance ratio test fits well to the (bootstrap-assisted) self-normalized testing approach of Reichold and Jentsch (2022) for hypotheses on the cointegrating vector.

[8] The limiting null distribution derived in Proposition 3.1 differs from the limiting null distribution derived in Breitung (2002, Prop. 3), reflecting the application of the variance ratio test to regression residuals rather than to an observed (potentially detrended) univariate time series.

Remark 3.2. *The variance ratio test applied to the residuals from the regression with GLS detrended data is defined as* $VR^{(GLS)} := T^{-2}\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\widehat{u}_s^{(GLS)}\right)^2 / \sum_{t=1}^{T}\left(\widehat{u}_t^{(GLS)}\right)^2$, *with* $\widehat{u}_t^{(GLS)}$ *as defined in Remark 3.1. Remarks 3.3 and 3.4 in Section 3.3.3 derive the limiting distribution of* $VR^{(GLS)}$ *both under the null hypothesis of no cointegration and under local alternatives and also provide some guidance on the choice of* $\bar{c}$.

3.3.3 Local Asymptotic Power

To analyzes the performance of the variance ratio test statistic under local alternatives, we set $\rho = \rho_T = 1 + c/T$, with $c \leq 0$, in (3.3). For $c = 0$, the regression errors $\{u_t\}_{t\in\mathbb{Z}}$ are integrated of order one, i. e., there is no cointegration between $\{y_t\}_{t\in\mathbb{Z}}$ and $\{x_t\}_{t\in\mathbb{Z}}$. Under the local alternative $c < 0$, the error process $\{u_t\}_{t\in\mathbb{Z}}$ is a near unit root process such that $T^{-1/2}u_{\lfloor rT\rfloor} \xrightarrow{w} \Omega_{\xi\cdot v}^{1/2} J_{\xi\cdot v}^{c}(r)$, $0 \leq r \leq 1$, where

$$J_{\xi\cdot v}^{c}(r) := \int_0^r e^{(r-s)c} d\left(W_{\xi\cdot v}(s) + \sqrt{R^2/(1-R^2)}\,\overline{W}_v(s)\right), \tag{3.14}$$

is an Ornstein-Uhlenbeck process, with $R^2 := \Omega_{\xi\xi}^{-1}\Omega_{\xi v}\Omega_{vv}^{-1}\Omega_{v\xi} = 1 - \Omega_{\xi\cdot v}/\Omega_{\xi\xi}$ and $\overline{W}_v(r) := m^{-1/2}1_m' W_v(r)$ (cf., e. g., Perron and Rodríguez, 2016, Lemma 5.1). The coefficient R^2 lies between zero and one and measures the squared long-run correlation between the regression errors $\{u_t\}_{t\in\mathbb{Z}}$ and the regressors $\{x_t\}_{t\in\mathbb{Z}}$. Pesavento (2004; 2007) shows analytically that R^2 is the only nuisance parameter affecting local asymptotic power of the ADF, MSB, and $\widehat{Z}_\alpha$ tests. The following propositions shows that R^2 is also the only nuisance parameter affecting local asymptotic power of the variance ratio test.

Proposition 3.3. *Let* $\{x_t\}_{t\in\mathbb{Z}}$ *and* $\{y_t\}_{t\in\mathbb{Z}}$ *be generated by* (3.1) *and* (3.2), *respectively and let* $\{u_t\}_{t\in\mathbb{Z}}$ *be generated by* (3.3) *with* $\rho = \rho_T = 1 + c/T$, *where* $c \leq 0$. *Let* $\{w_t\}_{t\in\mathbb{Z}}$ *satisfy Assumption 3.1. Then it holds for the variance ratio test statistic in cases D0, D1, and D2 that*

$$VR \xrightarrow{w} \mathcal{G}_{VR,c} := \frac{\int_0^1\left(\int_0^r \widetilde{J}_{\xi\cdot v}^{c,+}(s)ds\right)^2 dr}{\int_0^1\left(\widetilde{J}_{\xi\cdot v}^{c,+}(r)\right)^2 dr}, \tag{3.15}$$

where

$$\widetilde{J}_{\xi\cdot v}^{c,+}(r) := \widetilde{J}_{\xi\cdot v}^{c}(r) - \widetilde{W}_v(r)'\left(\int_0^1 \widetilde{W}_v(s)\widetilde{W}_v(s)'ds\right)^{-1}\int_0^1 \widetilde{W}_v(s)\widetilde{J}_{\xi\cdot v}^{c}(s)ds. \tag{3.16}$$

Proposition 3.3 shows that the limiting distribution of the variance ratio test statistic under local alternatives depends on the location parameter c, the number of regressors m and on the nuisance parameter R^2. Local asymptotic power of the variance ratio test at the nominal level α is given by the probability that $\mathcal{G}_{\mathrm{VR},c}$ is smaller than the α-quantile of the limiting null distribution of VR. For $c = 0$, $\mathcal{G}_{\mathrm{VR},c}$ coincides with the limiting null distribution given in Proposition 3.1, i. e., local asymptotic power at $c = 0$ is equal to α. To ease comparisons between the limiting distribution of the variance ratio test and the limiting distributions of the tests considered in Pesavento (2007) under local alternatives, rewrite $\mathcal{G}_{\mathrm{VR},c} = \left(\widetilde{\kappa}'\widetilde{C}_c\widetilde{\kappa}_c\right) / \left(\widetilde{\kappa}'\widetilde{A}_c\widetilde{\kappa}_c\right)$, where $\widetilde{C}_c := \int_0^1\left(\int_0^r \widetilde{W}_c(s)ds\right)\left(\int_0^r \widetilde{W}_c(s)ds\right)' dr$,

$\widetilde{A}_c := \int_0^1 \widetilde{W}_c(r)\widetilde{W}_c(r)'dr$, $\tilde{\kappa}_c' = [1, -\int_0^1 \widetilde{W}_v(r)'\widetilde{J}^c_{\xi\cdot v}(r)dr \left(\int_0^1 \widetilde{W}_v(r)\widetilde{W}_v(r)'dr\right)^{-1}]$, and $\widetilde{W}_c(r) := [\widetilde{J}^c_{\xi\cdot v}(r), \widetilde{W}_v(r)']'$.[9]

Remark 3.3. *In case of GLS detrending, it follows from Perron and Rodríguez (2016, Lemma 5.3) and similar arguments as used in Perron and Rodríguez (2016, Proof of Theorem 5.2) and in the proof of Proposition 3.1 that the variance ratio test statistic converges under the local alternative* $\rho = \rho_T = 1 + c/T$, *where* $c \leq 0$, *to*

$$VR^{(GLS)} \xrightarrow{w} \mathcal{G}^{(GLS)}_{VR,c}(\bar{c}) := \frac{\kappa_c^{(GLS)\prime} C_c^{(GLS)} \kappa_c^{(GLS)}}{\kappa_c^{(GLS)\prime} A_c^{(GLS)} \kappa_c^{(GLS)}}, \tag{3.17}$$

with $A_c^{(GLS)}$, $C_c^{(GLS)}$ *and* $\kappa_c^{(GLS)}$ *defined analogously to* $\widetilde{A}_c$, $\widetilde{C}_c$ *and* $\tilde{\kappa}_c$, *respectively:* $\widetilde{J}^c_{\xi\cdot v}(r)$ *and* $\widetilde{W}_v(r)$ *have to be replaced by* $J^c_{\xi\cdot v}(r)$ *and* $W_v(r)$, *respectively, in case D1 and with* $J^c_{\xi\cdot v}(r) - \left(\lambda J^c_{\xi\cdot v}(1) + 3(1-\lambda)\int_0^1 sJ^c_{\xi\cdot v}(s)ds\right) r$ *and* $W_v(r) - \left(\lambda W_v(1) + 3(1-\lambda)\int_0^1 sW_v(s)ds\right) r$, *respectively, in case D2, where* $\lambda := (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$ *and* $\bar{c}$ *as chosen in Remark 3.1. In case D1,* $\mathcal{G}^{(GLS)}_{VR,c}(\bar{c})$ *does not depend on* $\bar{c}$ *and coincides with* $\mathcal{G}_{VR,c}$ *as defined in Proposition 3.3 in case D0. The limiting null distribution of the variance ratio test based on GLS detrending is given by* $\mathcal{G}^{(GLS)}_{VR,0}(\bar{c})$.

Remark 3.4. *To provide some guidance on the choice of* $\bar{c}$, *we follow Perron and Rodríguez (2016, p. 93) and choose* $\bar{c}$ *such that* $P\left(\mathcal{G}^{(GLS)}_{VR,\bar{c}}(\bar{c}) < q_{0.05}(\bar{c})\right) = 0.5$ *when* $R^2 = 0.4$, *where* $q_\alpha(\bar{c})$ *satisfies* $P\left(\mathcal{G}^{(GLS)}_{VR,0}(\bar{c}) < q_\alpha(\bar{c})\right) = 0.05$.[10] *Table 3.3 in Appendix 3.7.1 displays the values of* $\bar{c}$ *in cases D1 and D2 for* $m = 1, \ldots, 5$. *Given these values of* $\bar{c}$, *Table 3.4 in Appendix 3.7.1 tabulates the corresponding critical values* $q_\alpha(\bar{c})$ *for the variance ratio test based on GLS detrended data.*

Figure 3.1 illustrates the local asymptotic power curve of the variance ratio test at the nominal 5% level (in cases D1 and D2 based on OLS detrended and GLS detrended data) for $m = 1$ and $R^2 \in \{0, 0.4, 0.8\}$. In case of GLS detrending, $\bar{c}$ is chosen as suggested in Table 3.3 in Appendix 3.7.1 – with results being qualitatively similar for other choices of $\bar{c}$. The figure also displays the local asymptotic power curves of the ADF test, the $\widehat{Z}_\alpha$ test, and the MSB test based on OLS detrended data derived in Pesavento (2007) and the local asymptotic power curve of the ADF and MSB tests based on GLS detrended data derived in Perron and Rodríguez (2016), with the value of $\bar{c}$ as suggested in Perron and Rodríguez (2016, Table 1).[11]

---

[9] In this notation, the limiting distributions of the ADF, $\widehat{Z}_\alpha$ and MSB statistics under local alternatives are given by $c\frac{\left(\tilde{\kappa}_c'\widetilde{A}_c\tilde{\kappa}_c\right)^{1/2}}{\left(\tilde{\kappa}_c' D\tilde{\kappa}_c\right)^{1/2}} + \frac{\tilde{\kappa}_c'\widetilde{B}_c\tilde{\kappa}_c}{\left(\tilde{\kappa}_c' D\tilde{\kappa}_c\right)^{1/2}\left(\tilde{\kappa}_c'\widetilde{A}_c\tilde{\kappa}_c\right)^{1/2}}$, $c + \frac{\tilde{\kappa}_c'\widetilde{B}_c\tilde{\kappa}_c}{\tilde{\kappa}_c'\widetilde{A}_c\tilde{\kappa}_c}$, and $\frac{\left(\tilde{\kappa}_c'\widetilde{A}_c\tilde{\kappa}_c\right)^{1/2}}{\left(\tilde{\kappa}_c' D\tilde{\kappa}_c\right)^{1/2}}$, respectively, with $\widetilde{B}_c := \int_0^1 \widetilde{W}_c(r)d\left((W_{\xi\cdot v}(s) + \sqrt{R^2/(1-R^2)}\overline{W}_v(s)\right)$, $D := \begin{bmatrix} 1+\bar{\delta}'\bar{\delta} & \bar{\delta}' \\ \bar{\delta} & I_m \end{bmatrix}$ and $\bar{\delta}' := \Omega_{\xi\cdot v}^{-1/2}\Omega_{\xi v}(\Omega_{vv}^{-1/2})'$, see Pesavento (2007, Theorem 1), where we corrected a typo in the limiting distribution of $\widehat{Z}_\alpha$. By definition, it holds that $\bar{\delta}'\bar{\delta} = R^2/(1-R^2)$.

[10] See also the discussion in Nielsen (2009, p. 1528).

[11] Perron and Rodríguez (2016) derive, among others, the limiting distributions of the ADF, $\widehat{Z}_\alpha$, and MSB tests based on GLS detrended data under local alternatives. However, the analytical expressions of the limiting distributions of the ADF test and the $\widehat{Z}_\alpha$ test (the $\widehat{Z}_{\hat{\rho}}$ test in their notation) in Theorem 5.2 contain typos. The square root should be removed from the numerator in the second term of the limiting distribution of the ADF test and from both the numerator and the denominator in the second term of the limiting distribution of the $\widehat{Z}_\alpha$ test (cf. Pesavento, 2007, Theorem 1). The results in Figure 3.1 are based on the corrected version of the limiting distribution of the ADF test.

As expected, local asymptotic power of all tests decreases when R^2 increases, with the power loss being least pronounced in case D0 and most pronounced in case D2. Figure 3.1 further reveals that local asymptotic power of the ADF test (in the presence of deterministics based on OLS detrended data), the $\widehat{Z}_\alpha$ test, and the MSB test is very similar throughout and considerably larger than local asymptotic power of the variance ratio test, irrespective of whether the variance ratio test is based on OLS detrended data or GLS detrended data.[12]

In cases D1 and D2, GLS detrending increases local asymptotic power of the variance ratio test for $R^2 = 0.8$, but this power improvement seems to be negligible when compared with local asymptotic power of the other tests considered here. On the contrary, for $R^2 = 0$ local asymptotic power of the variance ratio test based on GLS detrended data is much smaller than local asymptotic power of its OLS-based counterpart. For $R^2 = 0.4$, GLS detrending seems to lead to minor power improvements for small values of c, but as c moves further away from zero GLS detrending becomes disadvantageous. We conclude that in terms of local asymptotic power, the variance ratio test does not benefit from GLS detrending, which is in line with the findings in Breitung and Taylor (2003) and Nielsen (2009) for Breitung's (2002) unit root test applied to observed univariate time series. In contrast, GLS detrending improves local asymptotic power of the ADF and MSB tests considerably throughout, which is in line with the findings in Perron and Rodríguez (2016). Finally, note that results for other values of m are qualitatively similar, with local asymptotic power being lower overall for larger values of m.

The results clearly demonstrate that the ADF, $\widehat{Z}_\alpha$, and MSB tests outperform the variance ratio test in terms of local asymptotic power. These power losses might not come as a surprise, given the self-normalizing property of the variance ratio test and the fact that self-normalized tests are well known for having smaller local asymptotic power than their (semi-)parametric counterparts (cf., e. g., Shao, 2015). However, local asymptotic power analyses are unable to reveal finite sample effects of both different short-run dynamics in the DGP and tuning parameter choices. The next section complements the local asymptotic power analysis with a careful assessment of the performance of the tests in finite samples.

3.4 Finite Sample Performance

We generate data according to (3.1) and (3.2) with $m = 1$ regressor, i. e.,

$$x_t = \mu + x_{t-1} + v_t = x_0 + \mu t + \sum_{s=1}^{t} v_s, \tag{3.18}$$

$$y_t = d_t'\tau + x_t\beta + u_t, \tag{3.19}$$

[12] Hosseinkouchack (2014) analyzes local asymptotic power of Breitung's (2002) unit root test when applied to an observed (potentially OLS detrended) univariate time series and finds similar power losses.

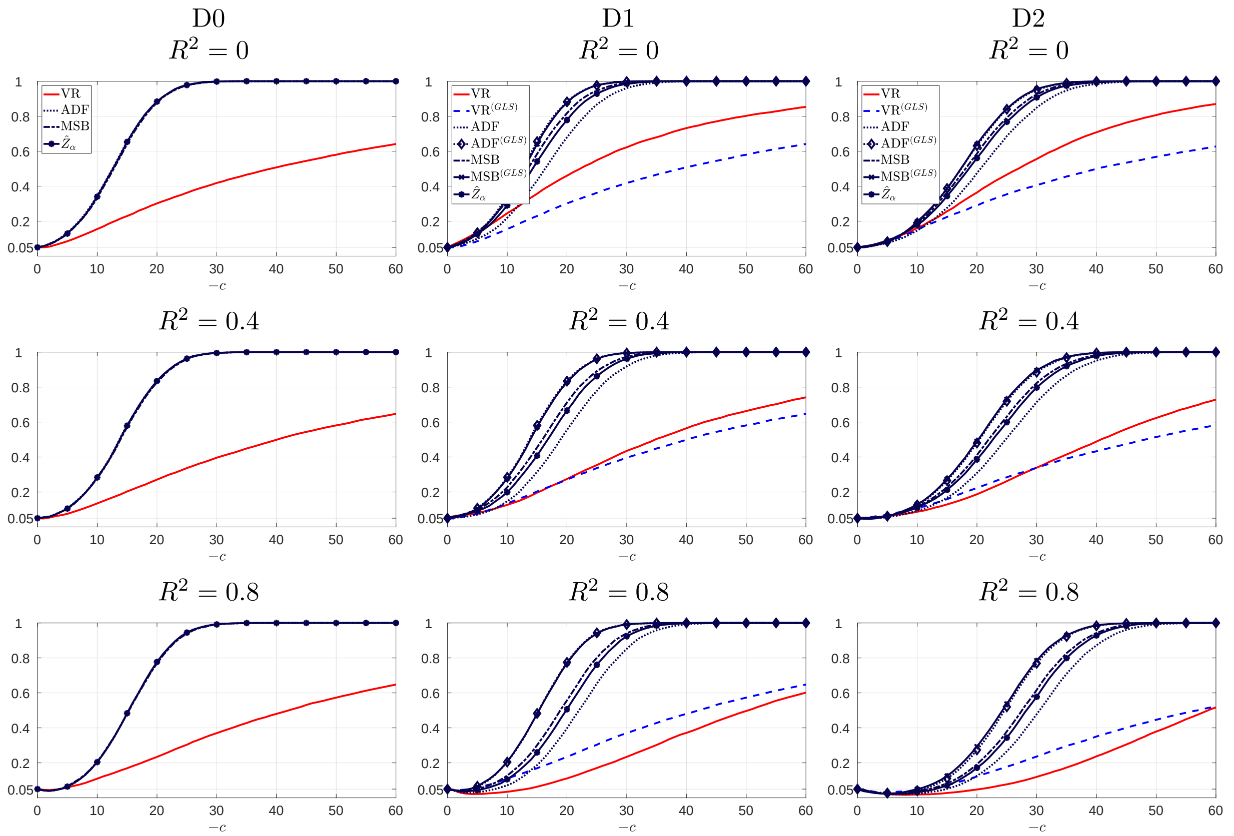


Figure 3.1: Asymptotic power of the tests at the nominal 5% level for H_0: $\rho = 1$ under the local alternative $\rho = \rho_T = 1 + c/T$ in cases D0 (first column), D1 (second column), and D2 (third column) for $m = 1$.

Note: The results are based on 10,000 Monte Carlo replications and standard Brownian motions are approximated by normalized partial sums of 10,000 i.i.d. standard normal random variables.

for $t = 1, \ldots, T$. The regression errors are generated as $u_t = \rho u_{t-1} + \xi_t$ using the following five different short-run dynamics:

$$\xi_t = \begin{cases} \varepsilon_t & \text{(IID)} \\ \phi\xi_{t-1} + \varepsilon_t & \text{(AR)} \\ \varepsilon_t - \theta\varepsilon_{t-1} & \text{(MA)} \\ \phi\xi_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{(ARMA)} \\ \sqrt{h_t}\varepsilon_t, \quad h_t = (1 - a_1 - a_2) + a_1\xi_{t-1}^2 + a_2 h_{t-1} & \text{(GARCH)}, \end{cases}$$

for $t = -99, \ldots, 0, 1, \ldots, T$.[13] The vectors $[\varepsilon_t, v_t]'$, are i.i.d. across t and follow a zero-mean bivariate normal distribution with covariance matrix $\Sigma := \begin{bmatrix} 1 & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & 1 \end{bmatrix}$. To ensure stationarity of ξ_t, the parameter values are restricted to $|\phi|, |\theta| < 1$ and $a_1 + a_2 < 1$, with $a_1, a_2 \geq 0$. In the IID and GARCH cases, the long-run covariance matrix of $[\xi_t, v_t]'$ is given by $\Omega_{\text{IID}} = \Omega_{\text{GARCH}} = \Sigma$, whereas

[13]We set $u_{-100} = \xi_{-100} = 0$ and $h_{-100} = 1$. The period $t = -99, \ldots, 0$ serves as a burn-in period to eliminate these starting-value effects. Section 3.4.3 analyzes the effect of a large initial value u_0 on the performance of the tests in finite samples.

in the ARMA case it is given by $\Omega_{\text{ARMA}} := \begin{bmatrix} \frac{(1-\theta)^2}{(1-\phi)^2} & \frac{(1-\theta)\sigma_{\varepsilon v}}{1-\phi} \\ \frac{(1-\theta)\sigma_{\varepsilon v}}{1-\phi} & 1 \end{bmatrix}$. In the AR and MA cases, the long-run covariance matrices Ω_{AR} and Ω_{MA} can be directly deduced from Ω_{ARMA} by setting θ and ϕ, respectively, equal to zero. Although the long-run covariance matrix of $[\xi_t, v_t]'$ differs across types of short-run dynamics, the only nuisance parameter affecting local asymptotic power of the residuals-based no-cointegration tests, i. e., R^2 (compare the discussion in Section 3.3.3), is equal to $\sigma^2_{\varepsilon v}$ in each case. The DGP thus allows to analyze the effects of different short-run dynamics in ξ_t on the performance of residuals-based no-cointegration tests in finite samples while controlling for effects of R^2.[14] In contrast, the local asymptotic power analysis in Section 3.3.3 only allows to assess the effect of R^2.

Remark 3.5. *Pesavento (2007, p. 127) and Perron and Rodríguez (2016, p. 99) state that local asymptotic power results serve as a useful indicator for the performance of residuals-based no-cointegration tests in finite samples. However, Pesavento (2007) only considers a DGP similar to our IID case, while Perron and Rodríguez (2016) allow the errors to be serially correlated – using a vector autoregression model of order one to generate $[\xi_t, v_t]'$ – but restrict the diagonal and off-diagonal elements of the long-run covariance matrix of $[\xi_t, v_t]'$ to one and a constant r, respectively, such that $R^2 = r^2$. Perron and Rodríguez (2016, p. 99) correctly point out that holding the long-run covariance matrix fixed implies that changes in the autoregressive parameters are compensated by changes in the covariance matrix of $[\xi_t, v_t]'$. Consequently, changes in the autoregressive parameters do not affect the performance of the tests in their DGP. In contrast, we shall see below that the DGP considered in this paper is able to detect severe effects of different short-run dynamics on the performance of residuals-based no-cointegration tests. The finite sample results in this section thus differ considerably from the local asymptotic power results in Section 3.3.3.*

Parameters are chosen as follows: In all cases, we set $\beta = 1$. Moreover, we set $\phi, \theta \in \{0.3, 0.6, 0.9\}$ in the AR and MA models, $(\phi, \theta) \in \{(0.3, 0.6), (0.3, 0.3), (0.6, 0.3)\}$ in the ARMA model, and $(a_1, a_2) \in \{(0.05, 0.94), (0.01, 0.98)\}$ in the GARCH model. In addition, we set $x_0 = 0$ and $\mu = 0$ in case D0, $x_0 = 1$, $\mu = 0$ and $\tau = 1$ in case D1, and $x_0 = 1$, $\mu = 1$ and $\tau = [1, 1]'$ in case D2. We present results for $R^2 \in \{0, 0.4, 0.8\}$ by choosing $\sigma_{\varepsilon v} \geq 0$ accordingly and $T \in \{100, 250\}$. All results are based on 5,000 Monte Carlo replications and all tests are carried out at the nominal 5% level.

We compare the variance ratio test with the ADF test, the MSB test, and the $\widehat{Z}_\alpha$ test in terms of empirical size and size-corrected power. In cases D1 and D2, the tests are based on OLS detrended data. In addition, we also consider the variance ratio test and the ADF test based on GLS detrended data, indicated by the superscript "(GLS)", where $\bar{c}$ is chosen as suggested in Table 3.3 in Appendix 3.7.1 and in Table 1 of Perron and Rodríguez (2016), respectively.

In contrast to the variance ratio test, the ADF, MSB, and $\widehat{Z}_\alpha$ tests require tuning parameter choices. The number of lags for the ADF test and the MSB test is selected using AIC. For the ADF test and its GLS-version, we also analyze the results based on a modified AIC (MAIC) criterion proposed in Ng and Perron (2001) for unit root testing in an observed univariate time

[14] Controlling for effects of R^2 allows to compare the finite sample results in this section with the local asymptotic power results obtained in Section 3.3.3.

series, taking into account further modifications suggested in Perron and Qu (2007).[15] We also analyze the performance of the GLS-version of the MSB test in combination with the MAIC. The use of MAIC is indicated by the superscript "$*$". For both information criteria, the minimal number of lags is zero and the maximal number of lags is restricted to not exceed $\lfloor 12(T/100)^{1/4} \rfloor$, the upper bound suggested in Perron and Qu (2007).[16] The $\widehat{Z}_\alpha$ test is based on a kernel estimator of a long-run variance parameter. We present results for the quadratic spectral (QS) kernel together with the corresponding data-dependent bandwidth-selection rule of Andrews (1991).[17] For convenience, Appendix 3.7.4 describes the construction of the ADF, MSB, and $\widehat{Z}_\alpha$ tests in more detail.[18]

3.4.1 Empirical Size

Table 3.1 presents empirical sizes of the tests under the null hypothesis of no cointegration ($\rho = 1$) in cases D1 and D2 for $T = 100$. The results directly reveal that the performance of the tests depends heavily on the short-run dynamics in ξ_t. In particular, size distortions in the MA and ARMA cases can be more severe than in the IID, AR, and GARCH cases. Moreover, relative to the IID case with $R^2 = 0$, changing the short-run dynamics in ξ_t can have much larger adverse effects on the performance of the tests than increasing R^2 to 0.8.

Focusing on the performance of the different tests in detail reveals that the variance ratio test is less size-distorted than the ADF and MSB tests. The variance ratio test often also outperforms the $\widehat{Z}_\alpha$ test, especially in the MA case with $\theta < 0.9$ and in the ARMA(0.3,0.6) case. GLS detrending worsens the performance of the variance ratio test, especially in case D2, and of the ADF test, but the GLS detrended version of the ADF test in combination with the MAIC performs relatively well. Similarly, the GLS detrended version of the MSB test in combination with the MAIC has much smaller size distortions than the MSB test, but it tends to be very conservative.

Comparing the ADF test with the ADF* test reveals that using the MAIC criterion is also advantageous under OLS detrending. In most cases, the variance ratio test and the ADF* test perform similarly, but the ADF* test has some performance advantages over the variance ratio test in the MA case with $\theta > 0.3$ and in the ARMA(0.3,0.6) case. These performance advantages are the more pronounced the larger R^2.

Increasing the sample size reduces the size distortions of the tests, especially for VR$^{(\mathrm{GLS})}$ in case D2, and makes the MSB$^{(\mathrm{GLS})*}$ less conservative, compare Table 3.6 in Appendix 3.7.2, which presents the results for $T = 250$. Beyond that, the results are qualitatively similar to those for

[15] Perron and Qu (2007) suggest to construct the ADF statistic based on GLS detrended data and to determine the number of lags in the auxiliary regression based on OLS detrended data.

[16] The Supplementary Material provides results based on BIC and MBIC, with MBIC defined analogously to MAIC (compare Section 3.7.4 in the Appendix). The tests based on MBIC perform similar to those based on MAIC. In contrast, the tests based on BIC are more prone to severe size distortions than those based on AIC, but often reveal some power advantages compared to their AIC based counterparts. These differences in size-corrected power, however, do not alter the overall picture emerging from the discussion in Section 3.4.2. In particular, the tests based on both BIC and MBIC suffer from similar power reversal problems as observed for the tests based on AIC and MAIC in Section 3.4.2.

[17] Using the Bartlett kernel instead of the QS kernel often leads to slightly larger size distortions.

[18] When the ADF and MSB tests are used in applications, it seems to be more popular to perform OLS detrending rather than GLS detrending and employing the AIC or BIC rather than the MAIC or MBIC. The following results thus also allow to assess whether practitioners should stick to these "default" choices or not.

$T = 100$. Overall, the findings also hold in case D0, although the variance ratio test seems to be rather conservative in this case even for $T = 250$, compare Table 3.5 in Appendix 3.7.2.

3.4.2 Size-Corrected Power

To analyze the finite sample properties of the tests under deviations from the null hypothesis, we generate data under the alternatives $\rho = \rho_T = 1 + c/T$ using an equidistant grid of 21 points over the interval $[0, 60]$ for the values of $-c$.[19] To account for the large performance differences between the tests in terms of empirical sizes under the null hypothesis ($c = 0$), the analysis focuses on size-corrected (empirical) power. To this end, test decisions are based on case-specific empirical critical values obtained from simulations under the null hypothesis rather than on asymptotic critical values. All size-corrected power curves thus start at the nominal 5% level.

Size-corrected power of the tests decreases when R^2 becomes larger. However, for all three deterministic specifications different short-run dynamics in ξ_t can have more adverse effects on the performance of the tests than increasing R^2 from zero to 0.8. Figure 3.2 presents the size-corrected power curves of the tests for the different short-run dynamics in case D2 with $T = 100$ and $R^2 = 0.4$.[20] For most short-run dynamics, the $\widehat{Z}_\alpha$ test performs best, while the MSB test performs worse. The ranking of the remaining tests is highly case dependent.

In general, the variance ratio test performs relatively well and much better than suggested by the local asymptotic power results. In particular, it outperforms the $\text{MSB}^{(\text{GLS})*}$ test and both the OLS and GLS versions of the ADF^* test – thus its biggest competitors under the null hypothesis – in the MA cases and in the ARMA(0.3,0.6). In the MA case with $\theta = 0.9$, the variance ratio test is even the most powerful test, but this result should be interpreted with caution, as all tests are heavily size-distorted in this case. In other cases, the variance ratio test is less powerful than its competitors for small deviations from the null hypothesis but outperforms the $\text{MSB}^{(\text{GLS})*}$ test and both the OLS and GLS versions of the ADF^* test for larger deviations from the null hypothesis.[21]

Importantly, for some short-run dynamics, the power curves of the $\text{MSB}^{(\text{GLS})*}$ test and both the OLS and GLS versions of the ADF^* test reveal power reversal problems. This phenomenon is well known in the unit root literature for the MSB and ADF tests in combination with the AIC criterion in the MA case with a large θ. However, the MAIC criterion has been introduced as a possible way to prevent this degeneracy (cf. the discussion in Perron and Qu, 2007). Comparing the power curves of the ADF^* and $\text{ADF}^{(\text{GLS})*}$ tests with those of the ADF and $\text{ADF}^{(\text{GLS})}$ tests reveals that using the MAIC criterion rather than the AIC criterion indeed prevents power reversal problems associated with the ADF test in the MA case with $\theta = 0.9$.[22] In other cases, however, the tests based on MAIC still suffer from power reversal problems (e. g., in the AR case with $\phi = 0.6$ and

[19] Note that the alternatives move closer to the null hypothesis as the sample size increases. Hence, for (very) large sample sizes, the size-corrected power results should match the local asymptotic power results in Figure 3.1, irrespective of the short-run dynamics in ξ_t.

[20] Results in case D0 and D1 are qualitatively similar. The Supplementary Material (Figures E.1 – E.18 and E.28 – E.45) provides figures similar to Figure 3.2 for all three deterministic specifications and all combinations of R^2 and T.

[21] For some short-run dynamics the power curves of all tests considered can decline slightly below α for small deviations from the null, especially for $R^2 = 0.8$. This is in line with the local asymptotic power curves in Figure 3.1.

[22] In the MA case with $\theta = 0.9$ the $\text{MSB}^{(\text{GLS})*}$ test still suffers from power reversal problems.

Table 3.1: Empirical sizes of the tests in cases D1 and D2 for $T = 100$.

			AR			MA			ARMA			GARCH	
R^2	Test	IID	0.3	0.6	0.9	0.3	0.6	0.9	(0.3,0.6)	(0.3,0.3)	(0.6,0.3)	(0.05,0.94)	(0.01,0.98)
Deterministic specification D1													
0	VR	0.05	0.04	0.03	0.01	0.07	0.16	0.72	0.10	0.05	0.03	0.05	0.05
	$VR^{(GLS)}$	0.12	0.11	0.08	0.03	0.15	0.26	0.82	0.19	0.12	0.09	0.12	0.12
	ADF	0.08	0.07	0.07	0.05	0.13	0.23	0.83	0.24	0.08	0.06	0.09	0.08
	$ADF^{(GLS)}$	0.11	0.10	0.09	0.06	0.17	0.26	0.70	0.28	0.11	0.09	0.11	0.11
	ADF^{*}	0.04	0.02	0.02	0.02	0.05	0.07	0.34	0.08	0.04	0.01	0.04	0.04
	$ADF^{(GLS)*}$	0.06	0.03	0.03	0.01	0.08	0.10	0.37	0.12	0.06	0.02	0.06	0.06
	MSB	0.11	0.14	0.14	0.15	0.14	0.21	0.77	0.22	0.11	0.14	0.12	0.11
	$MSB^{(GLS)*}$	0.02	0.02	0.03	0.03	0.04	0.05	0.27	0.08	0.02	0.01	0.03	0.02
	$\widehat{Z}_\alpha$	0.05	0.02	0.01	0.01	0.19	0.72	1.00	0.33	0.05	0.01	0.05	0.05
0.4	VR	0.05	0.04	0.03	0.01	0.07	0.19	0.78	0.11	0.05	0.03	0.05	0.05
	$VR^{(GLS)}$	0.12	0.11	0.09	0.05	0.16	0.29	0.85	0.20	0.12	0.09	0.12	0.12
	ADF	0.08	0.08	0.08	0.05	0.14	0.28	0.87	0.26	0.08	0.07	0.09	0.08
	$ADF^{(GLS)}$	0.11	0.11	0.10	0.07	0.17	0.29	0.75	0.28	0.11	0.09	0.11	0.11
	ADF^{*}	0.04	0.02	0.03	0.02	0.05	0.07	0.38	0.08	0.04	0.02	0.04	0.04
	$ADF^{(GLS)*}$	0.06	0.03	0.03	0.02	0.08	0.10	0.42	0.12	0.06	0.03	0.06	0.06
	MSB	0.12	0.14	0.14	0.14	0.14	0.24	0.82	0.25	0.12	0.14	0.12	0.12
	$MSB^{(GLS)*}$	0.03	0.02	0.03	0.03	0.04	0.05	0.30	0.08	0.03	0.02	0.03	0.03
	$\widehat{Z}_\alpha$	0.06	0.03	0.03	0.02	0.21	0.79	1.00	0.39	0.06	0.02	0.06	0.06
0.8	VR	0.05	0.05	0.06	0.01	0.09	0.31	0.87	0.15	0.05	0.05	0.05	0.05
	$VR^{(GLS)}$	0.12	0.12	0.14	0.06	0.18	0.43	0.93	0.26	0.12	0.12	0.12	0.12
	ADF	0.08	0.10	0.16	0.06	0.15	0.41	0.92	0.33	0.08	0.11	0.10	0.08
	$ADF^{(GLS)}$	0.11	0.13	0.20	0.10	0.19	0.40	0.79	0.35	0.11	0.15	0.12	0.11
	ADF^{*}	0.04	0.04	0.07	0.03	0.05	0.10	0.45	0.10	0.04	0.05	0.04	0.04
	$ADF^{(GLS)*}$	0.06	0.06	0.11	0.05	0.08	0.14	0.48	0.14	0.06	0.07	0.06	0.06
	MSB	0.11	0.13	0.16	0.12	0.16	0.35	0.89	0.32	0.11	0.14	0.12	0.12
	$MSB^{(GLS)*}$	0.03	0.03	0.06	0.03	0.04	0.06	0.36	0.09	0.03	0.04	0.03	0.03
	$\widehat{Z}_\alpha$	0.06	0.07	0.14	0.04	0.36	0.95	1.00	0.59	0.06	0.08	0.06	0.06
Deterministic specification D2													
0	VR	0.05	0.04	0.02	0.00	0.09	0.26	0.93	0.14	0.05	0.03	0.06	0.05
	$VR^{(GLS)}$	0.30	0.26	0.20	0.07	0.36	0.59	0.99	0.45	0.30	0.22	0.30	0.30
	ADF	0.10	0.10	0.08	0.05	0.19	0.35	0.91	0.35	0.10	0.08	0.10	0.10
	$ADF^{(GLS)}$	0.14	0.13	0.12	0.07	0.23	0.37	0.85	0.39	0.14	0.11	0.14	0.14
	ADF^{*}	0.03	0.01	0.02	0.01	0.06	0.09	0.51	0.11	0.03	0.01	0.04	0.03
	$ADF^{(GLS)*}$	0.06	0.02	0.03	0.01	0.08	0.12	0.54	0.15	0.06	0.01	0.06	0.06
	MSB	0.15	0.20	0.22	0.25	0.19	0.28	0.85	0.27	0.15	0.22	0.17	0.15
	$MSB^{(GLS)*}$	0.01	0.01	0.03	0.03	0.04	0.06	0.42	0.09	0.01	0.01	0.02	0.01
	$\widehat{Z}_\alpha$	0.04	0.01	0.00	0.00	0.24	0.87	1.00	0.43	0.04	0.00	0.05	0.04
0.4	VR	0.05	0.04	0.03	0.00	0.09	0.30	0.94	0.15	0.05	0.03	0.05	0.05
	$VR^{(GLS)}$	0.30	0.26	0.22	0.08	0.38	0.64	0.99	0.47	0.30	0.23	0.30	0.30
	ADF	0.11	0.10	0.09	0.05	0.20	0.40	0.92	0.38	0.11	0.09	0.11	0.11
	$ADF^{(GLS)}$	0.14	0.13	0.11	0.08	0.23	0.41	0.86	0.41	0.14	0.11	0.14	0.14
	ADF^{*}	0.04	0.01	0.02	0.01	0.06	0.10	0.48	0.11	0.04	0.01	0.04	0.04
	$ADF^{(GLS)*}$	0.06	0.02	0.03	0.01	0.09	0.14	0.51	0.16	0.06	0.01	0.06	0.06
	MSB	0.16	0.19	0.20	0.23	0.19	0.33	0.86	0.30	0.16	0.21	0.17	0.16
	$MSB^{(GLS)*}$	0.02	0.01	0.02	0.03	0.04	0.07	0.39	0.10	0.02	0.01	0.02	0.02
	$\widehat{Z}_\alpha$	0.04	0.01	0.01	0.00	0.28	0.93	1.00	0.48	0.04	0.01	0.05	0.05
0.8	VR	0.05	0.05	0.06	0.00	0.11	0.48	0.98	0.20	0.05	0.04	0.05	0.05
	$VR^{(GLS)}$	0.29	0.29	0.32	0.09	0.43	0.81	1.00	0.56	0.29	0.28	0.30	0.29
	ADF	0.11	0.11	0.15	0.04	0.22	0.57	0.93	0.47	0.11	0.11	0.11	0.11
	$ADF^{(GLS)}$	0.13	0.15	0.20	0.06	0.25	0.55	0.86	0.48	0.13	0.15	0.14	0.14
	ADF^{*}	0.04	0.03	0.05	0.01	0.06	0.14	0.45	0.14	0.04	0.03	0.04	0.04
	$ADF^{(GLS)*}$	0.06	0.05	0.08	0.01	0.09	0.18	0.48	0.18	0.06	0.05	0.06	0.06
	MSB	0.16	0.17	0.17	0.17	0.21	0.48	0.88	0.41	0.16	0.17	0.17	0.16
	$MSB^{(GLS)*}$	0.02	0.02	0.03	0.01	0.04	0.10	0.36	0.12	0.02	0.01	0.02	0.02
	$\widehat{Z}_\alpha$	0.05	0.05	0.09	0.00	0.46	1.00	1.00	0.70	0.05	0.05	0.05	0.05

Note: The superscripts "(GLS)" and "$*$" indicate GLS detrending instead of OLS detrending and the use of MAIC instead of AIC, respectively.

in the ARMA case with $\phi = 0.6$ and $\theta = 0.3$) and perform worse than those based on AIC (e. g., in the IID and GARCH cases).

With respect to preliminary data detrending, the results suggest that the GLS versions of the ADF and ADF* tests perform similarly to the OLS versions of the test, whereas the GLS version of the variance ratio test is often (considerably) less powerful than the OLS version.[23] On the other hand, the $\text{MSB}^{(\text{GLS})*}$ test is in most cases clearly more powerful than the MSB test.

Finally, increasing the sample size is beneficial for the power of the tests, but results for $T = 250$ are qualitatively similar, see, e. g., Figure 3.6 in Appendix 3.7.2 in case D2 with $R^2 = 0.4$. It is noteworthy that short-run dynamics in ξ_t even have adverse effects – including power reversal problems for the ADF tests and the $\text{MSB}^{(\text{GLS})*}$ test – on the power of the tests for larger sample sizes, especially in case D2 with $R^2 = 0.8$. However, as T increases, short-run dynamics in ξ_t become less pervasive such that the size-corrected power results approach the local asymptotic power results analyzed in Section 3.3.3, see, e. g., the results for $T = 1{,}000$ in case D2 with $R^2 = 0.4$ in Figure 3.7 in Appendix 3.7.2.[24] In general, we notice that for small to medium sample sizes, the finite sample performance of the tests are dominated by the short-run dynamics in ξ_t such that size-corrected power results deviate considerably from local asymptotic power results. Pesavento (2007) and Perron and Rodríguez (2016) do not detect these discrepancies between the finite sample performance of the tests and the local asymptotic power results because their DGPs only allow for very mild effects of short-run dynamics (compare the discussion in Remark 3.5). As the sample size increases, the short-run dynamics in ξ_t become less important and the performance of the tests is then dominated by R^2. In this case, size-corrected power results are more similar to local asymptotic power results. The sample size required for a sufficient degree of similarity between finite sample results and local asymptotic power results generally increases with the order of the deterministic component and R^2, i. e., the required sample size is much smaller in case D0 with $R^2 = 0$ than in case D2 with $R^2 = 0.8$.

To provide an overall assessment, it is important to note that residuals-based no-cointegration tests are used to avoid analyzing the relationship between stochastically trending (economic) variables in a spurious regression framework. For practitioners, small upward size distortions may thus have a higher weight in the overall assessment of the general performance of the tests than high power under the alternative. We thus conclude that both the variance ratio test and the ADF* test perform best among the tests considered in this simulation study and thus may be deemed useful for testing for cointegration in applications. However, practitioners should be aware of the individual specific shortcomings of the tests.[25]

[23] The local asymptotic power results suggest that the GLS version of the variance ratio test is generally more powerful than its OLS version for $R^2 = 0.8$. However, the finite sample results do not reveal a similar advantage of the GLS version over the OLS version in terms of size-corrected power, compare Figure 3.5 in Appendix 3.7.2 for the size-corrected power curves in case D2 for $T = 100$ and note that results in case D1 are similar, compare Figure E.6 in the Supplementary Material. For $T = 250$, GLS detrending even seems to be disadvantageous in cases D1 and D2, compare Figures E.15 and E.18 in the Supplementary Material.

[24] Figures E.19 – E.27 in the Supplementary Material show the results for $T = 1{,}000$ and all values of R^2 in cases D0, D1, and D2. Moreover, Figures E.46 – E.54 show the results for the ADF and MSB tests based on (M)BIC.

[25] To make the analysis more robust against individual specific shortcomings of the tests, practitioners could use a Fisher-type combination test or a "union-of-rejections" decision rule (cf. Bayer and Hanck, 2013).

3.4.3 A Large Initial Value u_0

Several contributions in related literature highlight strong effects of a large initial value u_0 on the power of unit root (see, e. g., Harvey *et al.*, 2009, and references therein) and residuals-based no-cointegration (see, e. g., Perron and Rodríguez, 2016) tests. To comment briefly on the effect of a large (in a well-defined sense) initial value u_0 on the power of the variance ratio test, we use the same set-up as before but generate initial values u_0 of order $T^{1/2}$.[26] To this end, we follow Harvey *et al.* (2009) and Perron and Rodríguez (2016) and generate u_0 as $u_0 = \lambda_u/(1-\rho_T^2)^{1/2}$, where $\rho_T = 1 + c/T$, $c < 0$, and λ_u is a fixed constant.[27] We compare the performance of the variance ratio test with the performance of the tests already analyzed in the previous subsections. To simplify the comparison with the size-corrected power curves discussed in Section 3.4.2, we again consider size-corrected power results of the tests, using the same empirical critical values as in the previous subsection.

Figure 3.3 presents the results for $c = -20$, $T = 100$ and $R^2 = 0.4$ in case D2. Again, the performance of the tests highly depends on the short-run dynamics in ξ_t. In the MA case with $\theta = 0.9$ and in the ARMA(0.3,0.6) case, size-corrected power of all tests decreases as λ_u moves away from zero, whereas in the AR case with $\phi = 0.9$, size-corrected power of the tests is rather unaffected by changes in λ_u. In the majority of cases, however, size-corrected power of the ADF and of the ADF* tests increases as λ_u moves away from zero, with clear performance advantages of ADF* over ADF, whereas power of the other tests decreases for larger initial values u_0.[28] These adverse effects are more pronounced for the GLS-versions of the tests than for their OLS counterparts, which fits well to the findings in Perron and Rodríguez (2016). Noticeably, the VR test is, except for the MA and ARMA(0.3,0.6) cases, rather unaffected by initial values u_0 corresponding to small to medium values of λ_u. Results for $T = 250$ and/or other choices of c and R^2 are qualitatively similar and we observe comparable effects also in case D1.

[26] It is clear from the proof of Proposition 3.1 that u_0 does not affect the distribution of the VR statistic under the null hypothesis in cases D1 and D2, neither in the limit nor in finite samples. In case D0, however, an initial value u_0 of order $T^{1/2}$ does affect the distribution of the VR statistic under the null hypothesis both in finite samples and in the limit. Since in applications an intercept is typically included in (3.2), we do not comment upon this issue any further.

[27] Effects in case u_0 is drawn from a normal distribution with mean zero and variance $\lambda_u^2/(1-\rho_T^2)$ are less pronounced.

[28] Replacing (M)AIC with (M)BIC in the construction of the ADF and MSB tests yields similar results.

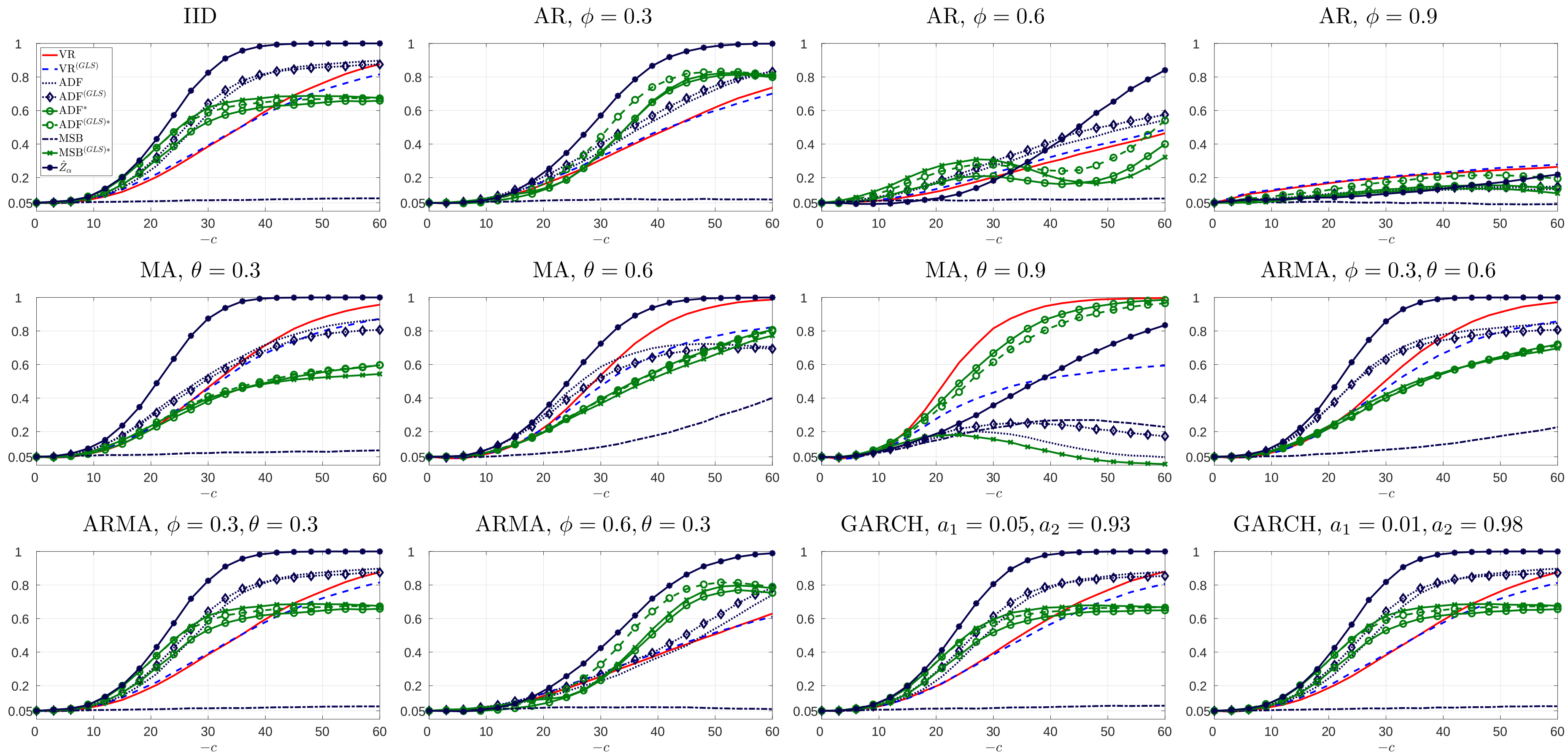


Figure 3.2: Size-corrected power of the tests at the nominal 5% level for H_0: $\rho = 1$ under the alternative $\rho = 1 + c/T$ in case D2, for $T = 100$ and $R^2 = 0.4$.

Note: The superscripts "(GLS)" and "$*$" indicate GLS detrending instead of OLS detrending and the use of MAIC instead of AIC, respectively.

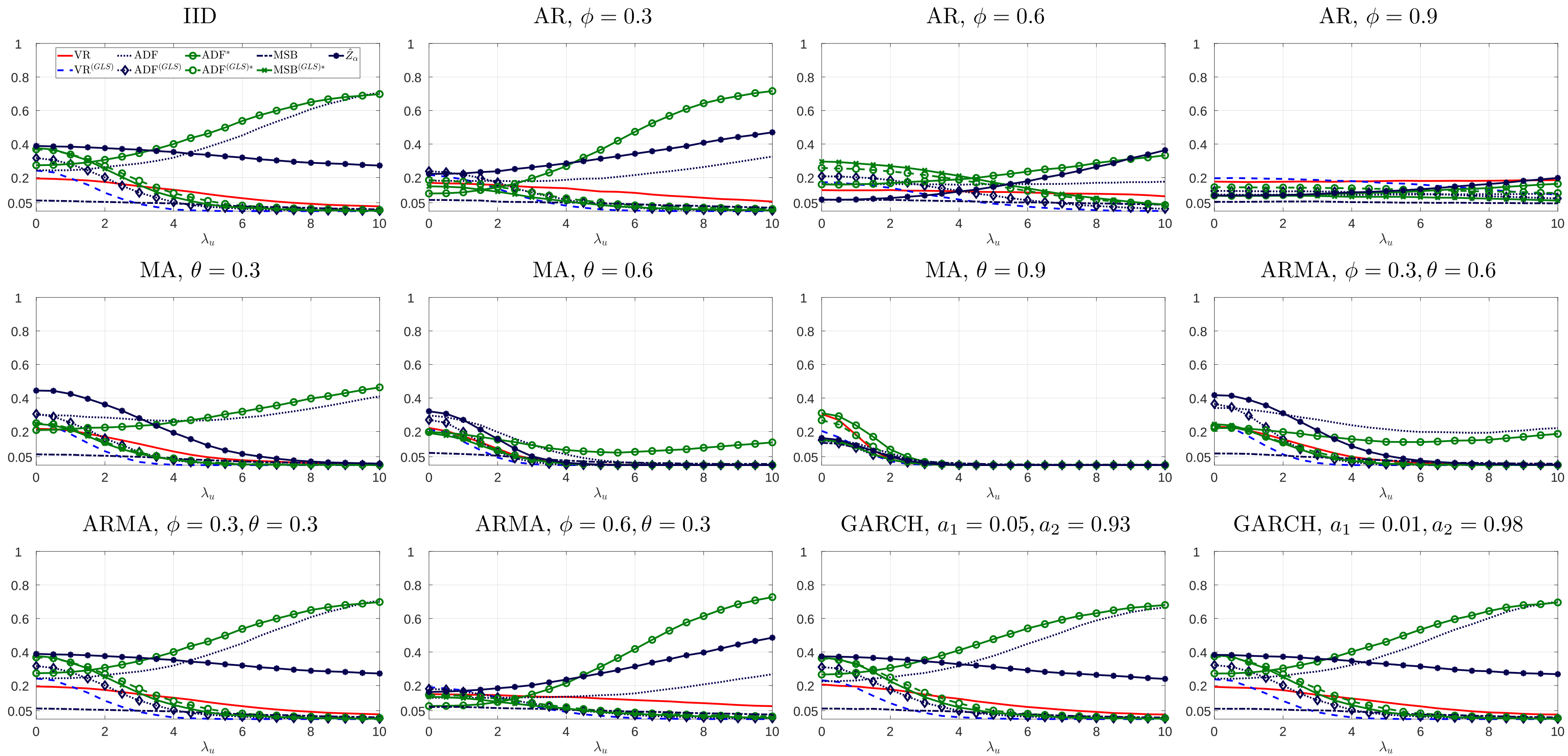


Figure 3.3: Size-corrected power of the tests at the nominal 5% level for H_0: $\rho = 1$ in the presence of a large non-random initial value u_0 under the alternative $\rho = 1 - 20/T$ in case D2, for $T = 100$ and $R^2 = 0.4$.

See note to Figure 3.2.

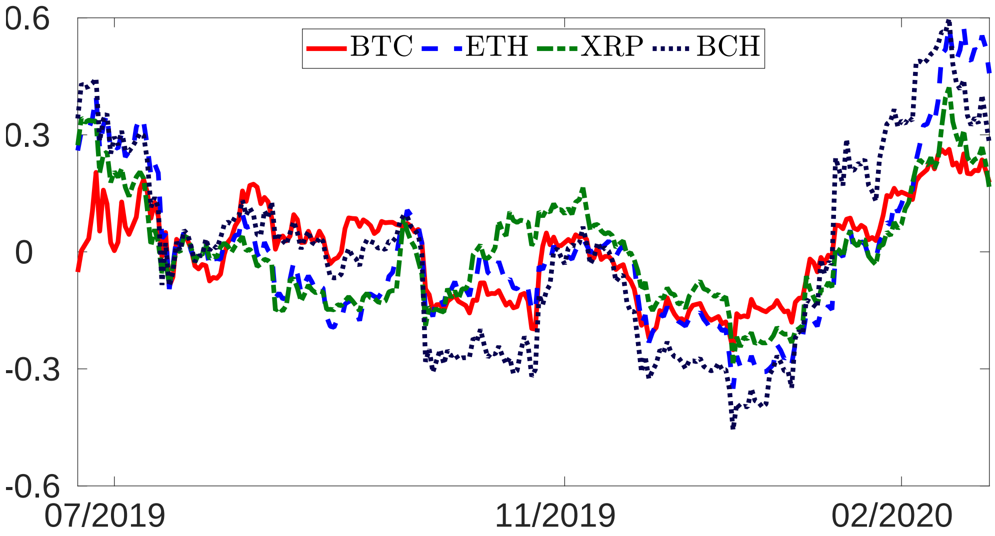


Figure 3.4: OLS detrended log prices of cryptocurrencies from June 21, 2019 to February 25, 2020.

3.5 Empirical Illustration

This illustration considers daily closing prices (in U.S. dollar) of cryptocurrencies. In particular, we test for cointegration between the four cryptocurrencies with the highest market capitalization (as of February 25, 2020, excluding stable coins), namely Bitcoin (BTC), Ethereum (ETH), XRP and Bitcoin Cash (BCH). We focus on the latest $T = 100$, $T = 200$ and $T = 250$ time points of the data set analyzed in detail in Keilbar and Zhang (2021), which ends in February 25, 2020.[29] Figure 3.4 shows the OLS detrended logarithms of the $T = 250$ daily prices of the four cryptocurrencies between June 21, 2019 and February 25, 2020. Keilbar and Zhang (2021) find evidence that the four series are integrated of order one. In addition, maybe apart from the beginning of the period, Figure 3.4 indicates a strong co-movement of the four series.

We choose Bitcoin (i. e., the cryptocurrency with the highest market capitalization) as the left-hand side variable and calculate the nine test statistics considered in the previous section allowing for a deterministic time trend (D2). Table 3.2 provides the results for the three periods. For each period,

Table 3.2: Realizations of test statistics

	VR	VR$^{(GLS)}$	ADF	ADF$^{(GLS)}$	ADF*	ADF$^{(GLS)*}$	MSB	MSB$^{(GLS)*}$	$\widehat{Z}_\alpha$
$T = 100$	**0.0010**	**0.0020**	**-4.5413**	-4.0909	**-4.5413**	-4.0909	0.1430	0.1465	-31.6678
$T = 200$	**0.0012**	0.0087	**-5.0965**	**-4.3508**	**-4.5956**	-3.9285	**0.1050**	0.1362	**-43.6046**
$T = 250$	0.0045	0.0420	**-5.2109**	-3.5599	**-5.3029**	-3.5599	0.1214	0.1486	**-42.7777**

Notes: **Bold** numbers indicate significance at the 5% level. The superscripts "(GLS)" and "$*$" indicate GLS detrending instead of OLS detrending and the use of MAIC instead of AIC, respectively. Except for the ADF$^{(GLS)}$ test and the MSB test in the $T = 200$ period, significance of results does not change when (M)BIC replaces (M)AIC. The $\widehat{Z}_\alpha$ test is based on the QS kernel. Using the Bartlett kernel instead leads to similar results.

[29]The data set is available on https://github.com/QuantLet/CryptoDynamics/blob/master/CryptoDynamics_Series/logprice.csv (accessed: September 4, 2022). Choosing the three (nested) periods is merely to analyze the relationship for different sample sizes similar to those used in Section 3.4. This should not be interpreted as a monitoring strategy.

test decisions at the 5% significance level are heterogeneous across tests. The ADF and ADF* tests reject the null hypothesis of no cointegration in all three periods, whereas the ADF$^{(\mathrm{GLS})*}$ test and the MSB$^{(\mathrm{GLS})*}$ test never reject. For the $T = 100$ and $T = 200$ periods, the test decisions of the VR test are in line with the those of the ADF and ADF* tests, but for the $T = 250$ period the VR test leads to an opposing result. Comparing the values of the test statistics between the $T = 200$ and the $T = 250$ periods reveals that the ADF and ADF* test statistics decrease (i. e., become more significant) as the sample size increases, whereas the remaining seven test statistics – including the VR statistic – increase (i. e., become less significant). Thus, it seems that a more detailed analysis is needed to decide whether the series are cointegrated in the $T = 250$ period or not. On the other hand, in light of the simulation results in Section 3.4, the results in the $T = 100$ and $T = 200$ periods seem to provide reliable evidence for the presence of cointegration.[30]

For investors, the presence of a cointegrating relationship between the four cryptocurrencies with the highest market capitalization clearly complicates diversification of their cryptocurrency portfolios. On the other hand, it allows them to use a cointegration-based trading strategy to increase profits (cf, e. g., Leung and Nguyen, 2019; Keilbar and Zhang, 2021).

3.6 Summary and Conclusions

This paper derives asymptotic theory for Breitung's (2002) nonparametric variance ratio unit root test when applied to regression residuals and analyzes its asymptotic and finite sample properties in this case. The results reveal that the variance ratio test has smaller local asymptotic power than its competitors. However, in finite samples, short-run dynamics in the errors can have severe effects on the performance of the tests both under the null hypothesis and under the alternative. In terms of empirical size, the variance ratio test and the ADF test based on a modified AIC (BIC) perform best and, in particular, outperform the ADF test based on the (unmodified) AIC (BIC). As both tests also perform relatively well in terms of size-corrected power, they may be deemed useful for testing for cointegration in applications. In particular, both test behave nicely in the important case where the regression errors have a moving average component with a small negative coefficient. However, practitioners should be aware that under some short-run dynamics the variance ratio test can be less powerful than its competitors under small deviations from the null hypothesis (but then quickly catches up), whereas the ADF test is prone to power reversal problems for larger deviations from the null hypothesis – whether or not the AIC (BIC) is modified. Finally, an empirical illustration testing for cointegration between daily prices of cryptocurrencies shows the usefulness of the variance ratio test in practice.

Future research could strive to make the ADF test (and the MSB test) more robust against power reversal problems. Moreover, it might be possible to increase local asymptotic power of the variance ratio test by extending Nielsen's (2009) family of unit root tests – which contains Breitung's (2002) test as a special case – to cointegration testing. However, this comes at the cost of introducing an index parameter to be chosen by the practitioner. A detailed simulation study then needs to

[30]Using different methods and a larger number of cryptocurrencies Keilbar and Zhang (2021) and Bykhovskaya and Gorin (2022b) also find evidence for cointegration in the cryptocurrency market.

assess whether increasing local asymptotic power of the variance ratio test is also beneficial for its finite sample performance both under the null hypothesis and under the alternative.

Acknowledgements

I thank Christoph Hanck, Carsten Jentsch and Martin Wagner for helpful comments.

3.7 Appendix

3.7.1 Values of $\bar{c}$ and Asymptotic Critical Values

Table 3.3: Values of $\bar{c}$ for the variance ratio GLS test

	m				
Deterministic specification	1	2	3	4	5
D1	-40.25	-46.25	-53.75	-55.75	-60.00
D2	-48.25	-55.25	-56.50	-65.00	-68.75

Notes: The values of $\bar{c}$ correspond to the local alternatives against which the variance ratio test based on GLS detrended data has asymptotic power equal to one-half at the nominal 5% level when $R^2 = 0.4$. The results are based on 10,000 Monte Carlo replications and standard Brownian motions are approximated by normalized partial sums of 10,000 i.i.d. standard normal random variables.

Table 3.4: α-quantiles of the limiting null distribution of the variance ratio test statistic

m	1.0%	2.5%	5.0%	7.5%	10.0%	15.0%
D0 (without detrending) and D1 with GLS detrending						
1	0.00487	0.00672	0.00908	0.01139	0.01364	0.01818
2	0.00367	0.00484	0.00619	0.00735	0.00863	0.01077
3	0.00258	0.00328	0.00422	0.00509	0.00597	0.00745
4	0.00207	0.00261	0.00327	0.00387	0.00446	0.00547
5	0.00158	0.00201	0.00256	0.00299	0.00342	0.00422
D1 with OLS detrending						
1	0.00344	0.00458	0.00579	0.00680	0.00772	0.00936
2	0.00242	0.00313	0.00379	0.00437	0.00491	0.00587
3	0.00175	0.00224	0.00278	0.00314	0.00349	0.00418
4	0.00141	0.00174	0.00211	0.00241	0.00267	0.00310
5	0.00112	0.00137	0.00164	0.00185	0.00204	0.00242
D2 with OLS detrending						
1	0.00166	0.00213	0.00259	0.00296	0.00328	0.00384
2	0.00130	0.00168	0.00201	0.00228	0.00253	0.00291
3	0.00106	0.00131	0.00159	0.00179	0.00197	0.00228
4	0.00092	0.00111	0.00130	0.00146	0.00159	0.00184
5	0.00077	0.00092	0.00110	0.00122	0.00132	0.00152
D2 with GLS detrending						
1	0.00363	0.00512	0.00668	0.00807	0.00926	0.01164
2	0.00274	0.00354	0.00468	0.00563	0.00649	0.00807
3	0.00220	0.00278	0.00354	0.00415	0.00468	0.00582
4	0.00165	0.00209	0.00267	0.00318	0.00363	0.00442
5	0.00133	0.00168	0.00214	0.00255	0.00287	0.00348

Notes: The variance ratio test is a left-tailed test rejecting the null hypothesis of no cointegration for realizations of the test statistic *smaller* than the α-quantile. m denotes the number of stochastic regressors in (3.2). Under GLS detrending, critical values in case D1 do not depend on $\bar{c}$, whereas critical values in case D2 depend on $\bar{c}$ and those reported here correspond to the values of $\bar{c}$ given in Table 3.3. Critical values are based on 10,000 Monte Carlo replications and standard Brownian motions are approximated by normalized partial sums of 10,000 i.i.d. standard normal random variables.

3.7.2 Additional Results

Table 3.5: Empirical sizes of the tests in case D0 for $T = 100$ and $T = 250$.

			AR			MA			ARMA			GARCH	
R^2	Test	IID	0.3	0.6	0.9	0.3	0.6	0.9	(0.3,0.6)	(0.3,0.3)	(0.6,0.3)	(0.05,0.94)	(0.01,0.98)
$T = 100$													
0	VR	0.02	0.01	0.01	0.01	0.02	0.03	0.17	0.02	0.02	0.01	0.02	0.02
	ADF	0.08	0.07	0.07	0.06	0.10	0.14	0.43	0.14	0.08	0.07	0.08	0.08
	ADF*	0.05	0.04	0.04	0.04	0.06	0.06	0.14	0.07	0.05	0.04	0.05	0.05
	MSB	0.04	0.04	0.04	0.05	0.05	0.07	0.33	0.07	0.04	0.04	0.04	0.04
	$\widehat{Z}_\alpha$	0.03	0.02	0.01	0.01	0.07	0.30	0.89	0.13	0.03	0.01	0.03	0.03
0.4	VR	0.01	0.01	0.01	0.01	0.01	0.03	0.17	0.02	0.01	0.01	0.01	0.01
	ADF	0.07	0.08	0.07	0.07	0.10	0.13	0.45	0.13	0.07	0.07	0.07	0.07
	ADF*	0.05	0.04	0.05	0.05	0.06	0.06	0.15	0.07	0.05	0.05	0.05	0.05
	MSB	0.03	0.04	0.04	0.04	0.04	0.07	0.34	0.07	0.03	0.04	0.03	0.03
	$\widehat{Z}_\alpha$	0.02	0.02	0.02	0.02	0.07	0.31	0.89	0.12	0.03	0.02	0.02	0.02
0.8	VR	0.01	0.01	0.01	0.01	0.01	0.03	0.18	0.01	0.01	0.01	0.01	0.01
	ADF	0.08	0.08	0.10	0.09	0.09	0.14	0.47	0.13	0.07	0.08	0.08	0.07
	ADF*	0.05	0.06	0.07	0.06	0.06	0.06	0.17	0.07	0.05	0.05	0.05	0.05
	MSB	0.02	0.03	0.03	0.04	0.03	0.07	0.37	0.06	0.02	0.03	0.02	0.02
	$\widehat{Z}_\alpha$	0.02	0.02	0.04	0.03	0.07	0.34	0.89	0.13	0.02	0.02	0.02	0.02
$T = 250$													
0	VR	0.02	0.02	0.02	0.02	0.02	0.03	0.14	0.03	0.02	0.02	0.02	0.02
	ADF	0.06	0.06	0.05	0.04	0.07	0.09	0.32	0.09	0.06	0.05	0.06	0.06
	ADF*	0.04	0.04	0.03	0.03	0.05	0.05	0.11	0.05	0.04	0.03	0.05	0.04
	MSB	0.03	0.03	0.03	0.03	0.04	0.05	0.17	0.06	0.03	0.03	0.04	0.03
	$\widehat{Z}_\alpha$	0.03	0.02	0.02	0.01	0.07	0.31	0.94	0.14	0.03	0.02	0.03	0.03
0.4	VR	0.02	0.02	0.02	0.02	0.02	0.03	0.16	0.02	0.02	0.02	0.02	0.02
	ADF	0.06	0.06	0.06	0.06	0.07	0.10	0.36	0.09	0.06	0.05	0.06	0.06
	ADF*	0.04	0.04	0.04	0.04	0.05	0.06	0.14	0.05	0.04	0.04	0.05	0.05
	MSB	0.02	0.03	0.03	0.03	0.03	0.04	0.21	0.05	0.03	0.03	0.03	0.03
	$\widehat{Z}_\alpha$	0.03	0.02	0.02	0.03	0.07	0.35	0.95	0.15	0.03	0.02	0.03	0.03
0.8	VR	0.01	0.01	0.01	0.02	0.02	0.03	0.18	0.02	0.01	0.01	0.01	0.01
	ADF	0.06	0.07	0.10	0.12	0.08	0.11	0.42	0.10	0.06	0.08	0.06	0.06
	ADF*	0.05	0.05	0.06	0.09	0.05	0.06	0.20	0.06	0.05	0.06	0.05	0.05
	MSB	0.02	0.02	0.05	0.07	0.03	0.03	0.28	0.04	0.02	0.03	0.02	0.02
	$\widehat{Z}_\alpha$	0.02	0.03	0.08	0.09	0.08	0.43	0.96	0.18	0.02	0.04	0.03	0.02

Note: The superscript "$*$" indicates the use of MAIC instead of AIC.

Table 3.6: Empirical sizes of the tests in cases D1 and D2 for $T = 250$.

			AR			MA			ARMA			GARCH	
R^2	Test	IID	0.3	0.6	0.9	0.3	0.6	0.9	(0.3,0.6)	(0.3,0.3)	(0.6,0.3)	(0.05,0.94)	(0.01,0.98)
Deterministic specification D1													
0	VR	0.05	0.05	0.04	0.02	0.06	0.10	0.56	0.08	0.05	0.04	0.06	0.05
	$\text{VR}^{(\text{GLS})}$	0.10	0.10	0.09	0.06	0.11	0.15	0.55	0.12	0.10	0.09	0.10	0.10
	ADF	0.06	0.06	0.05	0.04	0.08	0.12	0.61	0.12	0.06	0.05	0.07	0.06
	$\text{ADF}^{(\text{GLS})}$	0.06	0.06	0.05	0.04	0.08	0.12	0.43	0.13	0.06	0.05	0.07	0.06
	ADF^{*}	0.04	0.03	0.02	0.01	0.04	0.04	0.16	0.05	0.04	0.02	0.04	0.04
	$\text{ADF}^{(\text{GLS})*}$	0.04	0.04	0.03	0.01	0.05	0.05	0.16	0.06	0.04	0.02	0.04	0.04
	MSB	0.07	0.08	0.08	0.08	0.09	0.11	0.48	0.14	0.07	0.08	0.08	0.07
	$\text{MSB}^{(\text{GLS})*}$	0.03	0.03	0.03	0.02	0.04	0.03	0.04	0.05	0.03	0.02	0.04	0.03
	$\widehat{Z}_\alpha$	0.06	0.03	0.02	0.00	0.15	0.64	1.00	0.31	0.06	0.02	0.06	0.06
0.4	VR	0.05	0.05	0.05	0.03	0.06	0.12	0.66	0.08	0.05	0.05	0.06	0.05
	$\text{VR}^{(\text{GLS})}$	0.10	0.10	0.10	0.07	0.12	0.17	0.62	0.13	0.10	0.10	0.11	0.10
	ADF	0.06	0.06	0.06	0.05	0.08	0.14	0.71	0.14	0.06	0.05	0.07	0.06
	$\text{ADF}^{(\text{GLS})}$	0.07	0.07	0.07	0.06	0.09	0.14	0.49	0.14	0.07	0.06	0.08	0.07
	ADF^{*}	0.04	0.03	0.03	0.03	0.04	0.05	0.22	0.05	0.04	0.03	0.04	0.04
	$\text{ADF}^{(\text{GLS})*}$	0.05	0.04	0.04	0.03	0.05	0.06	0.22	0.06	0.05	0.03	0.04	0.05
	MSB	0.07	0.08	0.09	0.08	0.10	0.12	0.60	0.15	0.07	0.08	0.09	0.07
	$\text{MSB}^{(\text{GLS})*}$	0.04	0.04	0.04	0.03	0.05	0.03	0.07	0.05	0.04	0.03	0.03	0.03
	$\widehat{Z}_\alpha$	0.06	0.04	0.04	0.03	0.18	0.74	1.00	0.36	0.06	0.03	0.06	0.06
0.8	VR	0.05	0.05	0.06	0.06	0.07	0.19	0.83	0.10	0.05	0.06	0.06	0.05
	$\text{VR}^{(\text{GLS})}$	0.10	0.10	0.11	0.12	0.12	0.25	0.74	0.15	0.10	0.10	0.10	0.10
	ADF	0.06	0.07	0.14	0.13	0.09	0.21	0.89	0.17	0.06	0.09	0.07	0.06
	$\text{ADF}^{(\text{GLS})}$	0.07	0.08	0.15	0.18	0.10	0.20	0.61	0.17	0.07	0.10	0.08	0.07
	ADF^{*}	0.04	0.04	0.06	0.08	0.04	0.05	0.41	0.06	0.04	0.04	0.04	0.03
	$\text{ADF}^{(\text{GLS})*}$	0.04	0.05	0.08	0.12	0.05	0.07	0.39	0.07	0.04	0.06	0.04	0.04
	MSB	0.07	0.09	0.16	0.17	0.10	0.15	0.83	0.18	0.07	0.11	0.09	0.07
	$\text{MSB}^{(\text{GLS})*}$	0.03	0.04	0.07	0.11	0.04	0.02	0.18	0.05	0.03	0.05	0.04	0.03
	$\widehat{Z}_\alpha$	0.06	0.08	0.19	0.16	0.29	0.94	1.00	0.57	0.06	0.10	0.06	0.06
Deterministic specification D2													
0	VR	0.05	0.05	0.03	0.01	0.07	0.15	0.83	0.10	0.05	0.04	0.06	0.05
	$\text{VR}^{(\text{GLS})}$	0.13	0.12	0.11	0.05	0.15	0.23	0.65	0.18	0.13	0.11	0.14	0.13
	ADF	0.07	0.07	0.06	0.04	0.10	0.18	0.79	0.19	0.07	0.06	0.08	0.07
	$\text{ADF}^{(\text{GLS})}$	0.08	0.08	0.07	0.05	0.11	0.18	0.62	0.19	0.08	0.07	0.09	0.08
	ADF^{*}	0.04	0.03	0.03	0.01	0.04	0.05	0.24	0.06	0.04	0.01	0.04	0.04
	$\text{ADF}^{(\text{GLS})*}$	0.05	0.04	0.03	0.01	0.06	0.07	0.25	0.07	0.05	0.02	0.04	0.04
	MSB	0.09	0.11	0.12	0.14	0.12	0.15	0.68	0.19	0.09	0.12	0.10	0.09
	$\text{MSB}^{(\text{GLS})*}$	0.03	0.04	0.03	0.03	0.05	0.03	0.11	0.06	0.03	0.01	0.03	0.03
	$\widehat{Z}_\alpha$	0.05	0.03	0.01	0.00	0.21	0.83	1.00	0.43	0.05	0.01	0.06	0.06
0.4	VR	0.05	0.05	0.04	0.01	0.07	0.19	0.91	0.10	0.05	0.04	0.05	0.05
	$\text{VR}^{(\text{GLS})}$	0.13	0.12	0.11	0.07	0.15	0.25	0.69	0.19	0.13	0.11	0.13	0.13
	ADF	0.08	0.08	0.07	0.05	0.11	0.20	0.87	0.19	0.08	0.06	0.09	0.08
	$\text{ADF}^{(\text{GLS})}$	0.08	0.08	0.08	0.06	0.12	0.21	0.68	0.20	0.08	0.07	0.09	0.09
	ADF^{*}	0.04	0.03	0.03	0.02	0.04	0.05	0.29	0.06	0.04	0.02	0.04	0.04
	$\text{ADF}^{(\text{GLS})*}$	0.05	0.04	0.03	0.02	0.06	0.06	0.29	0.08	0.05	0.02	0.05	0.05
	MSB	0.08	0.11	0.11	0.11	0.12	0.16	0.79	0.20	0.08	0.11	0.11	0.09
	$\text{MSB}^{(\text{GLS})*}$	0.03	0.03	0.03	0.02	0.04	0.03	0.15	0.06	0.03	0.02	0.03	0.03
	$\widehat{Z}_\alpha$	0.06	0.03	0.03	0.01	0.25	0.91	1.00	0.51	0.06	0.02	0.06	0.06
0.8	VR	0.05	0.06	0.07	0.03	0.09	0.32	0.99	0.15	0.05	0.06	0.05	0.06
	$\text{VR}^{(\text{GLS})}$	0.13	0.13	0.15	0.13	0.16	0.35	0.73	0.22	0.13	0.13	0.13	0.13
	ADF	0.07	0.09	0.18	0.09	0.12	0.32	0.96	0.25	0.07	0.11	0.08	0.07
	$\text{ADF}^{(\text{GLS})}$	0.08	0.10	0.19	0.12	0.13	0.29	0.76	0.25	0.08	0.13	0.09	0.08
	ADF^{*}	0.04	0.04	0.07	0.04	0.04	0.06	0.44	0.06	0.04	0.05	0.04	0.04
	$\text{ADF}^{(\text{GLS})*}$	0.05	0.05	0.08	0.07	0.06	0.07	0.42	0.08	0.05	0.06	0.05	0.04
	MSB	0.09	0.11	0.19	0.12	0.13	0.22	0.92	0.25	0.09	0.13	0.10	0.09
	$\text{MSB}^{(\text{GLS})*}$	0.03	0.04	0.07	0.05	0.04	0.02	0.24	0.06	0.03	0.05	0.03	0.03
	$\widehat{Z}_\alpha$	0.06	0.08	0.21	0.07	0.42	1.00	1.00	0.76	0.06	0.10	0.06	0.05

Note: See note to Table 3.1.

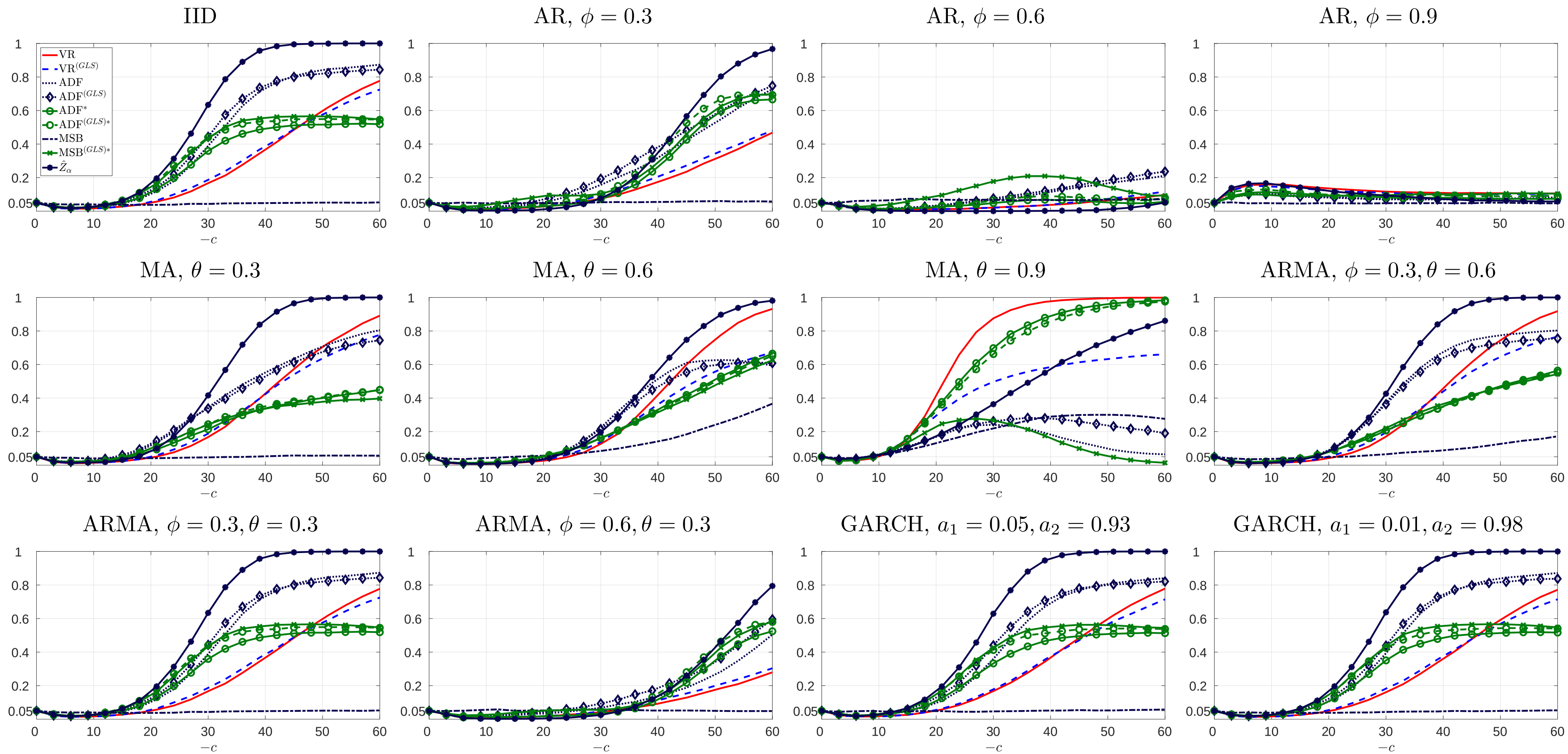


Figure 3.5: Size-corrected power of the tests at the nominal 5% level for H_0: $\rho = 1$ under the alternative $\rho = 1 + c/T$ in case D2, for $T = 100$ and $R^2 = 0.8$.

Note: See note to Figure 3.2.

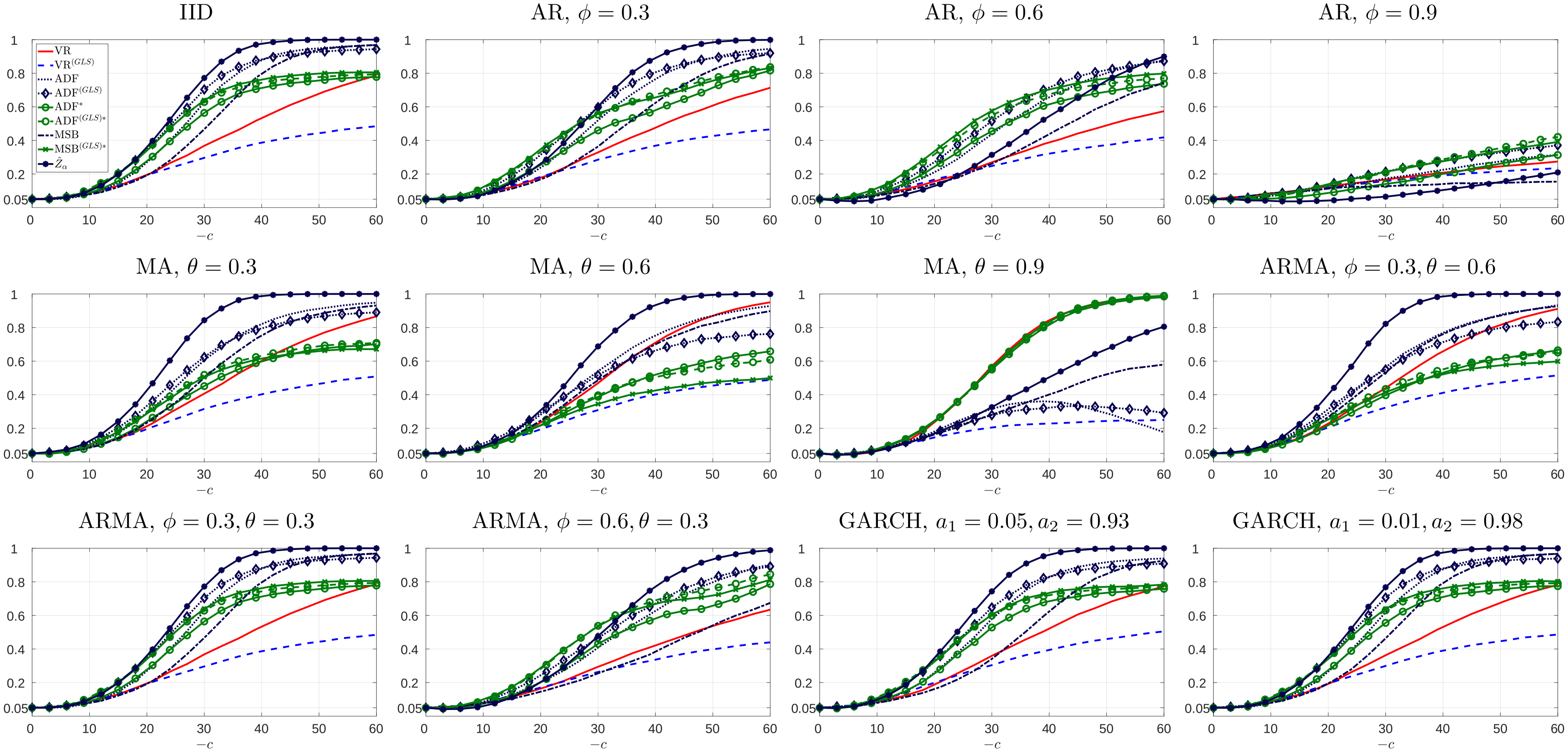


Figure 3.6: Size-corrected power of the tests at the nominal 5% level for H_0: $\rho = 1$ under the alternative $\rho = 1 + c/T$ in case D2, for $T = 250$ and $R^2 = 0.4$.

Note: See note to Figure 3.2.

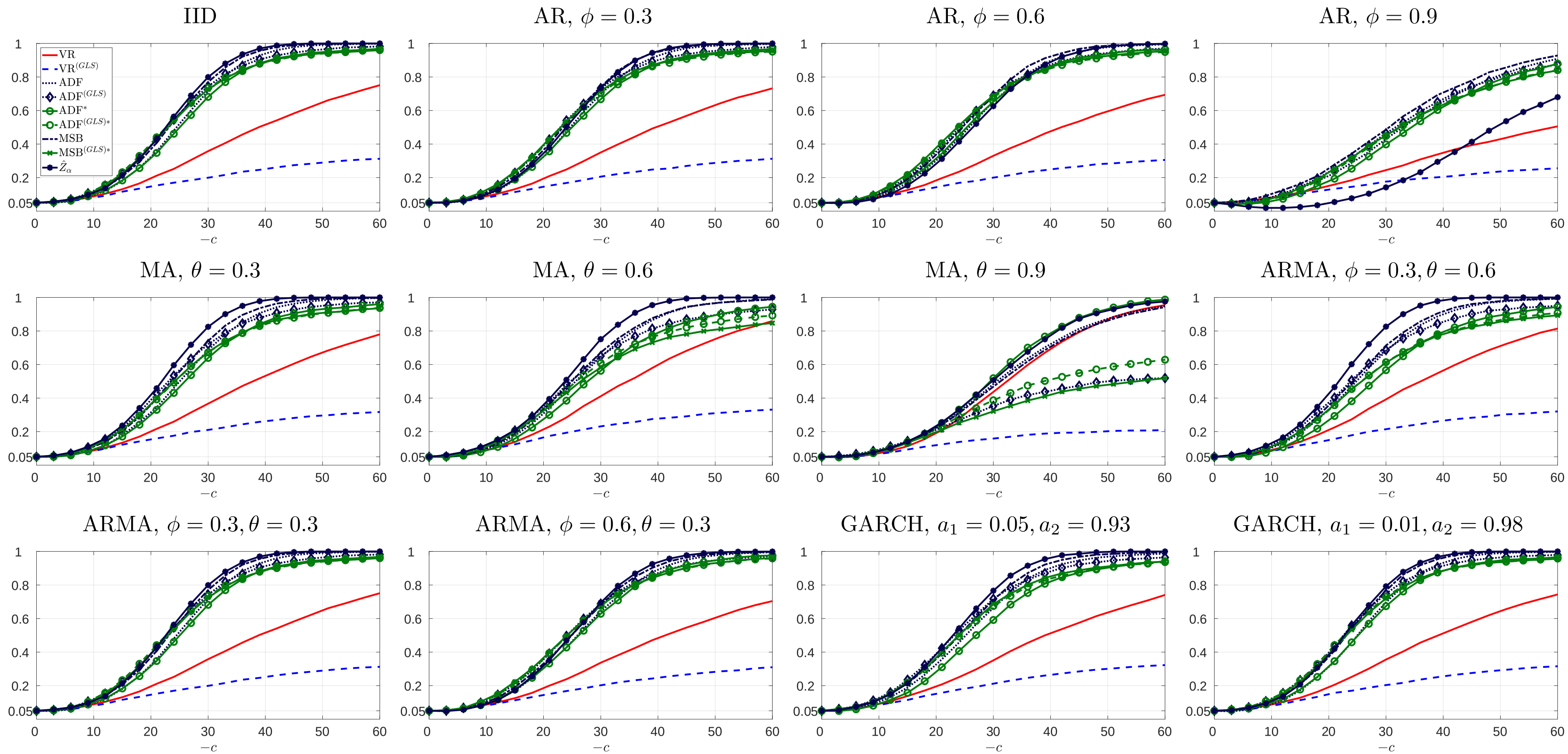


Figure 3.7: Size-corrected power of the tests at the nominal 5% level for H_0: $\rho = 1$ under the alternative $\rho = 1 + c/T$ in case D2, for $T = 1{,}000$ and $R^2 = 0.4$.

Note: See note to Figure 3.2.

3.7.3 Proofs

The proofs frequently use the fact that in cases D1 and D2 the deterministic component d_t fulfills

$$\lim_{T\to\infty} T^{1/2}A_D d_{\lfloor rT\rfloor} = D(r), \quad 0 \le r \le 1, \tag{3.20}$$

with $\int_0^1 D(r)D(r)'dr > 0$, where $A_D = T^{-1/2}$ and $D(r) = 1$ in case D1 and $A_D = \operatorname{diag}\left(T^{-1/2}, T^{-3/2}\right)$ and $D(r) = [1, r]'$ in case D2. The (transposed) potentially multivariate detrended stochastic process $\widetilde{P}(r)$, introduced in Section 3.3.1, can thus be written more generally as

$$\widetilde{P}(r)' = P(r)' - D(r)'\left(\int_0^1 D(s)D(s)'ds\right)^{-1}\int_0^1 D(s)P(s)'ds. \tag{3.21}$$

Proof of Proposition 3.1. Under the null hypothesis of no cointegration ($\rho = 1$), it holds that $u_t = u_0 + \sum_{s=1}^t \xi_s$. In cases D1 and D2, the regression errors in (3.9) are given by $\widetilde{u}_t = \sum_{s=1}^t \xi_s - d_t'\left(\sum_{s=1}^T d_s d_s'\right)^{-1}\sum_{s=1}^T d_s \sum_{l=1}^s \xi_l$. Under Assumption 3.1, we thus obtain

$$\begin{aligned}
T^{-1/2}\widetilde{u}_{\lfloor rT\rfloor} &= T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\xi_t - d_{\lfloor rT\rfloor}'\left(\sum_{t=1}^T d_t d_t'\right)^{-1}\sum_{t=1}^T d_t T^{-1/2}\sum_{s=1}^t \xi_s \\
&= T^{-1/2}\sum_{t=1}^{\lfloor rT\rfloor}\xi_t - \left(T^{1/2}A_D d_{\lfloor rT\rfloor}\right)'\left(T^{-1}\sum_{t=1}^T \left(T^{1/2}A_D d_t\right)\left(T^{1/2}A_D d_t\right)'\right)^{-1} \\
&\qquad\qquad \times T^{-1}\sum_{t=1}^T \left(T^{1/2}A_D d_t\right)T^{-1/2}\sum_{s=1}^t \xi_s \\
&\xrightarrow{w} B_\xi(r) - D(r)'\left(\int_0^1 D(s)D(s)'ds\right)^{-1}\int_0^1 D(s)B_\xi(s)ds \\
&= \widetilde{B}_\xi(r).
\end{aligned} \tag{3.22}$$

By construction, it holds that $B_\xi(r) = \Omega_{\xi\cdot v}^{1/2}W_{\xi\cdot v}(r) + \Omega_{\xi v}(\Omega_{vv}^{-1/2})'W_v(r)$, which implies $\widetilde{B}_\xi(r) = \Omega_{\xi\cdot v}^{1/2}\widetilde{W}_{\xi\cdot v}(r) + \Omega_{\xi v}(\Omega_{vv}^{-1/2})'\widetilde{W}_v(r)$. Analogously, it follows from (3.4) and the fact that OLS detrending annihilates x_0 (and μt) in x_t that

$$\begin{aligned}
T^{-1/2}\widetilde{x}_{\lfloor rT\rfloor}' &= T^{-1/2}x_{\lfloor rT\rfloor}' - d_{\lfloor rT\rfloor}'\left(\sum_{t=1}^T d_t d_t'\right)^{-1}\sum_{t=1}^T d_t T^{-1/2}x_t' \\
&= T^{-1/2}\left(\sum_{t=1}^{\lfloor rT\rfloor} v_t\right)' - d_{\lfloor rT\rfloor}'\left(\sum_{t=1}^T d_t d_t'\right)^{-1}\sum_{t=1}^T d_t T^{-1/2}\left(\sum_{s=1}^t v_s\right)' \\
&\xrightarrow{w} B_v(r)' - D(r)'\left(\int_0^1 D(s)D(s)'ds\right)^{-1}\int_0^1 D(s)B_v(s)'ds \\
&= \widetilde{B}_v(r)',
\end{aligned} \tag{3.23}$$

where it follows from $B_v(r) = \Omega_{vv}^{1/2} W_v(r)$ that $\widetilde{B}_v(r) = \Omega_{vv}^{1/2}\widetilde{W}_v(r)$. In case D0, it holds that, using the notation $\widetilde{P}(r) = P(r)$,

$$\begin{aligned} T^{-1/2}\widetilde{u}_{\lfloor rT \rfloor} &= T^{-1/2}u_0 + T^{-1/2}u_{\lfloor rT \rfloor} = T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor}\xi_t + o_{\mathbb{P}}(1) \\ &\xrightarrow{w} \widetilde{B}_\xi(r) = \Omega_{\xi\cdot v}^{1/2}\widetilde{W}_{\xi\cdot v}(r) + \Omega_{\xi v}(\Omega_{vv}^{-1/2})'\widetilde{W}_v(r), \end{aligned} \tag{3.24}$$

and

$$T^{-1/2}\widetilde{x}'_{\lfloor rT \rfloor} = T^{-1/2}x'_{\lfloor rT \rfloor} \xrightarrow{w} \widetilde{B}_v(r)' = \Omega_{vv}^{1/2}\widetilde{W}_v(r)'. \tag{3.25}$$

In cases D0, D1, and D2, it thus holds for the OLS residuals $\widehat{u}_t$ defined in (3.10) that

$$\begin{aligned} T^{-1/2}\widehat{u}_{\lfloor rT \rfloor} &= T^{-1/2}\widetilde{u}_{\lfloor rT \rfloor} - T^{-1/2}\widetilde{x}'_{\lfloor rT \rfloor}\left(T^{-1}\sum_{t=1}^{T}T^{-1/2}\widetilde{x}_t T^{-1/2}\widetilde{x}'_t\right)^{-1} T^{-1}\sum_{t=1}^{T}T^{-1/2}\widetilde{x}_t T^{-1/2}\widetilde{u}_t \\ &\xrightarrow{w} \widetilde{B}_\xi(r) - \widetilde{B}_v(r)'\left(\int_0^1 \widetilde{B}_v(s)\widetilde{B}_v(s)'ds\right)^{-1}\int_0^1 \widetilde{B}_v(s)\widetilde{B}_\xi(s)ds \\ &= \Omega_{\xi\cdot v}^{1/2}\widetilde{W}^{+}_{\xi\cdot v}(r), \end{aligned} \tag{3.26}$$

with $\widetilde{W}^{+}_{\xi\cdot v}(r)$ as defined in the main text. For the denominator of the variance ratio test statistic it directly follows that

$$T^{-2}\sum_{t=1}^{T}\widehat{u}_t^2 = T^{-1}\sum_{t=1}^{T}(T^{-1/2}\widehat{u}_t)^2 \xrightarrow{w} \Omega_{\xi\cdot v}\int_0^1\left(\widetilde{W}^{+}_{\xi\cdot v}(r)\right)^2 dr. \tag{3.27}$$

Analogously,

$$T^{-3/2}\sum_{t=1}^{\lfloor rT \rfloor}\widehat{u}_t = T^{-1}\sum_{t=1}^{\lfloor rT \rfloor}T^{-1/2}\widehat{u}_t \xrightarrow{w} \Omega_{\xi\cdot v}^{1/2}\int_0^r \widetilde{W}^{+}_{\xi\cdot v}(s)ds. \tag{3.28}$$

For the numerator of the variance ratio test statistic we thus obtain

$$\widehat{\eta}_T = T^{-1}\sum_{t=1}^{T}\left(T^{-3/2}\sum_{s=1}^{t}\widehat{u}_s\right)^2 \xrightarrow{w} \Omega_{\xi\cdot v}\int_0^1\left(\int_0^r \widetilde{W}^{+}_{\xi\cdot v}(s)ds\right)^2 dr, \tag{3.29}$$

Since the vector of numerator and denominator of the variance ratio test statistic can be expressed as a continuous functional of $T^{-1/2}\widehat{u}_{\lfloor rT \rfloor}$ up to an error of $o_{\mathbb{P}}(1)$, the weak convergence results in (3.27) and (3.29) hold jointly (cf., e.g., Phillips, 1987, Proof of Lemma 1). The limiting null distribution of the variance ratio test statistic stated in the proposition thus follows from the continuous mapping theorem and the fact that the scalar long-run variance parameter $\Omega_{\xi\cdot v} > 0$ in (3.27) and (3.29) cancels out. □

Proof of Proposition 3.2. Under the alternative of cointegration ($|\rho| < 1$), it holds that $\widehat{u}_t = \widetilde{u}_t - \widetilde{x}'_t(\widehat{\beta} - \beta)$, where $T(\widehat{\beta} - \beta) = O_{\mathbb{P}}(1)$, see, e.g., Phillips and Hansen (1990). The proof works similarly under all three deterministic specifications. We consider the case D0, where $\widetilde{u}_t = u_t$ and $\widetilde{x}_t = x_t$. For notational brevity, define $u_t^o := \sum_{j=0}^{t-1}\rho^j\xi_{t-j}$, such that $u_t = \rho^t u_0 + u_t^o$. For the denominator of the variance ratio test statistic it holds that

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2 &= T^{-1}\sum_{t=1}^{T}\left(\rho^t u_0 + u_t^o - x_t'(\widehat{\beta}-\beta)\right)^2 \\
&= u_0 T^{-1}\sum_{t=1}^{T}\left(\rho^2\right)^t + 2u_0T^{-1}\sum_{t=1}^{T}\rho^t\left(u_t^o - x_t'(\widehat{\beta}-\beta)\right) \\
&\quad + T^{-1}\sum_{t=1}^{T}\left(u_t^o - x_t'(\widehat{\beta}-\beta)\right)^2. \qquad (3.30)
\end{aligned}$$

Since $\sum_{t=0}^{\infty}\left(\rho^2\right)^t$ is a geometric series and $u_0 = O_{\mathbb{P}}(1)$ it follows that the first term in (3.30) is $o_{\mathbb{P}}(1)$. Next, note that

$$|T^{-1}\sum_{t=1}^{T}\rho^t\left(u_t^o - x_t'(\widehat{\beta}-\beta)\right)| \leq T^{-1}\sum_{t=1}^{T}|\rho^t u_t^o| + T^{-1}\sum_{t=1}^{T}|\rho^t\left(x_t'(\widehat{\beta}-\beta)\right)|. \qquad (3.31)$$

It follows from Markov's inequality, stationarity of u_t^o and the fact that $\sum_{t=0}^{\infty}|\rho|^t$ is a geometric series that $T^{-1}\sum_{t=1}^{T}|\rho^t u_t^o| = o_{\mathbb{P}}(1)$. From

$$T^{-1}\sum_{t=1}^{T}|\rho^t\left(x_t'(\widehat{\beta}-\beta)\right)| \leq T^{-3/2}\sum_{t=1}^{T}|\frac{x_t}{\sqrt{T}}||T(\widehat{\beta}-\beta)| = O_{\mathbb{P}}(T^{-1/2}) \qquad (3.32)$$

it thus follows that also the second term in (3.30) is $o_{\mathbb{P}}(1)$. Therefore,

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2 &= T^{-1}\sum_{t=1}^{T}\left(u_t^o - x_t'(\widehat{\beta}-\beta)\right)^2 + o_{\mathbb{P}}(1) \\
&= T^{-1}\sum_{t=1}^{T}(u_t^o)^2 - 2T^{-1}\left(T^{-1}\sum_{t=1}^{T}u_t^o x_t\right)' T\left(\widehat{\beta}-\beta\right) \\
&\quad + T^{-1}T\left(\widehat{\beta}-\beta\right)'\left(T^{-1}\sum_{t=1}^{T}\frac{x_t}{\sqrt{T}}\frac{x_t'}{\sqrt{T}}\right)T\left(\widehat{\beta}-\beta\right) + o_{\mathbb{P}}(1) \\
&= T^{-1}\sum_{t=1}^{T}(u_t^o)^2 + o_{\mathbb{P}}(1), \qquad (3.33)
\end{aligned}$$

since $T^{-1}\sum_{t=1}^{T}\frac{x_t}{\sqrt{T}}\frac{x_t'}{\sqrt{T}} = O_{\mathbb{P}}(1)$ by Assumption 3.1 in combination with the continuous mapping theorem and $T^{-1}\sum_{t=1}^{T}u_t^o x_t = O_{\mathbb{P}}(1)$ (see, e. g., Phillips and Hansen, 1990). By Assumption 3.1, $\{\xi_t\}_{t\in\mathbb{Z}}$ is strictly stationary and ergodic and $\mathbb{E}\left(\xi_t^2\right) < \infty$. This implies that $\{u_t^o\}_{t\in\mathbb{Z}}$ and thus also $\{(u_t^o)^2\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic (White, 2001, Theorem 3.35) and that $\mathbb{E}\left((u_t^o)^2\right) < \infty$. It thus follows from the law of large numbers for strictly stationary and ergodic time series (White, 2001, Theorem 3.34) that

$$T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2 = T^{-1}\sum_{t=1}^{T}(u_t^o)^2 + o_{\mathbb{P}}(1) \xrightarrow{p} \mathbb{E}((u_1^o)^2), \qquad (3.34)$$

where $0 < \mathbb{E}((u_1^o)^2) < \infty$.

Turning to the numerator of the variance ratio test statistic, we note that

$$T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \widehat{u}_t = T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} u_t^o - \left(T^{-1} \sum_{t=1}^{\lfloor rT \rfloor} \frac{x_t}{\sqrt{T}} \right)' T \left(\widehat{\beta} - \beta \right) + o_{\mathbb{P}}(1), \tag{3.35}$$

since $\sum_{t=0}^{\infty} \rho^t$ is a geometric series. The Beveridge-Nelson decomposition (Phillips and Solo, 1992) yields $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} u_t^o \xrightarrow{w} (1-\rho)^{-1} B_\xi(r)$. The second term in (3.35) is also $O_{\mathbb{P}}(1)$, with a limit that is different from $(1-\rho)^{-1} B_\xi(r)$. Hence, $T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \widehat{u}_t = O_{\mathbb{P}}(1)$, which implies that $T^{-2} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \widehat{u}_s \right)^2 = O_{\mathbb{P}}(1)$. In total, we thus have

$$\mathrm{VR} = T^{-1} \frac{T^{-2} \sum_{t=1}^{T} \left(\sum_{s=1}^{t} \widehat{u}_s \right)^2}{T^{-1} \sum_{t=1}^{T} \widehat{u}_t^2} = O_{\mathbb{P}}(T^{-1}), \tag{3.36}$$

as stated in the proposition. □

Proof of Proposition 3.3. Under the local alternative $\rho = \rho_T = 1 + c/T$, with $c \leq 0$, it holds that $T^{-1/2} u_{\lfloor rT \rfloor} \xrightarrow{w} \Omega_{\xi \cdot v}^{1/2} J_{\xi \cdot v}^c(r)$, $0 \leq r \leq 1$, with $J_{\xi \cdot v}^c(r)$ as defined in the main text (cf., e. g., Perron and Rodríguez, 2016, Lemma 5.1). In cases D1 and D2, it follows that $T^{-1/2} \widetilde{u}_{\lfloor rT \rfloor} \xrightarrow{w} \Omega_{\xi \cdot v}^{1/2} \widetilde{J}_{\xi \cdot v}^c(r)$, $0 \leq r \leq 1$, where

$$\widetilde{J}_{\xi \cdot v}^c(r) = J_{\xi \cdot v}^c(r) - D(r)' \left(\int_0^1 D(s) D(s)' ds \right)^{-1} \int_0^1 D(s) J_{\xi \cdot v}^c(s) ds. \tag{3.37}$$

Analogously, in case D0, it holds that, using the notation $\widetilde{P}(r) = P(r)$, $T^{-1/2} \widetilde{u}_{\lfloor rT \rfloor} \xrightarrow{w} \Omega_{\xi \cdot v}^{1/2} \widetilde{J}_{\xi \cdot v}^c(r) = \Omega_{\xi \cdot v}^{1/2} J_{\xi \cdot v}^c(r)$, $0 \leq r \leq 1$. The rest of the proof is similar to the proof of Proposition 3.1 and therefore omitted. □

Proof of Remark 3.3. Perron and Rodríguez (2016, Lemma 5.3) show that $\widetilde{u}_t^{(\mathrm{GLS})}$, as defined in Remark 3.1, fulfills $T^{-1/2} \widetilde{u}_{\lfloor rT \rfloor}^{(\mathrm{GLS})} \xrightarrow{w} \Omega_{\xi \cdot v}^{1/2} J_{\xi \cdot v}^{c,\mathrm{GLS}}(r)$, $0 \leq r \leq 1$, where $J_{\xi \cdot v}^{c,\mathrm{GLS}}(r)$ is given by $J_{\xi \cdot v}^c(r)$ in case D1 and by $J_{\xi \cdot v}^c(r) - \left(\lambda J_{\xi \cdot v}^c(1) + 3(1-\lambda) \int_0^1 s J_{\xi \cdot v}^c(s) ds \right) r$ in case D2, with λ as defined in the main text. The rest of the proof uses similar arguments as the proof of Theorem 5.2 in Perron and Rodríguez (2016) and the proof of Proposition 3.1 and is therefore omitted. □

3.7.4 Computation of the ADF, MSB and $\widehat{Z}_\alpha$ Tests

Test Statistics

Given the regression residuals $\widehat{u}_t$, $t = 1, \ldots, T$, as defined in (3.10), the ADF test statistic, the MSB test statistic, and the $\widehat{Z}_\alpha$ test statistic are defined as follows:[31]

- The ADF statistic is defined as the usual t-test statistic for testing $b_0 = 0$ in the auxiliary

[31] In case of GLS detrending, the test statistics are defined by replacing $\widehat{u}_t$ with $\widehat{u}_t^{(\mathrm{GLS})}$.

regression

$$\Delta \widehat{u}_t = b_0 \widehat{u}_{t-1} + \sum_{j=1}^{p} \pi_j \Delta \widehat{u}_{t-j} + r_{tp}, \tag{3.38}$$

$t = p+2, \dots, T$. The lag parameter p is determined by means of information criteria, compare the discussion in Section 3.7.4.

- Let $\widehat{\pi}_j$ denote the estimates of π_j obtained by estimating (3.38) with OLS and let $\widehat{r}_{tp}$ denote the corresponding residuals. The MSB statistic is then defined as

$$\text{MSB} := \left(\frac{T^{-2} \sum_{t=1}^{T} \widehat{u}_t^2}{\widehat{s}^2} \right)^{1/2}, \tag{3.39}$$

where $\widehat{s}^2 := \widehat{s}_{rp}^2 / (1 - \widehat{\pi}(1))^2$, with $\widehat{s}_{rp}^2 := T^{-1} \sum_{t=p+2}^{T} \widehat{r}_{tp}^2$ and $\widehat{\pi}(1) := \sum_{j=1}^{p} \widehat{\pi}_j$.

- To define the $\widehat{Z}_\alpha$ statistic, consider the auxiliary regression

$$\widehat{u}_t = \alpha \widehat{u}_{t-1} + k_t \tag{3.40}$$

$t = 2, \dots, T$. Let $\widehat{\alpha}$ and $\widehat{k}_t$ denote the OLS estimate of α and the corresponding OLS residuals, respectively. Define $s_k^2 := (T-1)^{-1} \sum_{t=2}^{T} \widehat{k}_t^2$ and

$$s_{Tb}^2 := s_k^2 + 2(T-1)^{-1} \sum_{h=1}^{b_T} \mathcal{K}\left(\frac{h}{b_T}\right) \sum_{t=h+2}^{T} \widehat{k}_t \widehat{k}_{t-h}, \tag{3.41}$$

where the kernel function $\mathcal{K}(\cdot)$ and the bandwidth parameter b_T fulfill some technical assumptions, see, e. g., Andrews (1991), Newey and West (1994) and Jansson (2002) for details. The $\widehat{Z}_\alpha$ statistic is then defined as

$$\widehat{Z}_\alpha := (T-1)(\widehat{\alpha} - 1) - \frac{1}{2} \left(s_{Tb}^2 - s_k^2 \right) \left((T-1)^{-2} \sum_{t=2}^{T} \widehat{u}_{t-1}^2 \right)^{-1}. \tag{3.42}$$

The three tests are left-tailed tests, rejecting the null hypothesis of no cointegration if the realization of the statistic is smaller than the corresponding critical value. Asymptotically valid critical values for the ADF and $\widehat{Z}_\alpha$ statistics in cases D0, D1, and D2 are tabulated in Phillips and Ouliaris (1990), whereas for the MSB statistic we use (unreported) critical values based on own simulations.[32]

Information Criteria

Implementing the ADF and MSB tests requires the specification of the lag parameter $0 \leq p \leq p_{\max}$ in the auxiliary regression (3.38). This is typically achieved by means of information criteria evaluated on exactly the same period $t = p_{\max}+2, \dots, T$ for each choice of p (Kilian and Lütkepohl,

[32] In case of GLS detrending, critical values for the ADF and MSB statistics are tabulated in Perron and Rodríguez (2016).

2017, p. 56). The AIC and the BIC are defined as

$$\mathrm{AIC}(p) \coloneqq \log(\widehat{s}^2_{rp_{\max}}) + \frac{2p}{T} \tag{3.43}$$

and

$$\mathrm{BIC}(p) \coloneqq \log(\widehat{s}^2_{rp_{\max}}) + \frac{p\log(T)}{T}, \tag{3.44}$$

respectively, where $\widehat{s}^2_{rp_{\max}} = T^{-1}\sum_{t=p_{\max}+2}^{T}\widehat{r}^2_{tp}$, with $\widehat{r}_{tp}$ denoting the OLS residuals in (3.38).[33] Ng and Perron (2001) and Perron and Qu (2007) propose a modified AIC (MAIC) criterion. Applied to the regression residuals $\widehat{u}_t$ – in the presence of deterministic components always based on OLS detrended data, even in case the test statistic is constructed using GLS detrended data – the MAIC becomes

$$\mathrm{MAIC}(p) \coloneqq \log((T-p_{\max})^{-1}T\widehat{s}^2_{rp_{\max}}) + \frac{2(p+\tau_T(p))}{T-p_{\max}}, \tag{3.45}$$

where $\tau_T(p) \coloneqq \left((T-p_{\max})^{-1}T\widehat{s}^2_{rp_{\max}}\right)^{-1}\widehat{b}_0^2\sum_{t=p_{\max}+2}^{T}\widehat{u}^2_{t-1}$, with $\widehat{b}_0$ denoting the OLS estimate of b_0 in (3.38). A similarly modified version of the BIC is then given by

$$\mathrm{MBIC}(p) \coloneqq \log((T-p_{\max})^{-1}T\widehat{s}^2_{rp_{\max}}) + \frac{\log(T-p_{\max})(p+\tau_T(p))}{T-p_{\max}}. \tag{3.46}$$

[33] We follow Kilian and Lütkepohl (2017, p. 56) and use $\widehat{s}^2_{rp_{\max}}$ rather than $(T-p)^{-1}T\widehat{s}^2_{rp_{\max}}$.

Conclusion

This cumulative dissertation proposes procedures to perform more reliable inference in different types of regressions involving stochastically trending variables. Although each chapter is devoted to a specific subfield of the cointegrating regression literature, the proposed methods share a common ground, which makes them suitable to be combined or extended to other settings involving stochastically trending variables. The chapter-specific conclusions at the end of each chapter have already discussed some promising and explicit directions for future research.

In addition, several other methodological challenges arise in the era of high-dimensional data. For example, it might be interesting to extend the nonparametric variance ratio test for cointegration, analyzed in Chapter 3, to settings where the number of integrated processes is allowed to increase with sample size T (potentially in the form of a system-based test rather than a test within a regression framework). A useful starting point seems to be the system-based test for the cointegration rank proposed in Breitung (2002, Section 5). To allow the dimension of the system to increase with sample size, the methodology proposed in Bykhovskaya and Gorin (2022a; 2022b) could be extended to this particular test statistic.

Another interesting direction of future research is the development of a LASSO-type estimator in so-called predictive (or even high-dimensional) cointegrating regressions that (i) selects the relevant regressors, (ii) estimates the corresponding coefficients consistently, and (iii) allows for standard asymptotic inference when testing restrictions on the coefficients corresponding to the relevant regressors. Lee *et al.* (2022) and Koo *et al.* (2020) have already proposed LASSO-type estimators fulfilling conditions (i) and (ii) in the predictive and high-dimensional cointegrating regression settings, respectively. However, in the presence of endogeneity, the limiting distributions of their estimators are contaminated by second order bias terms reflecting the dependence structure in the data. To eliminate the second order bias terms, it seems to be promising to extend the integrated modified OLS approach of Vogelsang and Wagner (2014), already employed in Chapter 1 to construct self-normalized test statistics, to predictive (or even high-dimensional) cointegrating regressions and combine the approach with the LASSO-type approaches of Lee *et al.* (2022) or Koo *et al.* (2020).

These interesting directions of future research in the contexts of predictive and high-dimensional cointegrating regressions are currently under investigation by the author of this dissertation and additional co-authors.

List of Tables

List of Figures

Bibliography

Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

Baxter, G. (1962). An Asymptotic Result for the Finite Predictor. *Mathematica Scandinavica* **10**, 137–144.

Bayer, C., Hanck, C. (2013). Combining Non-Cointegration Tests. *Journal of Time Series Analysis* **34**, 83–95.

Benati, L., Lucas Jr., R.E., Nicolini, J.P., Weber, W. (2021). International Evidence on Long-Run Money Demand. *Journal of Monetary Economics.* **117**, 43–63.

Breitung, J. (2002). Nonparametric Tests for Unit Roots and Cointegration. *Journal of Econometrics* **108**, 343–363.

Breitung, J., Taylor, A.M.R. (2003). Corrigendum to "Nonparametric Tests for Unit Roots and Cointegration [J. Econom. 108 (2002) 343–363]". *Journal of Econometrics* **117**, 401–404.

Brillinger, D.R. (1981). Time Series: Data Analysis and Theory. Holden-Day, Inc., San Francisco.

Brüggemann, R., Jentsch, C., and Trenkler, C. (2016). Inference in VARs with Conditional Heteroskedasticity of Unknown Form. *Journal of Econometrics* **191**, 69–85.

Brüggemann, R., Lütkepohl, H. (2005). Practical Problems with Reduced-rank ML Estimators for Cointegration Parameters and a Simple Alternative. *Oxford Bulletin of Economics and Statistics* **67**, 673–690.

Bühlmann, P. (1997). Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123–148.

Bykhovskaya, A., Gorin, V. (2022a). Cointegration in Large VARs. *Annals of Statistics* **50**, 1593–1617.

Bykhovskaya, A., Gorin, V. (2022b). Asymptotics of Cointegration Tests for High-Dimensional VAR(k). arXiv e-print 2202.07150v3.

Cai, Y., Shintani, M. (2006). On the Alternative Long-Run Variance Ratio Test for a Unit Root. *Econometric Theory* **22**, 347–372.

Caporale, G.M., Pittis, N. (2004). Estimator Choice and Fisher's Paradox: A Monte Carlo Study. *Econometric Reviews* **23**, 25–52.

Cavaliere, G., Nielsen, H.B., Rahbek, A. (2015). Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models. *Econometrica* **83**, 813–831.

Chang, Y., Park, J.Y., Song, K. (2006). Bootstrapping Cointegrating Regressions. *Journal of Econometrics* **133**, 703–739.

Choi, I., Kurozumi, E. (2012). Model Selection Criteria for the Leads-and-Lags Cointegrating Regression. *Journal of Econometrics* **169**, 224–238.

Dahlhaus, R., Kiss, I.Z., Neddermeyer, J.C. (2018). On the Relationship Between the Theory of Cointegration and the Theory of Phase Synchronization. *Statistical Science* **33**, 334–357.

Darvas, Z. (2008). Estimation Bias and Inference in Overlapping Autoregressions: Implications for the Target-Zone Literature. *Oxford Bulletin of Economics and Statistics* **70**, 1–22.

de Jong, R.M. (2002). Nonlinear Estimators with Integrated Regressors but Without Exogeneity. *Mimeo.*

de Jong, R.M., Wagner, M. (2022). Panel Cointegrating Polynomial Regression Analysis and the Environmental Kuznets Curve. *Econometrics and Statistics.* Forthcoming.

Dickey, D.A., Fuller, W.A. (1979). Distribution of the Estimator for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association* **74**, 427–431.

Doukhan, P., Wintenberger, O. (2007). An Invariance Principle for Weakly Dependent Stationary General Models. *Probability and Mathematical Statistics* **27**, 45–73.

Einmahl, U. (1987). A Useful Estimate in the Multidimensional Invariance Principle. *Probability Theory and Related Fields* **76**, 81–101.

Elliot, G., Rothenberg, T.J., Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica* **64**, 813 – 836.

Engle, R.F., Granger, C.W.J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* **55**, 251–276.

Gentle, J.E. (2007). Matrix Algebra. Springer, New York.

Granger, C.W.J., Newbold, P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics* **2**, 111–120.

Grossman, G.M., Krueger, A.B. (1993). Environmental Impacts of a North American Free Trade Agreement. In Garber, P. (Ed.) *The Mexico-US Free Trade Agreement*, 13–56, MIT Press, Cambridge.

Hannan, E.J., Deistler, M. (1988). The Statistical Theory of Linear Systems. Wiley, New York.

Harris, D., Poskitt, D.S. (2004). Determination of Cointegrating Rank in Partially Non-Stationary Processes via a Generalised von-Neumann Criterion. *The Econometrics Journal* **7**, 191–217.

Harvey, D.I., Leybourne, S.J., Taylor, A.M.R. (2009). Unit Root Testing in Practice: Dealing With Uncertainty Over The Trend and Initial Condition. *Econometric Theory* **25**, 587–636.

Holtz-Eakin, D., Selden, T.M. (1995). Stoking the Fires? CO_2 Emissions and Economic Growth. *Journal of Public Economics* **57**, 85–101.

Horn, R.A., Johnson, C.R. (2012). Matrix Analysis (2nd Ed.). Cambridge University Press, New York.

Hosseinkouchack, M. (2014). Local Asymptotic Power of Breitung's Test. *Oxford Bulletin of Economics and Statistics* **76**, 456–462.

Hwang, J., Sun, Y. (2018). Simple, Robust, and Accurate F and t Tests in Cointegrated Systems. *Econometric Theory* **34**, 949–984.

Hwang, J., Valdes, G. (2023). Low Frequency Cointegrating Regression with Local to Unity Regressors and Unknown Form of Serial Dependence. *Journal of Business & Economic Statistics.* Forthcoming.

Ibragimov, R., Phillips, P.C.B. (2008). Regression Asymptotics Using Martingale Convergence Methods. *Econometric Theory* **24**, 888–947.

Jansson, M. (2002). Consistent Covariance Matrix Estimation for Linear Processes. *Econometric Theory* **18**, 1449–1459.

Jensen, M.J. (2009). The Long-Run Fisher Effect: Can It Be Tested?. *Journal of Money, Credit and Banking* **41**, 221–231.

Jentsch, C., Politis, D.N., Paparoditis, E. (2015). Block Bootstrap Theory for Multivariate Integrated and Cointegrated Time Series. *Journal of Time Series Analysis* **36**, 416–441.

Jin, S., Phillips, P.C.B., Sun, Y. (2006). A New Approach to Robust Inference in Cointegration. *Economics Letters* **91**, 300–306.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **59**, 1551–1580.

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models. Oxford University Press, Oxford.

Keilbar, G., Zhang, Y. (2021). On Cointegration and Cryptocurrency Dynamics. *Digital Finance* **3**, 1–23.

Kiefer, N.M., Vogelsang, T.J. (2002). Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation. *Econometrica* **70**, 2093–2095.

Kiefer, N.M., Vogelsang, T.J., Bunzel, H. (2000). Simple Robust Testing of Regression Hypotheses. *Econometrica* **68**, 695–714.

Kilian, L., Lütkepohl, H. (2017). Structural Vector Autoregressive Analysis. Cambridge University Press, Cambridge.

Koo, B., Anderson, H.M., Seo, M.H., Yao, W. (2020). High-Dimensional Predictive Regression in the Presence of Cointegration. *Journal of Econometrics* **219**, 456–477.

Kreiss, J.-P. (1992). Bootstrap Procedures for AR(∞) Processes. In Bootstrapping and Related Techniques, Jöckel, K.H., Rothe, G., Sender, W. (eds.), Lecture Notes in Economics and Mathematical Systems, vol. 376. Springer, Heidelberg, 107–113.

Kremers, J.J.M., Ericsson, N.R., Dolado, J. (1992). The Power of Cointegration Tests. *Oxford Bulletin of Economics and Statistics* **54**, 325–348.

Kuznets, S. (1955). Economic Growth and Income Inequality. *American Economic Review* **45**, 1–28.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics* **54**, 159–178.

Labson, B.S., Crompton, P.L. (1993). Common Trends in Economic Activity and Metals Demand: Cointegration and the Intensity of Use Debate. *Journal of Environmental Economics and Management* **25**, 147–161.

Lee, J.H., Shi, Z., Gao, Z. (2022). On LASSO for Predictive Regression. *Journal of Econometrics* **229**, 322–349.

Leung, T., Nguyen, H. (2019). Constructing Cointegrated Cryptocurrency Portfolios for Statistical Arbitrage. *Studies in Economics and Finance* **36**, 581–599.

Li, H., Maddala, G.S. (1997). Bootstrapping Cointegrating Regressions. *Journal of Econometrics* **80**, 297–318.

Malenbaum, W. (1978). *World Demand for Raw Materials in 1985 and 2000.* New York: McGraw-Hill, EMJ Mining Information Services.

Merlevède, F., Peligrad, M., Utev, S. (2006). Recent Advances in Invariance Principles for Stationary Sequences. *Probability Surveys* **3**, 1–36.

Meyer, M., Kreiss, J.-P. (2015). On the Vector Autoregressive Sieve Bootstrap. *Journal of Time Series Analysis* **36**, 377–397.

Müller, U.K., Watson, M.W. (2013). Low-Frequency Robust Cointegration Testing. *Journal of Econometrics* **174**, 66–81.

Newey, W.K., West, K.D. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* **61**, 631–653.

Ng, S., Perron, P. (2001). Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. *Econometrica* **69**, 1519–1554.

Nielsen, M.Ø. (2009). A Powerful Test of the Autoregressive Unit Root Hypothesis Based on a Tuning Parameter Free Statistic. *Econometric Theory* **25**, 1515–1544.

Palm, F.C., Smeekes, S., Urbain, J.-P. (2010). A Sieve Bootstrap Test for Cointegration in a Conditional Error Correction Model. *Econometric Theory* **26**, 647–681.

Paparoditis, E. (1996). Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* **57**, 277 – 296.

Paparoditis, E., Politis, D.N. (2003). Residual-Based Block Bootstrap for Unit Root Testing. *Econometrica* **71**, 813–855.

Paparoditis, E., Politis, D.N. (2005). Bootstrap Hypothesis Testing in Regression Models. *Statistics & Probability Letters* **74**, 356–365.

Park, J.Y. (2002). An Invariance Principle for Sieve Bootstrap in Time Series. *Econometric Theory* **18**, 469–490.

Park, J.Y., Phillips, P.C.B. (2001). Nonlinear Regressions with Integrated Time Series. *Econometrica* **69**, 117–161.

Perron, P., Ng, S. (1996). Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties. *Review of Economic Studies* **63**, 435–463.

Perron, P., Qu, Z. (2007). A Simple Modification to Improve the Finite Sample Properties of Ng and Perron's Unit Root Tests. *Economics Letters* **94**, 12–19.

Perron, P., Rodríguez, G. (2016). Residual-Based Tests for Cointegration with Generalized Least-Squares Detrended Data. *The Econometrics Journal* **19**, 84–111.

Pesavento, E. (2004). Analytical Evaluation of the Power of Tests for Absence of Cointegration. *Journal of Econometrics* **122**, 349–384.

Pesavento, E. (2007). Residuals-Based Tests for the Null of No-Cointegration: An Analytical Comparison. *Journal of Time Series Analysis* **28**, 111–137.

Phillips, P.C.B. (1986). Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* **33**, 311–340.

Phillips, P.C.B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrica* **74**, 535–547.

Phillips, P.C.B. (2014). Optimal Estimation of Cointegrated Systems With Irrelevant Instruments. *Journal of Econometrics* **178**, 210–224.

Phillips, P.C.B. (2022). Estimation and Inference with Near Unit Roots. *Econometric Theory.* Forthcoming.

Phillips, P.C.B., Hansen, B.E. (1990). Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**, 99–125.

Phillips, P.C.B., Leirvik, T., Storelvmo, T. (2020). Econometric Estimates of Earth's Transient Climate Sensitivity. *Journal of Econometrics* **214**, 6–32.

Phillips, P.C.B., Loretan, M. (1991). Estimating Long Run Economic Equilibria. *Review of Economic Studies* **58**, 407–436.

Phillips, P.C.B., Magdalinos, T. (2009). Econometric Inference in the Vicinity of Unity. Singapore Management University, CoFie Working Paper, 7.

Phillips, P.C.B., Moon, H.R. (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica* **67**, 1057–1111.

Phillips, P.C.B., Moon, H.R., Xiao, Z. (2001). How to Estimate Autoregressive Roots near Unity. *Econometric Theory* **17**, 29–69.

Phillips, P.C.B., Ouliaris, S. (1990). Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* **58**, 165–193.

Phillips, P.C.B., Perron, P. (1988). Testing for Unit Roots in Time Series Regression. *Biometrika* **75**, 335–346.

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

Politis, D.N., Romano, J.P. (2004). The Stationary Bootstrap. *Journal of the American Statistical Association* **89**, 1303-1313.

Psaradakis, Z. (2001). On Bootstrap Inference in Cointegrating Regressions. *Economics Letters* **72**, 1–10.

Rad, H., Low, R.K.Y., Faff, R. (2016). The Profitability of Pairs Trading Strategies: Distance, Cointegration and Copula Methods. *Quantitative Finance* **16**, 1541–1558.

Reichold, K., Jentsch, C. (2022). A Bootstrap-Assisted Self-Normalization Approach to Inference in Cointegrating Regressions. arXiv e-print 2204.01373.

Reichold, K., Wagner, M. (2022). Cointegrating Polynomial Regressions with Integrated Regressors with Drift: Fully Modified OLS Estimation and Inference. *Mimeo.*

Rho, Y., Shao, X. (2019). Bootstrap-Assisted Unit Root Testing With Piecewise Locally Stationary Errors. *Econometric Theory* **35**, 142–166.

Said, S.E., Dickey, D.A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **71**, 599–608.

Saikkonen, P. (1991). Asymptotically Efficient Estimation of Cointegrating Regressions. *Econometric Theory* **7**, 1–21.

Shao, X. (2010a). A Self-Normalization Approach to Confidence Interval Construction in Time Series. *Journal of the Royal Statistical Society (Series B)* **72**, 343–366.

Shao, X. (2010b). The Dependent Wild Bootstrap. *Journal of the American Statistical Association* **105**, 218–235.

Shao, X. (2015). Self-Normalization for Time Series: A Review of Recent Developments. *Journal of the American Statistical Association* **110**, 1797–1817.

Shin, D.W., Hwang, E. (2013). Stationary Bootstrapping for Cointegrating Regressions. *Statistics and Probability Letters* **83**, 474–480.

Shin, Y., Schmidt, P. (1992). The KPSS Stationary Test as a Unit Root Test. *Economics Letters* **38**, 387–392.

Shintani, M. (2001). A Simple Cointegrating Rank Test Without Vector Autoregression. *Journal of Econometrics* **105**, 337–362.

Sims, C.A., Stock, J.H., Watson, M.W. (1990). Inference in Linear Time Series Models with some Unit Roots. *Econometrica* **58**, 113–144.

Stern, D.I. (2004). The Rise and Fall of the Environmental Kuznets Curve. *World Development* **32**, 1419–1439.

Stock, J.H. (1999). A Class of Tests for Integration and Cointegration. In. R.F. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, 137–167. Oxford University Press, Oxford.

Stock, J.H., Watson, M.W. (1993). A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* **61**, 783–820.

Svensson, L.E.O. (1992). An Interpretation of Recent Research on Exchange Rate Target Zones. *Journal of Economic Perspectives* **6**, 119–144.

van Giersbergen, N.P.A., Kiviet, J.F. (2002). How to Implement the Bootstrap in Static or Stable Dynamic Regression Models: Test Statistic Versus Confidence Region Approach. *Journal of Econometrics* **108**, 133–156.

Vogelsang, T.J., Wagner, M. (2014). Integrated Modified OLS Estimation and Fixed-b Inference for Cointegrating Regressions. *Journal of Econometrics* **178**, 741–760.

Wagner, M. (2015). The Environmental Kuznets Curve, Cointegration and Nonlinearity. *Journal of Applied Econometrics* **30**, 948–967.

Wagner, M., Grabarczyk, P., Hong, S.H. (2020). Fully Modified OLS Esimation and Inference for Seemingly Unrelated Cointegrating Polynomial Regressions and the Environmental Kuznets Curve for Carbon Dioxide Emissions. *Journal of Econometrics* **214**, 216–255.

Wagner, M., Hlouskova, J. (2010). The Performance of Panel Cointegration Methods: Results From a Large Scale Simulation Study. *Econometric Reviews* **29**, 182–223.

Wagner, M., Hong, S.H. (2016). Cointegrating Polynomial Regressions: Fully Modified OLS Estimation and Inference. *Econometric Theory* **32**, 1289–1315.

West, K.D. (1988). Asymptotic Normality, When Regressors Have a Unit Root. *Econometrica* **56**, 1397–1417.

Westerlund, J. (2005). New Simple Tests for Panel Cointegration. *Econometric Reviews* **24**, 297–316.

Westerlund, J. (2008). Panel Cointegration Tests of the Fisher Effect. *Journal of Applied Econometrics* **23**, 193–233.

White, H. (2001). Asymptotic Theory for Econometricians. Academic Press, San Diego.

Wu, W. (2007). Strong Invariance Principles for Dependent Random Variables. *Annals of Probability* **35**, 2294–2320.

Yandle, B., Bjattarai, M., Vijayaraghavan, M. (2004). Environmental Kuznets Curves: A Review of Findings, Methods, and Policy Implications. Research Study 02.1 update, PERC.

Zivot, E. (2000). The Power of Single Equation Tests for Cointegration When the Cointegrating Vector is Prespecified. *Econometric Theory* **16**, 407–439.

Eidesstattliche Versicherung und Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Dissertation ist bisher keiner anderen Fakultät vorgelegt worden. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt.

Dortmund, Mai 2023

Karsten Reichold