Unit-Root tests in high-dimensional panels

Wichert, Oliver

DOI:
10.26116/center-lis-2203

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Unit-Root Tests in High-Dimensional Panels

OLIVER WICHERT

TILBURG ♦ UNIVERSITY

Unit-Root Tests in High-Dimensional Panels

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op vrijdag 29 april 2022 om 10.00 uur door

Oliver Wichert

geboren te Hamburg, Duitsland

Promotor:	prof. dr. B.J.M. Werker (Tilburg University)
Copromotores:	dr. F.C. Drost (Tilburg University)
	dr. R. van den Akker (Tilburg University)
leden promotiecommissie:	prof. dr. D. Paindaveine (Vrije Universiteit Brussel)
	dr. O. Boldea (Tilburg University)
	prof. dr. B. Melenberg (Tilburg University)
	prof. dr. J.H.J. Einmahl (Tilburg University)

Acknowledgements

The completion of my PhD would not have been possible without the support and encouragement from a number of people. I would like to begin by thanking my supervisors, Bas Werker, Feico Drost, and Ramon van den Akker. Each of them helped and encouraged me in multiple ways. Ramon first introduced me to non-stationary time series and welcomed me into econometric research in general. Feico was always able to find even the most subtle mathematical issues, and often offered a potential solution. Bas gave me ideas on which projects would be most promising and always made sure I would not discard all of them. From each of them, I learnt a lot about conducting and writing research and I felt supported even in difficult times. I am grateful for your guidance, encouragement, and support.

I would also like to thank my committee, Davy Paindaveine, Otilia Boldea, Bertrand Melenberg, and John Einmahl for their helpful comments on my thesis, which have helped me improve it significantly. I would also like to thank Gaia Becheri for the great collaboration at the beginning of this project. I am also thankful to Ulrich Müller for welcoming me during my research visit and for our discussions, which really helped me with the final chapter of the thesis. Further, I would like to thank my former colleagues from the department of Econometrics and Operations Research, many of whom first taught me much of the basics underlying my research during my bachelor and research master studies; it was an honor working at the department with you.

In addition, there are a number of friends and colleagues who I am very thankful for. Thank you Laura, Santiago, Dorothee, Mirthe, Thijs and Sophie. Without you, my PhD years would not have been the same and you ensured that there were positive things even when the research wouldn't cooperate. I hope our friendship will last many years, despite going in different directions after our PhDs. I also want to thank my officemates Hao, Riley, Maciej, and Lkhagvaa; sharing an office made daily life much more fun, and I will miss our office conversations. Finally, climbing became my favorite activity during my PhD, and I quickly learned that it is an excellent activity to destress, and I am thankful for the friendships I made in the process. I would like to thank, in particular, Sophie, Aeron, Danny, and Maria Francesca.

Last but not least I would like to thank my family members for their support and encouragement. First of all my parents, who have always supported me in whatever I wanted to do. Thanks also to Elena and Omi. Finally, I would not have made it without my wife Maria. Thank you for always being there to uplift me and pushing me to do my best. Your love and support were essential for finishing this project.

Contents

Introduction

This thesis studies unit-root tests in large panels. 'Large' means here that both the number of panel units n as well as the number of time periods T are large; in our asymptotic analyses we typically assume that both n and T go to infinite jointly. Unit-Root Tests have long been a key research topic in both statistics and econometrics. Given a time series of observations, the aim of a unit-root test is to discriminate between stationary behavior and a stochastic trend.

The long-run properties of a stationary time series and that with a stochastic trend are vastly different. Based on a finite number of observations, however, it is challenging to differentiate between the two regimes. A time series can be stationary but look almost identical to a unit-root process in the short run. For example, consider the AR(1) process $\{y_t\}_{t=0}^{\infty}$ defined by $y_0 = 0$ and, for $t = 1, \ldots, T$,

$$y_t = \rho y_{t-1} + \epsilon_t, \tag{0.1}$$

where, say, ϵ_t is i.i.d. standard normally distributed. If $\rho = 1$, we have a unit-root process, while for $|\rho| < 1$ the process is stationary. However, whenever ρ is close to one, the first few observations of the process will have almost the same distribution as if ρ was actually equal to one. In the long run, on the other hand, the behavior under $\rho = 1$ is vastly different from that under $\rho < 1$. This is why unit-root tests have important policy implications as well as being an important ingredient for the statistical analysis of time-series data.

For example, the hypothesis of purchasing power parity is often tested by applying a unit-root test to the real exchange rate. Rejecting a unit-root in the real exchange rate means that the price levels of the two countries will roughly be the same in the long-run and is thus evidence for purchasing power parity. When real exchange rates exhibit unit-root behavior, on the other hand, price levels could diverge in the long run.

One approach to overcome the difficulty of distinguishing between stationary and unit-root series based on a limited number of time periods is to use multiple series at once. For example, when testing for purchasing power parity, one can exploit the real exchange rates of multiple countries jointly. That is, for each country $i = 1, \ldots, n$, one observes

$$y_{it} = \rho_i y_{i,t-1} + \epsilon_{it}. \tag{0.2}$$

By using n such series, the power of the unit root test can typically be improved at a rate of $\sqrt{n}$. Of course, this comes at the cost of imposing some homogeneity on the stationarity properties of the individual series. Typically, the null hypothesis is that all units have a unit root, while they are all stationary under the alternative. Sometimes, the units are allowed to have different values of ρ_i under the alternatives, but are on average stationary in the sense that the average ρ_i is smaller than one.

In Chapter 1, we relax the assumption that the units are on average stationary under alternatives and derive the power envelope for panel unit root tests where heterogeneous alternatives are modeled via zero-expectation random perturbations, i.e.,

$$\rho_i = 1 + \frac{h}{Tn^{1/4}} U_i, \quad h \leq 0, \tag{0.3}$$

for mean-zero perturbations U_i. While (0.3) is a common way to model heterogeneous alternatives, it is typically assumed that their mean is positive. We show that relaxing this assumption means that power gains are only possible at rate $n^{1/4}$. We obtain an asymptotically uniformly most powerful test and discuss how to proceed when one is agnostic about the expectation of the perturbations.

For the subsequent chapters, we go back to the more standard assumption of the U_i having a positive mean, or, for simplicity, assume $U_i = 1$ altogether. However, we relax one of the main limitations of the 'first-generation' panel unit-root tests as described in (0.2): the independence across units. In most applications, the panel units will not be independent of each other, but instead may depend on some common shocks. For example, the real exchange rates of different countries may all depend on events in the numeraire country or other global powers, as well as, for example, global health crises. In panel unit root tests, this is typically accounted for by adding unobserved common factors to the specification.

In Chapter 2, we reconsider the two prevalent approaches in the literature, that of Moon and Perron (2004), who specify a factor model for the innovations, and the PANIC setup proposed in Bai and Ng (2004), who test common factors and idiosyncratic deviations separately for unit roots. While these frameworks have been considered as completely different, we show that, in case of Gaussian innovations, testing for a unit-root in the observations a la Moon and Perron (2004) is asymptotically equivalent to the testing problem for the idiosyncratic parts in PANIC. Using Le Cam's theory of statistical experiments we derive an optimal test jointly in both setups. We show that the popular Moon and Perron (2004) and Bai and Ng (2010) tests only attain the power envelope in case there is no heterogeneity in the long-run variance of the idiosyncratic components. The new test is asymptotically uniformly most powerful irrespective of possible heterogeneity. Moreover, it turns out that for any test, satisfying a mild regularity condition, the size and local asymptotic power are the same under both data generating processes. Monte Carlo simulations corroborate our asymptotic results and document significant gains in finite-sample power if the variances of the idiosyncratic shocks differ substantially among the cross sectional units.

One way to phrase the results of Chapter 2 is that specifying a factor model in the innovations is equivalent to testing the idiosyncratic parts for a unit root in a component specification. However, nonstationarity in the observations may also be due to nonstationary factors, which would not be picked up by the unit-root tests considered in Chapter 2. Chapter 3, therefore studies unit-root tests for unobserved common factors in large panels. Recent panel unit-root tests typically allow for cross-sectional correlation due to common unobserved factors. As originally proposed in Bai and Ng (2004) ('PANIC'), unit-root tests are applied separately to the common factors and idiosyncratic deviations. While the testing problem for the idiosyncratic parts is in many cases well understood, the testing problem for the factors has received much less attention. Bai and Ng (2004) show that using principal component estimates in ADF tests does not change their properties. We generalize this result to other unit-root tests and other factor estimates, which can lead to higher finite sample powers. In particular, we show that a Kalman smoother imposing the null hypothesis to estimate the factors often has a simple closed-form solution that avoids the computational issues usually associated with such methods.

We also discuss the implications of including deterministic trends in the factor equation, i.e., having factors with non-zero mean innovations. This specification can be considered as an alternative to including individual deterministic trends for each unit. Although this leads to nontrivial powers closer to the unit root, we can again attain these powers based on estimated factors. In particular, we propose tests based on simple cross-sectional averages that are asymptotically uniformly most powerful. We derive the properties of these unit root tests in the presence of multiple potentially cointegrated factors and show that they can be interpreted as unit-root tests for the observations. The cross-sectional averaging approach can lead to higher powers than cointegration-rank based tests and does not require pre-estimation of the total number of factors.

The final Chapter 4 revisits the testing problem for the unobserved common factors, but exploits additional observed covariates that are known to be stationary to obtain higher powers. A typical macroeconomic example for such a covariate would, for example, be changes in the unemployment rate. The starting point is the popular PANIC framework and we analyze the potential power gains due to observing additional stationary covariates, focusing on panel unit-root tests that are robust to cross-sectional cointegration, i.e., tests for a unit root in the common unobserved factors. The stationary, observed covariates are assumed to be unit-specific but allowed to be cross-sectionally correlated. We differentiate two cases: one in which the contribution of the factor of interest to the covariance structure of the covariate can be perfectly identified, and a more general one, where the contribution of the factor innovations in the covariate equation is perturbed by another unobserved common shock.

In the former case, the inclusion of stationary covariates leads to vastly more powerful tests, entailing a faster convergence rate. We first analyze the problem for an observed factor, and show that the statistical experiment is locally asymptotically mixed normal (LAMN). This implies that no UMP test exists, but we obtain an asymptotically optimal invariant test. We demonstrate how to conduct valid inference also based on estimated factors. The improved rate allows us to compare different factor estimation schemes in terms of resulting asymptotic power. When implemented well, the asymptotic power of estimated factor based tests is relatively close to the observed-factor power envelope.

In the second case, the statistical problem is closely related to that of univariate unit-root tests with stationary factors that have been studied in Elliott and Jansson (2003) and Hansen (1995). We demonstrate that the original time-series experiment is locally asymptotically Brownian Functional (LABF) but converges to the better understood LAMN case as the contribution of the covariate grows to 1. Moreover, we show that the CADF test of Hansen (1995) becomes optimal invariant as the share of the variation explained by the covariate converges to unity. This explains why the tests of Hansen (1995) are competitive in terms of power to those of Elliott and Jansson (2003), in particular when the covariate is more important. We show that both the CADF tests and the point-optimal tests can also be implemented in a panel setting with unobserved common factors and that their optimality properties carry over to the panel setup.

Chapter 1

The power envelope of panel unit root tests in case stationary alternatives offset explosive ones[1]

Abstract

We derive the power envelope for panel unit root tests where heterogeneous alternatives are modeled via zero-expectation random perturbations. We obtain an asymptotically UMP test and discuss how to proceed when one is agnostic about the expectation of the perturbations.

[1] This chapter is based on Becheri, Drost, Van den Akker, and Wichert (2016).

1.1 Introduction

We start from the setup of Moon, Perron, and Phillips (2007), followed by Becheri, Drost, and Van den Akker (2015a), which study the asymptotic power envelope for the unit root testing problem in a Gaussian cross-sectionally independent panel where the observations Y_{it} for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, are generated by

$$
\begin{aligned}
Y_{it} &= m_i + Y_{it}^0, \\
Y_{it}^0 &= \rho_i Y_{it-1}^0 + \sigma_i \epsilon_{it},
\end{aligned}
$$

with m_i a deterministic observed fixed effect, $Y_{i0}^0 = 0$, and ε_{it} satisfying Assumption 1.1(a) below. Both papers assume the heterogeneous autoregression coefficients ρ_i to be generated according to the random coefficient structure $\rho_i = 1 + hU_i/(\sqrt{n}T)$ where $U_1, \ldots, U_n$ are i.i.d. unobserved random variables with mean 1 and unknown distribution. The results from Moon, Perron, and Phillips (2007) and Becheri, Drost, and Van den Akker (2015a) cannot be extended to the case where the perturbations have zero mean since the power envelopes would be flat (which intuitively means that there do not exist tests that can detect alternatives at the localizing rate $\sqrt{n}T$).

In this note we assume U_i to have mean zero and, more specifically, to satisfy Assumption 1.1(c) below and we reparameterize ρ_i as

$$
\rho_i = 1 + \frac{h}{Tn^\gamma} U_i, \quad h \leq 0, \tag{1.1}
$$

for some appropriate value of γ. Note that, U_i being unobserved, the sign of h is unidentified; thus there is no loss of generality in assuming $h \leq 0$.

Remark 1.1.1 *Alternatively, one could restate our local alternatives as* $\rho_i = 1 + U_i$ *where the variance of* U_i *is* $\frac{h^2}{T^2 n^{2\gamma}}, h \leq 0$*. This highlights the fact that the sign of* h *is not identified.*

Assumption 1.1 *(a) The innovations* ε_{it}*,* $i, t \in \mathbb{N}$*, are i.i.d.* $N(0, 1)$*.*

(b) The deterministic scale parameters σ_i *are positive, i.e.* $\sigma_i > 0$ *for* $i \in \mathbb{N}$*.*

(c) *The perturbations U_i, $i \in \mathbb{N}$, are i.i.d. with mean* 0 *and variance* 1*, have bounded support, and are independent of the idiosyncratic shocks ε_{it}, $i, t \in \mathbb{N}$. Moreover, the moment generating function of U_1 exists on an open interval containing* 0*.*

Throughout, we are interested in testing the unit root hypothesis

$$\mathrm{H}_0 : h = 0 \text{ versus } \mathrm{H}_a : h < 0. \tag{1.2}$$

Under the null hypothesis, each panel unit has a unit root whereas, under the alternative, there are both explosive and stationary time series $\{Y_{it},\, t \in \mathbb{N}\}$. Assumption 1.1 allows the U_i to have an atom at zero, so a random fraction of the time series $\{Y_{it},\, t \in \mathbb{N}\}$ might have a unit root under the alternative.

In this note we show that, under Assumption 1.1, the alternatives are contiguous to the null hypothesis if $\gamma = 1/4$. Note that this is a different rate than the one in Moon, Perron, and Phillips (2007) and Becheri, Drost, and Van den Akker (2015a) (where U_1 has expectation 1 and $\gamma = 1/2$). We derive the UMP test for (1.2) and we also compare this test to the UMP test for the setting where the expectation of U_1 is 1. The m_i and σ_i are treated as unknown nuisance parameters.

1.2 Main results

First we derive the limit experiment of the model where m_i and σ_i are known. This yields the power envelope for the testing problem (1.2). In Section 2.2, we prove *adaptivity* of our problem with respect to the nuisance parameters m_i and σ_i and propose an optimal test.

1.2.1 Limit experiment and Power envelope

In this section we assume the parameters m_i and σ_i to be known. The limit experiment for this model is given in Proposition 1.2.1.

Let $\mathbb{P}_h^{(n,T)}$ denote the law of $\mathbf{Y} := \{Y_{it}, i = 1, \dots, n, t = 1, \dots, T\}$, $\tilde{\mathbb{P}}_h^{(n,T)}$ the law of $\mathbf{Y}$ conditional on $U_1, \dots, U_n$, and $\mathbb{P}_u$ the law of $U_1, \dots, U_n$. Note that, thanks to Assumption 1.1(c), under the null, the law of $U_1, \ldots, U_n$ conditional on $\mathbf{Y}$ is still $\mathbb{P}_u$. Unless otherwise indicated, all expectations are taken under H_0.

In order to derive the limit experiment, we have to study the likelihood ratio of our model, that is $\mathrm{d}\mathbb{P}_h^{(n,T)}/\mathrm{d}\mathbb{P}_0^{(n,T)}$. To compute it, we use the following relation between likelihood ratios:

$$\frac{\mathrm{d}\mathbb{P}_h^{(n,T)}}{\mathrm{d}\mathbb{P}_0^{(n,T)}} = \mathbb{E}\left[\frac{\mathrm{d}\tilde{\mathbb{P}}_h^{(n,T)}}{\mathrm{d}\tilde{\mathbb{P}}_0^{(n,T)}} \,\middle|\, \mathbf{Y}\right], \tag{1.3}$$

where $\mathrm{d}\tilde{\mathbb{P}}_h^{(n,T)}/\mathrm{d}\tilde{\mathbb{P}}_0^{(n,T)}$ is the likelihood ratio of the model where both Y_{it} and U_i are observed.

Let $\Delta Y_{it} = Y_{it} - Y_{it-1}$ for $i = 1, \ldots n$ and $t = 1, \ldots, T$, and let us introduce the partial sum process $W_i^{(T)}$ as

$$W_i^{(T)}(u) := \frac{1}{\sqrt{T}\sigma_i} \sum_{t=1}^{[Tu]} \Delta Y_{it},$$

and define

$$X_i^{(T)} := \int_0^1 W_i^{(T)}(u-)\mathrm{d}W_i^{(T)}(u) \text{ and } J_i^{(T)} := \int_0^1 \left(W_i^{(T)}(u-)\right)^2 \mathrm{d}u, \tag{1.4}$$

where $W(u-) = \lim_{x \to u^-} W(x)$. The likelihood ratio $\mathrm{d}\tilde{\mathbb{P}}_h^{(n,T)}/\mathrm{d}\tilde{\mathbb{P}}_0^{(n,T)}$ can be easily computed thanks to Assumption 1.1(a) and it is given by

$$\frac{\mathrm{d}\tilde{\mathbb{P}}_h^{(n,T)}}{\mathrm{d}\tilde{\mathbb{P}}_0^{(n,T)}} = \prod_{i=1}^{n} \exp\left(U_i \frac{h}{n^{1/4}} X_i^{(T)} - \frac{h^2 U_i^2}{2\sqrt{n}} J_i^{(T)}\right). \tag{1.5}$$

In the following proposition, we make use of (1.3)–(1.5) to establish the LAN property for the model of interest under joint asymptotics $(T, n) \to \infty$, as in Becheri, Drost, and Van den Akker (2015a). The proof is postponed to the appendix.

Remark 1.2.1 *Note that* $(T, n) \to \infty$ *means that* $\min(T, n) \to \infty$.

Proposition 1.2.1 *Let Assumption* 1.1 *hold and put* $\gamma = 1/4$. *Then, under* $\mathbb{P}_0^{(n,T)}$ *as* $(T, n) \to \infty$,

$$\log \frac{\mathrm{d}\mathbb{P}_h^{(n,T)}}{\mathrm{d}\mathbb{P}_0^{(n,T)}} = h^2 \Delta_{n,T} - \frac{1}{2} h^4 J + o_p(1), \tag{1.6}$$

where $J = 5/8$ *and, under* $\mathbb{P}_0^{(n,T)}$

$$\Delta_{n,T} = \frac{1}{2} \sum_{i=1}^{n} \frac{(X_i^{(T)})^2 - J_i^{(T)}}{\sqrt{n}} \xrightarrow{d} N(0, J). \tag{1.7}$$

Moreover, under $\mathbb{P}_h^{(n,T)}$, $\Delta_{n,T} \xrightarrow{d} N(h^2 J, J)$ *as* $(T, n) \to \infty$.

Remark 1.2.2 *In Assumption 1.1 a), it might be possible to replace the Gaussian assumption by some milder conditions. Plausibly, the results of this note still hold if* $\epsilon_{i,t}$ *satisfy a functional central limit theorem for arrays that would ensure convergence of the partial sums to Wiener processes. However, this is beyond the aim of this note.*

Remark 1.2.3 *Note that the first moment of* U_i *being zero implies that the first term of the typical likelihood-ratio expansion drops out. The new central sequence mirrors the typical unit-root testing statistic* $X_i^{(T)}$, *but we now have to consider its square. At the same time, the contiguity rate is slower. One might wonder whether one could attain an even slower rate and even higher powers of* X *by setting additional moments to zero. However, if the second moment is also zero, the* U_i *have no cross-sectional variation anymore, likely necessitating an analysis that is not at all in the spirit of this chapter.*

This proposition and an application of Theorem 9.4 in Van der Vaart (2000) imply that the sequence of experiments $\{\mathbb{P}_h^{(n,T)} : h \in \mathbb{R}_-\}$ converges to the experiment $\{N(h^2 J, J) : h \in \mathbb{R}_-\}$ under $\mathbb{P}_h^{(n,T)}$.[2] Using the Asymptotic Representation Theorem,[3] we can thus obtain the (asymptotic) power envelope for testing hypothesis (1.2). The resulting power envelope is presented in the following corollary.

Corollary 1.2.1 *Let Assumption* 1.1 *hold,* $\gamma = 1/4$, $\alpha \in (0, 1)$, *and denote* $z_\alpha = \Phi^{-1}(1-\alpha)$. *Consider a test* $\varphi(Y_{11}, \ldots, Y_{nT})$ *of level* α *with power* $\pi_{n,T}(h)$.

---

[2] Note that Theorem 9.4 in Van der Vaart (2000) needs to be applied to the model where ρ_i is re-parameterized in terms of the local parameter $\tilde{h} = h^2$ as $\rho_i = 1 - \frac{\sqrt{\tilde{h}}}{Tn^{1/4}} U_i$.

[3] See, for example, Chapter 15 in Van der Vaart (2000).

Then, for all h, we have

$$\limsup_{(T,n)\to\infty} \pi_{n,T}(h) \leq \Phi\left(-z_\alpha + h^2\sqrt{J}\right), \tag{1.8}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Moreover, let

$$t_{n,T} = \frac{\Delta_{n,T}}{\sqrt{J}} = \sqrt{\frac{2}{5}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left((X_i^{(T)})^2 - J_i^{(T)}\right). \tag{1.9}$$

Then, for all h, the test $\psi_{n,T} = 1\{t_{n,T} \geq z_\alpha\}$ attains the upper bound (1.8) *uniformly in h.*

Remark 1.2.4 *Note that this test is semiparametrically optimal in the sense that the power envelope* (1.8) *does not depend on the distribution of the perturbations U_i.*

1.2.2 A feasible test

In this section we treat m_i and σ_i as unknown nuisance parameters. We show that the unit root testing problem is adaptive with respect to these parameters, that is the power envelope can still be attained when m_i and σ_i are unknown when $n/T \to 0$.[4] In fact, we can define a test whose (local and asymptotic) power achieves the power envelope (1.8) while being invariant with respect to m_i and where σ_i are estimated. This test is based on a feasible version of the central sequence $\Delta_{n,T}$, obtained by replacing σ_i^2, $i = 1, \ldots, n$, by

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=2}^{T}(\Delta Y_{it})^2.$$

Our test statistic $\hat{t}_{n,T}$ is thus defined on the basis of (1.9) as:

$$\hat{t}_{n,T} = \sqrt{\frac{2}{5}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\left(\frac{1}{T}\sum_{t=3}^{T}\left(\sum_{s=2}^{t-1}\frac{1}{\hat{\sigma}_i}\Delta Y_{is}\right)\frac{1}{\hat{\sigma}_i}\Delta Y_{it}\right)^2 - \frac{1}{T^2}\sum_{t=3}^{T}\frac{1}{\hat{\sigma}_i^2}\left(\sum_{s=2}^{t-1}\Delta Y_{is}\right)^2\right)$$

[4] The additional assumption on n and T is needed to handle an increasing number of nuisance parameters; this assumption is standard in the literature, see, for instance, Moon, Perron, and Phillips (2007) and Becheri, Drost, and Van den Akker (2015a).

$$= \sqrt{\frac{2}{5}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(\frac{\sigma_i^4}{\hat{\sigma}_i^4} (X_i^{(T)})^2 - \frac{\sigma_i^2}{\hat{\sigma}_i^2} J_i^{(T)} \right) - \frac{\sigma_i^4}{\hat{\sigma}_i^4} r_{i,T}^a + \frac{\sigma_i^2}{\hat{\sigma}_i^2} r_{i,T}^b \right], \tag{1.10}$$

where

$$\begin{aligned} r_{i,T}^a = & \frac{1}{\sigma_i^4} \left(\frac{1}{T} \sum_{t=2}^{T} \Delta Y_{i1} \Delta Y_{it} \right)^2 \\ & + \frac{2}{\sigma_i^4} \left(\frac{1}{T} \sum_{t=2}^{T} \Delta Y_{i1} \Delta Y_{it} \right) \left(\frac{1}{T} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \Delta Y_{is} \Delta Y_{it} \right), \text{ and} \\ r_{i,T}^b = & \frac{T-1}{T^2} \frac{\Delta Y_{i1}^2}{\sigma_i^2} + \frac{2}{T^2} \frac{\Delta Y_{i1}}{\sigma_i^2} \sum_{t=3}^{T} \sum_{s=2}^{t-1} \Delta Y_{is}. \end{aligned}$$

Note that $r_{i,T}^a$ and $r_{i,T}^b$ are remainder terms due to not observing $Y_{i0} = m_i$.

The following proposition proves that $\hat{t}_{n,T}$ is asymptotically equivalent to $t_{n,T}$ in the sense that they differ only for order $o_P(1)$ terms.

Proposition 1.2.2 *Let Assumption* 1.1 *hold and suppose* $n/T \to 0$. *Then we have, for all* $h \in \mathbb{R}$ *and under* $P_h^{(n,T)}$ *as* $(T, n) \to \infty$,

$$\hat{t}_{n,T} = t_{n,T} + o_P(1). \tag{1.11}$$

Remark 1.2.5 *From* (1.11) *and Corollary* 1.2.1, *it readily follows that the test* $\hat{\psi}_{n,T} = 1\{\hat{t}_{n,T} > z_\alpha\}$ *is asymptotically UMP.*

1.3 Testing for a unit root when $\mathbb{E}U_1$ is unknown

In practice, it may be difficult to determine whether some data were generated under the DGP introduced in Section 4.2, where $\mathbb{E}U_1 = 0$, or under the DGP considered in Moon, Perron, and Phillips (2007) and Becheri, Drost, and Van den Akker (2015a), where U_1 satisfies Assumption 1.2 below, i.e. $\mathbb{E}U_1 = 1$. In this section we address the problem of testing for a unit root while being agnostic about the first moment of U_1. For notational simplicity, we consider the test statistics $t_{n,T}$ and $\tau_{n,T}$ (introduced below) which, as in Section 1.2.1, rely on the nuisance parameters being known; the extension to their estimated, feasible counterparts is immediate as long as $n/T \to 0$.

Assumption 1.2 *The perturbations U_i, $i \in \mathbb{N}$, are i.i.d. with mean* 1 *and independent of the idiosyncratic shocks ε_{it}, $i, t \in \mathbb{N}$. Moreover, the moment generating function of U_1 exists on an open interval containing 0.*

Note that we need $\gamma = 1/2$ to ensure contiguity of the alternatives with respect to the null hypothesis under Assumption 1.2 (see Moon, Perron, and Phillips (2007) and Becheri, Drost, and Van den Akker (2015a)).

In Section 1.2, we have shown that optimal inference for the testing problem (1.2) can be based on $t_{n,T}$ when $\mathbb{E}U_1 = 0$. Becheri, Drost, and Van den Akker (2015a) shows that, if $\mathbb{E}U_1 = 1$, optimal inference can be based on

$$\tau_{n,T} = \frac{\sqrt{2}}{\sqrt{n}} \sum_{i=1}^{n} X_i^{(T)}.$$

Let us denote by $\mathbb{Q}_h^{(n,T)}$ the law of $\mathbf{Y}$ when the U_i satisfy Assumption 1.2 and ρ_i satisfies (1.1) with $\gamma = 1/2$. Clearly, since $\mathbb{Q}_0^{(n,T)} = \mathbb{P}_0^{(n,T)}$, the statistics $t_{n,T}$ and $\tau_{n,T}$ converge to a standard normal distribution under both $\mathbb{P}_0^{(n,T)}$ and $\mathbb{Q}_0^{(n,T)}$ (see Proposition 1.2.1 and Proposition 4.2 in Becheri, Drost, and Van den Akker (2015a)). This implies that both statistics are valid in terms of size for testing the unit root hypothesis (1.2) irrespective of the expectation of the U_1.

In what follows, we propose two tests based on the statistics $t_{n,T}$ and $\tau_{n,T}$ having power against $h < 0$ even when we do not know whether U_1 satisfies Assumption 1.1(c) or Assumption 1.2.

Lemma 1.3.1 provides the distribution of $t_{n,T}$ and $\tau_{n,T}$ under $\mathbb{P}_h^{(n,T)}$ and $\mathbb{Q}_h^{(n,T)}$. Its proof relies on a straightforward application of Le Cam's third lemma and can be found in the appendix.

Lemma 1.3.1 *Let Assumption* 1.1(a)-(b) *hold.*

(i) Let Assumption 1.1(c) *hold and* $\gamma = 1/4$. *Then, under* $\mathbb{P}_h^{(n,T)}$*, as* $(T, n) \to \infty$*,*

$$t_{n,T} \xrightarrow{d} N(h^2\sqrt{5/8}, 1) \quad \textit{and} \quad \tau_{n,T} \xrightarrow{d} N(h^2\sqrt{2/9}, 1). \tag{1.12}$$

(ii) *Let Assumption* 1.2 *hold and* $\gamma = 1/2$. *Then, under* $\mathbb{Q}_h^{(n,T)}$, *as* $(T,n) \to \infty$,

$$t_{n,T} \xrightarrow{d} N(h\sqrt{8/45}, 1) \quad \textit{and} \quad \tau_{n,T} \xrightarrow{d} N(h/\sqrt{2}, 1).$$

This result provides guidance on defining tests that do not rely on knowing the expectation of U_1 and it enables us to compute their (local and asymptotic) power.

From Lemma 1.3.1(i), we conclude that, under $\mathbb{P}_h^{(n,T)}$, one would reject for a large value of either test statistic. On the contrary, under $\mathbb{Q}_h^{(n,T)}$, one would reject for small values. This suggests that, when it is not known whether $\mathbb{E}U_1 = 0$ or $\mathbb{E}U_1 = 1$, one should reject for both large and small values of $t_{n,T}$ and $\tau_{n,T}$. Following this lead, we can define two tests having power both under $\mathbb{P}_h^{(n,T)}$ and $\mathbb{Q}_h^{(n,T)}$. Let us define the tests:

$$\varphi_{n,T} = 1 - 1\{-z_{\alpha/2} < t_{n,T} < z_{\alpha/2}\} \quad \text{and} \quad \tilde{\varphi}_{n,T} = 1 - 1\{-z_{\alpha/2} < \tau_{n,T} < z_{\alpha/2}\}.$$

From Lemma 1.3.1, we easily obtain the (local and asymptotic) powers of these tests which are presented in the following corollary.

Corollary 1.3.1 *Let Assumption* 1.1(a)-(b) *hold.*

(i) *Let Assumption* 1.1(c) *hold and* ρ_i *satisfy* (1.1) *with* $\gamma = 1/4$. *Then, under* $\mathbb{P}_h^{(n,T)}$,

$$\begin{aligned}&\lim_{(T,n)\to\infty}(1 - \mathbb{P}_h^{(n,T)}[-z_{\alpha/2} < t_{n,T} < z_{\alpha/2}])\\ &\quad = \Phi(-z_{\alpha/2} - h^2\sqrt{5/8}) + \Phi(-z_{\alpha/2} + h^2\sqrt{5/8}),\end{aligned}$$

and

$$\begin{aligned}&\lim_{(T,n)\to\infty}(1 - \mathbb{P}_h^{(n,T)}[-z_{\alpha/2} < \tau_{n,T} < z_{\alpha/2}])\\ &\quad = \Phi(-z_{\alpha/2} - h^2\sqrt{2/9}) + \Phi(-z_{\alpha/2} + h^2\sqrt{2/9})\end{aligned}$$

(ii) *Let Assumption* 1.2 *hold and* ρ_i *satisfy* (1.1) *with* $\gamma = 1/2$. *Then, under* $\mathbb{Q}_h^{(n,T)}$,

$$\lim_{(T,n)\to\infty}(1 - \mathbb{Q}_h^{(n,T)}[-z_{\alpha/2} < t_{n,T} < z_{\alpha/2}])$$

$$= \Phi(-z_{\alpha/2} - h\sqrt{8/45}) + \Phi(-z_{\alpha/2} + h\sqrt{8/45})$$

and

$$\lim_{(T,n)\to\infty} (1 - \mathbb{Q}_h^{(n,T)}[-z_{\alpha/2} < \tau_{n,T} < z_{\alpha/2}])$$
$$= \Phi(-z_{\alpha/2} - h/\sqrt{2}) + \Phi(-z_{\alpha/2} + h/\sqrt{2}),$$

Corollary 1.3.1(i) shows that under $\mathbb{P}_h^{(n,T)}$, the power of the test $\varphi_{n,T}$ is (asymptotically) higher than that of $\tilde{\varphi}_{n,t}$, while, from Corollary 1.3.1(ii) it follows that under $\mathbb{Q}_h^{(n,T)}$ the power of $\tilde{\varphi}_{n,t}$ is higher than that of $\varphi_{n,T}$. Furthermore, from Corollary 3.1 it is clear that neither $\varphi_{n,t}$ nor $\tilde{\varphi}_{n,T}$ is optimal. It is, however, important to note that the one-sided test $\psi_{n,T}$, which is optimal under $\mathbb{P}_h^{(n,T)}$, always has power less than the size α in the $\mathbb{Q}_h^{(n,T)}$-model. A similar remark applies to the test $1\{\tau < -z\}$, which is optimal under $\mathbb{Q}_h^{(n,T)}$, but has power less than α under $\mathbb{P}_h^{(n,T)}$. This implies that these tests are pretty useless when it is not possible to decide on the model $\mathbb{P}_h^{(n,T)}$ or $\mathbb{Q}_h^{(n,T)}$. Therefore, when it is not possible to determine under which DGP the data were generated, we recommend to use the two-sided tests.

1.A Proofs

1.A.1 Proof of Proposition 1.2.1

In the following, we first establish convergence (1.7), then we prove the expansion (1.6), and finally we establish the convergence result under the alternative. All probabilities and expectations are evaluated under $\mathbb{P}_0^{(n,T)}$ unless otherwise stated.

For $m = 2, \ldots, 8$, we introduce the random variables $K_{mi} = f_m(X_i^{(T)}, J_i^{(T)})$, $i = 1, \ldots, n$, where

$$f_2(x,j) = \tfrac{x^2-j}{2},\ f_3(x,j) = \tfrac{x^3-3xj}{6},\ f_4(x,j) = \tfrac{x^4-6x^2j+3j^2}{24},$$
$$f_5(x,j) = \tfrac{3xj^2-2x^3j}{24},\ f_6(x,j) = \tfrac{3x^2j^2-j^3}{48},\ f_7(x,j) = \tfrac{-xj^3}{48},\ f_8(x,j) = \tfrac{j^4}{384},$$

and $X_i^{(T)}$ and $J_i^{(T)}$ are as defined in (1.4). Note that these are approximations to the stochastic integrals $X_i = \int_0^1 W_i(u)\mathrm{d}W_i(u) = 1/2(W_i^2(1) - 1)$ and $J_i = \int_0^1 W_i^2(u)\mathrm{d}u$, based on independent Brownian motions W_i and that, for fixed m, the variables K_{mi}, $i = 1, \ldots, n$, are i.i.d.

Put $\mu_m^{(T)} = \mathbb{E}K_{m1}$, $\sigma_m^{(T)} = \sqrt{\mathrm{Var}(K_{m1})}$. Some tedious calculations show, as $T \to \infty$,

$$\mu_m^{(T)} = Ef_m(X_1^{(T)}, J_1^{(T)}) \quad \to \quad \mu_m = Ef_m(X_1, J_1) \text{ and}$$

$$\sigma_m^{(T)} = \sqrt{\text{Var}(f_m(X_1^{(T)}, J_1^{(T)}))} \quad \to \quad \sigma_m = \sqrt{\text{Var}(f_m(X_1, J_1))}.$$

Although it is not strictly necessary to demonstrate convergence to the moments of the limiting process (X_i, J_i), it provides some additional intuition why the sequences $\mu_m^{(T)}$ and $\sigma_m^{(T)}$ are bounded. Furthermore, the limits μ_m and σ_m can be easily obtained from Ito calculus. In particular, we obtain $\mu_2^{(T)} = \mu_2 = \mu_3^{(T)} = \mu_3 = 0$, $\mu_4^{(T)} \to \mu_4 = 0$, and $\sigma_2^{(T)} \to \sigma_2 = \sqrt{\frac{5}{8}}$. Using once again the Gaussianity of our innovations, it can be demonstrated that higher moments of K_{m1} are bounded as well.

The previous considerations on the moments of K_{m1} enable us to apply a Central Limit Theorem for a double array of random variables see p. 32 in Serfling (1980),

$$\frac{\sum_{i=1}^n (K_{mi} - \mu_m^{(T)})}{\sqrt{n}\sigma_m^{(T)}} \xrightarrow{d} N(0,1) \text{ and } \frac{1}{n}\sum_{i=1}^n (K_{mi}^2 - (\sigma_m^{(T)})^2 - (\mu_m^{(T)})^2) \xrightarrow{P} 0. \tag{1.A.1}$$

As $\mu_2^{(T)} = 0$ and $\sigma_2^{(T)} \to \sqrt{\frac{5}{8}}$, (1.A.1) establishes the limiting distribution (1.7) of the central sequence Δ_{nT} as well as a useful approximation to the Fisher information $J = \frac{5}{8}$, namely

$$\Delta_{nT} = \frac{1}{\sqrt{n}}\sum_{i=1}^n K_{2i} \xrightarrow{d} N(0, J) \text{ and } \frac{1}{n}\sum_{i=1}^n K_{2i}^2 \xrightarrow{P} J.$$

Next, we obtain the desired expansion of the loglikelihood ratio. Define $a_i = h\frac{U_i}{n^{1/4}}X_i^{(T)}$ and $b_i = -h^2\frac{U_i^2}{2\sqrt{n}}J_i^{(T)}$. From (1.3) and (1.5) and using the independence across i, we get

$$\log\frac{\mathrm{d}\mathbb{P}_h^{(n,T)}}{\mathrm{d}\mathbb{P}_0^{(n,T)}} = \sum_{i=1}^n \log\mathbb{E}(\exp(a_i + b_i) \mid \mathbf{Y}).$$

Expanding the exponential, we have, for some $0 \leq |\xi_{1,i}| = |\xi_1(U_i, n, T, X_i^{(T)}, J_i^{(T)})| \leq |a_i + b_i|$,

$$\log\frac{\mathrm{d}\mathbb{P}_h^{(n,T)}}{\mathrm{d}\mathbb{P}_0^{(n,T)}} = \sum_{i=1}^n \log\mathbb{E}\Big(1 + \sum_{k=1}^4 \frac{1}{k!}(a_i + b_i)^k + \frac{1}{5!}e^{\xi_{1,i}}(a_i + b_i)^5 \mid \mathbf{Y}\Big) = \sum_{i=1}^n \log(1 + L_i^{(T,n)}).$$

Recall U_i, with $\mathbb{E}U_i = 0$ and $\mathbb{E}U_i^2 = 1$, is independent of both $X_i^{(T)}$ and $J_i^{(T)}$ (see Assumption 1.1); hence

$$\begin{aligned} L_i^{(T,n)} &= \mathbb{E}\Big(\sum_{k=1}^4 \frac{1}{k!}(a_i + b_i)^k + \frac{1}{5!}e^{\xi_{1,i}}(a_i + b_i)^5 \mid \mathbf{Y}\Big) \\ &= h^2\frac{1}{\sqrt{n}}K_{2i} + \sum_{m=3}^8 h^m n^{-m/4}(\mathbb{E}U_1^m)K_{mi} + \frac{1}{120}\mathbb{E}(e^{\xi_{1,i}}(a_i + b_i)^5 \mid \mathbf{Y}). \end{aligned} \tag{1.A.2}$$

Using boundedness of moments and employing the following inequality for i.i.d. random variables due to Gumbel (1954),

$$\mathbb{E}\max_{i\leq n}|K_{mi}|^\ell \leq \mathbb{E}|K_{m1}|^\ell + \sqrt{\text{Var}(|K_{m1}|^\ell)}\frac{n-1}{\sqrt{2n-1}}, \ \ell > 0,$$

we obtain $\mathbb{E}\max_{i\leq n}|K_{mi}|^\ell = O(\sqrt{n})$. Therefore, the Markov inequality implies,

$$\max_{i\leq n}|K_{mi}| = o_p(n^\alpha) \text{ for any } \alpha > 0.$$

A similar reasoning shows $\zeta_{nT} = n^{1/5} \max_{i\leq n} \left\{ |\frac{h}{n^{1/4}} X_i^{(T)}| + \frac{h^2}{2\sqrt{n}} J_i^{(T)} \right\} = o_p(1)$. This implies that the final term of $L_i^{(T,n)}$ is asymptotically negligible. Indeed, using again a similar reasoning as before, we find that, for all $\epsilon > 0$, there exist n, T and k such that

$$\max_{i\leq n} |\mathbb{E}(e^{\xi_{1,i}}(a_i + b_i)^5 \mid \mathbf{Y})| \leq \frac{k}{n} e^{k\zeta_{nT}} \zeta_{nT}^5 = o_p(n^{-1}),$$

where k is a finite positive constant that depends on the support of U_1 (use Assumption 1.1(c)).

Collect the previous results and repeatedly use the Central Limit Theorem for a double array of random variables (Serfling (1980), p.32), to obtain

$$\mathbb{P}_0^{(n,T)}[\max_{i\leq n} |L_i^{(T,n)}| < \epsilon] \to 1, \tag{1.A.3}$$

$$\sum_{i=1}^n L_i^{(T,n)} = h^2 \frac{1}{\sqrt{n}} \sum_{i=1}^n K_{2i} + o_p(1) \xrightarrow{d} N(0, h^4 J), \tag{1.A.4}$$

$$\sum_{i=1}^n (L_i^{(T,n)})^2 = h^4 \frac{1}{n} \sum_{i=1}^n K_{2i}^2 + o_p(1) = \frac{5}{8} h^4 + o_p(1) \xrightarrow{p} h^4 J. \tag{1.A.5}$$

Subsequently, proceed with an expansion of the logarithm in the loglikelihood ratio, yielding

$$\log \frac{d\mathbb{P}_h^{(n,T)}}{d\mathbb{P}_0^{(n,T)}} = \sum_{i=1}^n \log(1 + L_i^{(T,n)}) = \sum_{i=1}^n L_i^{(T,n)} - \sum_{i=1}^n \frac{(L_i^{(T,n)})^2}{2} + \sum_{i=1}^n \frac{(L_i^{(T,n)})^3}{3(1+\xi_{2,i})^3},$$

for some $\xi_{2,i}$ between 0 and $L_i^{(T,n)}$. Since $|\xi_{2,i}| \leq |L_i^{(T,n)}| \leq \epsilon$ (with probability converging to one), we find the bound $\left| \sum_{i=1}^n \frac{(L_i^{(T,n)})^3}{3(1+\xi_{2,i})^3} \right| \leq \sum_{i=1}^n \frac{|L_i^{(T,n)}|^3}{3(1-\epsilon)^3} \leq \frac{\epsilon}{3(1-\epsilon)^3} \sum_{i=1}^n |L_i^{(T,n)}|^2$. Therefore, (1.A.3)–(1.A.5) establish the desired expansion.

Finally, an application of Le Cam's third lemma immediately yields convergence of the central sequence to a normal $N(hJ, J)$ distribution under the local alternatives.

1.A.2 Proof of Proposition 1.2.2

As we have shown that our model is LAN, contiguity is obtained from Le Cam's first lemma. Hence we only have to prove (1.11) under $\mathbb{P}_0^{(n,T)}$. In the remainder of this proof, all expressions, probabilities and expectations are evaluated under $\mathbb{P}_0^{(n,T)}$.

We have

$$\hat{t}_{n,T} - t_{n,T} = \sqrt{\frac{2}{5}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(r_{i,nT}^X - r_{i,nT}^J - \frac{\sigma_i^4}{\hat{\sigma}_i^4} r_{i,T}^a + \frac{\sigma_i^2}{\hat{\sigma}_i^2} r_{i,T}^b \right),$$

where

$$r_{i,nT}^X = \left(\frac{\sigma_i^4}{\hat{\sigma}_i^4} - 1 \right) (X_i^{(T)})^2 \quad \text{and} \quad r_{i,nT}^J = \left(\frac{\sigma_i^2}{\hat{\sigma}_i^2} - 1 \right) J_i^{(T)}.$$

Below, we analyze one term at the time and prove each one is $o_P(1)$. First we obtain some handy relationships on replacing the scale parameters σ_i by estimates. We obtain $\sum_{i=1}^n \left(\frac{\hat\sigma_i^2}{\sigma_i^2}-1\right)^2 = o_P(1)$ since

$$\mathbb{E}\sum_{i=1}^n \left(\frac{\hat\sigma_i^2}{\sigma_i^2}-1\right)^2 = \frac{2n}{T-1} \to 0.$$

Hence we also have $\min_{i\le n}\frac{\hat\sigma_i^2}{\sigma_i^2} \xrightarrow{P} 1$ and $\max_{i\le n}\frac{\hat\sigma_i^2}{\sigma_i^2} \xrightarrow{P} 1$. This also implies

$$\sum_{i=1}^n \left(\frac{\sigma_i^2}{\hat\sigma_i^2}-1\right)^2 \le \sum_{i=1}^n \left(\frac{\hat\sigma_i^2}{\sigma_i^2}-1\right)^2 / \min_{i\le n}\frac{\hat\sigma_i^4}{\sigma_i^4} = o_P(1),$$
$$\sum_{i=1}^n \left(\frac{\sigma_i^4}{\hat\sigma_i^4}-1\right)^2 \le \sum_{i=1}^n \left(\frac{\hat\sigma_i^2}{\sigma_i^2}-1\right)^2 \left(\max_{i\le n}\frac{\hat\sigma_i^2}{\sigma_i^2}+1\right)^2 / \min_{i\le n}\frac{\hat\sigma_i^8}{\sigma_i^8} = o_P(1).$$

Since the averages $\frac{1}{n}\sum_{i=1}^n (X_i^{(T)})^4$ and $\frac{1}{n}\sum_{i=1}^n (J_i^{(T)})^2$ are bounded in probability, the Cauchy-Schwarz inequality yields that the leading remainder terms due to $r_{i,nT}^X$ and $r_{i,nT}^J$ are negligible,

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n r_{i,nT}^X\right|^2 \le \sum_{i=1}^n \left(\frac{\sigma_i^4}{\hat\sigma_i^4}-1\right)^2 \frac{1}{n}\sum_{i=1}^n (X_i^{(T)})^4 = o_P(1),$$
$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n r_{i,nT}^J\right|^2 \le \sum_{i=1}^n \left(\frac{\sigma_i^2}{\hat\sigma_i^2}-1\right)^2 \frac{1}{n}\sum_{i=1}^n (J_i^{(T)})^2 = o_P(1).$$

Finally, we show that the remainder terms due to not observing the initial observations Y_{i0}, are negligible. Using $(a+b)^2 \le 2a^2+2b^2$ and Cauchy-Schwarz, we have

$$\begin{aligned}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\sigma_i^4}{\hat\sigma_i^4} r_{i,T}^a\right|^2 \le& 2\max_{i\le n}\frac{\sigma_i^8}{\hat\sigma_i^8}\sum_{i=1}^n \epsilon_{i1}^2\left(\frac{1}{T}\sum_{t=2}^T \epsilon_{it}\right)^2 \\ & \left(\frac{1}{n}\sum_{i=1}^n \epsilon_{i1}^2\left(\frac{1}{T}\sum_{t=2}^T \epsilon_{it}\right)^2 + \frac{4}{n}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=3}^T\sum_{s=2}^{t-1}\epsilon_{is}\epsilon_{it}\right)^2\right), \\ \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\sigma_i^2}{\hat\sigma_i^2} r_{i,T}^b\right|^2 \le& 2\max_{i\le n}\frac{\sigma_i^4}{\hat\sigma_i^4}\frac{1}{T}\sum_{i=1}^n \epsilon_{i1}^2\left(\frac{1}{nT}\sum_{i=1}^n \epsilon_{i1}^2 + \frac{4}{n}\sum_{i=1}^n\left(\frac{1}{T\sqrt{T}}\sum_{t=3}^T\sum_{s=2}^{t-1}\epsilon_{is}\right)^2\right).\end{aligned}$$

To obtain the desired negligibility of these two remaining terms, observe (take expectations and note the similarity to the proofs of the LAN theorem)

$$\sum_{i=1}^n \epsilon_{i1}^2\left(\frac{1}{T}\sum_{t=2}^T \epsilon_{it}\right)^2 = o_P(1), \quad \frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T}\sum_{t=3}^T\sum_{s=2}^{t-1}\epsilon_{is}\epsilon_{it}\right)^2 = O_P(1),$$
$$\frac{1}{T}\sum_{i=1}^n \epsilon_{i1}^2 = o_P(1), \quad \frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T\sqrt{T}}\sum_{t=3}^T\sum_{s=2}^{t-1}\epsilon_{is}\right)^2 = O_P(1).$$

1.A.3 Proof of Lemma 1.3.1

Note that the first part of the statement (i) and the second part of statement (ii) follow from Proposition 1.2.1 above and Proposition 4.2 of Becheri, Drost, and Van den Akker (2015a), respectively. The other two statements can also be obtained by a straightforward application of Le Cam's third lemma. To obtain the appropriate shifts under local alternatives, we calculate the covariance between the central sequences in both set-ups (see also Section 1.A.1)

$$\mathrm{Cov}\left(\frac{X_1^2 - J_1}{2}, X_1\right) = \mathbb{E}\frac{X_1^3 - 3X_1J_1}{6} + \frac{1}{3}\mathbb{E}X_1^3 = \mathbb{E}K_{31} + \frac{1}{3} = \frac{1}{3}.$$

To compute the distribution of $\tau_{n,T}$ under $\mathbb{P}_h^{(n,T)}$, we need to consider the (asymptotic) covariance between $\tau_{n,T}$ and the log-likelihood ratio $\log \mathrm{d}\mathbb{P}_h^{(n,T)}/\mathrm{d}\mathbb{P}_0^{(n,T)}$. Since, the central sequence $\Delta_{n,T}$ is multiplied by h^2 and the $\tau_{n,T}$-test has a factor $\sqrt{2}$ in front of the X_i, the shift under local alternatives $\mathbb{P}_h^{(n,T)}$ is $h^2\sqrt{2}/3$: $\tau_{n,T} \xrightarrow{d} N(h^2\sqrt{2/9}, 1)$.

Similarly, we compute the distribution of $t_{n,T}$ under $\mathbb{Q}_h^{(n,T)}$. To obtain the covariance between $t_{n,T}$ and the log-likelihood ratio $\log \mathrm{d}\mathbb{Q}_h^{(n,T)}/\mathrm{d}\mathbb{Q}_0^{(n,T)}$ note that, in the quadratic expansion of $\log \mathrm{d}\mathbb{Q}_h^{(n,T)}/\mathrm{d}\mathbb{Q}_0^{(n,T)}$ from Proposition 4.2 of Becheri, Drost, and Van den Akker (2015a), the central sequence is multiplied by h while the $t_{n,T}$-test has a factor $\sqrt{8/5}$. Hence, under $\mathbb{Q}_h^{(n,T)}$, we obtain a shift of $h\sqrt{8/5}/3$: $t_{n,T} \xrightarrow{d} N(h\sqrt{8/45}, 1)$.

This completes the proof of the lemma.

Chapter 2

Local Asymptotic Equivalence of the Bai and Ng (2004) and Moon and Perron (2004) Frameworks for Panel Unit Root Testing[1]

Abstract

This chapter considers unit root tests in dependent panels with a large cross-sectional and time dimension. We reconsider the two prevalent approaches in the literature, that of Moon and Perron (2004), who specify a factor model for the innovations, and the PANIC setup proposed in Bai and Ng (2004), who test common factors and idiosyncratic deviations separately for unit roots. While these frameworks have been considered as completely different, we show that, in case of Gaussian innovations, testing for a unit-root in the observations à la Moon and Perron (2004) is asymptotically equivalent to the testing problem for the idiosyncratic parts in PANIC. Using Le Cam's theory of statistical experiments we derive an optimal test jointly in both setups. We show that the popular

[1] Based on joint work with I.G. Becheri, F.C. Drost, and R. van den Akker.

Moon and Perron (2004) and Bai and Ng (2010) tests only attain the power envelope in case there is no heterogeneity in the long-run variance of the idiosyncratic components. The new test is asymptotically uniformly most powerful irrespective of possible heterogeneity. Moreover, it turns out that for any test, satisfying a mild regularity condition, the size and local asymptotic power are the same under both data generating processes. Monte Carlo simulations corroborate our asymptotic results and document significant gains in finite-sample power if the variances of the idiosyncratic shocks differ substantially among the cross sectional units.

Testing for unit roots is an important aspect of time series and panel data analysis. See, for example, the monographs Patterson (2011), Patterson (2012), and Choi (2015) for overviews. A well-known problem with univariate unit roots tests is their low power. In the last two decades, increased data availability led to the development of panel unit root tests that increase the statistical power by exploiting the cross-sectional data dimension. The "first generation" of panel unit root tests imposes the panel observations Z_{it} to be independent over panel units i. Surveys of this literature are provided by Banerjee (1999), Baltagi and Kao (2000), Choi (2006), Breitung and Pesaran (2008), and Westerlund and Breitung (2013). O'Connell (1998) and Gutierrez (2006) showed that presence of cross-sectional dependence typically leads to invalidity of "first generation tests". For this reason, a "second generation" of models and tests has been introduced.

This chapter considers two widely used setups for second generation panel unit root tests: the 'PANIC' framework of Bai and Ng (2004) and the framework of Moon and Perron (2004) ('MP').[2] To this end we introduce the following data generating process that covers both frameworks. The observations Z_{it}, $i = 1, \dots, n$ and $t = 1, \dots, T$, are assumed to be generated by the components specification

$$Z_{it} = m_i + Y_{it}, \tag{2.1}$$

[2] These setups are also popular in applied work, see, for example Carvalho and Júlio (2012) and Saldías (2013) for applications to testing Purchasing Power Parity and systemic risk.

$$Y_{it} = \sum_{k=1}^{K} \lambda_{ki} F_{kt} + E_{it}, \tag{2.2}$$

$$E_{it} = \rho E_{i,t-1} + \eta_{it}, \tag{2.3}$$

$$F_{kt} = \rho_k F_{k,t-1} + f_{kt}, \tag{2.4}$$

where λ_{ki} is the loading of *unobserved* factor F_{kt} on panel unit i, the m_i are fixed effects, and the innovations $\{\eta_{it}\}$ and $\{f_{kt}\}$ are assumed to be mutually independent, Gaussian, stationary time series. The innovations $\{\eta_{it}\}$ are idiosyncratic in the sense that they are cross-sectionally independent, i.e., the cross-sectional dependence in the panel is generated by the common factors. The number factors, K, is assumed to be deterministic and known.[3] Section 2.1.2 discusses the precise assumptions.

For $\rho_k = 1, k = 1, \ldots, K$, we obtain the PANIC framework and with $\rho_k = \rho, k = 1, \ldots, K$, we obtain the MP framework, in which we can also rewrite (2.2)-(2.4) as

$$Y_{it} = \rho Y_{i,t-1} + \varepsilon_{it} \text{ and } \varepsilon_{it} = \sum_{k=1}^{K} \lambda_{ki} f_{kt} + \eta_{it}. \tag{2.5}$$

Note that MP uses an autoregressive structure with the factors appearing in the innovations ε_{it} in (2.5), whereas the factors are part of the "mean specification", i.e. (2.2), in the PANIC setup. Consequently, the PANIC framework allows for non-stationarity of Z_{it} generated by the factors F_{kt} and for non-stationarity generated by the idiosyncratic components E_{it}, while the factors and the idiosyncratic components have the same order of integration in the MP framework. Following Bai and Ng (2010), Pesaran, Smith, and Yamagata (2013) and Westerlund (2015), when considering the PANIC framework we focus on testing for unit roots in the idiosyncratic components, i.e. $\mathrm{H}_0 : \rho = 1$ versus $\mathrm{H}_a : \rho < 1$. Note that, under the null hypothesis, the model equations of both models coincide. The two main restrictions on the DGP considered here are the absence of idiosyncratic deterministic trends and the assumption

[3] This number can be estimated consistently, so this makes no difference for the asymptotic analysis. See, for example, Section 2.3 in Moon and Perron (2004) and Section 5 in Bai and Ng (2010) for a discussion of this issue.

of Gaussian innovations $\{\eta_{it}\}$. As shown in Moon, Perron, and Phillips (2007), in the presence of deterministic trends the contiguity rate changes and thus an entirely separate analysis is required. Optimality theory extending beyond Gaussian innovations is not available even in a first-generation setting. Thus, we restrict ourselves to demonstrating that our proposed tests remain valid under deviations from Gaussianity.

This chapter offers four contributions. Firstly, we show that in case the nuisance parameters are known the MP experiment is *Locally Asymptotically Normal* (LAN) when $n, T \to \infty$ (jointly). This means that the limit experiment, in the Le Cam sense, is a simple Gaussian shift experiment; see, for example, Van der Vaart (2000). We further establish that the PANIC experiment for the idiosyncratic parts, in case of known nuisance parameters, is also LAN with *the same central sequence and Fisher information* as for the MP experiment.

Secondly, the LAN results imply that *for any test satisfying a mild regularity condition, it suffices to determine its asymptotic size and local power in one of the frameworks, since the same results automatically hold for the other one.* This appears to be a surprising result as the two frameworks, as well as tests and power analyses, have been considered to be completely different. To our best knowledge, the equivalence has not even been observed for the well-studied tests proposed in Moon and Perron (2004) and Bai and Ng (2010).

Thirdly, we derive the *local asymptotic power envelope.* The LAN results, which are based on known nuisance parameters, directly yield an upper bound, which is the same for PANIC and MP, to the local asymptotic power of unit root tests. We demonstrate that we can attain this upper bound also for the case the $O(n)$ nuisance parameters are unknown. In other words, we establish *adaptivity*: the obtained upper bound yields the local asymptotic power envelope. This result extends the work by Moon, Perron, and Phillips (2007), Becheri, Drost, and Van den Akker (2015a), Moon, Perron, and Phillips (2014), and Juodis and Westerlund (2018) on first generation frameworks, to the second generation. It turns out that the level of the local asymptotic power envelope only depends on the (local) deviation to the unit root. Thus, contrary to the asymptotic powers of existing tests, the level of the power envelope is not affected by the nuisance parameters. The power loss attributed by Moon and Perron (2004) and Westerlund (2015) to the heteroskedasticity in η_{it} is thus a feature of the test statistics under consideration, rather than of the MP and PANIC models.

Fourthly, we show that *the popular Moon and Perron (2004) and Bai and Ng (2010) tests are optimal only in case there is no heterogeneity in the long-run variances of the idiosyncratic components η_{it}. We propose a new test that is asymptotically uniformly most powerful* (irrespective of the presence of heterogeneity). Westerlund (2015) derived, via "triangular array asymptotics", the local asymptotic power function, only for the PANIC, framework, of the tests proposed in Bai and Ng (2010). Using our LAN results and Le Cam's third lemma, we provide a new and shorter derivation of these results and also derive the local asymptotic power functions of the tests proposed in Moon and Perron (2004). On comparing these power functions to the power envelope, it is seen that these tests are optimal only in case there is no heterogeneity in the long-run variances of the idiosyncratic components. Our new asymptotically UMP test is motivated by our LAN results. We report numerical asymptotic powers for commonly encountered amounts of heterogeneity and use Monte Carlo experiments to show that the new test also compares favorably in finite samples.

The chapter is organized as follows. Section 2.1 presents and discusses the precise assumptions we impose. Section 2.2 derives the common approximation to the local likelihood ratios in the two experiments and derives its limiting distribution. Section 2.3 introduces our new UMP test based on the limit experiment. Section 2.4 computes the local asymptotic power functions of the tests proposed in Moon and Perron (2004) and Bai and Ng (2010) and Section 2.5 compares their asymptotic and finite-sample power to those of the new UMP test. Section 2.6 concludes. All proofs are organized in several appendices.

2.1 Notation and Assumptions

2.1.1 Matrix notation

Before we introduce our assumptions, we introduce some notation in order to write the model in matrix form. We write I_n and I_T for identity matrices of dimension n and T, respectively, while ι denotes a T-vector of ones. Introduce the n-vectors $\lambda_k = (\lambda_{k1}, \ldots, \lambda_{kn})'$, $k = 1, \ldots, K$ and the $n \times K$ matrix $\Lambda = (\lambda_1, \ldots, \lambda_K)$. Collect the observations as $Z = (Z_{11}, Z_{12}, \ldots, Z_{1T}, \ldots, Z_{n1}, \ldots, Z_{nT})'$. We also write $Z_{-1} = (Z_{10}, Z_{11}, \ldots, Z_{1,T-1}, \ldots, Z_{n0}, \ldots, Z_{n,T-1})'$, $\Delta Z = Z - Z_{-1}$, and define ε, η, E, E_{-1}, ΔE, Y, Y_{-1}, and ΔY analogously. Write $m = (m_1, \ldots, m_n)'$, $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$, $i = 1, \ldots, n$, $f_k = (f_{k1}, \ldots, f_{kT})'$, $k = 1, \ldots, K$, and denote their corresponding covariance matrices by $\Sigma_{f,k} = \text{var}\, f_k \in \mathbb{R}^{T \times T}$ and

$$\Sigma_\eta = \text{diag}(\Sigma_{\eta,1}, \ldots, \Sigma_{\eta,n}), \text{ with } \Sigma_{\eta,i} = \text{var}\, \eta_i \in \mathbb{R}^{T \times T}.$$

The long-run variances of $\{f_{kt}\}$ and $\{\eta_{it}\}$, see Remark 2.1.3 below, are denoted by $\omega^2_{f,k}$ and $\omega^2_{\eta,i}$, respectively. In addition, we define the approximate long-run variances $\omega^2_{f,k,T} = \iota' \Sigma_{fk} \iota / T$ and $\omega^2_{\eta,i,T} = \iota' \Sigma_{\eta,i} \iota / T$. For a given T, these ignore the contribution of any autocovariances further than T apart. We will use the approximate long-run variances to simplify notation and the structure of our proofs. We add the subscript T to the approximate versions to emphasize the difference and define

$$\Omega_\eta = \text{diag}(\omega^2_{\eta,1,T}, \ldots, \omega^2_{\eta,n,T}) \text{ and } \Omega_F = \text{diag}(\omega^2_{f,1,T}, \ldots, \omega^2_{f,K,T}).$$

In addition to this 'vectorized' notation, it will also be useful to consider the observations as $T \times n$ matrices. Thus, let $\tilde{\eta} = (\eta_1, \ldots, \eta_n)$, and define $\tilde{\varepsilon}$, $\tilde{Y}$, $\tilde{Z}$, $\tilde{E}$, $\tilde{f} = (f_1, \ldots, f_K)$, and $\tilde{F}$ analogously. With this notation, (2.5) can be rewritten as

$$\tilde{\varepsilon} = \tilde{f} \Lambda' + \tilde{\eta}, \tag{2.6}$$

while for the vectorized versions we have

$$\varepsilon = \sum_{k=1}^{K} \lambda_k \otimes f_k + \eta.$$

Finally, we introduce the $T \times T$ matrix A by $A_{st} := 1$ if $s > t$ and 0 otherwise and we put $\mathcal{A} := I_n \otimes A \in \mathbb{R}^{nT \times nT}$, i.e.

$$A = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \dots & 1 & 0 \end{pmatrix} \text{ and } \mathcal{A} = \begin{pmatrix} A & 0_{T\times T} & \dots & 0_{T\times T} \\ 0_{T\times T} & A & \dots & 0_{T\times T} \\ \vdots & \ddots & \ddots & 0_{T\times T} \\ 0_{T\times T} & \dots & 0_{T\times T} & A \end{pmatrix}.$$

The matrix A can be considered a cumulative sum operator and premultiplying the vectorized panel with $\mathcal{A}$ takes the cumulative sum in the time direction for each panel unit. It is also related to 'approximate one-sided long-run variances', which we can define by $\delta_{\eta,i,T} = \mathrm{tr}[A\Sigma_{\eta,i}/T]$ and $\delta_{f,k,T} = \mathrm{tr}[A\Sigma_{f,k}/T]$. Note $A + A' = \iota\iota' - I_T$, so that, analogous to the long-run variances, we have $2\delta_{\eta,i,T} = \omega^2_{\eta,i,T} - \gamma_{\eta,i}(0)$.

2.1.2 Assumptions

Now we can formally state the full specifications of our DGPs in (2.1)–(2.4). The distributional assumptions on the time series of the factors $\{f_{kt}\}$ and idiosyncratic shocks $\{\eta_{it}\}$ are given in Assumption 2.1 and we formulate the assumptions on the (deterministic) factor loadings λ_{ki} in Assumption 2.2. Assumption 2.3 states the assumption on the initial values E_{i0} and F_{k0}. Assumption 2.4 specifies the joint asymptotics we consider in this chapter. Finally, Assumption 2.5 differentiates between the two setups discussed in Section 2.1.

Assumption 2.1

(a) Each factor innovation, indexed $k = 1, \dots, K$, is a zero-mean ergodic stationary time series $\{f_{kt}\}$ independent of the other factors and all idiosyncratic parts η_{it}. Its autocovariance function $\gamma_{f,k}$ satisfies

$$\sum_{m=-\infty}^{\infty} (|m| + 1)|\gamma_{f,k}(m)| < \infty$$

and is such that the variance of each factor innovation $\{f_{kt}\}$ is positive.

(b) *For each panel unit $i \in \mathbb{N}$, the idiosyncratic part $\{\eta_{it}\}$ is a Gaussian zero-mean stationary time series independent of the other idiosyncratic parts and all factors. The autocovariance function $\gamma_{\eta,i}$ satisfies*

$$\sup_{i\in\mathbb{N}} \sum_{m=-\infty}^{\infty} (|m|+1)|\gamma_{\eta,i}(m)| < \infty \tag{2.7}$$

and is such that the eigenvalues of the $T \times T$ covariance matrices are uniformly bounded away from zero, i.e., $\inf_{i,T} \lambda_{min}(\Sigma_{\eta,i}) > 0$.

Remark 2.1.1 *The Gaussianity of η_{it} facilitates a relatively easy proof of the LAN-results and it seems to be very difficult to generalize this assumption; even for first-generation frameworks no results on limit experiments and power envelopes are available yet for the non-Gaussian case. For the proposed asymptotically uniformly most powerful test, we stress that it is also valid (i.e., has correct asymptotic size, under suitable moment-conditions) in non-Gaussian settings.*

Remark 2.1.2 *The imposed restrictions on serial correlation are sometimes phrased in terms of spectral densities. Note that our assumption on the boundedness of the eigenvalues is implied by the spectral density being uniformly bounded away from zero (see, for example, Proposition 4.5.3 in Brockwell and Davis (1991)). Similarly, they are sometimes phrased in terms of linear processes on which analogous assumptions are imposed; see, for example, Assumption C in Bai and Ng (2004) and Assumption 2 in Moon and Perron (2004). Finally, note that a collection of causal ARMA processes satisfies Assumption 2.1 if the roots are uniformly bounded away from the unit-circle.*

Remark 2.1.3 *Note that, under Assumption 2.1, the long-run variances of the $\{\eta_{it}\}$, $\omega^2_{\eta,i}$, are also uniformly bounded and uniformly bounded away from zero.*[4] *Moreover, the one-sided long-run variances*

$$\delta_{\eta,i} = \sum_{m=1}^{\infty} \gamma_{\eta,i}(m) = \frac{1}{2}\left(\omega^2_{\eta,i} - \gamma_{\eta,i}(0)\right), \quad i \in \mathbb{N},$$

[4] The former directly follows from (2.7) whereas the latter follows from $\omega^2_{\eta,i} = \lim_{T\to\infty} \frac{1}{T}\iota'\Sigma_{\eta,i}\iota \geq \lim_{T\to\infty} \frac{1}{T}\lambda_{\min}(\Sigma_{\eta,i})\,\iota'\iota \geq \inf_{i,T} \lambda_{\min}(\Sigma_{\eta,i}) > 0$.

are also well-defined.

As already announced, we also need to impose some stability on the factor loadings λ_{ki}, which we assume to be fixed. Assumption 2.2 is standard in the literature, c.f. Assumption A in Bai and Ng (2004) or Assumption 6 in Moon and Perron (2004). It is commonly referred to as the factors being 'strong'.

Assumption 2.2 *There exists a positive definite $K \times K$ matrix Ψ_Λ such that $\lim_{n\to\infty} \frac{1}{n}\Lambda'\Lambda = \Psi_\Lambda$. Moreover, $\max_{k=1,\dots,K} \sup_{i\in\mathbb{N}} |\lambda_{ki}| < \infty$.*

For univariate time series it is known (see, for example, Müller and Elliott (2003)) that the initial value can have a non-negligible impact on the asymptotic behavior of unit root tests. Our assumption on the initial values is as follows.

Assumption 2.3 *We assume zero starting values: $E_{i0} = 0$ and $F_{k0} = 0$.*

We refer to Section 6.2 in Moon, Perron, and Phillips (2007) for a discussion on why relaxing initial conditions can be problematic in a panel context and do not pursue this issue further, except by noting that our tests are invariant with respect to the m_i.

Assumption 2.4 below specifies the asymptotic framework we consider throughout this chapter. We follow Moon and Perron (2004), Bai and Ng (2010), and Westerlund (2015) in considering large 'macro panels', where both n and T go to infinity, but T will be the larger dimension. We derive all our results using joint asymptotics, which yields more robust results than taking sequential limits where first $T \to \infty$ and subsequently $n \to \infty$.

Assumption 2.4 *We consider joint asymptotics, in the Phillips and Moon (1999) sense, with $n/T \to 0$.*

Finally, Assumption 2.5 below specifies that we either operate in the PANIC (case (a)) or in the MP (case (b)) framework. In the PANIC framework, we allow the long-run variance of the factor innovations to be zero, so that we consider both integrated and stationary factors. This is ruled out in the MP case, in which the factors have the same order of integration as the idiosyncratic parts.

Assumption 2.5 *One of the below holds:*

(a) For each factor $F_k, k = 1, \ldots, K$, we have $\rho_k = 1$, or,

(b) For each factor $k = 1, \ldots, K$, we have $\rho_k = \rho$. Moreover, $\{f_{kt}\}$ is Gaussian *and its long-run variance exists and is positive.*

2.2 Limit Experiment and Power Envelope

We phrase our hypotheses about ρ in (2.1)–(2.4) using the local parameterization

$$\rho = \rho^{(n,T)} = 1 + \frac{h}{\sqrt{n}T}. \tag{2.8}$$

As shown below, these rates indeed lead to contiguous alternatives, which allow us to obtain the (local) power of our tests.[5] The unit root hypothesis can be reformulated in terms of the “local parameter” h:

$$\mathrm{H}_0 : h = 0 \text{ versus } \mathrm{H}_a : h < 0.$$

Remark 2.2.1 *We do not allow for ‘heterogeneous alternatives’, i.e. we impose that ρ does not differ across panel units. This helps to unify the treatment of the two setups. Indeed, a more general MP framework, $Y_{it} = \rho_i Y_{i,t-1} + \varepsilon_{it}$, can no longer be rewritten in the PANIC form of* (2.1)*–*(2.4)*. Becheri, Drost, and Van den Akker (2015a) prove that, for the case without factors, unobserved heterogeneity in the autoregressive parameters has no impact on the power envelope or optimal tests. Therefore, in Section 2.5 we also investigate the performance of our tests in the presence of heterogeneous alternatives; those results seem to confirm their conclusion that there is no impact on power also for the general factor case.*

[5] Note that the rate depends favourably both on n and T. This can be interpreted as a ‘blessing of dimensionality’ that originally motivated the use of panel unit-root tests.

In this section we show that likelihood ratios related to the unit root hypothesis, for the MP and for the PANIC framework, exhibit the same local asymptotic expansion. For both setups, we consider the likelihood ratio for observing Z_{it} in case ρ is the only unknown parameter. Hence, the number of factors K, the factor loadings λ_{ki}, the autocovariance functions, and the fixed effects m_i are considered as known in this section. We will first show, for each model separately, that its likelihood ratio satisfies an expansion, under the null hypothesis, of the form

$$\log \frac{\mathrm{dP}_{h,n,T}}{\mathrm{dP}_{0,n,T}} = h\Delta_{n,T} - h^2 J/2 + o_p(1)$$

with Fisher-information $J = 1/2$. In Section 2.2.3, we consider the limiting distribution of their common central sequence $\Delta_{n,T}$ and will conclude that both experiments enjoy the LAN-property. This result allows us to treat the two setups jointly and to obtain three main results. Firstly, it yields an upper bound to the local asymptotic powers of tests (that are valid in case the nuisance parameters are unknown). Secondly, in Section 2.3 we propose a new test, valid in case the nuisance parameters are unknown, that attains this upper bound. This demonstrates that our test is locally asymptotically uniformly most powerful (UMP) and that the Gaussian MP and PANIC experiments are *adaptive* with respect to the nuisance parameters. Thirdly, the LAN results allow us to show that any test, satisfying a mild regularity condition, has the same, typically nonoptimal, local asymptotic power function under both data generating processes.

Remark 2.2.2 *For unit root problems in (univariate) time series, limit experiment theory has been exploited by, amongst others, Jansson (2008) and Zhou, Van den Akker, and Werker (2019). That limit experiment is of the Locally Asymptotically Brownian Functional (LABF) type for which asymptotically UMP tests do not exist. Also in our case, the central sequence could be written as an (approximate) stochastic integral. However, we obtain an additional sum across panel units. Combined with a CLT-type argument, but now in the more complicated joint (n, T)-convergence case, this sum is the intuition for the Gaussian limits we obtain in this panel setting.*

2.2.1 Expanding the likelihood in the PANIC setup

For the PANIC case we will assume, in this subsection, that the factors F_{kt} are observed. Just as for the nuisance parameters, we show in Section 2.3 that the resulting likelihood ratio can still be approximated by an observable version (up to a negligible term). This result implies that observing the factors will not lead to an increase in local asymptotic power *for the PANIC framework*. This appears to be a surprising result. Indeed, Moon, Perron, and Phillips (2014) derived the power envelope for a first-generation data generating process that basically corresponds to PANIC with observed factors. Our analysis implies that, for the PANIC framework, the same power envelope applies. We stress that for the MP setting the situation is different: Becheri, Drost, and Van den Akker (2015b) report higher powers in case factors are observed and Juodis and Westerlund (2018) show power gains when covariates correlated to the innovations are observed.

Denote the joint law of F and Z under Assumptions 2.1, 2.2 and 2.4 and Item (a) of Assumption 2.5 by $\mathrm{P}_{h,n,T}^{\mathrm{PANIC}}$. Using $\eta \sim N(0, \Sigma_\eta)$ and $\eta = \Delta E - hE_{-1}/(\sqrt{n}T)$, we obtain the log-likelihood ratio

$$\begin{aligned}\log \frac{\mathrm{dP}_{h,n,T}^{\mathrm{PANIC}}}{\mathrm{dP}_{0,n,T}^{\mathrm{PANIC}}} &= \frac{h}{\sqrt{n}T}\Delta E'\mathcal{A}'\Sigma_\eta^{-1}\Delta E - \frac{h^2}{2nT^2}\Delta E'\mathcal{A}'\Sigma_\eta^{-1}\mathcal{A}\Delta E \\ &=: h\Delta_{n,T}^{\mathrm{PANIC}} - \frac{1}{2}h^2 J_{n,T}^{\mathrm{PANIC}}.\end{aligned}$$

Note, from (2.6), $\Delta\tilde{E} = \Delta\tilde{Y} - \Delta\tilde{F}\Lambda'$, implying ΔE is indeed observable in this PANIC framework (with observed factors as considered here). Moreover, under $\mathrm{P}_{0,n,T}^{\mathrm{PANIC}}$, $\Delta E = \eta$. We now show that we can replace variances by long-run variances, to obtain a more tractable version of the central sequence and empirical Fisher information.

Lemma 2.2.1 *Suppose that Assumptions 2.1, 2.2 and 2.4 and Item (a) of Assumption 2.5 hold. Then we have, under* $\mathrm{P}_{0,n,T}^{PANIC}$, $(\Delta_{n,T}^{PANIC}, J_{n,T}^{PANIC}) = (\Delta_{n,T}, \frac{1}{2}) + o_p(1)$, *where*

$$\Delta_{n,T} = \frac{1}{\sqrt{n}T}\Delta E'\mathcal{A}'\Psi_\eta^{-1}\Delta E - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2}, \ \textit{with } \Psi_\eta^{-1} = \Omega_\eta^{-1} \otimes I_T.$$

Remark 2.2.3 *The simplified central sequence $\Delta_{n,T}$ is the result of substituting Σ_η^{-1} by Ψ_η^{-1}. To obtain the correct centering, a correction term involving the one-sided long-run variance is needed for each panel unit. This is analogous to the univariate case, see Elliott, Rothenberg, and Stock (1996), and arises due to the fact that, contrary to $\Sigma_\eta^{-1/2}\Delta E$, $\Psi_\eta^{-1/2}\Delta E$ exhibits serial correlation.*

2.2.2 Expanding the likelihood in the Moon and Perron (2004) setup

Let us denote the law of Z under Assumptions 2.1, 2.2 and 2.4 and Item (b) of Assumption 2.5 by $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$. Then the log-likelihood ratio of $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$ with respect to $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$ is given by, using $\varepsilon \sim N(0, \Sigma_\varepsilon)$ and $\varepsilon = \Delta Y - hY_{-1}/(\sqrt{n}T)$,

$$\begin{aligned}\log \frac{\mathrm{dP}^{\mathrm{MP}}_{h,n,T}}{\mathrm{dP}^{\mathrm{MP}}_{0,n,T}} &= \frac{h}{\sqrt{n}T}\Delta Y' \mathcal{A}' \Sigma_\varepsilon^{-1} \Delta Y - \frac{h^2}{2nT^2}\Delta Y' \mathcal{A}' \Sigma_\varepsilon^{-1} \mathcal{A} \Delta Y \\ &=: h\Delta^{\mathrm{MP}}_{n,T} - \frac{1}{2}h^2 J^{\mathrm{MP}}_{n,T}.\end{aligned}$$

In this more complicated model, we simplify the central sequence and also the Fisher information in two steps. The first is analogous to the approximation in the PANIC setup, i.e., we replace variances by long-run variances. Note that thanks to our independence assumptions, the $nT \times nT$ covariance matrix of the ε can be written as

$$\Sigma_\varepsilon = \operatorname{var} \varepsilon = \sum_{k=1}^{K} \left(\lambda_k \lambda_k' \otimes \Sigma_{f,k}\right) + \Sigma_\eta. \tag{2.9}$$

Replacing $\Sigma_{f,k}$ by $\omega^2_{f,k,T} I_T$ and $\Sigma_{\eta,i}$ by $\omega^2_{\eta,i,T} I_T$ in (2.9) we obtain the simplified versions of central sequence

$$\tilde{\Delta}^{\mathrm{MP}}_{n,T} := \frac{1}{\sqrt{n}T}\Delta Y' \mathcal{A}' \Psi_\varepsilon^{-1} \Delta Y - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_{\eta,i,T}}{\omega^2_{\eta,i,T}}, \tag{2.10}$$

where the $nT \times nT$ matrix Ψ_ε is defined by

$$\Psi_\varepsilon := \psi_\varepsilon \otimes I_T := \left(\Lambda \Omega_F \Lambda' + \Omega_\eta\right) \otimes I_T, \tag{2.11}$$

with $\Omega_\eta = \text{diag}(\omega^2_{\eta,1,T}, \dots, \omega^2_{\eta,n,T})$ and $\Omega_F = \text{diag}(\omega^2_{f,1,T}, \dots, \omega^2_{f,K,T})$. The following lemma demonstrates that applying these replacements to the central sequence and Fisher information do not affect their asymptotic behavior.

Lemma 2.2.2 *Suppose that Assumptions 2.1, 2.2 and 2.4 and Item (b) of Assumption 2.5 hold. Then we have, under* $\mathrm{P}^{MP}_{0,n,T}$, $(\Delta^{MP}_{n,T}, J^{MP}_{n,T}) = (\tilde{\Delta}^{MP}_{n,T}, \frac{1}{2}) + o_p(1)$.

Remark 2.2.4 *In the MP case, the covariance matrix that is approximated by long-run variances is not block diagonal. Therefore, contrary to Lemma 2.2.1, the proof of Lemma 2.2.2 exploits the Assumption that* $n/T \to 0$.

Exploiting the Sherman-Morrison-Woodbury formula we obtain

$$\Psi_\varepsilon^{-1} = \psi_\varepsilon^{-1} \otimes I_T = \left(\Omega_\eta^{-1} - \Omega_\eta^{-1}\Lambda\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Lambda'\Omega_\eta^{-1}\right) \otimes I_T. \quad (2.12)$$

Note that removing Ω_F^{-1} from (2.12) yields a projection matrix corresponding to 'projecting out the factors'. Thus, basing a central sequence on such a projection matrix would simplify approximating it based on observables by removing the need to estimate Ω_F^{-1} and, more importantly, by ensuring that the factors are projected out. The next lemma shows that using such a projection version ψ_ε^{*-1} of ψ_ε^{-1} in the central sequence does not change its asymptotic behaviour.

Lemma 2.2.3 *Suppose that Assumptions 2.1, 2.2 and 2.4 and Item (b) of Assumption 2.5 hold. Then we have, under* $\mathrm{P}^{MP}_{0,n,T}$, $\tilde{\Delta}^{MP}_{n,T} = \Delta^*_{n,T} + o_p(1)$, *where*

$$\Delta^*_{n,T} = \frac{1}{\sqrt{nT}}\Delta Y'\mathcal{A}'(\psi_\varepsilon^{*-1} \otimes I_T)\Delta Y - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega^2_{\eta,i,T}}, \text{ with} \quad (2.13)$$

$$\psi_\varepsilon^{*-1} = \Omega_\eta^{-1} - \Omega_\eta^{-1}\Lambda\left(\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Lambda'\Omega_\eta^{-1}. \quad (2.14)$$

2.2.3 Asymptotic normality

Having simplified each framework's central sequence and Fisher information separately, we are now ready to show that they are asymptotically equivalent and the central sequences converge to a normal distribution. We begin this section by showing that the central sequence in the MP framework is asymptotically equivalent to the one in the PANIC framework.

Lemma 2.2.4 *Suppose that Assumptions 2.1, 2.2, 2.4 and 2.5 hold. Then we have, under* $\mathrm{P}_{0,n,T}^{PANIC}$ *and* $\mathrm{P}_{0,n,T}^{MP}$*,* $\Delta_{n,T}^{*} = \Delta_{n,T} + o_p(1)$.

Finally, we consider the weak limit of the central sequence $\Delta_{n,T}$ (and therefore also of $\Delta_{n,T}^{*}$), showing that both experiments are locally asymptotically normal.

Proposition 2.2.1 *Suppose that Assumptions 2.1, 2.2, 2.4 and 2.5 hold. Then we have, under* $\mathrm{P}_{0,n,T}^{PANIC}$ *and* $\mathrm{P}_{0,n,T}^{MP}$*,* $\Delta_{n,T} \xrightarrow{d} N(0, J)$ *with* $J = \frac{1}{2}$.

Remark 2.2.5 *Under the null hypothesis, the model equations of both models coincide. Hence, the additional distributional Item (b) of Assumption 2.5 implies that under the null, the MP framework is a special case of the PANIC framework. Therefore, it is sufficient to show the desired convergence for* $\mathrm{P}_{0,n,T}^{PANIC}$. *This principle applies to all calculations under the hypothesis. As the central sequences are equal as well and thanks to the LAN result below, it even extends to many calculations under alternatives, through Le Cam's Third Lemma.*

Proposition 2.2.1 is an important result as it establishes that the unit root testing problem in both models is locally asymptotically normal, i.e., it is asymptotically equivalent to testing $h = 0$ against $h < 0$ based on one observation $X \sim N(Jh, J)$. This equivalence prescribes how to perform asymptotically optimal inference and yields the asymptotic local power envelope and the power functions of various test statistics: The asymptotic representation theorem (see, for example, Chapter 9 in Van der Vaart (2000)) implies that in our framework no unit root test can have higher power than the optimal test in the limit experiment. This best test is clearly rejecting for small values of X, leading to a power (for a level-α test) of $\Phi(\Phi^{-1}(\alpha) - J^{1/2}h)$. Thus, with $J = 1/2$, this constitutes the power envelope for our unit root testing problems:

Corollary 2.2.1 *Suppose that Assumptions 2.1, 2.2 and 2.4 and Item (a) of Assumption 2.5 hold. Let* $\phi_{n,T} = \phi_{n,T}(Z_{11}, \ldots, Z_{nT})$ *be a sequence of tests and denote their powers, under* $\mathrm{P}^{PANIC}_{h,n,T}$*, by* $\pi_{n,T}(h)$*. If the sequence* $\phi_{n,T}$ *is asymptotically of level* $\alpha \in (0,1)$*, i.e.* $\limsup_{n,T\to\infty} \pi_{n,T}(0) \le \alpha$*, we have, for all* $h \le 0$*,*

$$\limsup_{n,T\to\infty} \pi_{n,T}(h) \le \Phi\left(\Phi^{-1}(\alpha) - \frac{h}{\sqrt{2}}\right). \tag{2.15}$$

Replacing Item (a) of Assumption 2.5 by Item (b) of Assumption 2.5, the same bound applies to powers under $\mathrm{P}^{MP}_{h,n,T}$.

The above power envelope would be reached by any of our previously introduced central sequences.[6] In the next section we show that we can approximate these central sequences based on observables, yielding a feasible test that attains the asymptotic power envelope.

2.3 An Asymptotically UMP Test

In the previous section we derived a testing procedure that reaches the power envelope for the unit root testing problem. This test, however, is not feasible when the nuisance parameters are unknown. In this section, we demonstrate how to estimate the nuisance parameters to obtain a feasible version that also attains the power envelope. We provide a feasible version of $\Delta^*_{n,T}$, which is motivated by the likelihood ratio in the MP experiment. As (2.14) projects out the factors, basing our feasible version on $\Delta^*_{n,T}$ instead of $\Delta_{n,T}$ spares us the approximation of the idiosyncratic parts.

Recalling our LAN results in Section 2.2 and that the central sequences are asymptotically equivalent across the two setups (see Lemma 2.2.4) it is clear that a feasible version of $\Delta^*_{n,T}$ would be optimal. Therefore, we show that replacing all nuisance parameters with estimates does not change the limiting behavior of $\Delta^*_{n,T}$. Specifically, we need estimates $\hat{\Lambda}$ of the factor loadings, as

[6] This always holds in LAN experiments and follows from Le Cam's Third Lemma (see, for example, Chapter 6 in Van der Vaart (2000).

well as estimates $\hat{\delta}_{\eta,i}$ and $\hat{\omega}^2_{\eta,i}$ of the (one-sided) long-run variances of each idiosyncratic part. The feasible test statistic is then

$$\hat{\Delta}_{n,T} = \frac{1}{\sqrt{n}T}\sum_{t=2}^{T}\sum_{s=2}^{t-1}\Delta Z'_{\cdot,s}\hat{\psi}_{\varepsilon}^{-1}\Delta Z_{\cdot,t} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\hat{\delta}_{\eta,i}}{\hat{\omega}^2_{\eta,i}}, \text{ where} \tag{2.16}$$

$$\hat{\psi}_{\varepsilon}^{-1} := \hat{\Omega}_{\eta}^{-1} - \hat{\Omega}_{\eta}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Omega}_{\eta}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Omega}_{\eta}^{-1}. \tag{2.17}$$

Assumption 2.6 *Let $\hat{\delta}_{\eta,i}$, $\hat{\omega}^2_{\eta,i}$ and $\hat{\Lambda}$ be estimators of $\delta_{\eta,i}$, $\omega^2_{\eta,i}$ and Λ satisfying, under $\mathrm{P}^{MP}_{0,n,T}$ and $\mathrm{P}^{PANIC}_{0,n,T}$,*

1. *$\max_{i=1,\dots,n} \mathbb{E}|\hat{\delta}_{\eta,i} - \delta_{\eta,i}|^2 = o(1/n)$,*
2. *$\max_{i=1,\dots,n} \mathbb{E}|\hat{\omega}^2_{\eta,i} - \omega^2_{\eta,i}|^2 = o(1/n)$, and*
3. *for a $K \times K$ matrix H_K satisfying $\|H_K\|_F = O_p(1)$ and $\left\|H_K^{-1}\right\|_F = O_p(1)$, we have $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F = o_p(1)$.*

Under suitable restrictions on the bandwidth and the kernel, conditions Items 1 and 2 hold for kernel spectral density estimates; see Remark 2.9 in Moon, Perron, and Phillips (2014). Item 3, on the other hand, is stronger that the results in Moon and Perron (2004), so we show in Lemma 2.3.1 that it indeed holds under our assumptions.

Lemma 2.3.1 *Let $\bar{\Lambda}$ be $\sqrt{n}$ times the $n \times K$ matrix containing the K orthonormal eigenvectors corresponding to the K largest eigenvalues of $\frac{\Delta\tilde{Z}'\Delta\tilde{Z}}{nT}$. Take $\hat{\Lambda} = \frac{\Delta\tilde{Z}'\Delta\tilde{Z}}{nT}\bar{\Lambda}$. There exists a $K \times K$ matrix H_K such that, under $\mathrm{P}^{MP}_{0,n,T}$ and $\mathrm{P}^{PANIC}_{0,n,T}$, $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F = o_p(1)$ and both $\|H_K\|_F$ and $\left\|H_K^{-1}\right\|_F$ are $O_p(1)$.*

Remark 2.3.1 *These factor estimates are the same as those used in Moon and Perron (2004) and correspond to factor estimates based on classical principal component analysis. We adapt the proof of Moon and Perron (2004), who have demonstrated $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F = O_p(1)$, but we treat one term differently, see Remark 2.A.2.*

Remark 2.3.2 *The factors are only identified up to a 'rotation' H_K. Note that $\Delta^*_{n,T}$ is (indeed) invariant under such rotations, as ψ_{ε}^{*-1} also equals*

$$\Omega_{\eta}^{-1} - \Omega_{\eta}^{-1}\Lambda H_K\left(H_K'\Lambda'\Omega_{\eta}^{-1}\Lambda H_K\right)^{-1}H_K'\Lambda'\Omega_{\eta}^{-1}.$$

Lemma 2.3.2 *Under Assumptions 2.1, 2.2 and 2.4–2.6 we have, under* $\mathrm{P}^{MP}_{0,n,T}$ *and* $\mathrm{P}^{PANIC}_{0,n,T}$, $\hat{\Delta}_{n,T} = \Delta^*_{n,T} + o_p(1)$.

Although Lemma 2.3.2 only concerns adaptivity under the null hypothesis H_0, we can use Le Cam's First Lemma to obtain that, thanks to contiguity, also under $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$ or $\mathrm{P}^{\mathrm{PANIC}}_{h,n,T}$, $\hat{\Delta}_{n,T}$ has the same limiting distribution as $\Delta^*_{n,T}$, so that tests based on $\hat{\Delta}_{n,T}$ will be uniformly most powerful. Formally, the size and power properties of our optimal test follow from the following theorem.

Theorem 2.3.1 *Let* $t_{UMP} = \sqrt{2}\hat{\Delta}_{n,T}$. *Under Assumptions 2.1, 2.2 and 2.4–2.6 we have, under* $\mathrm{P}^{MP}_{h,n,T}$ *and* $\mathrm{P}^{PANIC}_{h,n,T}$,

$$t_{UMP} \xrightarrow{d} N\left(\frac{1}{\sqrt{2}}h, 1\right).$$

Rejecting H_0 *for* $t_{UMP} \leq \Phi^{-1}(\alpha)$, $\alpha \in (0,1)$, *leads to an asymptotic power of* $\Phi\left(\Phi^{-1}(\alpha) - \frac{h}{\sqrt{2}}\right)$, *implying that* t_{UMP} *is asymptotically uniformly most powerful.*

Remark 2.3.3 *The asymptotic size of our test can also be obtained under weaker assumptions not exploiting Gaussianity, see Remarks 2.A.1 and 2.A.3. In such a situation, our test is still valid although perhaps non-optimal. For optimal inference with non-Gaussian innovations a new analysis of the likelihood ratio would be needed, but this is not feasible here.*

Remark 2.3.4 *Note that the limiting distribution of* t_{UMP}, *both under the null hypothesis and under local alternatives, does not depend on the autocorrelations or the heterogeneity of the long-run variances. This shows that the decrease in asymptotic power attributed to these features, for example in Remark 2 of Westerlund (2015) was due to the specific tests under consideration rather than being a feature of the unit root testing problem.*

Remark 2.3.5 *Note that* $\hat{\Delta}_{n,T}$ *only involves differenced data, so that our test is invariant with respect to the incidental intercepts* m_i.

Here is one way to obtain the UMP test in practice:

1. Compute an estimator $\hat{K}$ of the number of common factors on the basis of the observations $\Delta Z_{\cdot t}$, $t = 2, \ldots, T$ using information criteria from Bai and Ng (2002).[7]

2. Use the observations $\Delta Z_{\cdot t}$, $t = 2, \ldots, T$, and $\hat{K}$ to determine the factor loadings $\hat{\Lambda}$ and the factor residuals $\hat{\eta}_{\cdot t}$, $t = 2, \ldots, T$, using principal components.

3. Determine estimates $\hat{\omega}^2_{\eta,i}$ of $\omega^2_{\eta,i}$ and $\hat{\delta}_{\eta,i}$ of $\delta_{\eta,i}$ from $\hat{\eta}_{\cdot t}$, $t = 2, \ldots, T$, using kernel spectral density estimates. Let $\hat{\Omega} = \text{diag}(\hat{\omega}^2_{\eta,1}, \ldots, \hat{\omega}^2_{\eta,n})$.

4. Calculate the estimated central sequence $\hat{\Delta}_{n,T}$ as in (2.16) and reject when $t_{\text{UMP}} = \sqrt{2}\hat{\Delta}_{n,T} \leq \Phi^{-1}(\alpha)$. Alternatively, based on small sample considerations, also estimate the empirical Fisher information

$$\hat{J}_{n,T} := \frac{1}{nT^2} \sum_{t=2}^{T} \sum_{s=2}^{t-1} \Delta Z'_{\cdot,s} \hat{\psi}_{\varepsilon}^{-1} \sum_{u=2}^{t-1} \Delta Z_{\cdot,u},$$

and reject the null hypothesis when $t^{\text{emp}}_{\text{UMP}} := \hat{\Delta}_{n,T} / \sqrt{\hat{J}_{n,T}} \leq \Phi^{-1}(\alpha)$.

Remark 2.3.6 *Although the uniformly most powerful test t_{UMP} does not require a complicated estimate of the known $J = 1/2$, it can be undersized in small samples, whereas the empirical version t^{emp}_{UMP} behaves very well in most DGPs, both in terms of size and power. Thus we recommend to use the t^{emp}_{UMP} in small samples. See Section 2.5 for details.*

2.4 Comparing Powers Across Tests and Frameworks

This section derives the asymptotic powers of commonly used tests in both the Moon and Perron (2004) and the Bai and Ng (2004) frameworks. Recall that the two frameworks are identical under the null hypothesis and write $P_{0,n,T}$ for both $\mathrm{P}^{\text{PANIC}}_{0,n,T}$ and $\mathrm{P}^{\text{MP}}_{0,n,T}$. We start by formalizing our observation that local powers are equal across the two frameworks.

[7] As $(n, T \to \infty)$, these criteria select the correct number of factors with probability one. Therefore, we can treat the number of factors as known in our asymptotic analyses.

Corollary 2.4.1 *Let $t_{n,T}$ be a test statistic that, under $P_{0,n,T}$, converges in distribution jointly with $\Delta_{n,T}$. Then, for all $x \in \mathbb{R}$, and all h,*

$$\lim_{(n,T\to\infty)} \mathrm{P}^{MP}_{h,n,T}[t_{n,T} \leq x] = \lim_{(n,T\to\infty)} \mathrm{P}^{PANIC}_{h,n,T}[t_{n,T} \leq x].$$

If, more specifically, $t_{n,T} \overset{P_{0,n,T}}{\to} N(\mu, \sigma^2)$ and if $t_{n,T}$ and $\Delta_{n,T}$ are jointly asymptotically normal under $P_{0,n,T}$ with asymptotic covariance $\sigma_{\Delta,t}$, its limiting distribution under local alternatives is given by

$$t_{n,T} \overset{\mathrm{P}^{PANIC}_{h,n,T}}{\to} N(\mu + h\sigma_{\Delta,t}, \sigma^2), \text{ and } t_{n,T} \overset{\mathrm{P}^{MP}_{h,n,T}}{\to} N(\mu + h\sigma_{\Delta,t}, \sigma^2).$$

Thus, rejecting for small values of $t_{n,T}$ leads to an asymptotic power for a level-α test of $\Phi(\Phi^{-1}(\alpha) - h\sigma_{\Delta,t}/\sigma)$ in both frameworks.

Once again, our result on the asymptotic equivalence of the two experiments allows us to obtain results for both frameworks at the same time. By demonstrating the joint normality under the null as in Corollary 2.4.1 we obtain simple proofs of the powers of commonly used tests in these frameworks, without ever relying on triangular array calculations. To show the elegance of this approach, we include here the full argument for the first part of this corollary. The second part follows immediately from a more specific version of Le Cam's third lemma, which directly prescribes the desired normal distribution under alternatives. We can use this simple way to obtain powers under local alternatives thanks to our LAN results of Section 2.2.

Denote the weak limit of $(t_{n,T}, \Delta_{n,T})$ under $P_{0,n,T}$ by (t, Δ). Thanks to our results in Section 2.2, both $(t_{n,T}, \frac{\mathrm{dP}^{\mathrm{PANIC}}_{h,n,T}}{\mathrm{dP}^{\mathrm{PANIC}}_{0,n,T}})$ and $(t_{n,T}, \frac{\mathrm{dP}^{\mathrm{MP}}_{h,n,T}}{\mathrm{dP}^{\mathrm{MP}}_{0,n,T}})$ converge in distribution to $(t, \exp(h\Delta - h^2/4))$. By a general form of Le Cam's third lemma, the distribution of $t_{n.T}$ under local alternatives only depends on this joint limiting law and is thus equal across the two frameworks (see Theorem 6.6 in Van der Vaart (2000)).

Remark 2.4.1 *The equality of powers across the two frameworks applies to the practically relevant case of the factors being unobserved. In the PANIC setting, observing the factors does not yield any additional power. This in sharp contrast to other data generating processes, used in the literature on panel unit* *roots, where observing factors or correlated covariates does yield additional power; see, for example, Pesaran, Smith, and Yamagata (2013), Becheri, Drost, and Van den Akker (2015b), and Juodis and Westerlund (2018).*

Before we apply Corollary 2.4.1 to derive asymptotic powers, we first describe the relevant test statistics in some detail. We focus on the tests proposed in Bai and Ng (2010) ('BN tests') and Moon and Perron (2004) ('MP tests'). Following these papers, we denote

$$\omega^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \omega_{\eta,i}^2, \;\; \phi^4 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left(\omega_{\eta,i}^2\right)^2, \;\; \delta = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \delta_{\eta,i}, \quad (2.18)$$

all assumed to be positive, and their estimated counterparts

$$\hat{\omega}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\omega}_{\eta,i}^2, \;\; \hat{\phi}^4 = \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\omega}_{\eta,i}^2\right)^2, \text{ and } \hat{\delta} = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}_{\eta,i}.$$

Finally, we define $\omega^4 = (\omega^2)^2$ and $\hat{\omega}^4 = (\hat{\omega}^2)^2$.

Both the MP and BN tests rely on a two stage procedure. In the first stage, the unobserved idiosyncratic innovations E are estimated. Subsequently, a pooled regression procedure is used to estimate the (pooled) autoregression parameter. This pooled estimator is then used to construct a t-test. The main difference between the MP and the BN procedures lies in the way the idiosyncratic innovations are estimated.

Bai and Ng (2010) propose to estimate the idiosyncratic errors E by the PANIC approach introduced in Bai and Ng (2004), which in turn relies on principal component analysis applied to the differences ΔY_{it}. Denoting this estimator of E_i by $\hat{E}_i$, the BN tests are

$$\begin{aligned}
P_a =& \frac{\sqrt{n}T(\hat{\rho}^+ - 1)}{\sqrt{2\hat{\phi}^4/\hat{\omega}^4}} \text{ and} \\
P_b =& \sqrt{n}T(\hat{\rho}^+ - 1)\sqrt{\frac{1}{nT^2}\sum_{i=1}^{n} \hat{E}_{-1,i}'\hat{E}_{-1,i}\frac{\hat{\omega}^2}{\hat{\phi}^4}}, \text{ where} \\
\hat{\rho}^+ =& \frac{\sum_{i=1}^{n} \hat{E}_{-1,i}'\hat{E}_i - nT\hat{\delta}}{\sum_{i=1}^{n} \hat{E}_{-1,i}'\hat{E}_{-1,i}}
\end{aligned}$$

is a bias-corrected pooled estimator for the autoregressive coefficients.

Remark 2.4.2 *Recall that t_{UMP}^{emp} is a modification of t_{UMP} that replaces the asymptotic Fisher Information $J = 1/2$, with its finite sample equivalent in the MP setup, $\tilde{J}_{n,T}^{MP}$. The resulting statistics can be considered a version of P_b: In the case of homogeneous long-run variances, inserting the true long-run variances into t_{UMP}^{emp} yields P_b. Conversely, t_{UMP}^{emp} is a version of P_b that takes into account the heterogeneity in the long-run variances.*

The MP tests are based on a different estimator of ρ. The idiosyncratic components E_i are estimated by projecting the data on the space orthogonal to the common factors. Let $\hat{\Lambda}$ be a consistent estimators for Λ as defined on p. 89–90 of Moon and Perron (2004), and $Y_{\cdot,t} = (Y_{1t}, \ldots, Y_{nt})'$. Then the MP test statistics are given by

$$\begin{aligned} t_a =& \frac{\sqrt{n}T(\rho_{\text{pool}}^{+} - 1)}{\sqrt{2\hat{\phi}^4/\hat{\omega}^4}}, \text{ and} \\ t_b =& \sqrt{n}T(\rho_{\text{pool}}^{+} - 1)\sqrt{\frac{1}{nT^2}\sum_{t=1}^{T} Y_{\cdot,t-1}' Q_{\hat{\gamma}} Y_{\cdot,t-1} \frac{\hat{\omega}^2}{\hat{\phi}^4}}, \text{ where} \\ \rho_{\text{pool}}^{+} =& \frac{\sum_{t=1}^{T} Y_{\cdot,t}' Q_{\hat{\gamma}} Y_{\cdot,t-1} - nT\hat{\delta}}{\sum_{t=1}^{T} Y_{\cdot,t-1}' Q_{\hat{\gamma}} Y_{\cdot,t-1}}, \text{ and } Q_{\hat{\gamma}} = I - \hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\Lambda}'. \end{aligned}$$

We are now ready to compute the asymptotic behaviour of the MP and BN tests under local alternatives by an application of Corollary 2.4.1. The power of the MP tests in the MP framework has been derived in Moon and Perron (2004) and that of the BN tests in the PANIC framework has been derived in Westerlund (2015). Given our LAN result, we can provide simple independent proofs of these results. These rely on the second part of Corollary 2.4.1; we demonstrate the required joint asymptotic normality in a supplementary appendix. More importantly, our approach also leads to new results, namely the asymptotic powers of the MP test in the PANIC framework and the asymptotic powers of the BN tests in the MP framework. In fact, those results can be considered an immediate consequence of the first part of Corollary 2.4.1 and the existing power results in the literature.

Proposition 2.4.1 *Suppose that Assumptions 2.1, 2.2 and 2.4–2.6 hold. Then, under* $\mathrm{P}_{h,n,T}^{PANIC}$ *or* $\mathrm{P}_{h,n,T}^{MP}$*, as* $(n, T \to \infty)$*, the test statistics* P_a, P_b, t_a*, and* t_b *all converge in distribution to a normal distribution with mean* $h\sqrt{\frac{\omega^4}{2\phi^4}}$ *and variance one. Rejecting for small values of any of these statistics leads to an asymptotic power for a level-*α *test of* $\Phi(\Phi^{-1}(\alpha) - h\sqrt{\frac{\omega^4}{2\phi^4}})$ *in both frameworks.*

Remark 2.4.3 *It turns out that the powers are equal, no matter which test statistic and which framework is considered. We have discussed in some detail that, for a given test, the equality of powers across frameworks is a general phenomenon. The fact that in each framework, the power of the MP tests is equal to that of the BN tests, on the other hand, is a 'coincidence'. Originally, the MP tests have been developed for the MP experiment, whereas the BN tests are designed for the PANIC experiment. It has been noted in Bai and Ng (2010) that the MP tests are valid in term of size in the PANIC setup for testing the idiosyncratic component of the innovation for a unit root but their (local and asymptotic) power in the PANIC framework has not been considered. More discussion on the use of the MP tests in the PANIC setup can be found in Bai and Ng (2010) and Gengenbach, Palm, and Urbain (2010). Similarly, to the best of our knowledge there are no studies on the local asymptotic power of the BN tests in the MP framework.*

The Cauchy-Schwarz inequality implies $\frac{\omega^4}{\phi^4} \leq 1$, thus Proposition 2.4.1 shows that, in general, the local asymptotic power of the MP and BN tests lies below the power envelope. In fact, they are all asymptotically UMP only when $\frac{\omega^4}{\phi^4} = 1$. This condition is satisfied when the long-run variances of the idiosyncratic shocks η_{it} are homogeneous across i. The proposed test t_{UMP} is asymptotically UMP irrespective of possible heterogeneity. In Section 2.5 we assess whether the asymptotic power gains, compared to the MP and BN tests, are also reflected in finite samples for realistic parametric settings.

2.5 Simulation results

This section reports the results of a Monte-Carlo study with three main goals: firstly, to assess the finite sample performance of our proposed test t_{UMP}, secondly, to see how the asymptotic equivalence between the Moon and Perron (2004) and PANIC setups is reflected in finite samples, and, finally, to check the robustness of our results to deviations from our assumptions.

2.5.1 The DGPs

We generate the data from (2.1)–(2.4) with $m_i = 0$.[8] Using sample sizes $n = 25, 50, 100$ and $T = n, 2n, 4n$, we simulate both the MP and the PANIC setups.[9] Recall that, for a local alternative h, we take $\rho = 1 + \frac{h}{\sqrt{nT}}$ in both setups. In the MP case we also set $\rho_k = \rho$, whereas in the PANIC case we set $\rho_k = 1$ under the null and all alternatives. The factor loadings Λ are drawn from a normal distribution with mean $K^{-1/2}$ and covariance matrix $K^{-1}I_K$.[10] Most of the simulations are run with $K = 1$ but we also explore what happens with more factors. Throughout this section we assume the number of factors to be known. For the innovation processes f_{kt} and η_{it} we examine Gaussian i.i.d., MA(1), and AR(1) processes. We fix the MA or AR parameter at 0.4 and set the variance such that the long-run variances of the f_{kt} equal one, and the long-run variance of the η_{it} is ω_i^2. The ω_i^2 are drawn i.i.d. from a lognormal distribution whose parameters are chosen to match different values of ω^4/ϕ^4 and a mean of one.[11]

8 Recall that our tests are invariant with respect to m_i.

9 While it is not clear that the $n/T \to 0$ asymptotics are a good approximation in the $T = n$ case, we consider this case to test the robustness of our results.

10 As done in Moon and Perron (2004), we scale by $\sqrt{K}$ to ensure the contribution of the factors is comparable across specifications.

11 Recall from Section 2.4 that the asymptotic relative efficiency of the existing tests compared to our UMP test depends on the heterogeneity of the long-run variances and more specifically on the ratio ω^4/ϕ^4. Therefore, the sample size at which it becomes worthwhile to estimate the heterogeneous long-run variances (i.e., use the asymptotically UMP tests suggested here) mainly depends on this ratio. We present simulation results for $\sqrt{\omega^4/\phi^4}$ between 0.6 and 1, where lower values indicate more heterogeneity.

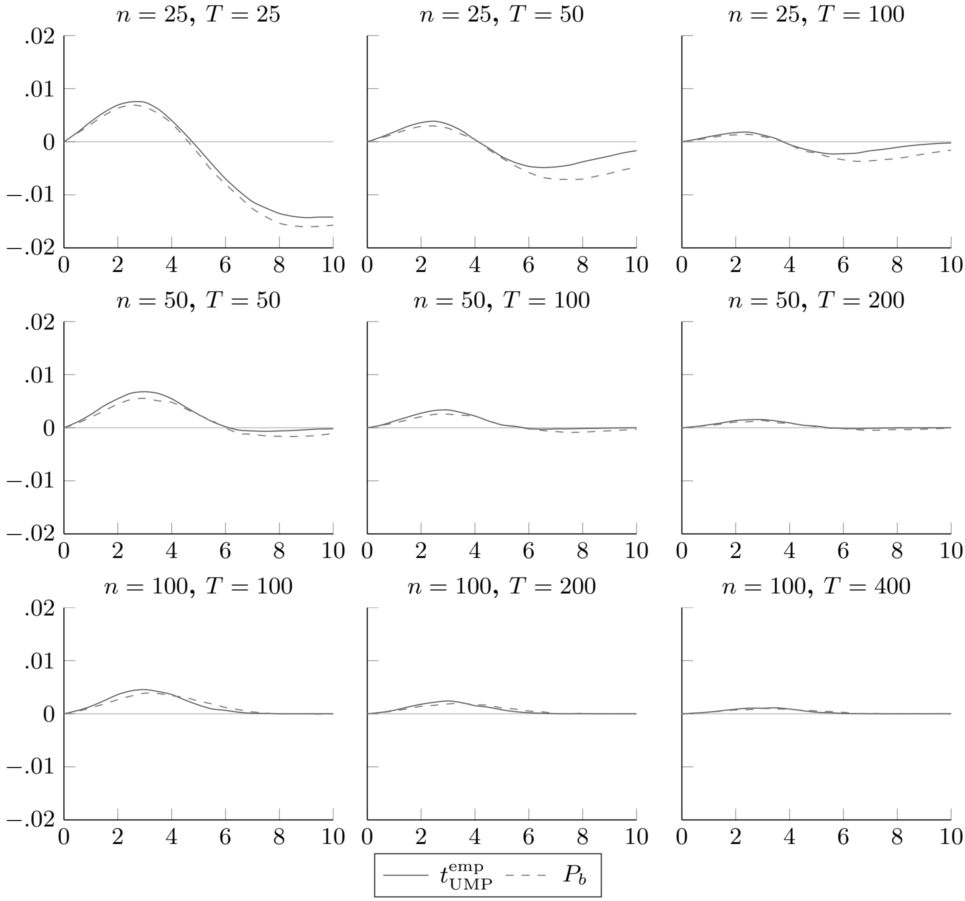


Figure 2.1: Difference between powers in the MP vs the PANIC framework as a function of $-h$ with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 1 000 000 replications.

2.5.2 The test statistics

In addition to the tests proposed in Section 2.3, t_{UMP} and $t_{\text{UMP}}^{\text{emp}}$, we consider the MP tests of Moon and Perron (2004) and the BN tests of Bai and Ng (2010). However, the powers and sizes of the (MP) t_b and (BN) P_b tests were very similar also in finite samples, so we only report results for P_b. We omit

A cursory look at a few typical applications reveals that these ratios are mostly between 0.6 and 0.8 and match the skewed nature of the lognormal distribution.

the comparison with P_a and t_a since they tend to show large biases in terms of size (see, for example, the Monte Carlo studies in Gengenbach, Palm, and Urbain (2010) and Bai and Ng (2010)).

The sizes of all considered tests are highly sensitive to estimation of the (one-sided) long-run variances. We have considered a variety of methods, for example, using a Bartlett or quadratic spectral kernel and selection of the bandwidth according to the Newey and West (1994) or the Andrews (1991) rule with/without various forms of prewhitening. Whereas the differences from using different kernels are small, the selection of both the bandwidth and the prewhitening are essential. Our preferred method employs a Bartlett kernel with prewhitening.[12] There is a size-power tradeoff between using the Andrews (1991) and the Newey and West (1994) bandwidth selection: The Andrews (1991) bandwidth leads to higher powers for the smallest sample sizes, but an oversized test when the innovations have a strong MA component. The decision which bandwidth to use thus depends on the preferences of the researcher. In this section, all results are based on the Andrews (1991) bandwidth. However, the sizes and powers based on the Newey and West (1994) bandwidth can be found in a supplementary appendix.

2.5.3 Sizes

Table 2.1 reports the sizes of our tests for the baseline DGP based on the Andrews bandwidth. Many other specifications can be found in the supplemental appendix. Recall that the sizes depend considerably on how the long-run variances are estimated. Using the method described above, the sizes of $t_{\mathrm{UMP}}^{\mathrm{emp}}$ reasonable across most DGPs and generally comparable to those of P_b. t_{UMP}, on the other hand, is undersized in many specifications, so that we focus on its empirical version $t_{\mathrm{UMP}}^{\mathrm{emp}}$ in the remainder. Only in the MA(1) example, both $t_{\mathrm{UMP}}^{\mathrm{emp}}$ and P_b are oversized ($t_{\mathrm{UMP}}^{\mathrm{emp}}$ is more oversized for the smallest sample sizes and marginally less oversized in the larger ones). Thus, when a strong MA component is suspected, we recommend to use tests based on the Newey

12 As in Moon, Perron, and Phillips (2014), the prewhitening model is selected based on the BIC between four simple ARMA models.

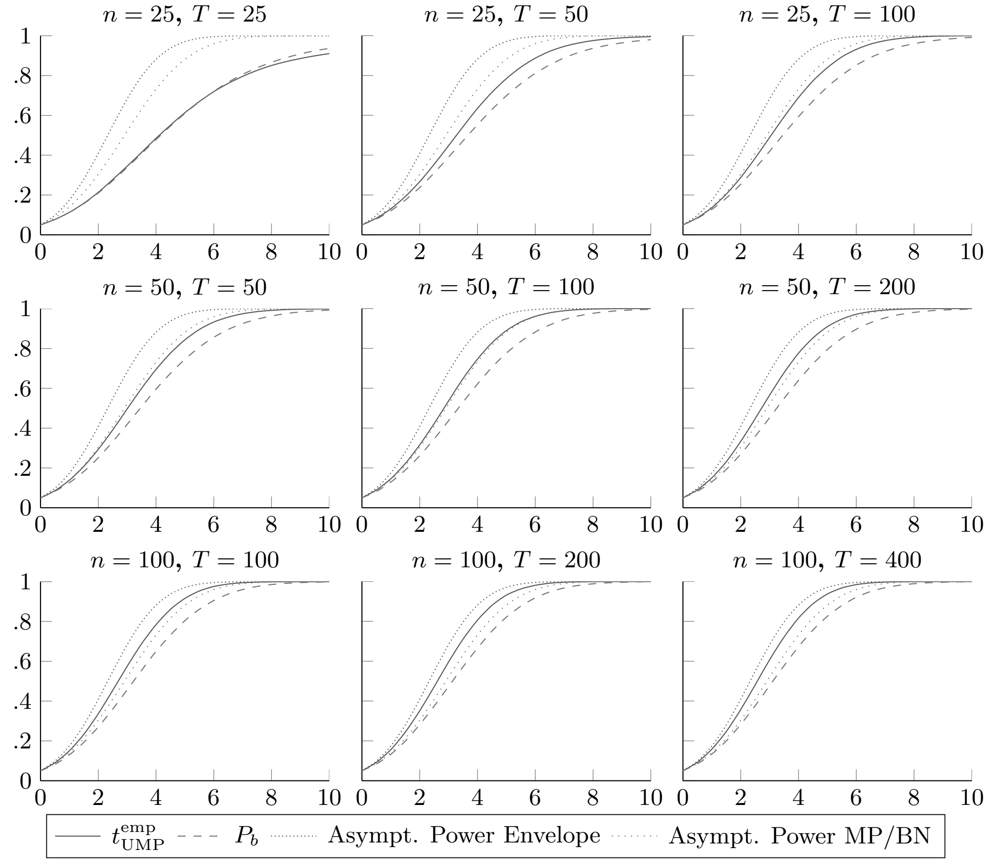


Figure 2.2: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

and West (1994) bandwidth. Generally, the Newey and West (1994) bandwidth provides better sizes, especially in the MA case. However, small sample powers are slightly lower. Both sizes and powers based on the Newey and West (1994) bandwidth can be found in a supplementary appendix.

2.5.4 Powers

We start this subsection by investigating the finite-sample differences between the MP and the PANIC setups. Recall that we have shown that the asymp-

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.6	2.8	3.1	1.8	4.5	4.2	2.2	7.0	5.6
25	50	0.6	1.4	4.7	4.0	1.7	4.9	3.6	3.1	8.9	6.2
25	100	0.6	1.8	5.5	4.6	2.3	6.1	4.1	3.9	10.1	6.7
50	50	0.6	2.0	4.3	3.7	2.5	4.5	3.5	5.3	9.9	6.6
50	100	0.6	2.6	5.1	4.2	2.9	5.2	3.7	6.1	11.0	7.0
50	200	0.6	2.9	5.5	4.6	3.4	5.9	4.1	5.3	9.2	6.1
100	100	0.6	3.2	5.0	4.2	3.3	4.9	3.8	9.1	13.1	8.2
100	200	0.6	3.6	5.3	4.5	3.7	5.3	4.1	7.0	10.0	6.6
100	400	0.6	3.6	5.3	4.5	4.3	6.1	4.5	4.9	7.1	5.1
25	25	0.8	0.9	3.1	3.5	1.8	4.3	4.7	2.4	6.7	6.4
25	50	0.8	1.8	5.1	4.6	1.7	4.4	4.0	3.1	8.3	7.2
25	100	0.8	2.3	5.8	5.2	2.2	5.3	4.6	3.9	9.3	7.8
50	50	0.8	2.4	4.6	4.2	2.4	4.2	4.2	5.1	9.3	8.3
50	100	0.8	3.0	5.4	4.8	2.6	4.6	4.3	5.9	10.1	8.5
50	200	0.8	3.3	5.7	5.2	3.1	5.2	4.7	5.0	8.4	7.1
100	100	0.8	3.5	5.1	4.6	3.1	4.4	4.4	8.7	12.3	10.4
100	200	0.8	3.8	5.5	5.0	3.3	4.7	4.5	6.6	9.2	7.9
100	400	0.8	3.9	5.5	5.1	3.9	5.5	5.0	4.7	6.6	5.9
25	25	1.0	1.0	3.3	3.9	1.9	4.3	5.4	2.4	6.5	7.2
25	50	1.0	2.0	5.2	5.1	1.7	4.2	4.5	3.2	8.1	8.2
25	100	1.0	2.6	6.0	5.8	2.1	5.0	5.1	3.9	8.9	8.8
50	50	1.0	2.5	4.7	4.6	2.4	4.0	5.0	5.2	9.2	10.1
50	100	1.0	3.1	5.4	5.2	2.6	4.4	4.8	5.8	9.8	10.0
50	200	1.0	3.4	5.7	5.6	3.0	5.0	5.1	4.9	8.2	8.1
100	100	1.0	3.6	5.2	4.9	3.0	4.2	4.9	8.6	12.1	12.6
100	200	1.0	3.9	5.5	5.3	3.2	4.6	4.9	6.5	9.0	9.1
100	400	1.0	4.0	5.6	5.5	3.8	5.3	5.5	4.6	6.4	6.4
Mean abs. dev. from 5%			2.3	0.6	0.6	2.3	0.5	0.6	1.4	4.1	2.7

Table 2.1: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Andrews Bandwidth.

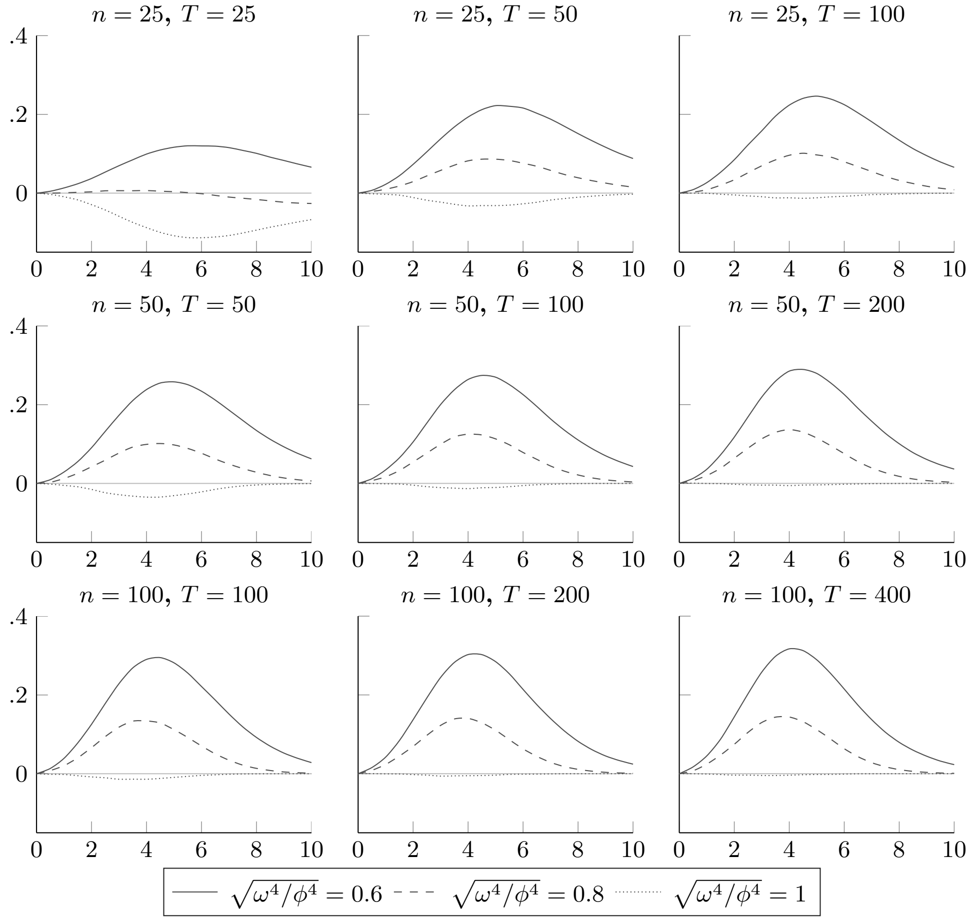


Figure 2.3: (Size-corrected) power gains from using $t_{\text{UMP}}^{\text{emp}}$ over P_b for varying values of $\sqrt{\omega^4/\phi^4}$ and sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts. Based on 100 000 replications.

totic, local power functions are the same and that (under some regularity conditions) all tests have the same asymptotic power in the MP framework as they do in the PANIC framework. Figure 2.1 compares the powers of $t_{\text{UMP}}^{\text{emp}}$ and P_b across the two frameworks.[13] Indeed, also in small samples the powers are very similar. Moreover, both a larger n and a larger T contribute to re-

[13] All figures show size-corrected powers, i.e., powers based on exact simulation-based critical values.

duce the difference. When the factor is stationary under the hypothesis, the difference is considerably smaller still. Noting the small scale on the y axis in these plots, in the remainder we will only present results for the PANIC framework, as the lines would otherwise be mostly indistinguishable.

We now turn to comparing the performance of the UMP tests to existing ones. As discussed in Section 2.3, we need to estimate the individual long-run variance of each idiosyncratic part in order to attain the power envelope. Of course, this becomes easier with a larger time series dimension and is more beneficial when the long-run variances differ substantially between series.

Figure 2.2 presents the baseline power results for a medium amount of heterogeneity ($\sqrt{\omega^4/\phi^4} = 0.8$). It is evident that even for relatively small samples using the optimal test pays off: except for $n = T = 25$, the power of $t_{\mathrm{UMP}}^{\mathrm{emp}}$ is uniformly higher than that of P_b.

Next, Figure 2.3 presents the power difference between the optimal test and P_b for varying degrees of heterogeneity. As expected, the higher the amount of heterogeneity, the more beneficial it is to use the optimal test, also in finite samples. In the case of perfect homogeneity, the losses from estimating individual long-run variances are minor, except for the $n = T = 25$ case.

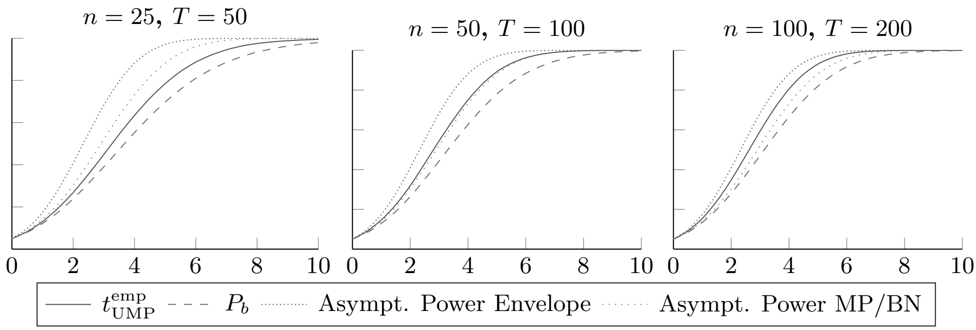


Figure 2.4: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Innovations drawn from a t_5 distribution. Note that the power envelopes refer to the Gaussian experiment. Based on 100 000 replications.

Finally, Figure 2.4 investigates powers when the innovations have a t_5 distribution, rather than the Gaussian one underlying our optimality results. Finite sample powers are not significantly different in this case. In the supplemental appendix, we investigate the effects of serial correlation and multiple factors. Qualitatively, the power results are not affected by these variations in the DGP. We also consider the robustness of our results to further deviations of our assumptions: we consider the power against heterogeneous alternatives and further investigate the effects of non-Gaussian innovations.

2.6 Conclusion and Discussion

This chapter shows that the MP and PANIC frameworks are equivalent, for unit root testing, from a local and asymptotic point of view. Using the underlying LAN-result, the local asymptotic power envelope for the MP and PANIC frameworks readily follows. We show that the tests proposed in Moon and Perron (2004) and Bai and Ng (2010) only attain this bound in case the long-run variances of the idiosyncratic component are homogeneous. We develop an asymptotically uniformly most powerful test; a Monte Carlo study demonstrates that this test also improves on existing tests for finite-samples.

To obtain the local and asymptotic equivalence of the MP and PANIC frameworks, we need to impose some restrictions. First, we assume that the driving innovations are Gaussian. Second, we impose the deviations to the unit root, under the alternative hypothesis, to be the same for all panel units. And third, we do not allow for (incidental) trends. The Gaussianity facilitates a relatively easy proof of the LAN-result and it seems to be rather difficult to generalize this assumption; even for first-generation frameworks no results are available yet. For the proposed asymptotically uniformly most powerful test, we stress that Gaussianity is not required for its validity. In view of Becheri, Drost, and Van den Akker (2015a) we do not expect that imposing constant deviations to the unit root, under the alternative hypothesis, affects our main results. The Monte Carlo results seem to confirm this conjecture for finite-samples. To allow for incidental trends the proper strategy seems to be to first determine the maximal invariant (i.e. determine which part of the observations is invariant with respect to the incidental trends), and to analyze if the resulting maximal invariant satisfies a LAN-result (yielding the power envelope). On basis of Moon, Perron, and Phillips (2007) we expect that the reduction of the data to the maximal invariant will result in a different localizing rate compared to the situation in which there no incidental trends. This indicates that the generalization to incidental trends really requires a separate analysis.

2.A Detailed Proofs

2.A.1 Preliminaries

This section present some preliminary results that are heavily exploited in the proofs of our main results.

First, we recall some elementary results from linear algebra (throughout we only consider real matrices); see, e.g., Lütkepohl (1996) and Magnus and Neudecker (1999). Let $\mathrm{tr}[C]$ denote the trace of a square, real matrix C and let $\lambda_{\min}(C)$ (and $\lambda_{\max}(C)$) denote the minimal (maximal) eigenvalue of a symmetric, real matrix C. For any real matrix C, let $\|C\|_F = \sqrt{\mathrm{tr}[C'C]} = \|C'\|_F$ denote its Frobenius norm, while $\|C\|_{\mathrm{spec}} = \sqrt{\lambda_{\max}(C'C)} = \|C'\|_{\mathrm{spec}}$ denotes its spectral norm. Recall $\|C\|_{\mathrm{spec}} \leq \|C\|_F$.

The inequality $\|CD\|_F \leq \|C\|_{\mathrm{spec}} \|D\|_F$ is immediate from Raleigh's quotient. It follows that the Frobenius is submultiplicative, $\|CD\|_F \leq \|C\|_F \|D\|_F$. Moreover, the identity $\|C \otimes D\|_F = \|C\|_F \|D\|_F$ easily follows from the alternative interpretation of the Frobenius norm being the square-root of the sum of all squared individual matrix entries. Finally, we note that for square matrices $\langle C, D\rangle_F = \mathrm{tr}[C'D]$ defines an inner product, so we have the Cauchy-Schwarz inequality $|\mathrm{tr}[C'D]| \leq \|C\|_F \|D\|_F$.

Next, we present a general lemma on approximating variances with long-run variances. The results we present in this appendix are the main keys to many proofs in Section 2.2. Moreover, they may be of general interest.

Lemma 2.A.1 *Consider an indexed collection of stationary time series $\{X_t^{(h)}\}$, $h \in \mathcal{H}$. Denote the $T \times T$ covariance matrix of $(X_1^{(h)}, \ldots, X_T^{(h)})$ by Σ_h, the m-th autocovariance of $\{X_t^{(h)}\}$ by $\gamma_h(m)$, and its long run variance by $\omega_h^2 < \infty$. Also write $\omega_{h,T}^2 = \iota'\Sigma_h\iota/T$. If $\sup_{h\in\mathcal{H}} \sum_{m=-\infty}^{\infty}(|m|+1)|\gamma_h(m)| < \infty$, then*

1. $\sup_{h\in\mathcal{H}} |\omega_{h,T}^2 - \omega_h^2| = O(T^{-1})$,
2. $\sup_{h\in\mathcal{H}} \left\|A'(\Sigma_h - \omega_h^2 I_T)\right\|_F + \sup_{h\in\mathcal{H}} \left\|A(\Sigma_h - \omega_h^2 I_T)\right\|_F = O(\sqrt{T})$,
3. $\sup_{h\in\mathcal{H}} \left\|A'(\Sigma_h - \omega_{h,T}^2 I_T)\right\|_F + \sup_{h\in\mathcal{H}} \left\|A(\Sigma_h - \omega_{h,T}^2 I_T)\right\|_F = O(\sqrt{T})$,
4. $\sup_{h\in\mathcal{H}} \|A'\Sigma_h\|_F + \sup_{h\in\mathcal{H}} \|A\Sigma_h\|_F = O(T)$.

Proof Item 1 follows from $\omega_{h,T}^2 = \frac{1}{T}\sum_{m<T}(T-|m|)\gamma_h(m)$ and $\omega_h^2 = \sum_{m=-\infty}^{\infty}\gamma_h(m)$, so

$$|\omega_{h,T}^2 - \omega_h^2| = \left|\frac{1}{T}\sum_{m=-\infty}^{\infty}(\min(|m|,T)\gamma_h(m)\right|,$$

which is indeed $O(T^{-1})$ uniformly in h.

For Item 2, tedious but elementary calculations yield

$$\begin{aligned}
\left\|A(\Sigma_h - \omega_h^2 I_T)\right\|_F^2 &= \left\|A'(\Sigma_h - \omega_h^2 I_T)\right\|_F^2 \\
&= \sum_{s=1}^{T}\sum_{t=1}^{T}\left(\sum_{m=s-t+1}^{T-t}\gamma_h(m) - \omega_h^2 1_{s<t}\right)^2 \\
&= \sum_{s=1}^{T-1}\left(\sum_{t=1}^{s}\left(\sum_{m=s+1}^{T}\gamma_h(m-t)\right)^2\right. \\
&\quad \left. + \sum_{t=s+1}^{T}\left(\sum_{m=-\infty}^{s}\gamma_h(m-t) + \sum_{m=T+1}^{\infty}\gamma_h(m-t)\right)^2\right) \\
&= \sum_{s=1}^{T-1}\sum_{t=1}^{T-s}\left(\left(\sum_{m=s}^{T-t}\gamma_h(m)\right)^2 + \left(\sum_{m=s}^{\infty}\gamma_h(m) + \sum_{m=t}^{\infty}\gamma_h(m)\right)^2\right) \\
&\leq 5T\sum_{s=1}^{T}\left(\sum_{m=s}^{\infty}|\gamma_h(m)|\right)^2 \\
&\leq 5T\left(\sum_{m=-\infty}^{\infty}|\gamma_h(m)|\right)\sum_{m=1}^{\infty}\min(m,T)|\gamma_h(m)|.
\end{aligned}$$

Taking suprema, Item 2 follows immediately from this bound. Item 3 follows by combining the first two parts and $\|A\|_F = \sqrt{\frac{T(T-1)}{2}} = O(T)$. The order on $\|A\|_F$ also yields

$$\begin{aligned}
\sup_{h\in\mathcal{H}}\left\|A'\Sigma_h\right\|_F \leq& \sup_{h\in\mathcal{H}}\left\|A'(\Sigma_h - \omega_h^2 I_T)\right\|_F + \sup_{h\in\mathcal{H}}\omega_h^2\left\|A'\right\|_F \\
=& O(\sqrt{T}) + O(1)O(T).
\end{aligned}$$

Again, the second part of Item 4 is analogous. ■

Recall the covariance matrices Σ_η and Σ_ε and their rough approximations Ψ_η and Ψ_ε defined in Lemma 2.2.1 and (2.11), respectively. The following three lemmas use Lemma 2.A.1 to show that these approximations do work well when considering partial sums.

Lemma 2.A.2 *Under Assumption 2.1, $\left\|\Sigma_\eta^{-1}\right\|_{spec}$, $\left\|\Psi_\eta^{-1}\right\|_{spec}$, $\left\|\Sigma_\varepsilon^{-1}\right\|_{spec}$, and $\left\|\Psi_\varepsilon^{-1}\right\|_{spec}$ are all $O(1)$ as $n, T \to \infty$.*

Proof Note that $\Sigma_\varepsilon - \Sigma_\eta$ and $\Psi_\varepsilon - \Psi_\eta$ are positive semidefinite. Hence $\lambda_{\min}(\Sigma_\varepsilon) \geq \lambda_{\min}(\Sigma_\eta) \geq \inf_{i,T}\lambda_{\min}(\Sigma_{\eta,i}) > 0$ and, using Remark 2.1.3 (Remark 2.1.3) and Item 1 of Lemma 2.A.1,

$$\lambda_{\min}(\Psi_\varepsilon) \geq \lambda_{\min}(\Psi_\eta) = \lambda_{\min}(\Omega_\eta \otimes I_T) = \min_{i=1,\ldots,n}\omega_{\eta,i,T}^2$$

$$\geq \inf_{i\in\mathbb{N}} \omega_{\eta,i}^2 - \sup_{i\in\mathbb{N}} |\omega_{\eta,i,T}^2 - \omega_{\eta,i}^2| \to \inf_{i\in\mathbb{N}} \omega_{\eta,i}^2 > 0.$$

This shows the boundedness of all four norms. ∎

Lemma 2.A.3 *Under Assumption 2.1 we have, as $n, T \to \infty$,*

$$\left\|\mathcal{A}'\left(\Sigma_\eta - \Psi_\eta\right)\right\|_F + \left\|\mathcal{A}\left(\Sigma_\eta - \Psi_\eta\right)\right\|_F = O(\sqrt{nT}) = o(\sqrt{n}T).$$

Proof Using block diagonality and Lemma 2.A.1, we obtain the bound

$$\begin{aligned}\left\|\mathcal{A}'\left(\Sigma_\eta - \Psi_\eta\right)\right\|_F^2 &= \sum_{i=1}^n \left\|A'(\Sigma_{\eta,i} - \omega_{\eta,i,T}^2 I_T)\right\|_F^2 \\ &\leq n \sup_{i\in\mathbb{N}} \left\|A'(\Sigma_{\eta,i} - \omega_{\eta,i,T}^2 I_T)\right\|_F^2 = O(nT).\end{aligned}$$

The other part is analogous; every $\mathcal{A}'$ and A' are replaced by $\mathcal{A}$ and A, respectively. ∎

Lemma 2.A.4 *Under Assumptions 2.1, 2.2 and 2.4 we have, as $n, T \to \infty$,*

$$\left\|\mathcal{A}'\left(\Sigma_\varepsilon - \Psi_\varepsilon\right)\right\|_F + \left\|\mathcal{A}\left(\Sigma_\varepsilon - \Psi_\varepsilon\right)\right\|_F = O(n\sqrt{T}) = o(\sqrt{n}T).$$

Proof From the definitions of Σ_ε and Ψ_ε we obtain

$$\mathcal{A}'\left(\Sigma_\varepsilon - \Psi_\varepsilon\right) = \sum_{k=1}^K \mathcal{A}'\left(\lambda_k\lambda_k' \otimes \left(\Sigma_{f,k} - \omega_{f,k,T}^2 I_T\right)\right) + \mathcal{A}'\left(\Sigma_\eta - \Omega_\eta \otimes I_T\right),$$

which yields the bound $\left\|\mathcal{A}'\left(\Sigma_\varepsilon - \Psi_\varepsilon\right)\right\|_F \leq I + II$ with

$$I = \sum_{k=1}^K \left\|\left(\lambda_k\lambda_k' \otimes A'\left(\Sigma_{f,k} - \omega_{f,k,T}^2 I_T\right)\right)\right\|_F \text{ and } II = \left\|\mathcal{A}'\left(\Sigma_\eta - \Omega_\eta \otimes I_T\right)\right\|_F.$$

Part II is already treated in Lemma 2.A.3. For part I, again using Lemma 2.A.1, we get a slightly weaker bound since for the factor part there is no block diagonality:

$$\begin{aligned}I &= \sum_{k=1}^K \left\|\lambda_k\lambda_k'\right\|_F \left\|A'\left(\Sigma_{f,k} - \omega_{f,k,T}^2 I_T\right)\right\|_F \\ &\leq \sum_{k=1}^K \lambda_k'\lambda_k \left\|A'\left(\Sigma_{f,k} - \omega_{f,k,T}^2 I_T\right)\right\|_F = O(n\sqrt{T}) = o(\sqrt{n}T).\end{aligned}$$

The proof for $\left\|\mathcal{A}\left(\Sigma_\varepsilon - \Psi_\varepsilon\right)\right\|_F$ is analogous. ∎

We now present a general weak convergence result for partial sums using joint asymptotics. Proposition 2.2.1 is a special case of Lemma 2.A.5 with $a_{i,n,T} = 1$. We provide Lemma 2.A.5 in general terms here as it might be of independent interest and we also use it in the proof of Proposition 2.4.1 to demonstrate the joint convergence of P_a and the local likelihood ratio.

Lemma 2.A.5 *Let $a_{i,n,T}$ be a bounded sequence of non-random numbers and assume that $\frac{1}{n}\sum_{i=1}^{n} a_{i,n,T}^2 \to \alpha$. Then, under $\mathrm{P}_{0,n,T}^{MP}$ or $\mathrm{P}_{0,n,T}^{PANIC}$, as $(n, T \to \infty)$,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{a_{i,n,T}}{\omega_{\eta,i,T}^2}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\eta_{is}\eta_{it}-\delta_{\eta,i}\right)\xrightarrow{d} N(0,\alpha/2).$$

Proof First consider the case of $a_{i,n,T}$ being identically equal to one and observe that this implies convergence of $\Delta_{n,T}$. Recall $A + A' = \iota\iota' - I_T$ and $2\delta_{\eta,i,T} = \omega_{\eta,i,T}^2 - \gamma_{\eta,i}(0)$, hence, with $\omega_{\eta,i,T}^2 = \frac{1}{T}\iota'\Sigma_{\eta,i}\iota$,

$$\begin{aligned}\Delta_{n,T} =&\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\frac{1}{\omega_{\eta,i,T}^2}\eta_i'\frac{A+A'}{2}\eta_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2}\\ =&\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\left(\left(\frac{\iota'\eta_i}{\sqrt{T}\omega_{\eta,i,T}}\right)^2-1\right)-\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\omega_{\eta,i,T}^2}\left(\frac{1}{T}\eta_i'\eta_i-\gamma_{\eta,i}(0)\right).\end{aligned}$$

Observe that $X_{i,T} := \frac{\iota'\eta_i}{\sqrt{T\omega_{\eta,i,T}^2}} \sim N(0,1)$ and are independent across $i \in \mathbb{N}$. Thus, for each T, $\frac{1}{\sqrt{2n}}\sum_{i=1}^{n}(X_{i,T}^2-1)$ has the same distribution as $\frac{1}{\sqrt{2n}}\sum_{i=1}^{n}(X_i^2-1)$, where $X_i^2 \overset{iid}{\sim} \chi^2(1)$. Therefore, as the latter converges to a standard normal distribution as $n \to \infty$ (CLT), so does the former under joint limits. Thus, the first, leading term converges in distribution to $N(0, 1/2)$.

Asymptotic negligibility of the second, mean-zero term follows from

$$\begin{aligned}\sup_i \operatorname{var}\left(\frac{1}{T}\eta_i'\eta_i\right) =&\frac{2}{T^2}\sup_i \operatorname{tr}[\Sigma_{\eta,i}^2] = \frac{2}{T^2}\sup_i \|\Sigma_{\eta,i}\|_F^2\\ =&\frac{2}{T}\sup_i\left|\sum_{m=-(T-1)}^{T-1}\left(1-\frac{|m|}{T}\right)\gamma_{\eta,i}^2(m)\right| = O(T^{-1}).\end{aligned}$$

For general $a_{i,n,T}$ we can apply a double array CLT, see 1.9.3 in Serfling (1980), to the first (slightly adapted) term in the expansion. The Lindeberg condition is readily verified since we have a weighted sum of i.i.d. centered χ^2 variables. Asymptotic negligibility of the second remainder term follows from the boundedness condition on the $a_{i,n,T}$. ∎

Remark 2.A.1 *We can obtain the same conclusion without requiring Gaussian innovations: as long as the Lindeberg condition holds, for example thanks to higher moment conditions, the same Theorem 1.9.3 of Serfling (1980) applies.*

We conclude this subsection by taking care of important terms that appear repeatedly in the remainder.

Lemma 2.A.6 *Suppose that Assumptions 2.1-2.4 hold. Then, under $\mathrm{P}_{0,n,T}^{MP}$ or $\mathrm{P}_{0,n,T}^{PANIC}$ and as $n, T \to \infty$, we have*

1. $\left\|\left(\frac{1}{n}\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right\|_F = O(1)$,
2. $\left\|\sum_{t=2}^{T}\eta_{\cdot,t}\right\|_F = O_p(\sqrt{nT})$,

3. $\left\|\sum_{t=2}^{T} f_{\cdot,t}\right\|_F = O_p(\sqrt{T})$,
4. $\|\iota'\tilde{\eta}\Omega_\eta^{-1}\Lambda\|_F = O_p(\sqrt{nT})$, *and*
5. $\|\tilde{\eta}\Omega_\eta^{-1}\Lambda\|_F = O_p(\sqrt{nT})$.

Proof For Item 1, recall that K is fixed, so that the norm we consider is irrelevant. As

$$\Lambda'\Omega_\eta^{-1}\Lambda = \sum_{i=1}^{n} \frac{1}{\omega_{\eta,i,T}^2}\lambda_i\lambda_i' \geq \frac{1}{\sup_{i\in\mathbb{N}}\omega_{\eta,i,T}^2}\sum_{i=1}^{n}\lambda_i\lambda_i',$$

the smallest eigenvalue of $\Lambda'\Omega_\eta^{-1}\Lambda$ is larger than that of $\Lambda'\Lambda$. Thus,

$$\left\|\left(\frac{1}{n}\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right\|_{\text{spec}} \leq \sup_{i\in\mathbb{N}}\omega_{\eta,i,T}^2\left\|\left(\frac{1}{n}\Lambda'\Lambda\right)^{-1}\right\|_{\text{spec}} \to \sup_{i\in\mathbb{N}}\omega_i^2\left\|\Psi_\Lambda^{-1}\right\|_{\text{spec}} < \infty,$$

thanks to Assumptions 2.1 and 2.2.

Item 2 follows from

$$\mathbb{E}\left\|\sum_{t=1}^{T}\eta_{\cdot,t}\right\|_F^2 = \mathbb{E}\left\|\tilde{\eta}'\iota\right\|_F^2 = \iota'\mathbb{E}\tilde{\eta}\tilde{\eta}'\iota = \iota'\sum_{i=1}^{n}\mathbb{E}\eta_i\eta_i'\iota = T\sum_{i=1}^{n}\omega_{\eta,i,T}^2 = O(nT).$$

Note that the expectation of $\left\|\sum_{t=2}^{T}\eta_{\cdot,t}\right\|_F^2$ is given by $(T-1)\sum_{i=1}^{n}\omega_{\eta,i,T-1}^2$ and is thus of the same order. Item 3 can be obtained along a similar line of proof.

For Item 4, note $\mathbb{E}\tilde{\eta}'\iota\iota\tilde{\eta} = T\Omega_\eta$, so that

$$\begin{aligned}\mathbb{E}\left\|\iota'\tilde{\eta}\Omega_\eta^{-1}\Lambda\right\|_F^2 &= \operatorname{tr}\mathbb{E}[\tilde{\eta}'\iota\iota\tilde{\eta}]\Omega_\eta^{-1}\Lambda\Lambda'\Omega_\eta^{-1}\\ &= T\operatorname{tr}\Lambda\Lambda'\Omega_\eta^{-1} \leq T\|\Lambda\|_F^2\left\|\Omega_\eta^{-1}\right\|_{\text{spec}} = O(nT).\end{aligned}$$

Item 5 follows similarly from $\mathbb{E}\eta_{\cdot,t}\eta_{\cdot,t}' = \operatorname{diag}(\gamma_{\eta,1}(0),\ldots,\gamma_{\eta,n}(0)) =: D$, so

$$E\left\|\tilde{\eta}\Omega_\eta^{-1}\Lambda\right\|_F^2 = \operatorname{tr}(\Lambda'\Omega_\eta^{-1}\sum_{t=1}^{T}\mathbb{E}[\eta_{\cdot,t}\eta_{\cdot,t}']\Omega_\eta^{-1}\Lambda) \leq T\|\Lambda\|_F^2\left\|\Omega_\eta^{-1}\right\|_{\text{spec}}^2\|D\|_{\text{spec}},$$

which is indeed $O(nT)$ thanks to Assumptions 2.1 and 2.2. ■

2.A.2 Proofs of Section 2.2

Proof of Lemma 2.2.1

Proof In the following all probabilities and expectations are evaluated under $\mathrm{P}_{0,n,T}^{\text{PANIC}}$. To obtain the desired result, we consider the difference between the two central sequences $\Delta_{n,T} - \Delta_{n,T}^{\text{PANIC}}$ and the difference between the two Fisher informations $J_{n,T}^{\text{PANIC}} - \frac{1}{2}$. We show that expectations and variances of both differences converge to zero, implying L_2 convergence.

Part A: Under the null, $\Delta E = \eta$ and hence

$$\Delta_{n,T} - \Delta_{n,T}^{\text{PANIC}} = \frac{1}{\sqrt{nT}}\eta'\mathcal{A}'(\Psi_\eta^{-1} - \Sigma_\eta^{-1})\eta - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2}.$$

We first show that the difference has mean zero. We have, using $\operatorname{tr}(\mathcal{A}) = 0$ and block diagonality of Σ_η,

$$\begin{aligned}
\mathbb{E}[\Delta_{n,T} - \Delta_{n,T}^{\text{PANIC}}] &= \frac{1}{\sqrt{nT}} \operatorname{tr}(\mathcal{A}'(\Psi_\eta^{-1} - \Sigma_\eta^{-1})\Sigma_\eta) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} \\
&= \frac{1}{\sqrt{nT}} \operatorname{tr}(\mathcal{A}'\Psi_\eta^{-1}\Sigma_\eta) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} \\
&= \frac{1}{\sqrt{nT}} \operatorname{tr}((\Omega_\eta{}^{-1} \otimes A')\Sigma_\eta) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} \\
&= \frac{1}{\sqrt{n}} \frac{1}{T} \sum_{i=1}^{n} \frac{1}{\omega_{\eta,i,T}^2} \operatorname{tr}\left[A'\Sigma_{\eta,i}\right] - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} = 0,
\end{aligned}$$

as $\operatorname{tr}\left[A'\Sigma_{\eta,i}\right] = T\delta_{\eta,i,T}$.

To show that the variance of $\Delta_{n,T}^{\text{PANIC}} - \Delta_{n,T}$ goes to zero, observe

$$\begin{aligned}
nT^2 \operatorname{var}(\Delta_{n,T}^{\text{PANIC}} - \Delta_{n,T}) &= \operatorname{var}(\eta' C_\eta \eta) && \text{(2.A.1)} \\
&= \operatorname{tr}[C_\eta\Sigma_\eta C_\eta\Sigma_\eta] + \operatorname{tr}[C_\eta\Sigma_\eta C_\eta'\Sigma_\eta] && \text{(2.A.2)} \\
&\leq \|C_\eta\Sigma_\eta\|_F^2 + \|C_\eta\Sigma_\eta\|_F \|\Sigma_\eta C_\eta\|_F , && \text{(2.A.3)}
\end{aligned}$$

with $C_\eta = \mathcal{A}'(\Psi_\eta^{-1} - \Sigma_\eta^{-1})$. Hence, it suffices to show $\|C_\eta\Sigma_\eta\|_F = o(\sqrt{n}T)$ and $\|\Sigma_\eta C_\eta\|_F = o(\sqrt{n}T)$. Since Ψ_η^{-1} and $\mathcal{A}'$ commute, we obtain

$$\|C_\eta\Sigma_\eta\|_F = \left\|\mathcal{A}'\Psi_\eta^{-1}(\Sigma_\eta - \Psi_\eta)\right\|_F \leq \left\|\Psi_\eta^{-1}\right\|_{\text{spec}} \left\|\mathcal{A}'(\Sigma_\eta - \Psi_\eta)\right\|_F ,$$

which is indeed $o(\sqrt{n}T)$ by Lemmas 2.A.2 and 2.A.3. For $\|\Sigma_\eta C_\eta\|_F$, we first have to approximate $\mathcal{A}\Sigma_\eta$ with $\mathcal{A}\Psi_\eta$ before we can use the commutativity as above:

$$\begin{aligned}
\|\Sigma_\eta C_\eta\|_F &\leq \|\Psi_\eta C_\eta\|_F + \left\|C_\eta'(\Sigma_\eta - \Psi_\eta)\right\|_F \\
&= \left\|\mathcal{A}'(\Sigma_\eta - \Psi_\eta)\Sigma_\eta^{-1}\right\|_F + \left\|\left(\Psi_\eta^{-1} - \Sigma_\eta^{-1}\right)\mathcal{A}(\Sigma_\eta - \Psi_\eta)\right\|_F \\
&\leq \left\|\Sigma_\eta^{-1}\right\|_{\text{spec}} \left\|\mathcal{A}'(\Psi_\eta - \Sigma_\eta)\right\|_F \\
&\quad + \left(\left\|\Psi_\eta^{-1}\right\|_{\text{spec}} + \left\|\Sigma_\eta^{-1}\right\|_{\text{spec}}\right) \|\mathcal{A}(\Sigma_\eta - \Psi_\eta)\|_F = o(\sqrt{n}T).
\end{aligned}$$

Part B: First, we show that the expectation of $J_{n,T}^{\text{PANIC}}$ converges to $\frac{1}{2}$. We have

$$\begin{aligned}
nT^2\mathbb{E}J_{n,T}^{\text{PANIC}} &= \operatorname{tr}\left[\mathcal{A}'\Sigma_\eta^{-1}\mathcal{A}\Sigma_\eta\right] = \operatorname{tr}\left[\mathcal{A}'\Psi_\eta^{-1}\mathcal{A}\Sigma_\eta\right] - \operatorname{tr}\left[\mathcal{A}'C_\eta'\Sigma_\eta\right] \\
&= \operatorname{tr}[\mathcal{A}'\mathcal{A}] + \operatorname{tr}[\mathcal{A}'\Psi_\eta^{-1}\mathcal{A}(\Sigma_\eta - \Psi_\eta)] - \operatorname{tr}[\Sigma_\eta C_\eta\mathcal{A}].
\end{aligned}$$

This implies that the leading term is $\frac{1}{2}nT^2$, since the final two terms are $o(nT^2)$: use the arguments already presented in Part A together with the relation between the trace and the Frobenius norm and

$$\frac{1}{nT^2} \|\mathcal{A}\|_F^2 = \frac{1}{nT^2} \operatorname{tr}\left[\mathcal{A}'\mathcal{A}\right] = \frac{1}{T^2} \operatorname{tr}\left[A'A\right] = \frac{T(T-1)}{2T^2} \to \frac{1}{2}.$$

Next, we show that the variance converges to zero. By the arguments in (2.A.1), with $D_\eta = \mathcal{A}'\Sigma_\eta^{-1}\mathcal{A}$,

$$n^2T^4 \operatorname{var}(J_{n,T}^{\text{PANIC}}) \leq 2\left\|\Sigma_\eta D_\eta\right\|_F^2 .$$

The required order is now easily verified, since

$$\begin{aligned}\left\|\Sigma_\eta D_\eta\right\|_F &\leq \left\|\mathcal{A}'\Psi_\eta^{-1}\mathcal{A}\Sigma_\eta\right\|_F + \left\|\Sigma_\eta C_\eta \mathcal{A}\right\|_F \\ &\leq \left\|\mathcal{A}'\mathcal{A}\right\|_F + \left\|\mathcal{A}'\Psi_\eta^{-1}\mathcal{A}(\Sigma_\eta - \Psi_\eta)\right\|_F + \left\|\Sigma_\eta C_\eta \mathcal{A}\right\|_F\end{aligned}$$

and $\left\|\mathcal{A}'\mathcal{A}\right\|_F = \sqrt{n}\left\|A'A\right\|_F \leq \sqrt{n}\left\|A\right\|_F^2 = \sqrt{n}T(T-1)/2.$ ∎

Proof of Lemma 2.2.2

Proof In the following all probabilities and expectations are evaluated under $\mathrm{P}_{0,n,T}^{\text{MP}}$. The proof of this lemma follows the idea of the proof of Lemma 2.2.1 by considering means and variances. The proof that $J_{n,T}^{\text{MP}}$ converges to $\frac{1}{2}$ in L_2 is almost identical to its counterpart in the proof of Lemma 2.2.1: just replace η by ε, Σ_η by Σ_ε, C_η by C_ε etc. The same replacements yield that the variance of $\tilde{\Delta}_{n,T}^{\text{MP}} - \Delta_{n,T}^{\text{MP}}$ converges to zero, by applying them to the arguments starting at (2.A.1). We are left to show that the expectation of $\tilde{\Delta}_{n,T}^{\text{MP}} - \Delta_{n,T}^{\text{MP}}$ converges to zero. This remaining expectation is more complicated since the variance matrices Σ_ε and Ψ_ε have additional terms due to the presence of unobservable factors.

Recall, under $\mathrm{P}_{0,n,T}^{\text{MP}}$, $\Delta Y = \varepsilon$ and note

$$\tilde{\Delta}_{n,T}^{\text{MP}} - \Delta_{n,T}^{\text{MP}} = \frac{1}{\sqrt{n}}\left(\frac{1}{T}\varepsilon'\mathcal{A}'\left(\Psi_\varepsilon^{-1} - \Sigma_\varepsilon^{-1}\right)\varepsilon - \sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2}\right).$$

Thus, we have

$$\begin{aligned}\mathbb{E}[\tilde{\Delta}_{n,T}^{\text{MP}} - \Delta_{n,T}^{\text{MP}}] &= \frac{1}{\sqrt{n}T}\operatorname{tr}[\mathcal{A}'\Psi_\varepsilon^{-1}\Sigma_\varepsilon] - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} \\ &= \frac{1}{\sqrt{n}T}\operatorname{tr}[\mathcal{A}'\Psi_\eta^{-1}\Sigma_\eta] - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_{\eta,i,T}}{\omega_{\eta,i,T}^2} \\ &+ \frac{1}{\sqrt{n}T}\sum_{k=1}^{K}\operatorname{tr}\left[\psi_\varepsilon^{-1}\lambda_k\lambda_k' \otimes A'\Sigma_{f,k}\right] \\ &+ \frac{1}{\sqrt{n}T}\operatorname{tr}\left[\left((\psi_\varepsilon^{-1} - \Omega_\eta^{-1}) \otimes A'\right)\Sigma_\eta\right] =: I + II + III.\end{aligned}$$

In the proof of Lemma 2.2.1 we have established that the first term equals zero. Therefore, the current proof is complete once we show the final two terms converge to zero.

Convergence to zero of II follows from $\frac{1}{T}\operatorname{tr}(A'\Sigma_{f,k}) = \delta_{f,k,T} = O(1)$ in combination with

$$\sum_{k=1}^{K}\operatorname{tr}\left[\psi_\varepsilon^{-1}\lambda_k\lambda_k'\right] = \operatorname{tr}[\Lambda'\psi_\varepsilon^{-1}\Lambda] = \operatorname{tr}\left[\Omega_F^{-1} - \Omega_F^{-1}\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Omega_F^{-1}\right] \qquad (2.A.4)$$

$$\leq \operatorname{tr}\left[\Omega_F^{-1}\right] = \sum_{k=1}^{K} \frac{1}{\omega_{f,k,T}^2} \to \sum_{k=1}^{K} \frac{1}{\omega_{f,k}^2} < \infty. \tag{2.A.5}$$

Convergence to zero of III follows from

$$|III| \leq \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\Omega_\eta^{-1}\Lambda\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Lambda'\Omega_\eta^{-1}\right)_{i,i} |\operatorname{tr}\left[A'\Sigma_{\eta,i}\right]| \tag{2.A.6}$$

$$\leq \frac{1}{\sqrt{nT}} \operatorname{tr}(\Omega_\eta^{-1}\Lambda\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Lambda'\Omega_\eta^{-1}) \sup_i |\operatorname{tr}\left[A'\Sigma_{\eta,i}\right]|. \tag{2.A.7}$$

$$\leq \frac{1}{\sqrt{nT}} \|\Lambda\|_F^2 \left\|\Omega_\eta^{-1}\right\|_{\text{spec}}^2 \left\|\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right\|_{\text{spec}} \sup_i |\operatorname{tr}\left[A'\Sigma_{\eta,i}\right]|. \tag{2.A.8}$$

Observe $\sup_i \operatorname{tr}[A'\Sigma_{\eta,i}] = O(T)$ by Item 4 of Lemma 2.A.1. From Assumption 2.2 we get $\|\Lambda\|_F = O(\sqrt{n})$ and

$$\begin{aligned} n\left\|\left(\Omega_F^{-1} + \Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right\|_{\text{spec}} &= \left\|\left(\frac{1}{n}\Omega_F^{-1} + \frac{1}{n}\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right\|_{\text{spec}} \\ &= \lambda_{min}^{-1}(\frac{1}{n}\Omega_F^{-1} + \frac{1}{n}\Lambda'\Omega_\eta^{-1}\Lambda) \leq \lambda_{min}^{-1}(\frac{1}{n}\Lambda'\Omega_\eta^{-1}\Lambda) \\ &\leq \lambda_{min}^{-1}(\frac{1}{n}\Lambda'\Lambda) \sup_{i\in\mathbb{N}} \omega_{\eta,i,T}^2 \to \lambda_{min}^{-1}(\Psi_\Lambda) \sup_{i\in\mathbb{N}} \omega_{\eta,i}^2 < \infty. \end{aligned}$$

A combination of these observations with the penultimate display yields $III = o(1)$. ∎

Proof of Lemma 2.2.3

Proof We have

$$\begin{aligned} |\Delta_{n,T}^* - \tilde{\Delta}_{n,T}^{\text{MP}}| &= \frac{1}{\sqrt{nT}} |\operatorname{tr}(A\tilde{\varepsilon}(\psi_\varepsilon^{*\,-1} - \psi_\varepsilon{}^{-1})\tilde{\varepsilon}')| \\ &\leq \frac{1}{\sqrt{nT}} \left\|\psi_\varepsilon^{*\,-1} - \psi_\varepsilon{}^{-1}\right\|_F \left\|\tilde{\varepsilon}' A \tilde{\varepsilon}\right\|_F. \end{aligned}$$

We consider each norm separately. We have

$$\begin{aligned} \left\|\psi_\varepsilon^{*\,-1} - \psi_\varepsilon{}^{-1}\right\|_F &\leq \left\|(\Lambda'\Omega_\eta^{-1}\Lambda + \Omega_F)^{-1} - (\Lambda'\Omega_\eta^{-1}\Lambda)^{-1}\right\|_{\text{spec}} \left\|\Omega_\eta^{-1}\right\|_{\text{spec}}^2 \|\Lambda\|_F^2 \\ &= O(n^{-2})O(1)O(n) = O(n^{-1}), \end{aligned}$$

as $\|\Lambda\|_F = O(\sqrt{n})$ by Assumption 2.2, $\left\|\Omega_\eta^{-1}\right\|_{\text{spec}} = O(1)$ by Assumption 2.1, and

$$\begin{aligned} & n\left\|(\Lambda'\Omega_\eta^{-1}\Lambda + \Omega_F)^{-1} - (\Lambda'\Omega_\eta^{-1}\Lambda)^{-1}\right\|_{\text{spec}} \\ &= \left\|\left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n} + \frac{\Omega_F}{n}\right)^{-1} - \left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n}\right)^{-1}\right\|_{\text{spec}} \\ &= \left\|-\left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n} + \frac{\Omega_F}{n}\right)^{-1} \frac{\Omega_F}{n} \left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n}\right)^{-1}\right\|_{\text{spec}} \\ &\leq \left\|\frac{\Omega_F}{n}\right\|_{\text{spec}} \left\|\left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n} + \frac{\Omega_F}{n}\right)^{-1}\right\|_{\text{spec}} \left\|\left(\frac{\Lambda'\Omega_\eta^{-1}\Lambda}{n}\right)^{-1}\right\|_{\text{spec}}, \end{aligned}$$

which is $O(n^{-1})$: the second norm converges to the third, which is O(1) by Item 1 of Lemma 2.A.6. For $\|\tilde{\varepsilon}' A\tilde{\varepsilon}\|_F$, we note that $\|\tilde{\varepsilon}' A\tilde{\varepsilon}\|_F = \left\|\tilde{\varepsilon}' \frac{A+A'}{2}\tilde{\varepsilon}\right\|_F$ and recall that $A + A' = \iota\iota' - I_T$, so that

$$2\left\|\tilde{\varepsilon}' A\tilde{\varepsilon}\right\|_F = \left\|\tilde{\varepsilon}'(\iota\iota' - I_T)\tilde{\varepsilon}\right\|_F \leq \left\|\iota'\tilde{\varepsilon}\right\|_F^2 + \|\tilde{\varepsilon}\|_F^2 = O_p(nT), \tag{2.A.9}$$

as $\|\tilde{\varepsilon}\|_F \leq \|\Lambda\|_F \left\|\tilde{f}\right\|_F + \|\tilde{\eta}\|_F = O(\sqrt{n})O_p(\sqrt{T}) + O_p(\sqrt{nT})$ and, using Items 2 and 3 of Lemma 2.A.6, a similar bound holds for $\|\iota'\tilde{\varepsilon}\|_F$. Conclude that the central sequence difference is $O_p(n^{-1/2})$. ■

Proof of Lemma 2.2.4

Proof As ψ_ε^{*-1} projects out the factors, we have

$$\begin{aligned}
\Delta^*_{n,T} - \Delta_{n,T} &= \frac{1}{\sqrt{nT}}\operatorname{tr}(A\tilde{\varepsilon}\psi_\varepsilon^{*-1}\tilde{\varepsilon}') - \frac{1}{\sqrt{nT}}\operatorname{tr}(A\tilde{\eta}\Omega_\eta{}^{-1}\tilde{\eta}') \\
&= \frac{1}{\sqrt{nT}}\operatorname{tr}(A\tilde{\eta}(\psi_\varepsilon^{*-1} - \Omega_\eta{}^{-1})\tilde{\eta}').
\end{aligned}$$

Note that for a symmetric matrix B,

$$\operatorname{tr}(A\tilde{\eta}B\tilde{\eta}') = \operatorname{tr}(\tilde{\eta}B\tilde{\eta}'A') = \operatorname{tr}(A'\tilde{\eta}B\tilde{\eta}') = \operatorname{tr}\left(\frac{A+A'}{2}\tilde{\eta}B\tilde{\eta}'\right),$$

so, as ψ_ε^{*-1} and Ω_η are symmetric and $A + A' = \iota\iota' - I_T$, we have

$$\begin{aligned}
|\operatorname{tr}(A\tilde{\eta}(\psi_\varepsilon^{*-1} - \Omega_\eta{}^{-1})\tilde{\eta}')| &= \frac{1}{2}|\operatorname{tr}((\iota\iota' - I_T)\tilde{\eta}(\psi_\varepsilon^{*-1} - \Omega_\eta{}^{-1})\tilde{\eta}')| \\
&\leq |\operatorname{tr}(\iota'\tilde{\eta}(\psi_\varepsilon^{*-1} - \Omega_\eta{}^{-1})\tilde{\eta}'\iota)| + |\operatorname{tr}(\tilde{\eta}(\psi_\varepsilon^{*-1} - \Omega_\eta{}^{-1})\tilde{\eta}')| \\
&\leq \left\|(\Lambda'\Omega_\eta^{-1}\Lambda)^{-1}\right\|_F \left(\|\iota'\tilde{\eta}\Omega_\eta^{-1}\Lambda\|_F^2 + \|\tilde{\eta}\Omega_\eta^{-1}\Lambda\|_F^2\right) \\
&= O(n^{-1})(O_p(nT) + O_p(nT)) = O_p(T),
\end{aligned}$$

using Items 1, 4 and 5 of Lemma 2.A.6. ■

Proof of Proposition 2.2.1

Proof Apply Lemma 2.A.5 with $a_{i,n,T} = 1$ for all i, n, T. ■

2.A.3 Proofs of Section 2.3

Proof of Lemma 2.3.1

Remark 2.A.2 *The proof follows along similar lines as that of Moon and Perron (2004). By treating the norm of $\tilde{\eta}'\tilde{\eta}$ differently, we obtain, under the assumptions of this chapter, $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F = o_p(1)$ instead of the $O_p(1)$ obtained by Moon and Perron (2004). In particular, we exploit $\|\tilde{\eta}'\tilde{\eta}\|_{spec} = o_p(\sqrt{n}T)$, whereas Moon and Perron (2004) only use $\|\tilde{\eta}'\tilde{\eta}\|_F = O_p(\sqrt{n}T)$.*

PROOF As Moon and Perron (2004), we take $H_K = \frac{\tilde{f}'\tilde{f}}{T}\frac{\Lambda'\bar{\Lambda}}{n}$. First note that from the definitions of H_K and $\hat{\Lambda}$ and using $\tilde{\varepsilon} = \tilde{f}\Lambda' + \tilde{\eta}$ we have

$$\hat{\Lambda} - \Lambda H_K = \frac{1}{nT}(\tilde{\varepsilon}'\tilde{\varepsilon} - \Lambda\tilde{f}'\tilde{f}\Lambda')\bar{\Lambda} = \frac{1}{nT}(\tilde{\eta}'\tilde{f}\Lambda' + \Lambda\tilde{f}'\tilde{\eta} + \tilde{\eta}'\tilde{\eta})\bar{\Lambda}, \tag{2.A.10}$$

so that

$$\left\|\Lambda H_K - \hat{\Lambda}\right\|_F \leq \frac{\left\|\tilde{\eta}'\tilde{f}\Lambda'\bar{\Lambda}\right\|_F}{nT} + \frac{\left\|\Lambda\tilde{f}'\tilde{\eta}\bar{\Lambda}\right\|_F}{nT} + \frac{1}{nT}\left\|\tilde{\eta}'\tilde{\eta}\bar{\Lambda}\right\|_F \tag{2.A.11}$$

$$\leq 2\sqrt{\frac{n}{T}}\frac{\left\|\tilde{\eta}'\tilde{f}\right\|_F}{\sqrt{nT}}\frac{\|\Lambda\|_F}{\sqrt{n}}\frac{\left\|\bar{\Lambda}\right\|_F}{\sqrt{n}} + \frac{1}{nT}\left\|\tilde{\eta}'\tilde{\eta}\right\|_{\text{spec}}\left\|\bar{\Lambda}\right\|_F. \tag{2.A.12}$$

By the definition of $\bar{\Lambda}$, $\left\|\bar{\Lambda}\right\|_F = \sqrt{nK} = O(\sqrt{n})$. We have

$$\begin{aligned}
\mathbb{E}\left\|\tilde{\eta}'\tilde{f}\right\|_F^2 =& \mathbb{E}\sum_{k=1}^K\sum_{i=1}^n\left(\sum_{t=1}^T f_{kt}\eta_{it}\right)^2 \\
=& \sum_{k=1}^K\sum_{i=1}^n\sum_{t=1}^T\sum_{s=1}^T \gamma_{\eta,i}(t-s)\gamma_{f,k}(t-s) \\
\leq& Mn\sum_{k=1}^K\sum_{t=1}^T\sum_{s=1}^T |\gamma_{f,k}(t-s)| \\
=& Mn\sum_{k=1}^K\sum_{m=-(T-1)}^{T-1}(T-|m|)|\gamma_{f,k}(m)| = O(nT),
\end{aligned}$$

for some finite constant M, using that, thanks to Assumption 2.1, $\gamma_{\eta,i}(t-s)$ is bounded uniformly in i and $t-s$. Thus, each term of the first summand in (2.A.12) is $O_p(1)$.

Finally, we consider the second summand, which is treated differently from Moon and Perron (2004). We obtain $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F = o_p(1)$ if we can indeed show that $\|\tilde{\eta}'\tilde{\eta}\|_{\text{spec}} = o_p(\sqrt{n}T)$ (Moon and Perron (2004) only use $\|\tilde{\eta}'\tilde{\eta}\|_F = O_p(\sqrt{n}T)$). For this, note that $\frac{1}{T}\tilde{\eta}'\tilde{\eta} = \frac{1}{T}\sum_{t=1}^T \tilde{\eta}_{\cdot,t}\tilde{\eta}_{\cdot,t}'$, which can be considered an approximation to $\Gamma_\eta := \text{diag}(\gamma_{\eta,1}(0),\ldots,\gamma_{\eta,n}(0))$, the $n\times n$ cross-sectional covariance matrix of the η. From Assumption 2.1, $\|\Gamma_\eta\|_{\text{spec}} < \infty$. We now show that indeed the approximation works. Using Isserlis' Theorem to write $\mathbb{E}[\eta_{i,t}^2\eta_{i,s}^2] = 2\gamma_{\eta,i}(t-s)^2 + \mathbb{E}[\eta_{i,t}^2]\mathbb{E}[\eta_{i,s}^2]$, we have

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^T\tilde{\eta}_{\cdot,t}\tilde{\eta}_{\cdot,t}' - \Gamma_\eta\right\|_F^2 =& \sum_{i=1}^n\sum_{j=1}^n\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^T\eta_{i,t}\eta_{j,t} - \mathbb{E}[\eta_{i,t}\eta_{j,t}]\right)^2 \\
=& \sum_{i=1}^n\sum_{j=1}^n\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\mathbb{E}[\eta_{i,t}\eta_{j,t}\eta_{i,s}\eta_{j,s}] - \mathbb{E}[\eta_{i,t}\eta_{j,t}]\mathbb{E}[\eta_{i,s}\eta_{j,s}] \\
=& \sum_{i=1}^n\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T 2\gamma_{\eta,i}(t-s)^2 \\
&+ \sum_{i\neq j}\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T\gamma_{\eta,i}(t-s)\gamma_{\eta,j}(t-s)
\end{aligned}$$

$$= O(n/T) + O(n^2/T).$$

Conclude that the difference in Frobenius norm is $O_p(n/\sqrt{T})$.

Remark 2.A.3 *Note that, even without Gaussianity, this conclusion holds as long as the long-run variances of the $\{\eta_{i,t}^2\}$ are uniformly bounded.*

Thus,

$$\begin{aligned}\|\tilde{\eta}'\tilde{\eta}\|_{\text{spec}} \leq & \left\| \sum_{t=1}^{T} \tilde{\eta}_{\cdot,t}\tilde{\eta}'_{\cdot,t} - T\Gamma_\eta \right\|_F + \|T\Gamma_\eta\|_{\text{spec}} \\ =& O_p(n\sqrt{T}) + O(T) = o_p(\sqrt{n}T).\end{aligned}$$

Finally, we show the boundedness properties of H_K. First note that

$$\|H_K\|_F \leq \frac{\left\|\tilde{f}'\tilde{f}\right\|_F}{T} \frac{\|\Lambda\|_F}{\sqrt{n}} \frac{\|\bar{\Lambda}\|_F}{\sqrt{n}} = O_p(1).$$

To show boundedness of the inverse, we will show that the limiting eigenvalues of H_K are positive. Introduce $\Gamma_f := \operatorname{diag}(\gamma_{f,1}(0), \ldots, \gamma_{f,K}(0))$, the $K \times K$ covariance matrix of the f, and write

$$\left\| H_K - \Gamma_f \frac{\Lambda'\bar{\Lambda}}{n} \right\|_{\text{spec}} \leq \left\| \frac{\Lambda'\bar{\Lambda}}{n} \right\|_F \left\| \frac{\tilde{f}'\tilde{f}}{T} - \Gamma_F \right\|_F = O_p(1)o_p(1),$$

where the latter follows from Assumption 2.1. As Γ_F has full rank, it is sufficient to show that the eigenvalues of $\frac{\Lambda'\bar{\Lambda}}{n}$ are bounded away from zero. $\bar{\Lambda}$ is defined through the eigenvectors of $\tilde{\varepsilon}'\tilde{\varepsilon}/(nT)$. As the eigenvalues of $\tilde{\varepsilon}'\tilde{\varepsilon}$ are closely related to those of $\Lambda\tilde{f}'\tilde{f}\Lambda'$, we can use this relation to learn about the rank of $\Lambda'\bar{\Lambda}$. Formally, define D to be the $K \times K$ matrix with the K largest eigenvalues of $\tilde{\varepsilon}'\tilde{\varepsilon}/(nT)$. Then, from the definition of $\bar{\Lambda}$,

$$D = \frac{\bar{\Lambda}'}{\sqrt{n}} \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{nT} \frac{\bar{\Lambda}}{\sqrt{n}}.$$

Recalling some of the above results we obtain

$$\left\| \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{nT} - \frac{\Lambda\tilde{f}'\tilde{f}\Lambda'}{nT} \right\|_{\text{spec}} = o_p(n^{-1/2}), \tag{2.A.13}$$

so that

$$D = \frac{\bar{\Lambda}'}{\sqrt{n}} \frac{\Lambda\tilde{f}'\tilde{f}\Lambda'}{nT} \frac{\bar{\Lambda}}{\sqrt{n}} + o_p(n^{-1/2}) = \frac{\bar{\Lambda}'\Lambda}{n}\Gamma_f \frac{\Lambda'\bar{\Lambda}}{n} + o_p(1).$$

As the Kth largest eigenvalue of $\tilde{\varepsilon}'\tilde{\varepsilon}/(nT)$ is bounded away from zero (using (2.A.13) the nonzero limiting eigenvalues are given by those of $\Psi_\Lambda\Gamma_F$, a product of two rank K matrices), so must the limit of $\frac{\Lambda'\bar{\Lambda}}{n}$ and thus H_K. ■

Proof of Lemma 2.3.2

PROOF First note that

$$\mathbb{E}\left\|\hat{\Omega}_\eta - \Omega_\eta\right\|_F^2 = \sum_{i=1}^n \mathbb{E}(\hat{\omega}_{\eta,i}^2 - \omega_{\eta,i}^2)^2 \leq n \max_{i=1,\ldots,n} \mathbb{E}(\hat{\omega}_{\eta,i}^2 - \omega_{\eta,i}^2)^2 = o(1),$$

from Assumption 2.6. Thus both $\left\|\hat{\Omega}_\eta - \Omega_\eta\right\|_F$ and $\left\|\hat{\Omega}_\eta - \Omega_\eta\right\|_{\text{spec}}$ are $o_p(1)$. Together with Assumption 2.1 this also implies, with probability converging to one,

$$0 < \frac{\inf_{i\in\mathbb{N}} \omega_{\eta,i}^2}{2} < \min_{i=1,\ldots,n} \hat{\omega}_{\eta,i}^2 \leq \max_{i=1,\ldots,n} \hat{\omega}_{\eta,i}^2 < 2\sup_{i\in\mathbb{N}} \omega_{\eta,i}^2 < \infty. \tag{2.A.14}$$

Therefore, $\left\|\hat{\Omega}_\eta^{-1}\right\|_{\text{spec}} = O_p(1)$, so that finally also $\left\|\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\right\|_F$ and $\left\|\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\right\|_{\text{spec}}$ are $o_p(1)$. Similarly, we note for the one-sided long-run variances that $\sum_{i=1}^n (\hat{\delta}_{\eta,i} - \delta_{\eta,i})^2 = o_p(1)$ follows from Assumption 2.6, so that, along the same lines, we obtain $\max_{i=1,\ldots,n} \hat{\delta}_{\eta,i} = O_p(1)$.

We split the central sequence difference in three parts: one for replacing ψ_ε^* with $\hat{\psi}_\varepsilon$, one to take care of the initial value, and one for estimating the correction term. Thus $\hat{\Delta}_{n,T} - \Delta_{n,T}^* = I - II - III$, with

$$I = \frac{1}{\sqrt{nT}} \operatorname{tr}(A'\tilde{\varepsilon}\left(\hat{\psi}_\varepsilon^{-1} - \psi_\varepsilon^{*-1}\right)\tilde{\varepsilon}')$$

$$II = \frac{1}{\sqrt{nT}} \sum_{t=2}^T \varepsilon_{\cdot,1}' \hat{\psi}_\varepsilon^{-1} \varepsilon_{\cdot,t}$$

$$III = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\hat{\delta}_{\eta,i}}{\hat{\omega}_{\eta,i}^2} - \frac{\delta_{\eta,i}}{\omega_{\eta,i}^2}\right).$$

For part I, insert (2.14) and (2.17) to find

$$\begin{aligned}
|I| &= \frac{1}{\sqrt{nT}} |\operatorname{tr}(\tilde{\varepsilon}' A' \tilde{\varepsilon}(\hat{\psi}_\varepsilon^{-1} - \psi_\varepsilon^{*-1}))| \\
&\leq \frac{1}{\sqrt{nT}} |\operatorname{tr}(\tilde{\varepsilon}' A' \tilde{\varepsilon}(\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1})| \\
&+ \frac{1}{\sqrt{nT}} \left|\operatorname{tr}\left(\hat{\Lambda}'\hat{\Omega}_\eta^{-1}\tilde{\varepsilon}'A'\tilde{\varepsilon}\hat{\Omega}_\eta^{-1}\hat{\Lambda}\left(\hat{\Lambda}'\hat{\Omega}_\eta^{-1}\hat{\Lambda}\right)^{-1} - \Lambda'\Omega_\eta^{-1}\tilde{\varepsilon}'A'\tilde{\varepsilon}\Omega_\eta^{-1}\Lambda\left(\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\right)\right| \\
&\leq \frac{1}{\sqrt{nT}} |\operatorname{tr}(\tilde{\varepsilon}' A' \tilde{\varepsilon}(\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1})| \\
&+ \frac{1}{\sqrt{nT}} \left\|\hat{\Omega}_\eta^{-1}\hat{\Lambda}\left(\hat{\Lambda}'\hat{\Omega}_\eta^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\Lambda\left(\Lambda'\Omega_\eta^{-1}\Lambda\right)^{-1}\Lambda'\Omega_\eta^{-1}\right\|_F \left\|\tilde{\varepsilon}'A'\tilde{\varepsilon}\right\|_F.
\end{aligned}$$

As $\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}$ is diagonal, the first summand is bounded by (using Cauchy-Schwarz)

$$\frac{1}{\sqrt{nT}} \left(\sum_{i=1}^n (\varepsilon_i' A \varepsilon_i)^2\right)^{1/2} \left\|\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\right\|_F = \frac{1}{\sqrt{nT}} O_p(\sqrt{n}T) o_p(1) = o_p(1).$$

For II, we have

$$\begin{aligned}\sqrt{n}TII &\leq \left\|\hat{\psi}_{\varepsilon}^{-1}\right\|_{\text{spec}} \|\varepsilon_{\cdot,1}\|_F \left(\left\|\iota'\tilde{\varepsilon}\right\|_F + \|\varepsilon_{\cdot,1}\|_F\right)\\ &= O_p(1)O_p(\sqrt{n})(O_p(\sqrt{nT}) + O_p(\sqrt{n}) = O_p(n\sqrt{T}),\end{aligned}$$

where $\|\varepsilon_{\cdot,1}\|_F \leq \|\Lambda\|_F \left\|\tilde{f}_{\cdot,1}\right\|_F + \|\tilde{\eta}\cdot,1\|_F = O_p(\sqrt{n})$ and $\left\|\hat{\psi}_{\varepsilon}^{-1}\right\|_{\text{spec}} = O_p(1)$ follows from Assumption 2.6 and Item 2 of Lemma 2.A.7 implying

$$\begin{aligned}\left\|\hat{\psi}_{\varepsilon}^{-1} - \psi_{\varepsilon}^{*-1}\right\|_{\text{spec}} &= O_p(n^{-1/2}) \text{ and}\\ \left\|\psi_{\varepsilon}^{*-1}\right\|_{\text{spec}} &\leq \|\Omega_{\eta}^{-1}\|_{\text{spec}} + \|\Omega_{\eta}^{-1}\|_{\text{spec}}^2 \|\Lambda\|_F^2 \left\|\left(\Lambda'\Omega_{\eta}^{-1}\Lambda\right)^{-1}\right\|_F\\ &= O(1) + O(1)O(n)O(n^{-1}) = O(1),\end{aligned}$$

using Assumptions 2.1 and 2.2 and Item 1 of Lemma 2.A.6. We conclude that $II = O_p\left(\frac{\sqrt{n}}{\sqrt{T}}\right) = o_p(1)$.

Finally, we obtain for III:

$$\begin{aligned}III =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{1}{\omega_{\eta,i}^2}(\hat{\delta}_{\eta,i} - \delta_{\eta,i}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\hat{\delta}_{\eta,i}}{\hat{\omega}_{\eta,i}^2\omega_{\eta,i}^2}(\omega_{\eta,i}^2 - \hat{\omega}_{\eta,i}^2)\\ \leq& \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\omega_{\eta,i}^2)^2}\right)^{1/2}\left(\sum_{i=1}^{n}(\hat{\delta}_{\eta,i} - \delta_{\eta,i})^2\right)^{1/2}\\ &+ \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\delta}_{\eta,i}^2}{(\hat{\omega}_{\eta,i}^2\omega_{\eta,i}^2)^2}\right)^{1/2}\left(\sum_{i=1}^{n}(\hat{\omega}_{\eta,i}^2 - \omega_{\eta,i}^2)^2\right)^{1/2},\end{aligned}$$

which is indeed $o_p(1)$ thanks to the observations at the beginning of this proof. ∎

2.A.4 Auxiliary Lemmas

Lemma 2.A.7 *Consider the factor estimates and the H_K from Lemma 2.3.1. Then, under Assumptions 2.1, 2.2 and 2.4–2.6, under $\mathrm{P}_{0,n,T}^{MP}$ or $\mathrm{P}_{0,n,T}^{PANIC}$ and as $n, T \to \infty$, we have*

1. $\left\|\left(\hat{\Lambda}'\hat{\Omega}_{\eta}^{-1}\hat{\Lambda}\right)^{-1} - \left(H_K'\Lambda'\Omega_{\eta}^{-1}\Lambda H_K\right)^{-1}\right\|_F = o_p(n^{-3/2})$, *and*
2. $\left\|\hat{\Omega}_{\eta}^{-1}\hat{\Lambda}\left(\hat{\Lambda}'\hat{\Omega}_{\eta}^{-1}\hat{\Lambda}\right)^{-1}\hat{\Lambda}'\hat{\Omega}_{\eta}^{-1} - \Omega_{\eta}^{-1}\Lambda\left(\Lambda'\Omega_{\eta}^{-1}\Lambda\right)^{-1}\Lambda'\Omega_{\eta}^{-1}\right\|_F = o_p(n^{-1/2})$.

Proof We start by noting that $\left\|H_K'\Lambda'\Omega_{\eta}^{-1}\Lambda H_K - \hat{\Lambda}'\hat{\Omega}_{\eta}^{-1}\hat{\Lambda}\right\|_F = o_p(\sqrt{n})$: the terms for approximating the loadings are negligible thanks to $\left\|\Lambda H_K - \hat{\Lambda}\right\|_F$ (Lemma 2.3.1) and $\left\|\Omega_{\eta}^{-1} - \hat{\Omega}_{\eta}^{-1}\right\|_{\text{spec}}$ being $o_p(1)$ in combination with H_K being bounded and $\|\Omega_{\eta}^{-1}\|_{\text{spec}} = O(1)$. The term due to approximating the long-run variances, $H_K'\Lambda'(\Omega_{\eta}^{-1} - \hat{\Omega}_{\eta}^{-1})\hat{\Lambda}$, can again be treated using Cauchy-Schwarz: ignoring H_K, its (k,l)th entry is given by

$$\sum_{i=1}^{n}\lambda_{ik}\hat{\lambda}_{il}((\hat{\omega}_{\eta,i}^2)^{-1} - (\omega_{\eta,i}^2)^{-1}) \leq \left(\sum_{i=1}^{n}\lambda_{ik}\hat{\lambda}_{il}\right)^{1/2}\left\|\Omega_{\eta}^{-1} - \hat{\Omega}_{\eta}^{-1}\right\|_F$$

$$=O_p(\sqrt{n})o_p(1),$$

thanks to the discussion at the beginning of Section 2.A.3.

Next, we have that

$$\left\| \frac{1}{n} H_K' \Lambda' \Omega_\eta^{-1} \Lambda H_K \right\|_F \leq \|H_K\|_F^2 \frac{\|\Lambda\|_F^2}{n} \left\|\Omega_\eta^{-1}\right\|_{\text{spec}} = O(1),$$

and

$$\begin{aligned}
\lambda_{\min}\left(\frac{1}{n} H_K' \Lambda' \Omega_\eta^{-1} \Lambda H_K\right) &= \left\| H_K^{-1} \left(\frac{1}{n}\Lambda' \Omega_\eta^{-1} \Lambda\right)^{-1} (H_K')^{-1} \right\|_{\text{spec}}^{-1} \\
&\geq \left\|H_K^{-1}\right\|_F^{-2} \left\| \left(\frac{1}{n}\Lambda' \Omega_\eta^{-1} \Lambda\right)^{-1} \right\|_{\text{spec}}^{-1},
\end{aligned}$$

which is bounded away from zero thanks to $\left\|H_K^{-1}\right\|_F$ being bounded and Item 1 of Lemma 2.A.6. Thus, we can restrict attention to a compact subset of the invertible matrices on $\mathbb{R}^K$, on which the matrix inverse is uniformly continuous. Therefore, $\left\| \frac{1}{n} H_K' \Lambda' \Omega_\eta^{-1} \Lambda H_K - \frac{1}{n} \hat{\Lambda}' \hat{\Omega}_\eta^{-1} \hat{\Lambda} \right\|_F = o_p(n^{-1/2})$ implies the same for $\left\| \left(\frac{1}{n} H_K' \Lambda' \Omega_\eta^{-1} \Lambda H_K\right)^{-1} - \left(\frac{1}{n} \hat{\Lambda}' \hat{\Omega}_\eta^{-1} \hat{\Lambda}\right)^{-1} \right\|_F$.

For Item 2, let $a = \Omega_\eta^{-1} \Lambda H_K$ and $b = \left(H_K' \Lambda' \Omega_\eta^{-1} \Lambda H_K\right)^{-1}$ and define $\hat{a} = \hat{\Omega}_\eta^{-1} \hat{\Lambda}$ and $\hat{b} = \left(\hat{\Lambda}' \hat{\Omega}_\eta^{-1} \hat{\Lambda}\right)^{-1}$ analogously. Thus

$$\begin{aligned}
&\left\| \hat{\Omega}_\eta^{-1} \hat{\Lambda} \left(\hat{\Lambda}' \hat{\Omega}_\eta^{-1} \hat{\Lambda}\right)^{-1} \hat{\Lambda}' \hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1} \Lambda \left(\Lambda' \Omega_\eta^{-1} \Lambda\right)^{-1} \Lambda' \Omega_\eta^{-1} \right\|_F \\
&= \left\| \hat{a}\hat{b}\hat{a}' - aba' \right\|_F \\
&\leq \|\hat{a} - a\|_F \left\|\hat{b}\right\|_F \|\hat{a}\|_F + \|a\|_F \left\|\hat{b} - b\right\|_F \|\hat{a}\|_F + \|a\|_F \|b\|_F \|\hat{a} - a\|_F .
\end{aligned}$$

From Assumption 2.2 and H_K being bounded it follows that $\|b\|_F = O_p(n^{-1})$ and in combination with Assumption 2.1 we obtain

$$\|a\|_F \leq \left\|\Omega_\eta^{-1}\right\|_{\text{spec}} \|\Lambda\|_F \|H_K\|_F = O_p(\sqrt{n}).$$

From Item 1, $\left\|\hat{b} - b\right\|_F = o_p(n^{-3/2})$ so that also $\left\|\hat{b}\right\|_F = O_p(n^{-1})$. Finally, we have

$$\begin{aligned}
\|\hat{a} - a\|_F &\leq \left\|\hat{\Omega}_\eta^{-1} - \Omega_\eta^{-1}\right\|_{\text{spec}} \left\|\hat{\Lambda}\right\|_F \|H_K\|_F + \left\|\Omega_\eta^{-1}\right\|_{\text{spec}} \left\|\hat{\Lambda} - \Lambda H_K\right\|_F \\
&= o_p(1) O_p(\sqrt{n}) O_p(1) + O(1) o_p(1) = o_p(\sqrt{n}),
\end{aligned}$$

where $\left\|\hat{\Lambda} - \Lambda H_K\right\|_F = o_p(1)$ by Lemma 2.3.1. Combining all these results indeed yields the correct rate. ■

Proof (Independent proof of Proposition 2.4.1) Here we demonstrate the joint asymptotic normality required to apply the second part of Corollary 2.4.1. We divide the proof into two parts. In Part A, we prove the theorem for P_a while in Part B we discuss t_a. We omit the proofs concerning P_b and t_b as they follow along the same lines.

Part A: First, we establish the joint convergence, under $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$ and $\mathrm{P}^{\mathrm{PANIC}}_{0,n,T}$, of P_a and the local likelihood ratio. As already hinted at in Remark 2.2.5, the results in Sections 2.2.1 and 2.2.2 imply that we only have to show this convergence once to get the powers in both experiments, as both likelihood ratios are asymptotically equivalent and the models coincide under the hypothesis. Having established this joint convergence, an application of Le Cam's third lemma will lead to the asymptotic distribution of P_a under $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$ and $\mathrm{P}^{\mathrm{PANIC}}_{h,n,T}$.

Specifically, Lemmas 2.2.1 and 2.2.4 imply that the limiting distributions of $\left(P_a, \log \frac{\mathrm{dP}^{\mathrm{PANIC}}_{h,n,T}}{\mathrm{dP}^{\mathrm{PANIC}}_{0,n,T}}\right)$ and $\left(P_a, \log \frac{\mathrm{dP}^{\mathrm{MP}}_{h,n,T}}{\mathrm{dP}^{\mathrm{MP}}_{0,n,T}}\right)$ are equal to that of $\left(P_a, h\Delta_{n,T} - \frac{1}{4}h^2\right)$, under $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$ and $\mathrm{P}^{\mathrm{PANIC}}_{0,n,T}$. From Lemma 1 and Lemma 2 in Bai and Ng (2010) we see that P_a is adaptive with respect to the estimation of nuisance parameters while Lemma A.2 in Moon and Perron (2004) shows that $\frac{1}{nT^2}\sum_{i=1}^n E'_{i,-1}E_{i,-1}$ converges in probability to $\frac{1}{2}\omega^2$. Therefore, P_a is asymptotically equivalent to $\tilde{P}_a = \frac{\frac{1}{\sqrt{n}T}\sum_{i=1}^n E'_{i,-1}\Delta E_i - \frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_{\eta,i}}{\sqrt{\phi^4/2}}$.

Under $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$ or $\mathrm{P}^{\mathrm{PANIC}}_{0,n,T}$, we can compute the asymptotic distribution of all possible linear combinations of $\tilde{P}_a$ and $\Delta_{n,T}$ by an application of Lemma 2.A.5. For all α, β in $\mathbb{R}$, we find, using $a_{i,n,T} = \alpha\frac{\omega^2_{\eta,i,T}}{\sqrt{\phi^4/2}} + \beta$ in Lemma 2.A.5,

$$\alpha\tilde{P}_a + \beta\Delta_{n,T} \xrightarrow{d} N\left(0, \left(\alpha^2 + \alpha\beta\sqrt{\frac{2\omega^4}{\phi^4}} + \frac{\beta^2}{2}\right)\right).$$

Thus, the Cramér-Wold theorem and the asymptotic equivalence of P_a and $\tilde{P}_a$, yield, still under $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$ or $\mathrm{P}^{\mathrm{PANIC}}_{0,n,T}$,

$$(P_a, \Delta_{n,T}) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{\frac{\omega^4}{2\phi^4}} \\ \sqrt{\frac{\omega^4}{2\phi^4}} & 1/2 \end{pmatrix}\right).$$

Equivalently,

$$\left(P_a, \log\frac{\mathrm{dP}_{h,n,T}}{\mathrm{dP}_{0,n,T}}\right) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ -\frac{1}{4}h^2 \end{pmatrix}, \begin{pmatrix} 1 & h\sqrt{\frac{\omega^4}{2\phi^4}} \\ h\sqrt{\frac{\omega^4}{2\phi^4}} & 1/2h^2 \end{pmatrix}\right).$$

Applying Le Cam's third lemma, we obtain $P_a \xrightarrow{d} N\left(h\sqrt{\frac{\omega^4}{2\phi^4}}, 1\right)$ under $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$ or $\mathrm{P}^{\mathrm{PANIC}}_{h,n,T}$.

Part B: As far as t_a is concerned, we recall that t_a is adaptive with respect to the estimation of nuisance parameters (see proofs of Theorem 2a) and b) in Moon and Perron (2004)) and that $\frac{1}{nT^2}\sum_{t=1}^T Y'_{\cdot,t-1}Q_\gamma Y_{\cdot,t-1}$ converges in probability to $\frac{1}{2}\omega^2$ under $\mathrm{P}^{\mathrm{MP}}_{0,n,T}$. Thus, t_a is asymptotically equivalent to

$$\tilde{t}_a = \frac{\frac{1}{\sqrt{n}T}\sum_{i=1}^n Y'_{\cdot,t-1}Q_\Lambda\Delta Y_{\cdot,t-1} - \sqrt{n}\sum_{i=1}^n \delta_{\eta,i}}{\sqrt{\phi^4/2}}.$$

Moreover, we have

$$\frac{1}{\sqrt{n}T}\sum_{t=1}^T Y'_{\cdot,t}Q_\Lambda\Delta Y_{\cdot,t-1} = \frac{1}{\sqrt{n}T}\sum_{t=1}^T E'_{\cdot,t}Q_\Lambda\Delta E_{\cdot,t-1}$$

$$= \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} E'_{\cdot,t} \Delta E_{\cdot,t-1} - \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} E'_{\cdot,t} \Lambda(\Lambda'\Lambda)^{-1} \Lambda \Delta E_{\cdot,t-1}$$
$$= \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} E'_{-1,i} \Delta E_i + o_p(1),$$

where the last equality follows from the proof of Lemma 2 c) in Moon and Perron (2004). Therefore, t_a is asymptotically equivalent to $\tilde{P}_a$. Thus, following the same steps as in Part A, we find $t_a \xrightarrow{d} N\left(h\sqrt{\frac{\omega^4}{2\phi^4}}, 1\right)$ under $\mathrm{P}^{\mathrm{MP}}_{h,n,T}$ or $\mathrm{P}^{\mathrm{PANIC}}_{h,n,T}$. ■

2.B Additional Monte-Carlo Results

In this supplement we present sizes and powers for additional DGPs and additional long-run variance estimates. The first subsection provides sizes and powers for additional DGPs. In the second subsection, we consider the same DGPs as in Sections 2.B.1 and 2.5, but with long-run variances estimated using the Newey and West (1994) bandwidth.

Tables 2.4–2.6 are analogous to Tables 2.1–2.3. Figures 2.11–2.19 are analogous to Figures 2.1–2.3 and 2.5–2.10. In general, the sizes for the MA case are slightly better controlled with the Newey and West (1994) bandwidth, at the expense of slightly lower power for small sample sizes.

2.B.1 Sizes and Powers in Additional DGPs

First, Figures 2.5 and 2.6 consider the powers in the presence of MA and AR serial correlation, respectively. The results are similar to those for i.i.d innovations. Figure 2.7 shows the results when the factor innovations are overdifferenced, i.e., the factor is stationary under the hypothesis. The powers appear to be unaffected. Figure 2.8 considers the case of the dependence being generated by three factors, with the corresponding sizes reported in Table 2.2. For very small sample sizes, powers of both tests are affected, but generally the results are similar also here.

We now consider deviations from our assumptions. Figure 2.9 reports the size-corrected powers of our tests against heterogeneous alternatives of the form

$$\rho_i = 1 + \frac{hU_i}{\sqrt{nT}}, \tag{2.B.1}$$

where the U_i are i.i.d. random variables with mean one. We draw the U_i from a Uniform(0.2,1.8) distribution. Once again, the finite-sample behaviour does not appear to be affected significantly, for both small and large samples.

Finally, we consider non-Gaussian innovations. Figure 2.10 reports size corrected powers with the innovations drawn from a t distribution with five degrees of freedom. The corresponding sizes are reported in Table 2.3. Also here, the conclusions remain the same.

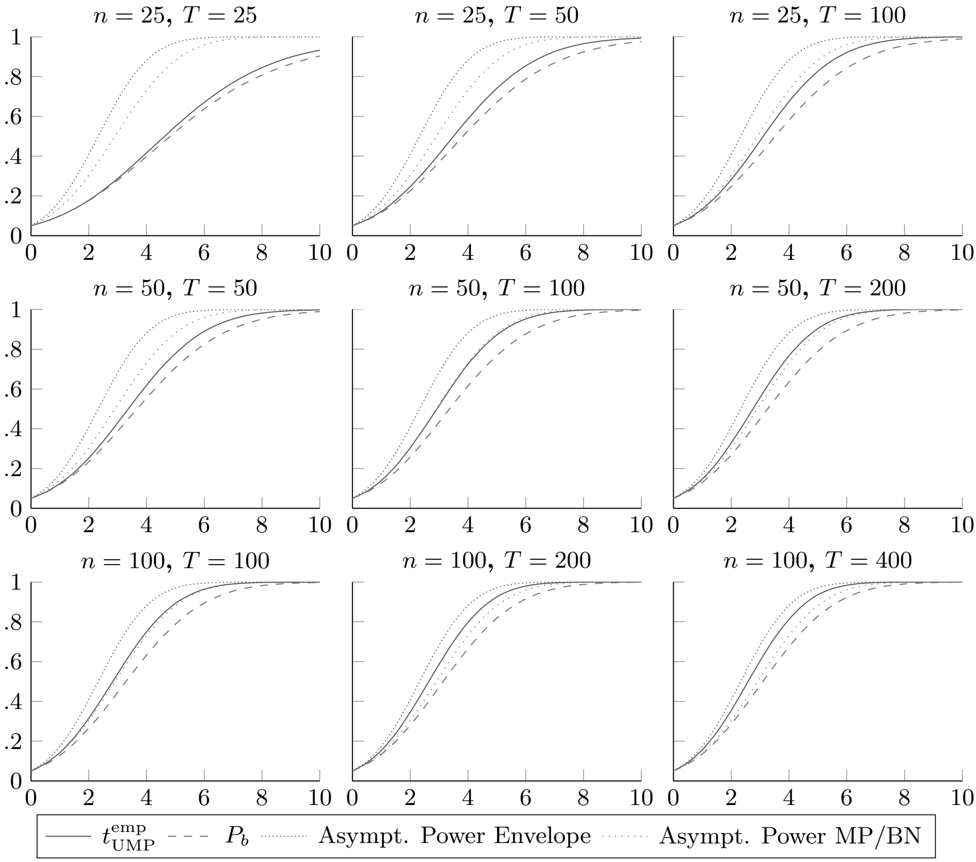


Figure 2.5: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with MA factor innovations and MA idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

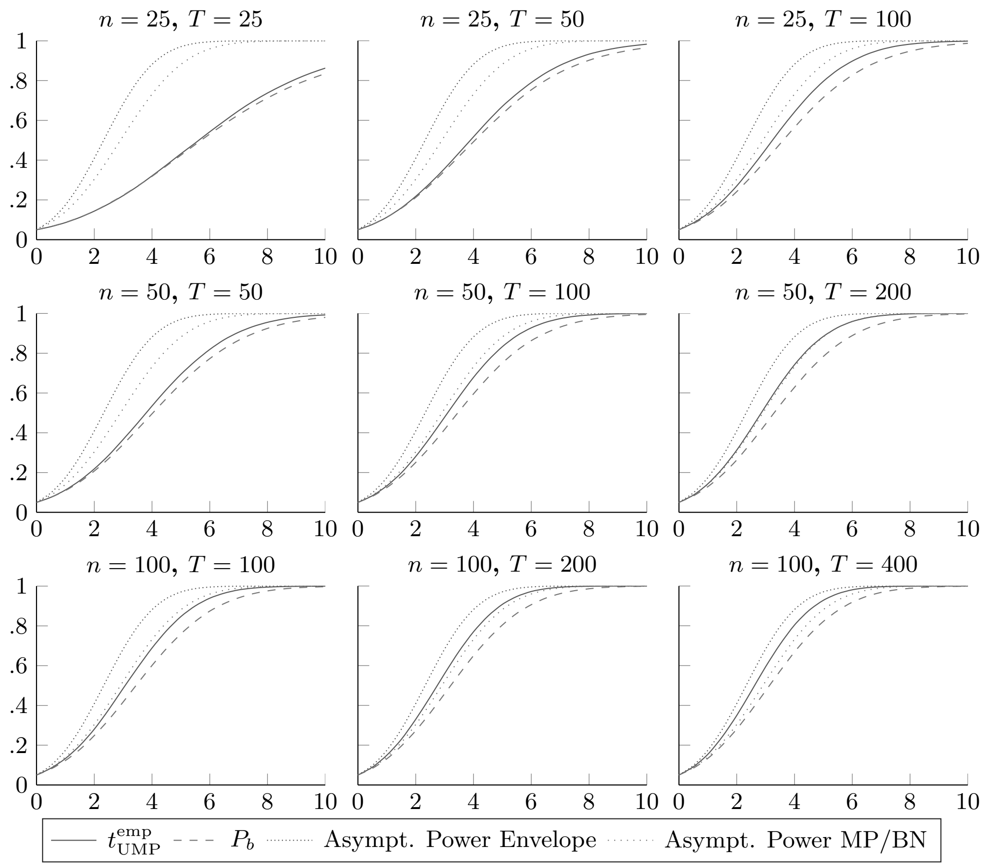


Figure 2.6: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with AR factor innovations and AR idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

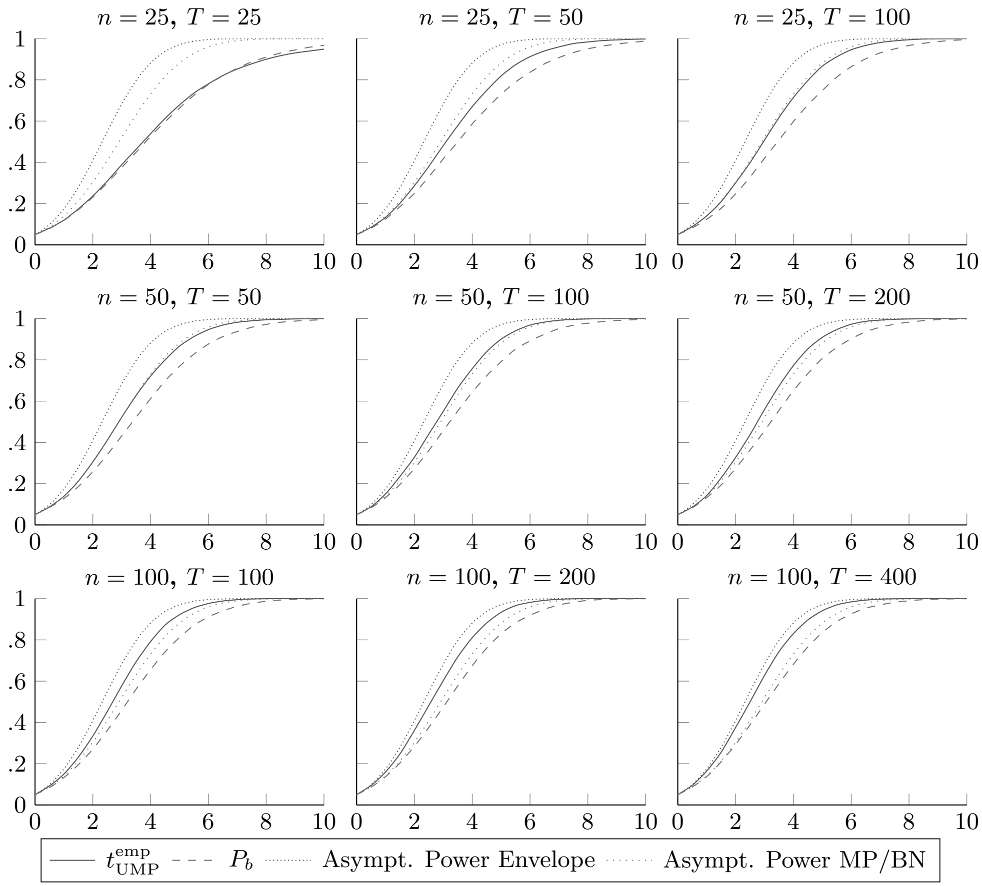


Figure 2.7: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with overdifferenced i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. The factor is stationary. Based on 100 000 replications.

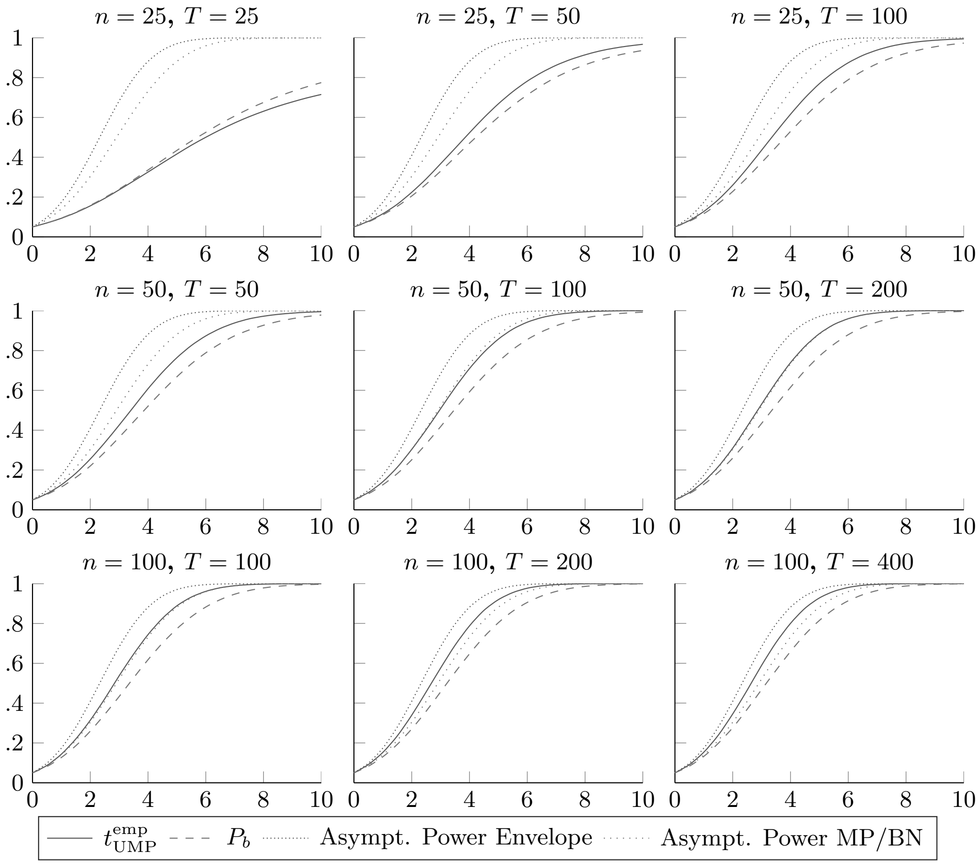


Figure 2.8: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Dependence based on three factors. Based on 100 000 replications.

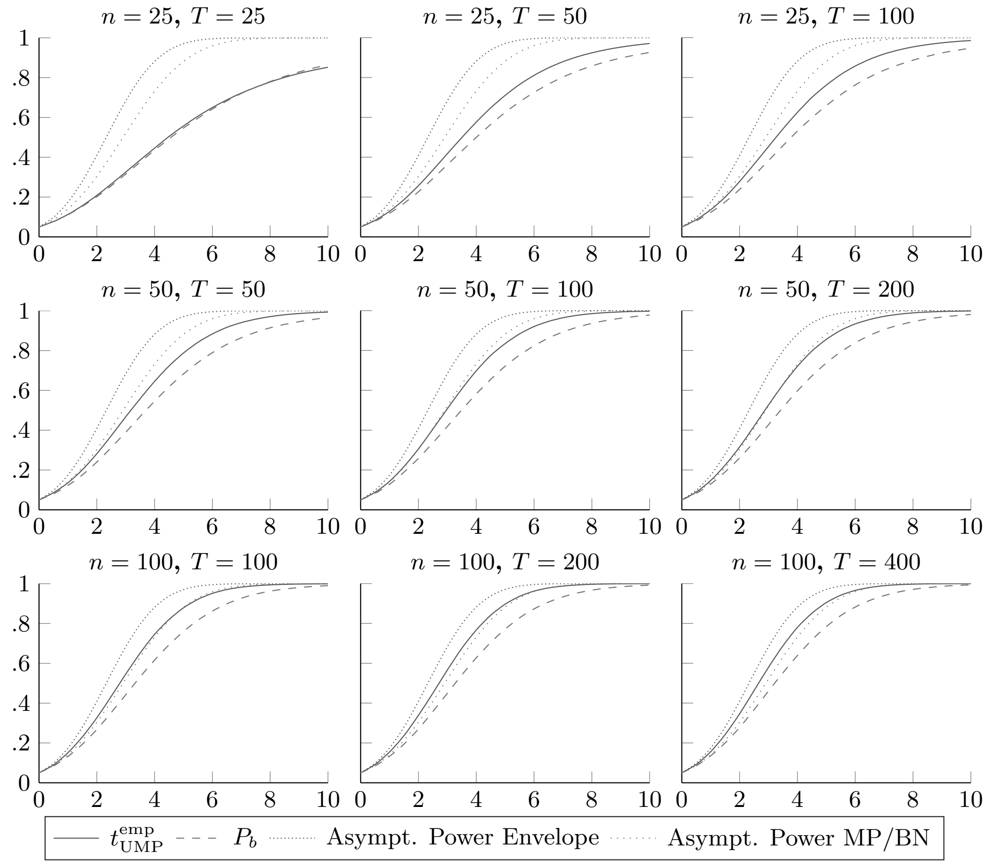


Figure 2.9: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Alternatives drawn from a Uniform(0.2,1.8) distribution. Based on 100 000 replications.

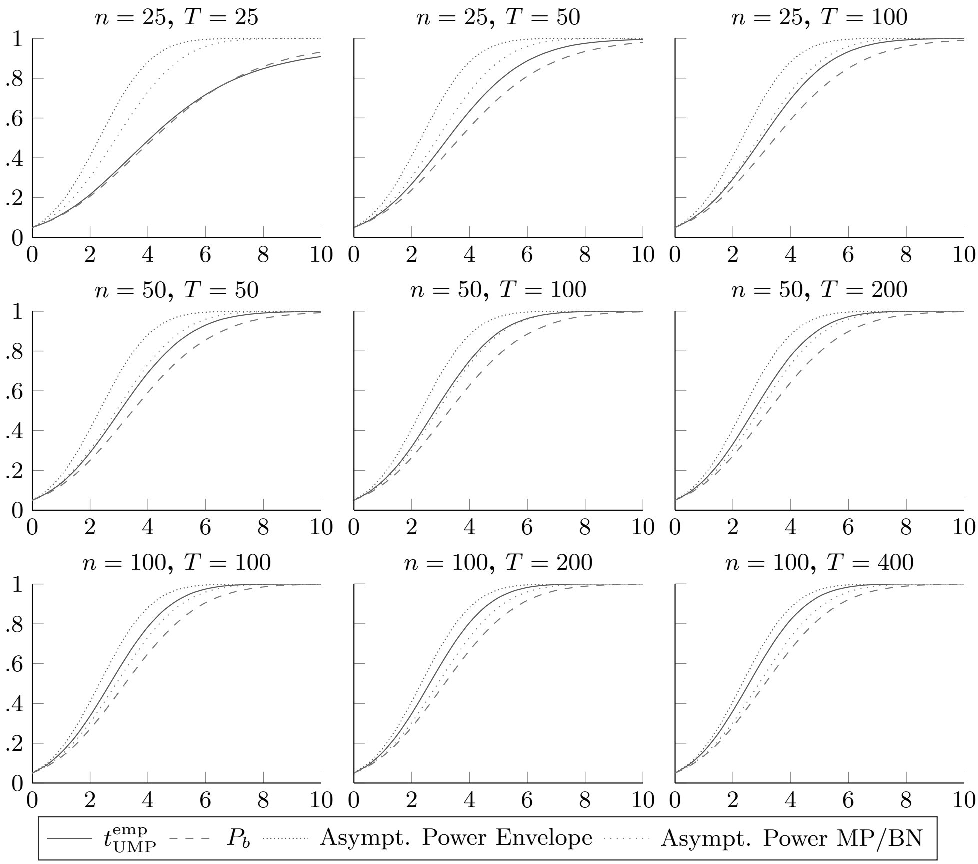


Figure 2.10: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Innovations drawn from a t_5 distribution. Note that the power envelopes refer to the Gaussian experiment. Based on 100 000 replications.

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.8	3.9	5.9	4.6	10.5	9.5	3.9	10.8	9.6
25	50	0.6	1.4	5.7	6.6	3.1	8.2	6.4	4.2	12.0	9.4
25	100	0.6	1.8	6.5	7.1	3.5	9.3	6.5	5.1	13.7	9.9
50	50	0.6	1.7	4.4	4.7	4.8	8.2	5.6	6.8	12.9	8.4
50	100	0.6	2.1	5.1	5.1	4.3	8.0	4.8	7.4	14.0	8.4
50	200	0.6	2.4	5.5	5.4	4.6	8.5	5.0	6.4	11.9	7.3
100	100	0.6	2.9	5.0	4.6	5.4	7.8	4.7	11.3	16.6	9.3
100	200	0.6	3.1	5.2	4.8	5.0	7.4	4.5	8.5	12.5	7.4
100	400	0.6	3.3	5.3	5.0	5.7	8.3	4.9	6.0	8.9	5.7
25	25	0.8	1.0	3.7	5.2	4.9	9.8	9.6	4.1	10.0	9.5
25	50	0.8	1.9	5.7	6.0	2.8	6.7	6.0	4.0	10.1	9.0
25	100	0.8	2.5	6.6	6.6	2.9	7.0	6.0	4.7	11.1	9.5
50	50	0.8	2.4	5.0	5.0	4.5	7.1	6.5	6.7	11.4	9.9
50	100	0.8	3.0	5.6	5.5	3.6	6.2	5.3	6.8	11.7	9.6
50	200	0.8	3.3	6.0	5.8	3.7	6.3	5.3	5.7	9.7	8.1
100	100	0.8	3.6	5.4	5.0	4.6	6.3	5.7	10.2	14.2	11.6
100	200	0.8	3.8	5.6	5.3	4.0	5.6	5.0	7.4	10.4	8.6
100	400	0.8	3.9	5.6	5.4	4.4	6.2	5.4	5.2	7.3	6.4
25	25	1.0	1.2	4.0	5.2	5.1	9.6	10.2	4.4	9.8	10.1
25	50	1.0	2.4	6.0	6.1	2.8	6.2	6.3	4.1	9.6	9.5
25	100	1.0	3.1	7.0	6.8	2.8	6.2	6.1	4.8	10.4	10.1
50	50	1.0	2.9	5.3	5.4	4.5	6.8	7.7	6.6	10.9	11.5
50	100	1.0	3.4	5.9	5.7	3.4	5.6	5.8	6.7	10.9	10.9
50	200	1.0	3.8	6.2	6.1	3.4	5.5	5.6	5.6	9.0	8.9
100	100	1.0	3.9	5.6	5.3	4.4	5.9	6.6	9.9	13.6	13.9
100	200	1.0	4.1	5.7	5.5	3.7	5.1	5.4	7.2	9.9	9.9
100	400	1.0	4.2	5.8	5.7	4.1	5.6	5.7	5.0	6.8	6.8
Mean abs. dev. from 5%			2.3	0.8	0.6	1.0	2.2	1.2	1.7	6.1	4.2

Table 2.2: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Andrews Bandwidth, three factors.

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.7	2.9	3.3	2.0	4.9	4.6	2.3	7.2	5.9
25	50	0.6	1.4	4.7	4.2	1.8	5.0	3.7	3.2	9.1	6.4
25	100	0.6	1.8	5.5	4.7	2.3	6.1	4.2	3.9	10.1	6.8
50	50	0.6	2.0	4.3	3.7	2.6	4.6	3.6	5.3	10.0	6.8
50	100	0.6	2.6	5.1	4.3	2.9	5.3	3.8	6.1	10.9	7.0
50	200	0.6	2.9	5.4	4.5	3.4	5.9	4.1	5.3	9.2	6.1
100	100	0.6	3.2	5.0	4.2	3.3	4.9	3.8	9.1	13.2	8.2
100	200	0.6	3.6	5.3	4.4	3.6	5.3	4.0	6.9	9.9	6.7
100	400	0.6	3.7	5.4	4.6	4.4	6.2	4.5	4.9	7.1	5.2
25	25	0.8	0.9	3.1	3.5	2.0	4.5	4.9	2.4	6.8	6.5
25	50	0.8	1.8	5.0	4.6	1.7	4.5	4.1	3.1	8.3	7.2
25	100	0.8	2.3	5.9	5.2	2.2	5.3	4.6	3.9	9.3	7.7
50	50	0.8	2.3	4.6	4.2	2.4	4.2	4.3	5.2	9.4	8.3
50	100	0.8	3.0	5.4	4.8	2.6	4.7	4.3	5.9	10.1	8.5
50	200	0.8	3.3	5.7	5.2	3.0	5.2	4.7	5.0	8.4	7.2
100	100	0.8	3.5	5.2	4.7	3.1	4.4	4.4	8.7	12.4	10.4
100	200	0.8	3.8	5.5	5.0	3.3	4.7	4.5	6.6	9.3	7.9
100	400	0.8	3.9	5.5	5.1	3.9	5.5	5.0	4.7	6.5	5.9
25	25	1.0	1.0	3.3	3.8	2.0	4.4	5.6	2.5	6.7	7.3
25	50	1.0	2.0	5.2	5.1	1.7	4.2	4.5	3.3	8.1	8.2
25	100	1.0	2.6	6.0	5.8	2.2	5.1	5.1	3.9	9.0	8.9
50	50	1.0	2.5	4.7	4.6	2.4	4.1	5.0	5.1	9.1	10.0
50	100	1.0	3.1	5.4	5.2	2.6	4.4	4.8	5.8	9.9	10.0
50	200	1.0	3.5	5.8	5.6	3.0	5.0	5.2	4.9	8.1	8.1
100	100	1.0	3.6	5.3	4.9	3.0	4.3	5.0	8.6	12.1	12.6
100	200	1.0	3.9	5.5	5.2	3.2	4.6	4.9	6.4	9.0	9.0
100	400	1.0	4.1	5.6	5.5	3.8	5.3	5.4	4.6	6.3	6.4
Mean abs. dev. from 5%			2.3	0.6	0.6	2.2	0.5	0.6	1.4	4.1	2.8

Table 2.3: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Andrews Bandwidth, t-distribution with five degrees of freedom.

2.B.2 Finite-Sample Results with the Newey and West (1994) Bandwidth

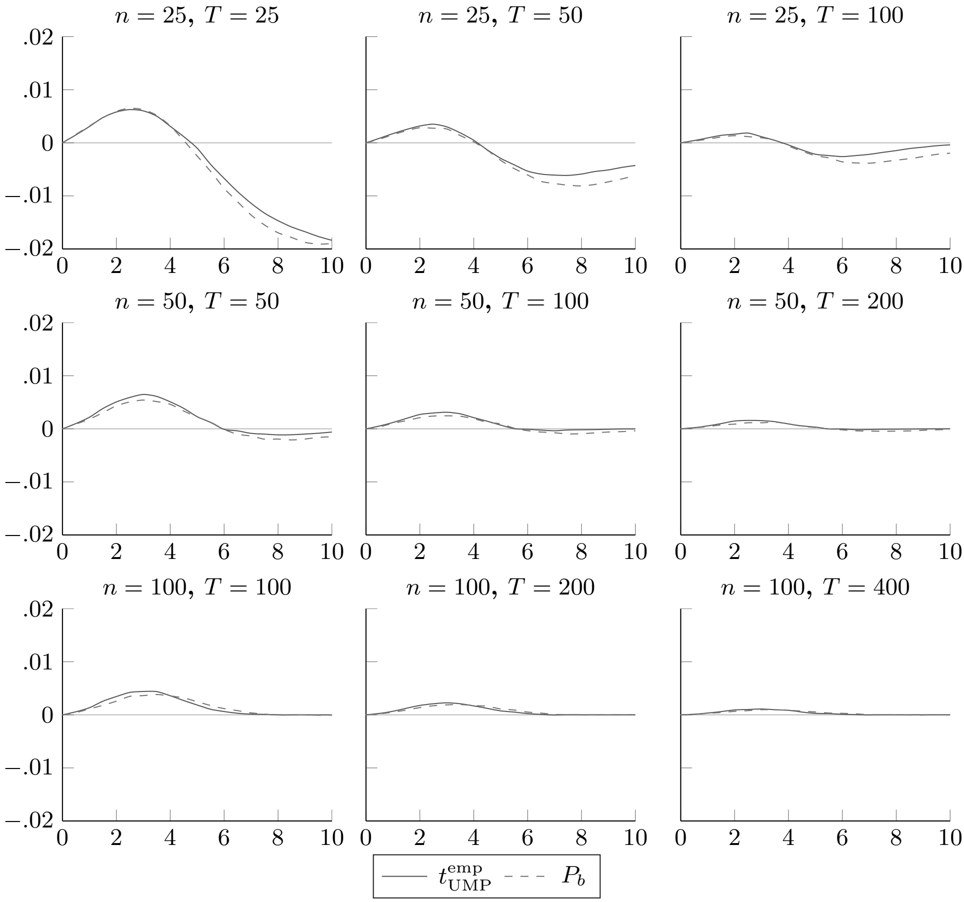


Figure 2.11: Difference between powers in the MP vs the PANIC framework as a function of $-h$ with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 1 000 000 replications.

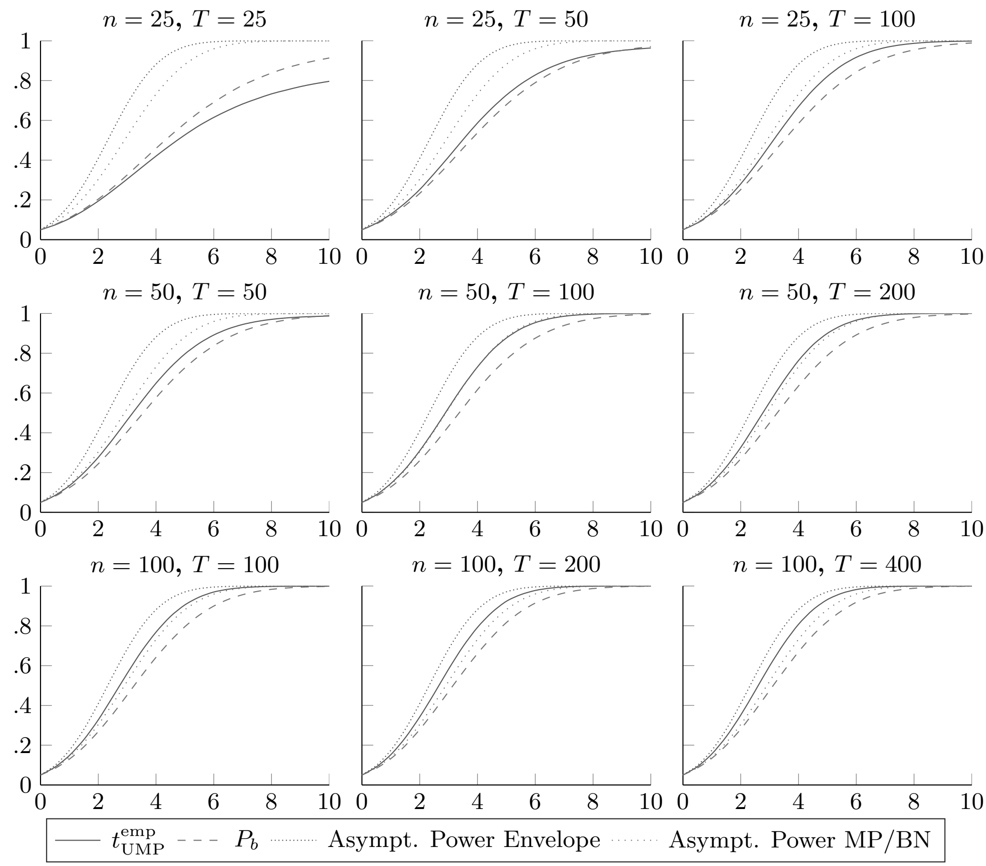


Figure 2.12: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

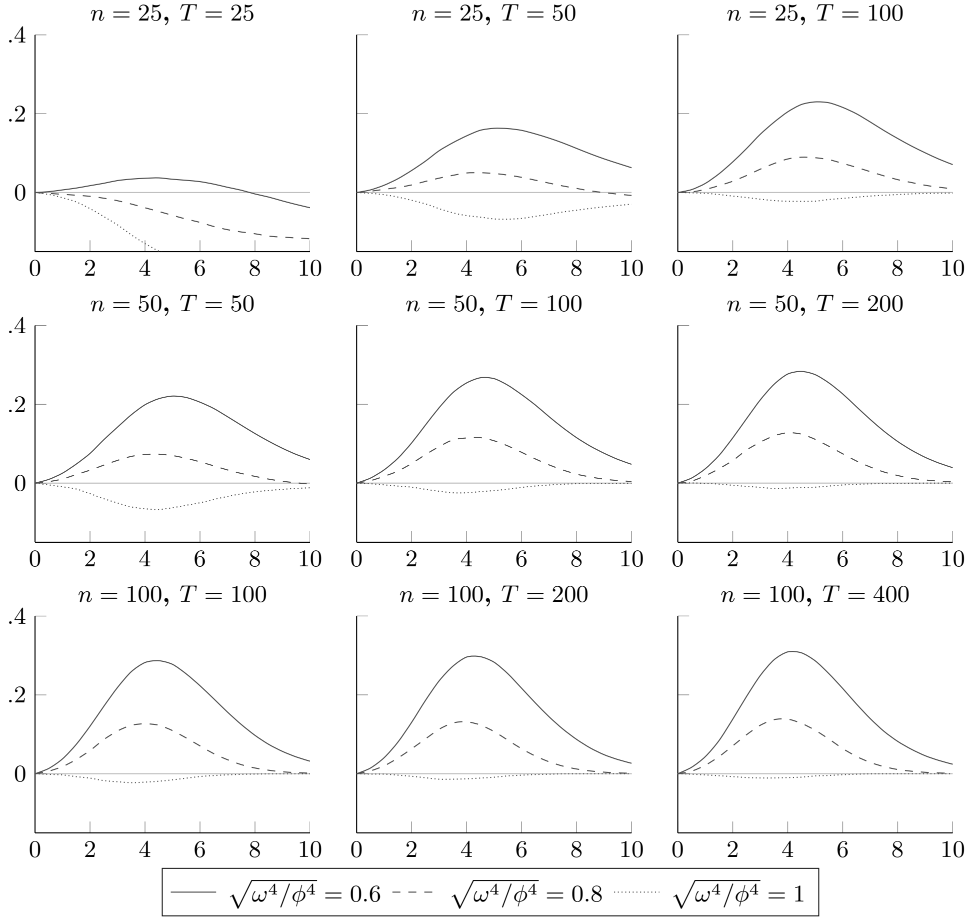


Figure 2.13: (Size-corrected) power gains from using $t_{\text{UMP}}^{\text{emp}}$ over P_b for varying values of $\sqrt{\omega^4/\phi^4}$ and sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts. Based on 0 replications.

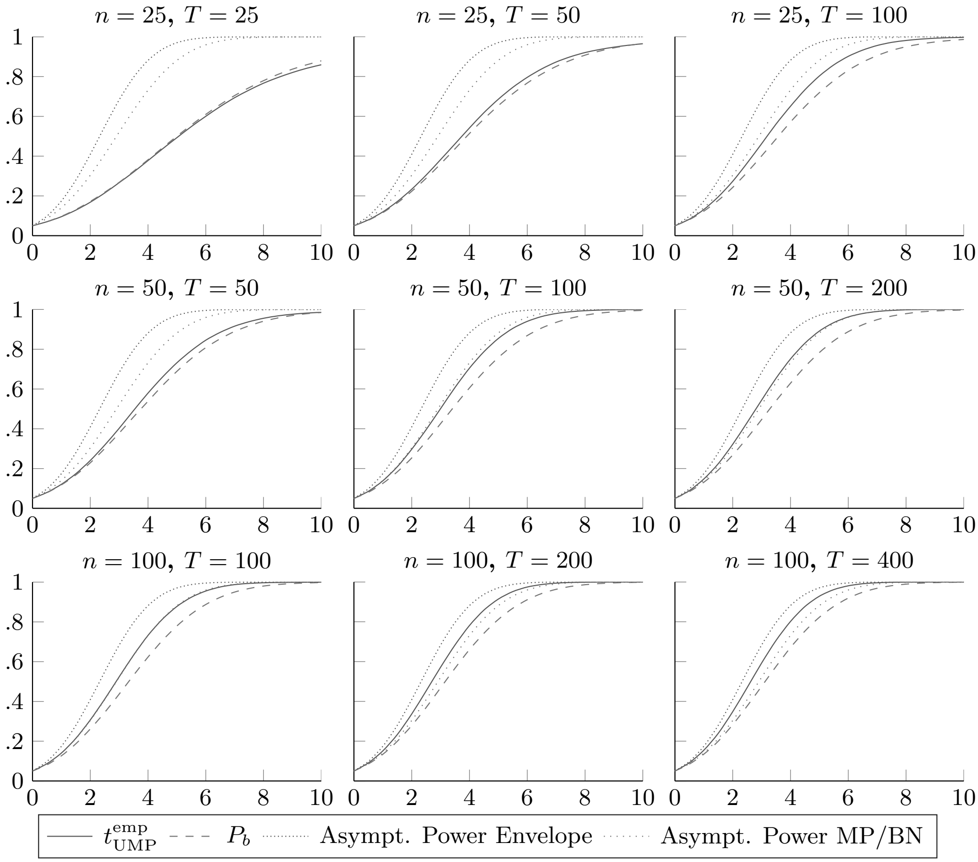


Figure 2.14: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with MA factor innovations and MA idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

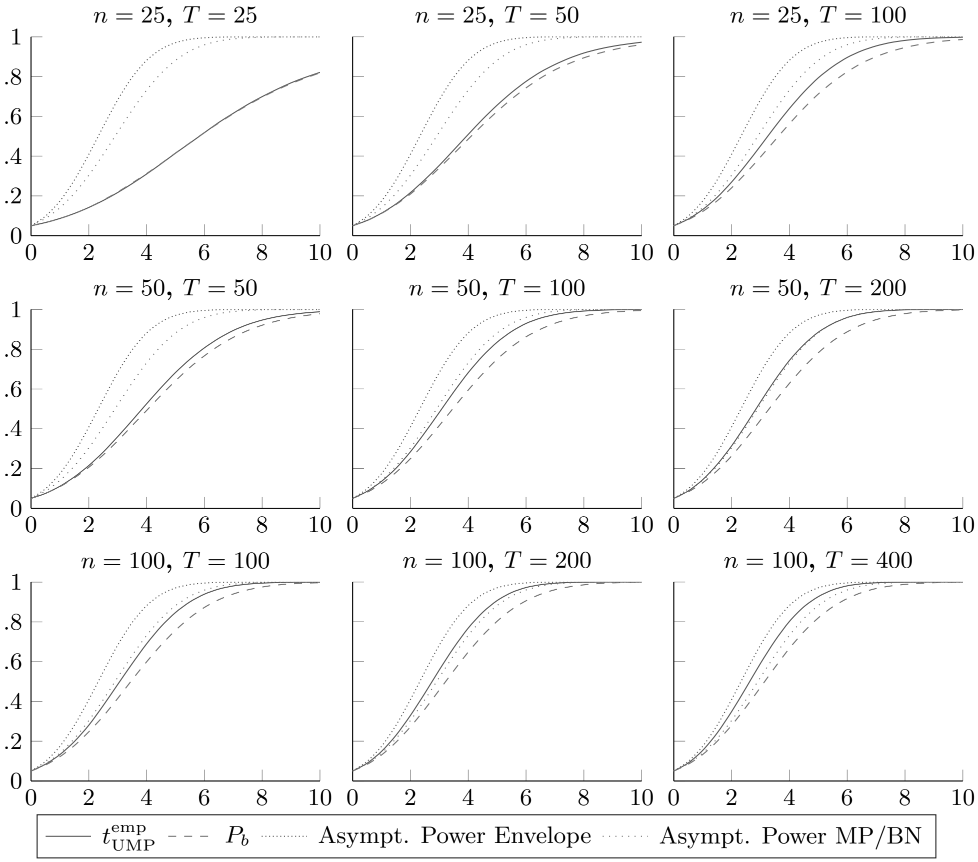


Figure 2.15: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with AR factor innovations and AR idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

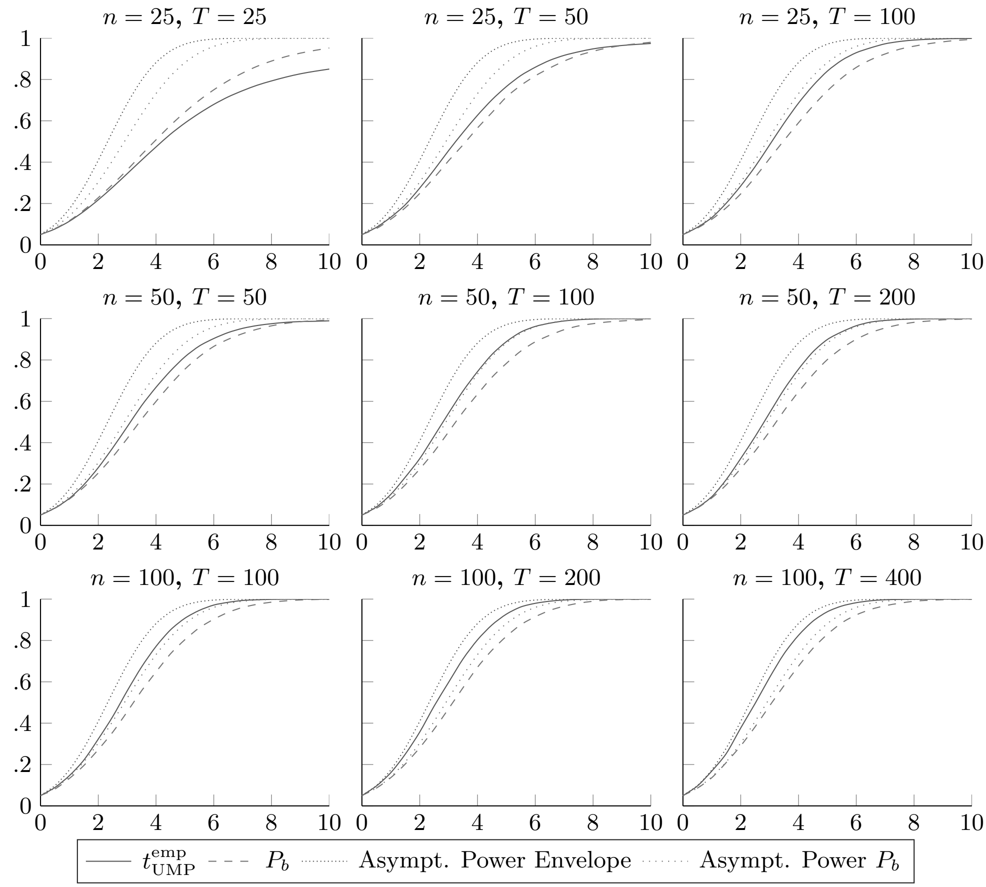


Figure 2.16: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with overdifferenced i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

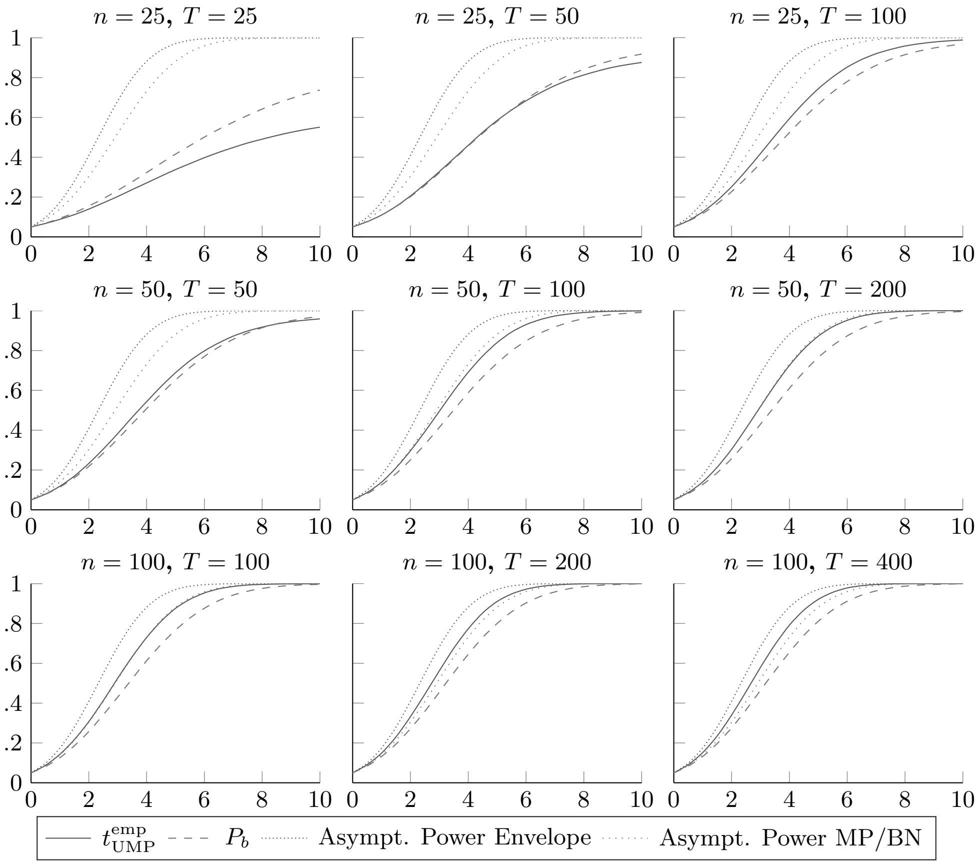


Figure 2.17: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Dependence based on three factors. Based on 100 000 replications.

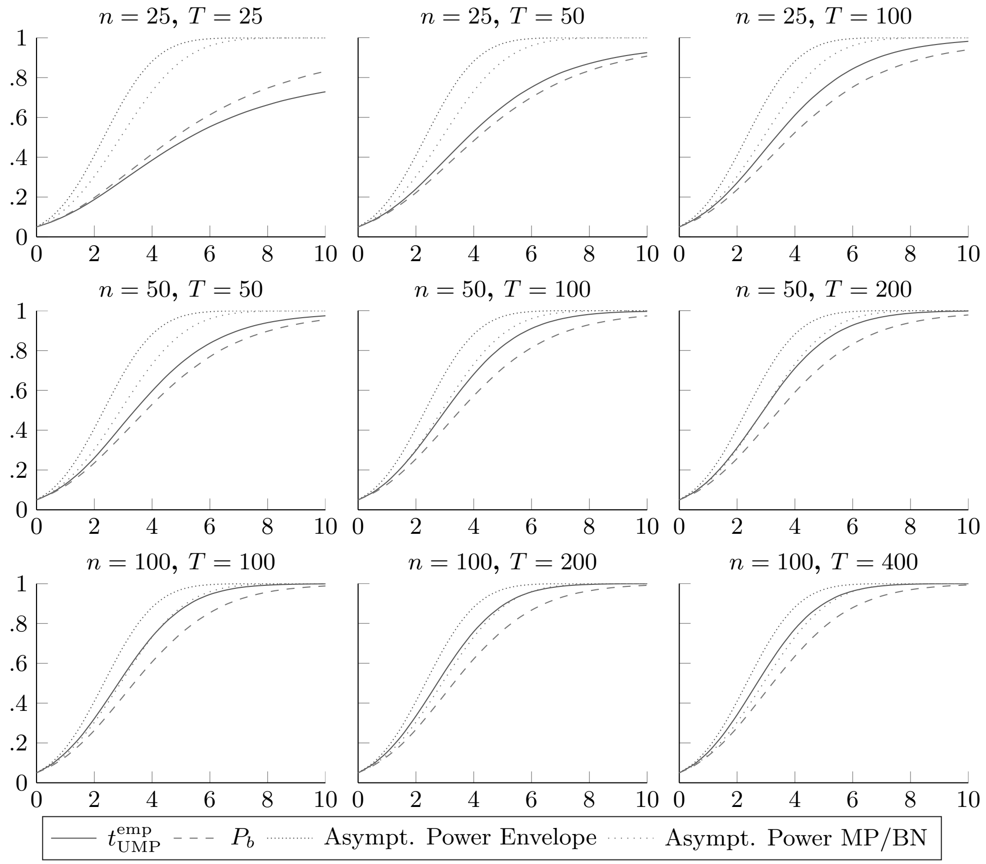


Figure 2.18: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Alternatives drawn from a Uniform(0.2,1.8) distribution. Based on 100 000 replications.

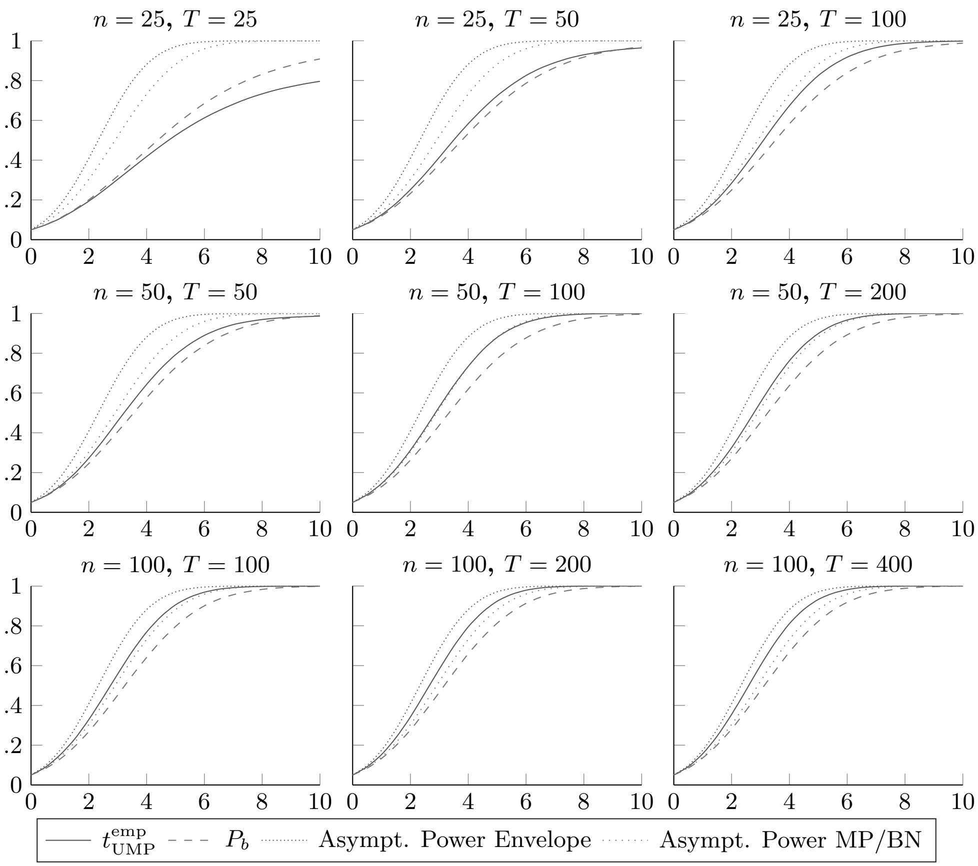


Figure 2.19: Size-corrected power of unit-root tests as a function of $-h$ for varying sample sizes in the PANIC framework with i.i.d. factor innovations and i.i.d. idiosyncratic parts and $\sqrt{\omega^4/\phi^4} = 0.8$. Based on 100 000 replications.

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.3	1.2	1.5	1.3	3.4	3.6	1.0	3.3	3.6
25	50	0.6	0.6	2.5	2.3	1.4	4.2	3.1	1.3	4.1	3.3
25	100	0.6	1.3	4.2	3.6	2.3	6.0	4.0	2.2	6.3	4.4
50	50	0.6	0.9	2.1	1.9	2.1	3.9	3.0	1.9	3.8	3.1
50	100	0.6	1.9	3.8	3.1	2.9	5.3	3.6	3.1	6.0	4.2
50	200	0.6	2.4	4.6	3.7	3.4	6.0	3.9	2.8	5.1	3.6
100	100	0.6	2.3	3.7	2.8	3.4	5.1	3.6	4.1	6.1	4.3
100	200	0.6	2.9	4.4	3.5	3.8	5.5	3.8	3.2	4.7	3.4
100	400	0.6	3.2	4.8	3.9	4.2	6.0	4.1	3.1	4.6	3.5
25	25	0.8	0.4	1.3	1.7	1.4	3.2	4.1	1.1	3.2	4.1
25	50	0.8	0.9	2.8	2.6	1.4	3.7	3.4	1.4	3.9	3.7
25	100	0.8	1.7	4.6	4.0	2.1	5.3	4.4	2.3	5.9	5.0
50	50	0.8	1.2	2.4	2.1	2.0	3.6	3.6	1.9	3.7	3.7
50	100	0.8	2.2	4.2	3.4	2.6	4.7	4.1	3.1	5.6	4.8
50	200	0.8	2.8	4.9	4.2	3.1	5.3	4.4	2.7	4.7	4.0
100	100	0.8	2.6	3.9	3.0	3.2	4.6	4.2	4.0	5.8	5.1
100	200	0.8	3.2	4.6	3.8	3.5	4.9	4.2	3.0	4.4	3.7
100	400	0.8	3.5	5.0	4.3	3.9	5.4	4.6	3.0	4.3	3.8
25	25	1.0	0.5	1.5	1.9	1.4	3.3	4.8	1.1	3.2	4.5
25	50	1.0	1.1	3.0	2.9	1.4	3.6	3.9	1.4	3.9	4.2
25	100	1.0	2.0	4.8	4.5	2.1	5.0	4.9	2.4	5.7	5.6
50	50	1.0	1.3	2.5	2.2	2.0	3.5	4.2	2.0	3.6	4.4
50	100	1.0	2.4	4.2	3.7	2.6	4.5	4.6	3.1	5.5	5.4
50	200	1.0	2.9	5.0	4.4	3.0	5.0	4.8	2.8	4.7	4.4
100	100	1.0	2.7	4.0	3.1	3.1	4.4	4.7	3.9	5.7	5.7
100	200	1.0	3.3	4.8	3.9	3.4	4.8	4.5	3.0	4.3	3.9
100	400	1.0	3.7	5.1	4.5	3.8	5.3	4.9	3.0	4.2	3.9
Mean abs. dev. from 5%			3.0	1.3	1.8	2.4	0.7	0.9	2.5	0.9	0.9

Table 2.4: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Newey Bandwidth.

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.5	1.5	3.3	3.0	7.2	7.9	1.6	5.0	6.3
25	50	0.6	0.6	2.6	4.2	2.4	6.7	5.6	1.6	5.6	5.5
25	100	0.6	1.1	4.7	5.7	3.4	9.1	6.2	2.9	8.7	6.9
50	50	0.6	0.6	1.7	2.6	3.9	6.8	4.8	2.4	5.0	4.3
50	100	0.6	1.3	3.5	3.8	4.2	7.9	4.6	3.7	7.6	5.1
50	200	0.6	1.8	4.4	4.4	4.6	8.6	4.7	3.3	6.6	4.4
100	100	0.6	1.9	3.4	3.2	5.5	7.9	4.5	5.0	7.8	4.9
100	200	0.6	2.4	4.1	3.8	5.2	7.7	4.2	3.7	5.8	3.8
100	400	0.6	2.8	4.6	4.2	5.7	8.3	4.5	3.7	5.7	3.8
25	25	0.8	0.5	1.5	2.8	3.2	6.9	8.1	1.8	4.6	6.2
25	50	0.8	0.8	2.8	3.6	2.2	5.5	5.2	1.6	4.7	5.0
25	100	0.8	1.8	5.1	5.2	2.8	6.9	5.8	2.8	7.0	6.3
50	50	0.8	1.0	2.3	2.6	3.7	5.9	5.6	2.5	4.7	4.9
50	100	0.8	2.1	4.2	4.0	3.6	6.2	5.1	3.6	6.4	5.6
50	200	0.8	2.7	5.0	4.7	3.7	6.4	5.0	3.1	5.5	4.6
100	100	0.8	2.5	3.9	3.3	4.7	6.5	5.4	4.7	6.9	5.8
100	200	0.8	3.1	4.7	4.0	4.1	5.9	4.7	3.3	4.9	4.0
100	400	0.8	3.5	5.0	4.5	4.3	6.1	4.9	3.3	4.7	4.0
25	25	1.0	0.7	1.7	2.7	3.4	6.8	8.7	1.9	4.7	6.6
25	50	1.0	1.1	3.2	3.6	2.2	5.1	5.4	1.8	4.6	5.2
25	100	1.0	2.3	5.5	5.3	2.7	6.2	5.9	2.9	6.7	6.6
50	50	1.0	1.3	2.7	2.7	3.7	5.7	6.6	2.6	4.6	5.6
50	100	1.0	2.6	4.5	4.0	3.4	5.6	5.6	3.5	6.1	6.1
50	200	1.0	3.2	5.4	4.8	3.4	5.6	5.3	3.1	5.1	4.8
100	100	1.0	2.9	4.2	3.4	4.5	6.1	6.3	4.7	6.6	6.6
100	200	1.0	3.4	4.9	4.1	3.8	5.3	5.0	3.3	4.7	4.3
100	400	1.0	3.8	5.3	4.7	4.0	5.5	5.2	3.2	4.4	4.1
Mean abs. dev. from 5%			3.1	1.3	1.2	1.3	1.6	0.8	2.0	1.0	0.8

Table 2.5: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Newey Bandwidth, three factors.

			i.i.d.			AR(1)			MA(1)		
n	T	$\sqrt{\omega^4/\phi^4}$	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b	t_{UMP}	$t_{\text{UMP}}^{\text{emp}}$	P_b
25	25	0.6	0.3	1.2	1.7	1.5	3.7	4.0	1.1	3.4	3.8
25	50	0.6	0.7	2.5	2.4	1.5	4.3	3.2	1.3	4.2	3.5
25	100	0.6	1.3	4.2	3.6	2.3	6.0	4.0	2.3	6.4	4.5
50	50	0.6	0.9	2.1	1.9	2.2	4.0	3.0	1.9	3.9	3.2
50	100	0.6	1.9	3.9	3.1	2.9	5.3	3.6	3.1	6.0	4.2
50	200	0.6	2.4	4.6	3.7	3.4	6.0	3.9	2.8	5.2	3.6
100	100	0.6	2.3	3.7	2.9	3.5	5.1	3.6	4.1	6.1	4.3
100	200	0.6	2.9	4.4	3.4	3.8	5.5	3.8	3.2	4.7	3.4
100	400	0.6	3.3	4.9	3.9	4.3	6.1	4.1	3.2	4.6	3.5
25	25	0.8	0.4	1.3	1.7	1.5	3.5	4.3	1.1	3.3	4.1
25	50	0.8	0.9	2.8	2.6	1.4	3.8	3.5	1.4	3.9	3.7
25	100	0.8	1.7	4.6	4.0	2.1	5.2	4.4	2.3	5.9	5.0
50	50	0.8	1.1	2.4	2.0	2.1	3.6	3.6	2.0	3.7	3.8
50	100	0.8	2.2	4.2	3.4	2.7	4.7	4.2	3.1	5.6	4.9
50	200	0.8	2.8	4.9	4.1	3.0	5.3	4.4	2.7	4.8	4.1
100	100	0.8	2.6	4.0	3.0	3.2	4.6	4.2	4.0	5.8	5.1
100	200	0.8	3.2	4.7	3.8	3.5	4.9	4.2	3.0	4.4	3.7
100	400	0.8	3.5	5.0	4.3	3.9	5.5	4.6	3.0	4.3	3.7
25	25	1.0	0.5	1.4	1.8	1.5	3.4	4.9	1.2	3.2	4.6
25	50	1.0	1.0	3.0	2.9	1.4	3.6	3.8	1.5	3.9	4.2
25	100	1.0	2.0	4.9	4.4	2.1	5.0	4.9	2.4	5.8	5.6
50	50	1.0	1.3	2.5	2.2	2.1	3.6	4.3	2.0	3.6	4.3
50	100	1.0	2.4	4.3	3.6	2.6	4.5	4.6	3.1	5.5	5.5
50	200	1.0	3.0	5.0	4.5	3.0	5.0	4.8	2.7	4.7	4.4
100	100	1.0	2.7	4.0	3.1	3.1	4.5	4.7	3.9	5.7	5.7
100	200	1.0	3.3	4.7	3.9	3.4	4.8	4.5	2.9	4.2	3.9
100	400	1.0	3.7	5.2	4.6	3.7	5.2	4.9	2.9	4.1	3.8
Mean abs. dev. from 5%			3.0	1.3	1.8	2.4	0.7	0.8	2.5	0.9	0.9

Table 2.6: Sizes (in percent) of nominal 5% level tests with no heterogeneity in the alternatives. Based on 1 000 000 replications. Newey Bandwidth, t-distribution with five degrees of freedom.

Chapter 3

Panel Unit-Root Tests Under Cross-Sectional Cointegration[1]

Abstract

We study unit-root tests for unobserved common factors in large panels. Recent panel unit-root tests typically allow for cross-sectional correlation due to common unobserved factors. As originally proposed in Bai and Ng (2004) ('PANIC'), unit-root tests are applied separately to the common factors and idiosyncratic deviations. While the testing problem for the idiosyncratic parts is in many cases well-understood, the testing problem for the factors has received much less attention. Bai and Ng (2004) show that using principal component estimates in ADF tests does not change their properties. We generalize this result to other unit-root tests and other factor estimates, which can lead to higher finite sample powers. In particular, we show that a Kalman smoother imposing the null hypothesis to estimate the factors often has a simple closed-form solution that avoids the computational issues usually associated with other methods.

We also discuss the implications of including deterministic trends in the factor equation, i.e., having factors with non-zero mean innovations. This specification can be considered as an alternative to including individual deterministic trends for each unit. Although this leads to nontrivial powers closer to the unit-root, we can again attain these powers based on

[1] Based on joint work with B.J.M. Werker.

estimated factors. In particular, we propose tests based on simple cross-sectional averages that are asymptotically uniformly most powerful. We derive the properties of these unit-root tests in the presence of multiple potentially cointegrated factors and show that they can be interpreted as unit-root tests for the observations. The cross-sectional averaging approach can lead to higher powers than cointegration-rank based tests and does not require pre-estimation of the total number of factors.

3.1 Introduction

For some years now, panel unit-root tests have been developed that are robust to cross-sectional correlation, see, e.g., Breitung and Pesaran (2008) for a review. In the presence of strong cross-sectional correlation (i.e., when the eigenvalues of the cross sectional covariance matrix are not bounded), this is modeled by assuming a factor structure, see, for example, Bai and Ng (2004), Breitung and Das (2007), Moon and Perron (2004), Pesaran (2007), and Phillips and Sul (2003).[2] The PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components) approach of Bai and Ng (2004) that tests separately for unit roots in common factors and idiosyncratic components has become a frequently used method of conducting panel unit root tests. Instead of a unit-root test for the observations, the two components are tested separately for a unit root. This approach allows the common factos and idiosyncratic parts to have different orders of integration.

The vast majority of follow-up papers have focused on the testing problem for the idiosyncratic components.[3] However, in many cases the testing problem for the idiosyncratic parts is just as important. Firstly, this is the

---

[2] Applications include O'Connell (1998), Papell (2006), and Silva, Hadri, and Tremayne (2009).

[3] A notable exception is the working paper Barigozzi and Trapani (2018), who consider both the number of (nonstationary) factors and number of factors with deterministic trends. However, their focus is different, as they attempt to consistently estimate the number of factors under very general conditions. The cost of this is that they do not consider local-to-unity specifications and thus cannot study systematically the asymptotic local power of their tests.

case whenever the main interest is the stationarity of the observations themselves and one does not assume that the factors and idiosyncratic parts are either both stationary or both nonstationary[4]: When the idiosyncratic parts are stationary, the stationarity of the common factor determines the long-run behavior of the observations. For example, when the variable of interest is to be used in a regression setting, rejecting a unit-root in the idiosyncratic parts is not sufficient to avoid spurious regression issues – unless nonstationarity in the other variables is exclusively due to the same nonstationary common factor.

Secondly, one may be directly interested in the source of potential nonstationarity, i.e., whether the factors, the idiosyncratic parts, or both are stationary. This may change the interpretation of the results. For example, if one were to find that all non-stationarity in stock prices comes from common factors, this may signal inefficient markets despite the stock prices being nonstationary. Thus, the problem of testing the factors for a unit root has to be solved as well, in particular when the idiosyncratic parts turn out to be stationary. Providing a better understanding of this unit-root testing problem for the unobserved factors is the goal of this chapter.

Our first contribution is to show that in many cases estimated factors can be inserted into univariate unit-root tests without their asymptotic power being affected. In Section 3.3 we consider unit-root tests for the common factors

[4] Some unit-root tests have been developed for the observations, however, Breitung and Das (2007) show that when these tests are evaluated under DGPs where the order of integration of the factors differs from that of the idiosyncratic parts, these tests do not attain close to nominal size even in large samples. As shown by Wichert et al. (2019), this is due to the fact that the Moon and Perron (2004) tests are equivalent to the tests that only test the idiosyncratic components. Consider the case of a dependent panel with a common stochastic trend but stationary idiosyncratic component. In this case, the observations for each panel unit are nonstationary. However, the commonly used panel unit-root tests would falsely reject, as they implicitly test only for a unit root in the idiosyncratic components. On the other hand, a unit-root test that is robust to this cross-sectional cointegration would essentially boil down to a unit root test for the common factor, as this testing problem is harder. Based on the commonly employed tests for the factors, this would negate many of the power gains that one sought out panel data for in the first place.

in the original PANIC setup, i.e., the setup most commonly used in practice. In the single factor case, when principal component factor estimates based on first-differenced observations are inserted into ADF statistics, their asymptotic distribution (and thus power) is the same as when based on observed factors. This was shown when the PANIC approach was originally introduced in Bai and Ng (2004). We generalize this result to likelihood-based unit-root tests such as the commonly used Elliott, Rothenberg, and Stock (1996) tests and other factor estimates, for example principal components estimated in levels as proposed in Bai (2004) and generalized principal components as in Choi (2017). These alternative approaches are attractive, as ADF tests require ad-hoc specifications of lag length and, if the idiosyncratic components are known to be stationary, differencing makes the principal component estimates less efficient.[5] Moreover, when the factor estimates are to be used for a unit-root test, one can impose the unit-root in the estimation stage. We show that a Kalman smoother that takes into account the joint distribution of the factors and the observations under the null hypothesis also leads to correct asymptotic sizes and powers. This can be exploited for finite-sample gains, but also in iterative procedures to directly identify the factor of interest. We also develop a computationally simple closed-form solution of the Kalman smoother and relate it to existing factor estimates.

Our second contribution, in Section 3.4, reconsiders the unit-root testing problem for the factors, but with non-zero mean factor innovations, i.e., deterministic trends in the factors. Throughout the unit-root literature, various specifications with regards to how deterministic trends, regressors, and unobserved factors enter have been considered. As mentioned earlier, we consider

[5] In practice, of course one does not know the true DGP. However, the unit root tests for the factors considered here will have power in T^{-1} neighbourhoods of the unit root, whereas tests for the idiosyncratic parts have power in $T^{-1}n^{-1/2}$ or $T^{-1}n^{-1/4}$ neighbourhoods around unity. Thus, for testing the factors for a unit root, the order of integration of the idiosyncratic parts can be considered as known and this knowledge can be exploited to obtain better factor unit-root tests. Moreover, Banerjee, Marcellino, and Masten (2017) argues that for most economic time series the idiosyncratic parts are likely to be stationary.

only a component specification for the factors, as the testing problem with a factor structure in the innovation is well understood. However, we pay attention to the role of deterministic trends and regressors entering either in a component specification or as part of the innovations in the factor specification. Allowing for deterministic trends is important in many applications of unit-root tests. However, specifying individual-specific trends greatly reduces the power of panel unit root tests: Instead of $T^{-1}n^{-1/2}$ neighbourhoods of the unit root, testing with idiosyncratic trends leads to power only in $T^{-1}n^{-1/4}$ neighbourhoods, see Moon, Perron, and Phillips (2007). Depending on the application and the hypothesis of interest, specifying trends in the factor equation can be an attractive alternative that suits the data.

With observed factors, the deterministic trend in the factor equation would lead to sizeable power gains, see Hallin, Van den Akker, and Werker (2011, 2016), who obtain power in $T^{-3/2}$ neighbourhoods of the unit root. We show that even with unobserved factors the gains due to nonzero-mean innovations in the factor equation can be fully realized. First, we show that simple cross-section averages can successfully estimate factors that have deterministic trends. Moreover, we do not only consider a single factor, but discuss what exactly is estimated by cross-section averages in the presence of multiple potentially cointegrated factors. In particular, we show that if the factors are estimated using cross-sectional averages, the asymptotic size and power of our unit root tests are unaffected when additional stationary factors are present. Also, enlarging our null hypothesis to allow for more than one common stochastic trend does not lead to any size distortions and, as expected, has a positive effect on the attainable local asymptotic powers.

Often, the most relevant question will be whether the observations are stationary or not, rather than what the exact number of common stochastic trends will be. We propose tests that nearly attain the power envelope for the unit root testing problem and have correct size also if there are multiple stochastic trends present. These tests have more power than the cointegration tests previously considered, and, importantly, they do not require estimation of the number of factors.[6]

The chapter is structured as follows: Section 3.2 introduces the basic setup and issues common to all specifications. Section 3.3 discusses alternative unit-root tests in the most frequently considered model with zero-mean innovations and Section 3.4 considers deterministic trends in the factor equation. Section 3.5 presents simulation results for all specifications. Proofs are relegated to Section 3.A.

3.2 Setup

Throughout this chapter, we consider the factor model of Bai and Ng (2004),

$$Y_{it} = \sum_{k=1}^{r} \lambda_{ki} F_{kt} + E_{it}, \tag{3.1}$$

$$E_{it} = \rho_E E_{i,t-1} + \eta_{it}, \tag{3.2}$$

$$F_{kt} = \rho_k F_{k,t-1} + \mu_k + f_{kt}, \tag{3.3}$$

with λ_{ki} the loading of the (unobserved) factor $\{F_{kt}\}$ on panel unit i, and $r \in \mathbb{N}$ being the fixed and known number of factors. The $\{\eta_{it}\}$ and $\{f_{kt}\}$ are zero-mean idiosyncratic and common shocks, respectively. For both the factors and idiosyncratic parts we assume zero starting values. For ease of notation suppose that $1 \geq \rho_1 \geq \rho_2 \geq \cdots \geq \rho_r \geq 0$. Our starting point is the most commonly found framework where $\mu_k = 0$, i.e., the factor innovations have zero mean. We consider the null hypothesis $\rho_1 = 1$, with alternatives $\rho_1 = 1 + \frac{h}{T^\nu}$, and $\nu \in \{1, 3/2\}$ depending on whether the factors contain a deterministic trend, i.e., on whether $\mu_1 \neq 0$. In case no trend is specified, $\nu = 1$ implies alternatives contiguous to the null hypothesis, see Proposition 3.3.1. With a possible deterministic trend we have power even at $\nu = 3/2$, see Theorem 3.4.1. Throughout, we consider large panels in the sense that both n and T go to infinity. This is standard in this literature. We require n-asymptotics in order

6 These are typically selected based on information criteria, see, for example Bai and Ng (2002). However, it is known that in finite samples these often select the maximum number of factors and can thus be of limited use in practice.

to successfully estimate the unobserved common factors. Thus $n \to \infty$ is assumed although n does not appear in the local alternatives – after all one cannot do better than the observed-factor benchmark.

Collect the panel units in the $T \times n$ matrices Y, E, and η. Also, write Y_i, E_i, and η_i for their ith columns, respectively. Introduce the $n \times r$ matrix Λ to contain the factor loadings $\lambda_{i,k}$ and write $\lambda_k = (\lambda_{1,k}, \dots, \lambda_{n,k})'$. Let $F_k = \{F_{k,t}\}_{t=1}^T$ and define f_k analogously. Finally, collect the factors in $F = (F_1, \dots, F_r)$ and $f = (f_1, \dots, f_r)$. Also, denote by F_{-1} and $\hat{F}_{-1}$, $(F_0, \dots, F_{T-1})'$ and $(\hat{F}_0, \dots, \hat{F}_{T-1})'$, respectively. Note that with this notation, we have $Y = F\Lambda' + E$. The $T \times T$ covariance matrices of f, η_i, are denoted by Σ_f and $\Sigma_{\eta,i}$, respectively, with long-run variances ω_f^2, $\omega_{\eta,i}^2$ and autocovariance functions γ_k and γ_i, respectively.

For a matrix A, let $\|A\|_F$ denote its Frobenius norm and $\|A\|_{\text{spec}}$ its spectral norm. By $\to$ we denote convergence, of real-valued sequences, $\Rightarrow$ denotes convergence in distribution, and $(n, T \to \infty)$ refers to n and T going to infinity jointly as in Phillips and Moon (1999).

For now, we make no assumptions on the idiosyncratic parts η, but instead assume the existence of certain estimates of $F_{1,t}$ that are available under various conditions on η. We first establish a sufficient condition for the estimated factors to yield adaptive likelihood-ratio tests. Later, we show that several factor estimates available in the literature satisfy this condition. To enable us to write out likelihood ratios we impose the following assumption.

Assumption 3.1 *The factor innovations f_k are a stationary Gaussian time-series with mean zero and variance one, independent of the idiosyncratic parts η and satisfying $\sum_{m=0}^{\infty}(|m|+1)\gamma_k(m) < \infty$.*

The zero mean assumption will be relaxed in the next section and the unit variance is necessary for identification; it allows us to estimate the factor up to its sign. The normality assumption could probably be relaxed at the expense of slightly more complicated likelihood ratios (as in Jansson (2008)), but is in line with the literature on panel unit-root tests even in much simpler settings. We also impose the standard assumption of strong factors.

Assumption 3.2 *The factors are strong, i.e., there exists a positive definite matrix* Ψ_Λ *such that* $\lim_{n\to\infty}\frac{1}{n}\Lambda'\Lambda \to \Psi_\Lambda$.

3.3 Likelihood-Based Tests and Varying Factor Estimates

The goal of this section is to illustrate some features of the likelihood ratio in the experiment where $\mu_k = 0$. In a simple setting, we formally show that panel unit-root tests that are robust to cross-sectional cointegration, can (only) have power in T^{-1} neighbourhoods of the unit-root. This is one motivation for considering an alternative framework in Section 3.4.

The case of $\mu_k = 0$ is the case most commonly encountered in the literature and originally proposed by Bai and Ng (2004). In case of a single factor, they show that after estimating the factor by principal components these can be used in ADF tests as if the factor was observed. We start by generalizing these results to likelihood-based test statistics and various factor estimates.

In the spirit of applying existing unit-root tests to estimated factors, we first consider the experiment where both Y and F are observed and recall the local likelihood ratio. We consider local alternatives of the form

$$\rho_{1,T} = 1 + \frac{h}{T}, h \in \mathbb{R}_- \tag{3.4}$$

and rephrase our hypotheses as $H_0 : h = 0$ vs. $H_A : h < 0$. Let $\tilde{P}_{h,n,T}$ be the joint law of Y and F under (3.1)–(3.4), write $P_{h,n,T}$ for the marginal law of Y and $\tilde{P}_{h,T}$ for the marginal law of F. Like the existing results for ADF tests, we restrict ourselves to the single factor case in this section. When multiple factors are present, typically cointegration-based methods are employed that are beyond the scope of this chapter. However, we show in Sections 3.4.2 and 3.4.3 that such an approach is not necessary for factors with deterministic trends as the distribution of our proposed test statistics does not change when additional stationary factors are present.

Assumption 3.3 *We have a single factor* ($r = 1$) *with unit variance and* $\mu_1 = 0$.

Writing out a local likelihood ratio in the experiment where both Y and F are observed yields, under Assumptions 3.1 and 3.2,

$$\log \frac{\mathrm{d}\tilde{P}_{h,n,T}}{\mathrm{d}\tilde{P}_{0,n,T}} = -hS_T - \frac{1}{2}h^2 H_T, \tag{3.5}$$

with

$$S_T = \frac{1}{T}\Delta F' \Sigma_f^{-1} F_{-1} \text{ and } H_T = \frac{1}{T^2} F_{-1}' \Sigma_f^{-1} F_{-1}. \tag{3.6}$$

As expected, if the factor is observed there is no additional information in Y, i.e., this is the same likelihood ratio as for only observing F and F constitutes a sufficient statistic in this experiment.

$$\frac{\mathrm{d}\tilde{P}_{h,n,T}}{\mathrm{d}\tilde{P}_{0,n,T}} = \frac{\mathrm{d}\tilde{P}_{h,T}}{\mathrm{d}\tilde{P}_{0,T}}.$$

This allows us to use the well-known results for the first-order autoregression, i.e.,

$$(S_T, H_T) \overset{\tilde{P}_{0,T}}{\Rightarrow} \left(\int_0^1 W(r)\,\mathrm{d}W(r), \int_0^1 W^2(r)\,\mathrm{d}r \right) =: (S, H), \tag{3.7}$$

where W is a standard Brownian Motion, see, e.g., Elliott, Rothenberg, and Stock (1996). It turns out that given sufficiently good estimates $\hat{F}_{1,t}$ of the factor, replacing $F_{1,t}$ by $\hat{F}_{1,t}$ in (3.6) leads to an asymptotically negligible difference between (S_T, H_T) and the estimated counterparts. We now show under which conditions S_T and H_T can be approximated based on observing Y alone. This enables likelihood-based inference for the unobserved factor. We first establish a sufficient condition for the estimated factor to yield adaptive likelihood-ratio tests. In Sections 3.3.1–3.3.4 we show that several factor estimates available in the literature satisfy this condition.

Proposition 3.3.1 *Let $\hat{F}_{1,t}$ satisfy*

$$MSE_T := \frac{1}{T}\sum_{t=1}^{T} (\hat{F}_{1,t} - R_{n,T}F_{1,t})^2 = o_{\tilde{P}_{0,n,T}}(1) \tag{3.8}$$

for some $R_{n,T}$ satisfying $\|R_{n,T}\|_F = O_{\tilde{P}_{0,n,T}}(1)$ and $\left\|R_{n,T}^{-1}\right\|_F = O_{\tilde{P}_{0,n,T}}(1)$, and assume $\{\hat{F}_T^2/T\}_{T\in\mathbb{N}}$ is uniformly integrable. Let $\hat{\omega}_f^2$ be a consistent estimator *of the long-run variance of $\Delta\hat{F}$ and define $\hat{S}_T$ and $\hat{H}_T$ analogous to* (3.6) *by*

$$\hat{S}_T = \frac{1}{T\hat{\omega}_f^2}\sum_{t=2}^{T}\Delta\hat{F}_{1,t}\hat{F}_{1,t-1} - \frac{\hat{\delta}_f}{\hat{\omega}_f^2} \text{ and } \hat{H}_T = \frac{1}{T^2\hat{\omega}_f^2}\sum_{t=2}^{T}\hat{F}_{1,t-1}^2, \tag{3.9}$$

where $\hat{\delta}_f = \frac{\hat{\omega}_f^2 - 1}{2}$. Then, under Assumptions 3.1–3.3, $(\hat{S}_T, \hat{H}_T) = (S_T, H_T) + o_{\tilde{P}_{h,n,T}}(1)$ as $(n, T \to \infty)$.

The proof is provided in Section 3.A.1. Note that as we have normalized the factor innovations to have unit variance and $\hat{\delta}_f$ approximates the one-sided long-run variance $\delta = \frac{\omega_f^2 - \sigma_f^2}{2}$.

Remark 3.3.1 *We use joint convergence as in Phillips and Moon (1999), where both n and T go to infinity together without any particular relation between the two; in this section we also do not require any restrictions on the rates. These joint limits also imply sequential ones, but restricting ourselves to sequential results would in our case entail potentially misleading results: When first $n \to \infty$ then $T \to \infty$, we would conclude that any test statistic is adaptive, as long as we plug in a factor estimate that is consistent for large n. Therefore, the additional requirements of joint asymptotics are essential in this problem.*

Remark 3.3.2 *Bai and Ng (2002) remark that for estimating the number of factors, it is the 'average convergence' of the factor estimates as in* (3.8) *that is needed, rather than uniform convergence. Proposition 3.3.1 implies that, for a given factor, the same remark applies to judging its stationarity.*

Certainly, not observing the factor cannot make the testing problem easier. Also, note that neither does observing Y in addition to F, since the likelihood ratios are the same. So, not surprisingly, the power envelope for the univariate testing problem derived in Elliott, Rothenberg, and Stock (1996) is an upper bound for testing an unobserved factor for stationarity. It is also attainable in the same way as in the univariate case. Firstly, it is attainable pointwise: in the model with an observed factor, the Neyman-Pearson Lemma applies (see Elliott, Rothenberg, and Stock (1996)), so that likelihood-ratio tests are most powerful against specific alternatives. Therefore, against specific alternatives, we can construct tests with the same asymptotic power without observed factors.

It is well-known that no UMP tests exist, see e.g. Jeganathan (1995). However, in the univariate setup, 'nearly efficient' tests in the sense that their power is very close to the power envelope, have been suggested. Among these are the p_T tests from Elliott, Rothenberg, and Stock (1996), but also, for example, the likelihood-ratio tests Jansson and Nielsen (2012). Also these tests can be adapted based on (3.8).

Different factor estimates have been shown to satisfy Condition (3.8) under a variety of assumptions. In particular, Bai (2004) shows it is satisfied for the principal component estimator, Bai and Ng (2004) show it is satisfied for principal components estimated in differences and Choi (2017) shows it is satisfied for a generalized principal component estimator. We proceed to recall these results and demonstrate their compatibility with our assumptions.[7]

3.3.1 Level Principal Components

To successfully estimate the factor using principal components estimated in levels,[8] we need to assume that $|\rho_E| < 1$, i.e., we have stationary idiosyncratic parts; see Onatski and Wang (2019) for a discussion on the problems of level principal components in combination with integrated idiosyncratic parts. Bai (2004) discusses in detail the properties of level principal component estimators in large panels with nonstationary factors. Lemma 1 in that paper states that $MSE_T = O_{P_0}\left(\frac{1}{\min\{n,T^2\}}\right)$, implying (3.8). In Bai (2004) these are shown under four assumptions, A–D. Assumption A is indeed satisfied in our setup thanks to Assumption 3.1: The first part (moments of the factor innovation) is trivially satisfied due to the normality assumption. Part two just states the convergence of H_T which is certainly satisfied under Assumption 3.1, with

7 In the below we focus on verifying (3.8) for the existing factor estimates. To formally apply Proposition 3.3.1 to these estimates one should also verify the uniform integrability requirement.

8 That is, based on the estimated covariance matrix of Y rather than ΔY.

$\Omega_{uu} = 1$. Similarly, the law of iterated logarithm is standard here. Part four is also satisfied easily, since $F_{1,0} = 0$. Assumption B–D in Bai (2004) need to be imposed; they concern only factor loadings and idiosyncratic errors.

3.3.2 Generalized Principal Components

Choi (2017) introduces generalized principal components for nonstationary factors. These generalized principal components take the heteroskedasticity of the idiosyncratic parts into account and can thus be more efficient. Our condition (3.8) is verified in Lemma A.1 in the appendix to Choi (2017); it states the same as the result in Bai (2004). The assumptions on both the idiosyncratic errors and the f_t are high level but examples are given. Assumption 1 concerns the idiosyncratic innovations and needs to be imposed, as do Assumptions 2(i) to 2(iii), which concern the factor loadings. The functional central limit theorem in Assumption 2(iv) holds with $\Phi_F = 1$ based on our Assumption 3.1. Assumption 3(i) and 3(ii)(a) also concern the idiosyncratic parts.

3.3.3 Difference Principal Components

Difference principal components proceed by estimating ΔF based on the estimated covariance matrix of ΔY and taking cumulative sums. Our condition in (3.8) is stated in levels, making it hard to verify for factor estimates that are based on differences. However, the only such factor estimate we are aware of are the difference principal components suggested in Bai and Ng (2004) and the two conclusions of Proposition 3.3.1 regarding $\hat{S}_T$ and $\hat{H}_T$ are verified therein as Lemma B.(iv) and Lemma B.(ii), respectively. Therefore, our conclusions concerning power envelopes and point-optimal tests apply also to difference principal components, under the assumptions of Bai and Ng (2004).

3.3.4 The Kalman smoother

Recently, methods that use the time-series properties of the data in addition to postulated cross-sectional correlation structure have been developed, see, for example Doz, Giannone, and Reichlin (2011) and Poncela, Ruiz, and Miranda

(2020). As we are interested in the unit-root testing problem, and thus know the time-series properties of the factor under the hypothesis, the use of these methods is particularly appropriate. A popular approach for this is to employ the 'Kalman smoother' factor estimate, which is given by $\tilde{F}^{KS} = \mathrm{E}[F_1|Y]$, i.e., the expectation of the factor given all observations $Y_{it}, i = 1, \dots n, t = 1, \dots T$.

As common in this literature, we now also require Gaussianity of the factor innovations. Moreover, we focus on the case where the $nT \times nT$ covariance matrix of the innovations η can be written as a Kronecker-product, i.e., each panel unit may have a different (long-run) variance, as long as the correlation structure is the same across panel units. We focus on the case $\rho_E = 0$; the theory for nonstationary idiosyncratic parts could be developed analogously. Moreover, we impose the hypothesis of an integrated factor. Recall that (3.8) only concerns the behavior under the null hypothesis; convergence to zero under local alternatives is then implied by contiguity.

Assumption 3.4 *The idiosyncratic innovations η are normally distributed with mean zero and covariance matrix $\Omega_\eta \otimes \Sigma_\eta$, where Ω_η is a diagonal matrix whose diagonal entries are bounded and bounded away from zero and Σ_η is a covariance matrix of a stationary time series with summable autocorrelations. The factor innovations f are normally distributed with mean zero and covariance matrix Σ_f with spectral density bounded and bounded away from zero. Finally, $\rho_E = 0$.*

Under Assumption 3.4, we can rewrite the desired conditional expectation as

$$\tilde{F}^{KS} = \mathrm{E}[F_1|Y] = \Sigma_{F,Y}\Sigma_Y^{-1}\operatorname{vec}(Y), \tag{3.10}$$

where $\Sigma_{F,Y}$ is the $T \times nT$ 'cross-covariance matrix' of F_1 and Y, that is $(\Sigma_{F,Y})_{t,(i-1)T+s} = \mathrm{Cov}(F_{1,t}, Y_{is})$. To our knowledge, the Kalman smoother has not been studied in the presence of nonstationary factors. However, in the case of equal correlation structure among the idiosyncratic errors, it is relatively straightforward to obtain condition (3.8). The key insight is that the inverse $nT \times nT$ covariance matrix of Y does not have to be computed, as for the Kalman smoother we only need the inverse premultiplied by $\Sigma_{F,Y}$. Let $\tilde{A}$ denote a cumulative sum operator, a $T \times T$ matrix with ones on and below the diagonal and zeros on and above, so that $F = \tilde{A}\Delta F$. More explicitly, we have[9]

$$\tilde{F}^{KS} = (\lambda' \otimes \tilde{A}\Sigma_f \tilde{A}')(\lambda\lambda' \otimes \tilde{A}\Sigma_f \tilde{A}' + \Omega_\eta \otimes \Sigma_\eta)^{-1} \operatorname{vec}(Y) \tag{3.11}$$

$$= \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f \tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda} \right)^{-1} Y\Omega_\eta^{-1} \frac{\lambda}{\lambda'\Omega_\eta^{-1}\lambda}. \tag{3.12}$$

This insight has three main consequences. First, it greatly reduces the computational burden when implementing the Kalman smoother. Often, the Kalman smoother is employed in an iterative procedure to jointly estimate other nuisance parameters. However, repeatedly computing the $nT \times nT$ inverse is computationally prohibitive. (3.12), on the other hand, only involves a $T \times T$ inverse. Second, the formulation disentangles the cross-sectional and time-series manipulation of the observations, aiding interpretation of the procedure. Indeed, with i.i.d. innovations and homoskedasticity, $\tilde{F}^{KS} = \left(I_T + \frac{(\tilde{A}\tilde{A}')^{-1}}{\lambda'\lambda} \right)^{-1} \tilde{F}^{OLS}$, where $\tilde{F}^{OLS}$ is the estimate of a least squares regression of Y on λ. As expected, the Kalman smoother leaves intact the cross-sectional correlation structure of principal components/OLS, but re-weighs the estimates in the time direction. In the more general case, we can similarly relate the Kalman smoother estimate to a GLS estimate, i.e.,

$$\tilde{F}^{KS} = \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f \tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda} \right)^{-1} \tilde{F}^{GLS}.$$

Note that with unobserved factor loadings, the GLS estimate corresponds to the Generalized Principal Components estimate of Section 3.3.2. This insight also allows a relatively straightforward proof of the following proposition, implying that the Kalman smoother can be used for panel unit-root tests.

Proposition 3.3.2 *Let $(\hat{\lambda}, \Delta\hat{F})$ be GPC estimates of (λ, F) and let $\hat{\Sigma}_f$, $\hat{\Sigma}_\eta$, and $\hat{\Omega}_\eta$ be consistent estimators of Σ_f, Σ_η, and Ω_η, respectively. Let $\hat{F}^{KS} = \left(I_T + \hat{\Sigma}_\eta \frac{(\tilde{A}\hat{\Sigma}_f \tilde{A}')^{-1}}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}} \right)^{-1} Y\hat{\Omega}_\eta^{-1} \frac{\hat{\lambda}}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}}$. Under Assumptions 3.1–3.4, we have*

$$MSE_T = \frac{1}{T}\sum_{t=1}^{T} (\hat{F}_t^{KS} - F_{1,t})^2 = o_{\tilde{P}_{0,n,T}}(1). \tag{3.13}$$

[9] For a proof of this relation see Section 3.A.1.

Remark 3.3.3 *Although we have imposed Assumption 3.4, joint Gaussianity of the factor and idiosyncratic innovations is only necessary for* $\hat{F}_t^{KS}$ *to be the Kalman smoother, i.e., a conditional expectation. However,* (3.13) *can be shown under weaker conditions. In particular, the proof in Section 3.A.1 uses only the MSE condition of the GPC estimator as well as boundedness of the spectral norms of the estimated covariance matrices.*

3.4 Factors with Deterministic Trends

Having formally shown that in the standard setting cross-sectional cointegration robust unit-root tests cannot have a faster convergence rate than tests based on a single time series, we now consider an alternative setup that allows for more powerful tests: we consider factor innovations with non-zero mean, i.e., $\mu_k \neq 0$. Hallin, Van den Akker, and Werker (2011) study such a univariate autoregressive model with deterministic trends under the hypothesis, treating the trend as a nuisance parameter. For Gaussian innovations, Equation (14) in Hallin, Van den Akker, and Werker (2011) shows that the asymptotically optimal test for the unit root hypothesis for an observed factor F_1 is based on

$$\mathcal{T}(\Delta F_{1,T}) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{t}{T+1} - \frac{1}{2} \right) \Delta F_{1,t}. \tag{3.14}$$

For contiguous alternatives of the form

$$(\rho_{1,T}, \mu_{1,T}) = \left(1 + \frac{h_1}{T^{3/2}}, \mu_1 + \frac{h_2}{\sqrt{T}} \right), \tag{3.15}$$

where $\rho_{1,T}$ and $\mu_{1,T}$ refer to the autoregressive parameter and the trend, respectively, they show that the model is locally asymptotically normal (LAN).

Note that the presence of the trend makes it easier to identify ρ in the sense that now the local alternatives are closer to the hypothesis. Once again, we show that observing the factors 'does not make a difference' (asymptotically) and suggest how to estimate them.

The literature on estimating factors in the presence of both a stochastic and a deterministic trend is very limited. The papers by Bai (2004), Bai and Ng (2004), and Choi (2017) do not allow for deterministic trends. In an unpublished working paper, Maciejowska (2010) shows that indeed the solutions of Bai (2004) do not work in the presence of a deterministic trend but that the method of principal components can still be used to get consistent estimates. However, it turns out that much simpler estimates are sufficient for optimal unit-root tests. In particular, we show that replacing $\Delta F_{1,T}$ with scaled cross-section averages in (3.14) leads to asymptotically uniformly most powerful unit-root tests.

In Section 3.4.1, we again consider a single factor and show how to implement an asymptotically UMP unit-root test. In Section 3.4.2, we allow for multiple factors, with at most one of them being nonstationary. We show that the asymptotic distributions of the test we have proposed in Section 3.4.1 is not affected by additional stationary factors. Finally, in Section 3.4.3, we allow for multiple potentially nonstationary factors. This implies that the observations are nonstationary, so we do not want to reject the panel unit-root hypothesis. We show that, also in the presence of multiple nonstationary factors, the asymptotic size of the proposed test does not exceed its nominal level.

3.4.1 A Single Factor with Trend

We now consider the case where $\mu_1 \neq 0$. Once again, we are interested in the unit-root hypothesis, i.e., testing $H_0 : \rho = 1$, based on observing Y. Under the hypothesis of a unit-root, this corresponds to the presence of both a deterministic and a stochastic trend, whereas under the alternative neither is present.

We allow for both stationary and integrated idiosyncratic parts. Moreover, we do not require Gaussianity at this stage, relaxing Assumption 3.1. At the same time, we impose restrictions on the idiosyncratic parts that allows us to estimate the factors without relying on external estimates.

Assumption 3.5 *The idiosyncratic parts $\{E_i\}$ are cross-sectionally independent. They also have mean zero and start at zero, i.e., $E_{i,0} = 0$, $\mathrm{E}E_{i,t} = 0$, we have $\mathrm{E}[(\Delta E_{i,t})^2] = \sigma_e^2 < \infty$ and one of the following holds:*

1. *The idiosyncratic parts are covariance stationary, or*

2. *The first differences of the idiosyncratic parts, $\Delta E_{i,t}$, are covariance stationary with summable autocovariances, i.e., we have*

$$\sum_{s=0}^{\infty} |\mathrm{E}[\Delta E_{i,t}\Delta E_{i,t-s}]| < \infty.$$

Assumption 3.6 *The factor innovations are covariance stationary and ergodic with $\mathrm{E}f_{1,t} = 0$ and $\mathrm{E}f_{1,t}^2 = 1$ and independent of the idiosyncratic innovations $E_{\cdot,T}$.*

The assumption of summable autocovariances in Assumption 3.5(2) implies that the variance of E_{iT} is of the same order as that of a random walk. The assumption is satisfied, for example, for near epoch dependent processes on a mixing process, as demonstrated in Theorem 17.7 in Davidson (1994) or in particular a stationary ARMA process.

Under the alternatives (3.15), let $\tilde{P}_{h_1,h_2,n,T}$ be the joint law of Y and F_1, write $\tilde{P}_{h_1,h_2,T}$ for the marginal law of F_1 and write $P_{h_1,h_2,n,T}$ for the marginal law of Y. As in Section 3.3, the likelihood ratio of the joint law of Y and F_1 equals the likelihood ratio of the marginal law of F_1 alone, i.e.,

$$\frac{\mathrm{d}\tilde{P}_{h_1,h_2,n,T}}{\mathrm{d}\tilde{P}_{\tilde{h}_1,\tilde{h}_2,n,T}} = \frac{\mathrm{d}\tilde{P}_{h_1,h_2,T}}{\mathrm{d}\tilde{P}_{\tilde{h}_1,\tilde{h}_2,T}}, \tag{3.16}$$

for all $h_1, h_2, \tilde{h}_1, \tilde{h}_2$. This is due to the independence between idiosyncratic and factor innovations from Assumption 3.6 ensuring that the distribution of $Y_{i,t}$ conditional on F_t and the past does not depend on the local parameters. Because of (3.16), the asymptotically optimal test for the panel model with observed factors is based on (3.14) as well. In fact, the same LAN result as in Proposition 2.1 in Hallin, Van den Akker, and Werker (2011) holds, which, analogous to Lemma 3.A.1 implies mutual contiguity of local alternatives around the unit root; see also Remark 2.1 in Hallin, Van den Akker, and Werker (2011). In our setting, where the factors are unobserved, we will show that $\mathcal{T}(\Delta F_{1,T})$ can be approximated with Y-measurable estimates up to an $o_p(1)$ term, implying that the more complicated unobserved factor model is adaptive.

Lemma 3.4.1 *Write* $\overline{\Delta Y}_{\cdot,t} = \frac{1}{n}\sum_{i=1}^{n} \Delta Y_{i,t}$ *and let* $\Delta\hat{F}_{1,t} = \overline{\Delta Y}_{\cdot,t}/s(\overline{\Delta Y}_{\cdot,t})$*, where* $s^2(\cdot)$ *denotes the sample variance. Then, under Assumptions 3.2, 3.5 and 3.6,* $\mathcal{T}(\Delta F_{1,T}) - \mathcal{T}(\Delta\hat{F}_{1,T}) = o_{\tilde{P}_{h_1,h_2,n,T}}(1)$ *as* $(n,T)\to\infty$ *jointly.*

The proof is provided in Section 3.A.2. Lemma 3.4.1 allows us to use many of the results in Hallin, Van den Akker, and Werker (2011) as summarized in Proposition 3.4.1. In the spirit of the existing panel unit-root tests, this chapter does not consider rank-based tests. Instead, we rely on their statistics that are valid and optimal under normality.[10] Therefore, we strengthen Assumption 3.6:

Assumption 3.7 *The factor innovations are independent of the idiosyncratic ones and satisfy* $\mu_1 \neq 0$ *as well as* $f_{1,t} \overset{iid}{\sim} N(0,1)$.

Theorem 3.4.1 *Let* $\tau_{n,T}$ *be the test that rejects iff* $\sqrt{12}\mathcal{T}(\Delta\hat{F}_{1,T}) \geq \Phi_{1-\alpha}$*, where* $\Phi_{1-\alpha}$ *is the* $1-\alpha$ *percentile of the standard normal distribution. Under Assumptions 3.2, 3.5 and 3.7, this test is an asymptotically uniformly most powerful level-*α *test for a unit-root in the unobserved factors: For any asymptotic level* α *test* $t_{n,T}$ *let* $\pi_{t_{n,T}}(h_1) = P_{h_1,0,n,T}[t_{n,T}$ *rejects*$]$ *be the power function. Then*

$$\lim_{(n,T)\to\infty} \pi_{t_{n,T}}(h_1) \leq \lim_{(n,T)\to\infty} \pi_{\tau_{n,T}}(h_1) = 1 - \Phi\left(\Phi_{1-\alpha} - \frac{h_1\mu_1}{\sqrt{12}}\right),$$

for all $h_1 \geq 0$*, where* $\Phi(\cdot)$ *refers to the standard normal CDF.*

Proof These results follow immediate from those in Hallin, Van den Akker, and Werker (2011), noting that any power function we can get with tests based on Y only we can also get based on (Y,F). Thus, if a test is optimal among on (Y,F) based tests, and it is just based on Y, it must also be optimal among those tests. ∎

[10] The rank-based versions of these test statistics require estimation of weights that depend on the levels of the factor of interest. While these estimates can be obtained in case the idiosyncratic parts are stationary, this is problematic if the idiosyncratic parts are integrated.

3.4.2 Additional Stationary Factors

We now extend the model from Section 3.4.1 to allow for multiple factors. Here, many approaches are possible. Our choice of model and hypothesis are motivated by our ultimate goal of conducting optimal unit-root tests for the observations $Y_{i,t}$ under cross-unit cointegration. The observations have a unit root if and only if *at least one* of the factors is nonstationary. In Section 3.4.2, our null hypothesis is that exactly one of the factors is integrated, with the alternative being all of them are stationary.

Since different factors usually represent completely different economic series, it makes little sense to impose any homogeneity on intercepts or autoregressive parameters. Our model therefore reads

$$Y_{i,t} = \sum_{k=1}^{r} \lambda_k F_{k,t} + E_{i,t}, i = 1, \dots, n, t = 1, \dots, T \tag{3.17}$$

$$F_{k,t} = \rho_{k,T} F_{k,t-1} + f_{k,t} + \mu_{k,T}, F_{k,0} = 0, k = 1, \dots, r, t = 1, \dots, T. \tag{3.18}$$

For the first factor, we use the same local alternatives as in Section 3.4.1, i.e., we impose (3.15). Under these alternatives, we write, just as in Section 3.4.1, $\tilde{P}_{h_1,h_2,n,T}$ for the joint law of Y and F_T, $\tilde{P}_{h_1,h_2,T}$ for the marginal law of $F_{1,T}$ and write $P_{h_1,h_2,n,T}$ for the marginal law of Y. Our null hypothesis is $\rho_{1,T} = 1$, treating the other autoregressive parameters as nuisance parameters under the assumption that the other factors are stationary, see Assumption 3.9 below. Since (3.16) also holds in this setting, once again an optimal unit-root test for the model in (3.17) and (3.18) with observed factors would be based on (3.14). Again, it turns out we can draw on Hallin, Van den Akker, and Werker (2011), by showing that even in the presence of other factors, we can estimate $F_{1,T}$ well enough to approximate the test statistic in (3.14) up to an $o_p(1)$ term.

The following assumption allows us to approximate the test statistic under the null hypothesis. We maintain Assumption 3.2 on the loadings of the first factor as well as Assumption 3.5 on the idiosyncratic parts from Section 3.4.1. Assumption 3.8 is completely analogous to Assumption 3.6.

Assumption 3.8 *The factor innovations are covariance stationary and ergodic with* $\mathrm{E}f_{k,t} = 0$ *and* $\mathrm{E}f_{k,t}^2 = 1$ *and independent of the idiosyncratic innovations* $E_{\cdot,T}$ *and the other factors.*

Assumption 3.9 *The number of factors is fixed at* $r < \infty$. *At most one of them has a unit root, i.e.,* $\rho_{k,T} = \rho_k < 1, k = 2, \ldots, r$. *The loadings of the stationary factors are bounded, i..e,* $\bar{\lambda}_{k,\cdot} < M$ *for some* $M \in \mathbb{R}$, $k = 2, \ldots, r, n \in \mathbb{N}$.

The following lemma, proved in Section 3.A.2, shows that the adaptivity result is robust to the presence of additional stationary factors.

Lemma 3.4.2 *Let* $\hat{\bar{\lambda}}_1^2$ *be a consistent estimator of the long run variance of* $\overline{\Delta Y}_{\cdot,t}$ *and let* $\Delta\hat{F}_{1,t} = \overline{\Delta Y}_{\cdot,t}/\hat{\bar{\lambda}}_1$. *Then, under Assumptions 3.2, 3.5, 3.8 and 3.9,* $\mathcal{T}(\Delta F_{1,T}) - \mathcal{T}(\Delta\hat{F}_{1,T}) = o_{\tilde{P}_{0,0,n,T}}(1)$ *as* $(n, T) \to \infty$ *jointly.*

Remark 3.4.1 *Once more we can use cross-section averages, however, now we have to take more care to scale them correctly. We have, under the null hypothesis,*

$$\begin{aligned}\Delta\bar{Y}_{\cdot,t} =& \sum_{k=1}^{r} \bar{\lambda}_{k,\cdot}\Delta F_{k,t} + \Delta\bar{E}_{\cdot,t} \\ =& \bar{\lambda}_{1,\cdot}(u_{1,t} + \mu_1) + \sum_{k=2}^{r} \bar{\lambda}_{k,\cdot}\left(\mu_k\rho_k^{t-1} + f_{k,t} + (1 - \frac{1}{\rho_k})\sum_{s=1}^{t-1}\rho_k^{t-s}u_{k,s}\right),\end{aligned}$$

so that the stationary factors indeed contribute something to the variance of the differenced cross-sectional averages. Therefore, we cannot use the method from Section 3.4.1 to scale our estimator. However, since the differenced stationary factors are over-differenced, they do not impact the long-run variance, so that scaling by the long-run variance correctly scales the nonstationary factor. Since the model in this section generalizes the single-factor model of Section 3.4.1, the long-run variance approach also works with a single factor. Simulations in Section 3.5 compare the performance of the two approaches in finite samples.

Lemma 3.4.2 implies that we can also extend the results of Proposition 3.4.1 to the case of multiple factors. In particular, we get optimal unit-root tests for the observations in case the idiosyncratic parts are stationary and we have at most one nonstationary factor. As before, we need to assume normality of the factor innovations for the optimality result, i.e., impose Assumption 3.7.

Theorem 3.4.2 *Consider a cross-sectionally cointegrated panel generated by* (3.17) *satisfying Assumptions 3.2, 3.5 and 3.7–3.9. Let* $\Delta\hat{F}_{1,t} = \overline{\Delta Y}_{\cdot,t}/\hat{\bar{\lambda}}_1$*. For the problem of testing* $H_0 : Y_{i,t}$ *are integrated for most* i *against* H_A : *all* $Y_{i,t}$ *are stationary, the test* $\tilde{\tau}_{n,T}$ *that rejects iff* $\sqrt{12}\mathcal{T}((\Delta\hat{F}_{1,t})_{t=1}^{T}) \geq \Phi_{1-\alpha}$ *is asymptotically uniformly most powerful.*

Remark 3.4.2 *The statement 'for most i' relates to the problem that some units might not load on the nonstationary factor. Therefore, the precise null hypothesis would depend on which assumptions are made concerning the number of* $\lambda_{1,i}$ *that can be zero. For example, if we assume that, for some* $\epsilon > 0$, $|\lambda_{1,i}| > \epsilon$ *for a share* c *of the panel units* i*, the hypothesis would be that at least a share* c *of the panel units have a unit root.*

3.4.3 Multiple Common Trends

In the previous section we have derived an optimal test for testing one common trend against none. What we really want, however, is a test for 'at least one' trend against no trend, i.e., a unit root test for the observations without the restriction that there is at most one common trend.

Enlarging the null hypothesis in this way complicates optimal inference significantly, since we can no longer immediately rely on the results for optimal univariate unit root tests in Hallin, Van den Akker, and Werker (2011). Even if we could perfectly estimate each nonstationary factor, it is unclear how to conduct optimal inference based on those multiple factors. However, we can still demonstrate that the test $\tilde{\tau}_{n,T}$ proposed in Theorem 3.4.2 is robust to multiple trends, in the sense that that it still has the correct size if more than one factor is nonstationary. The assumptions Assumptions 3.10–3.13 adapt Assumptions 3.2 and 3.7–3.9. Once again we maintain Assumption 3.5.

Assumption 3.10 *The number of factors is fixed at* $r < \infty$*. The first* $r_1 \leq r$ *of them have a unit root.*

Assumption 3.11 *The factor loadings are deterministic and satisfy* $\sum_{k=1}^{r_1} \bar{\lambda}_k^2 = \sigma^2 \neq 0$*, where* $\bar{\lambda}_k = \lim_{n\to\infty} \bar{\lambda}_{k,\cdot} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \lambda_{k,i}$*.*

Assumption 3.12 *The factor innovations are covariance stationary and ergodic with* $\mathrm{E} f_{k,t} = 0$ *and* $\mathrm{E} f_{k,t}^2 = 1$ *and independent of the idiosyncratic innovations and each other.*

Assumption 3.13 *The nonstationary factors satisfy* $f_{k,t} \overset{iid}{\sim} N(0,1)$*,* $k \leq r_1$*.*

Proposition 3.4.1 *Reconsider the model* (3.17)–(3.18) *but now under Assumptions 3.5 and 3.10–3.13. Denote the law of* Y *under this DGP by* P_{r_1}*. For* $r_1 \geq 1$*, we have*

$$\lim_{(n,T)\to\infty} P_{r_1}[\tilde{\tau}_{n,T} \textit{ rejects}] = \alpha$$

for the test $\tilde{\tau}_{n,T}$ *from Theorem 3.4.2.*

Remark 3.4.3 *With multiple nonstationary factors, the notation* $\Delta\hat{F}_{1,t}$ *on which* $\tilde{\tau}_{n,T}$ *is based, may be misleading, since* $\Delta\hat{F}_{1,t}$ *in fact estimates a certain linear combination of the nonstationary factors. This linear combination will not be stationary under the hypothesis, since the factors are independent and thus not cointegrated. Therefore, intuitively, multiple nonstationary factors do not create a problem for detecting nonstationarity in the observations.*

Note that the problem of multiple nonstationary factors is not only a problem under the null hypothesis. Of course, against a fixed alternative the power will be the same as before, since under a fixed alternative all factors are stationary. However, the power, and even size under *local* alternatives merits another look here.

We have seen that neither the presence of stationary nor the presence of additional nonstationary factors changes the distribution of our test statistic under the hypothesis. However, adding multiple heterogeneous local-to unity factors, i.e., localizing the autoregressive parameters of more than one factor at once changes the distribution both under the null of at least one integrated factor and under the local alternatives:

Proposition 3.4.2 *Reconsider the model* (3.17)–(3.18) *but now assuming* $\rho_k = 1 + \frac{h_{1,k}}{T^{3/2}}$ *for* $k \leq r_1$ *and* $\rho_k < 1 - \epsilon$ *for* $k > r_1$ *and some* $\epsilon > 0$. *Under Assumptions 3.5 and 3.10–3.13, we have* $\mathcal{T}(\Delta \hat{F}_{1,T}) \to N\left(\frac{1}{12\sigma}\sum_{k=1}^{r_2} \bar{\lambda}_k h_{1,k}\mu_k, \frac{1}{12}\right)$.

Proof Since local alternatives of the form considered here are still contiguous, it is sufficient to show the desired convergence for $\mathcal{T}(\Delta G_T)$. Note that $\mathcal{T}$ is linear, so that

$$\mathcal{T}(\Delta G_T) = \frac{1}{\sigma}\sum_{k=1}^{r_2} \bar{\lambda}_k \mathcal{T}(\Delta F_{k,T}).$$

The result now follows from Theorem 2.2 in Hallin, Van den Akker, and Werker (2011), together with independence of the factors. ■

Remark 3.4.4 *Note that even if one* $h_{1,k}$ *equals zero, i.e., we should not reject the hypothesis of nonstationary observations, the test statistic will have nonzero mean unless all* $h_{1,k}$ *equal zero. In this situation,* $\tilde{\tau}_{n,T}$ *will not have correct asymptotic size. The problem is that the 'locally stationary' factors are close enough to one to be estimated as part of the cross-section average, while being far enough from one to change the distribution of our test statistic. Of course, the problem of incorrect size vanishes if we impose some homogeneity under the null on the local parameters, e.g.,* $h_{1,k} = h_1, k = 1, \ldots, r_1$.

3.5 Finite Sample Performance

In this section, we investigate to what extent our asymptotic results remain valid in finite samples.

3.5.1 Factors without Trends

Here we consider the setup from Section 3.3, where we confirm that estimating the factors using the approaches of Bai (2004), Bai and Ng (2004), and Choi (2017) leads to adaptive unit root tests, as do tests based on the Kalman filter. We use different unit-root tests, applied to estimated or observed factors:

1. Dickey-Fuller t-tests (labeled 'ADF'): This is one of the most commonly-used unit-root tests, based on OLS statistics in the univariate autoregression, see Dickey and Fuller (1979). We report three different kinds of ADF statistics, corresponding to a regression without intercept (no extra label), to one with an intercept under the alternative (labeled 'Drift'), and to one with a time trend (labeled 'Trend'). Since, in Section 3.3, we assumed that there is no intercept, all three specifications are valid in the sense that they produce asymptotically correctly sized tests when based on observed factors.

2. Test for the number of common trends (labeled 'Q-Test'): These tests establish, for a given multivariate time series, the number of underlying common trends, see Stock and Watson (1988). As suggested in Bai and Ng (2004), we apply these tests to the (observed or estimated) factors, instead of the observations Y in order to analyze the stationarity of common and idiosyncratic parts separately. If more than one common trend is suspected, these tests have the advantage of being able to determine the number of underlying trends through sequential testing. However, since we consider at most one trend and are interested in unit-root tests, we only conduct the 'final stage', i.e., test for a single trend. We use the version based on a first-order VAR, i.e., the Q_c test in the notation of Bai and Ng (2004).

3. Point-optimal tests from Elliott, Rothenberg, and Stock (1996) (labeled 'ERS'): These are likelihood-ratio tests against a specific alternative. We choose a fixed alternative of $h = 10$.

These three different classes of tests are applied to different factor estimates. For reference, we also consider the infeasible estimator, labeled 'observed', where the above tests are based on observed factors. The factors have been estimated in four different ways.

1. Level Principal Components: This is the principal component estimator discussed in Section 3.3.1 and studied in detail in Bai (2004). In the notation of that paper, we use the estimate $\bar{F}^k$, which is based on the

$n \times n$ matrix $X'X$. It does in fact, for the performance of the test, make a difference which principal components solution is used. The solution $\bar{F}^k$ is easy to compute for large values of T and has good size properties.

2. Difference Principal Components. This is the estimator originally proposed in Bai and Ng (2004) and discussed in Section 3.3.3. Here, we use the eigenvectors of XX', as suggested in Bai and Ng (2004) and again note that using another principal components solution seriously affects the tests. We also note that the relevant eigenvectors of the $T \times T$ matrix XX' are equal to the left singular vectors of the $T \times n$ matrix X, enabling us to do simulations also for large T.

3. Cross Section Averages: Here we estimate the factors using cross-sectional averages scaled as in Section 3.4.1.

4. Cross Section Averages Multivariate: Here we estimate the factors using cross-sectional averages scaled as in Section 3.4.2.

Table 3.1 shows the empirical sizes of various nominal 5% level tests for the unit root hypothesis of a single factor without a trend based on these factor estimates. All three ADF tests are almost perfectly sized when based on observed factors, except for the zero-mean ADF test for $T = 50$ (see 'ADF Observed', 'ADF Observed (Trend)','ADF Observed (Drift)' in Table 3.1). Estimating the factors with principal components also results in reasonably well-sized tests, but we do note that level principal components consistently outperform principal components based on differenced data, as advocated in Bai and Ng (2004), especially for small T (see 'ADF Difference', 'ADF Level', 'ADF Difference (Trend)','ADF Level (Trend)','ADF Difference (Drift)','ADF Level (Drift)'). This is expected, since our idiosyncratic parts are stationary. The behavior of the p_T tests from Elliott, Rothenberg, and Stock (1996), labelled 'ERS', are very similar to the ADF tests.

To illustrate the virtue of the factor-based approach, we have added a test just based on a single cross-section unit (see 'Univariate ADF', 'Univariate ADF (Trend)', 'Univariate ADF (Drift)'). This test appears to be completely useless, both in terms of size and power. The reason might be that the factor loading will relatively often be close to zero, so that the test cannot detect the nonstationarity.

We have also included tests based on the factor estimates suggested in Section 3.4. Whereas we have demonstrated that the different principal component estimators will lead to asymptotically correctly sized tests for all four tests, we are unsure about the asymptotic behavior of the average-based tests. In finite samples, they generally lead to oversized ADF and Q tests (see 'ADF Average Mult.', 'ADF Average Mult. (Trend)', 'ADF Average Mult. (Drift)', 'Q Test Average', 'Q Test Average Mult.').

We conclude that, in terms of size, estimating the factors by principal components indeed is almost as good as using observed ones, even for sample sizes as small as $T = 50, n = 40$.

Now we consider the power in finite samples. In Figure 3.1, we compare the power of the ADF tests to other unit root tests. The nearly efficient likelihood-ratio tests show a very similar performance to the Dickey-Fuller ones, but now even for small sample sizes the differences between observed and estimated factors are negligible. This also holds for the ERS and Q tests, however, these have considerably lower power. Whereas the Q-tests are significantly less powerful than the nearly efficient tests for all sample sizes, the ERS tests can perform even worse in small samples but on the other hand are nearly efficient in large samples. We conclude that all considered unit root tests work well with both level and difference estimated principal components even in small samples.

3.5.2 Factors with Trends

In this section we turn to the DGP from Section 3.4 and the tests suggested for it. Again, we are interested in the difference between observed and estimated factor-based tests. We also want to see how the factor estimators suggested in Sections 3.4.1 and 3.4.2 compare to each other and existing estimators.

Table 3.2 shows the empirical sizes in this setup. The only ADF tests that are valid under the null hypothesis are those with a trend term, so only these are included. Their power is close to the nominal power if the factors

T	50	100	200	500	1000
ADF Observed	10.4	5.0	5.2	5.2	5.0
ADF Difference	6.5	6.0	5.9	5.7	5.4
ADF Level	5.4	5.4	5.5	5.6	5.4
ADF Average Mult.	15.1	9.3	9.3	9.7	9.4
Univariate ADF	48.2	49.9	50.2	50.8	51.4
ADF Observed (Trend)	5.1	5.0	4.8	4.9	4.9
ADF Difference (Trend)	6.9	6.4	6.0	6.1	5.7
ADF Level (Trend)	5.6	5.7	5.6	6.0	5.6
ADF Average Mult. (Trend)	11.0	11.7	12.1	12.3	12.4
Univariate ADF (Trend)	62.1	66.6	67.7	68.5	69.8
ADF Observed (Drift)	4.3	5.0	5.2	5.2	5.1
ADF Difference (Drift)	6.3	5.8	5.8	5.6	5.6
ADF Level (Drift)	5.5	5.6	5.4	5.5	5.5
ADF Average Mult. (Drift)	7.3	9.1	10.0	10.0	9.8
Univariate ADF (Drift)	42.6	56.3	57.2	57.8	58.5
Q Test Observed	3.6	6.1	6.7	6.5	6.5
Q Test Difference	5.3	7.3	7.4	6.9	6.6
Q Test Level	4.1	6.7	7.2	6.8	6.6
Q Test Average	9.6	11.9	11.0	9.1	8.2
Q Test Average Mult.	9.6	11.9	11.0	9.1	8.2
ERS	1.5	2.4	3.7	4.9	4.9
ERS Difference	1.3	2.3	3.7	4.8	4.9
ERS Level	1.4	2.3	3.7	4.8	4.9
Average ERS	1.2	2.2	3.7	4.8	5.0
Average ERS Mult.	1.2	2.2	3.7	4.8	5.0

Table 3.1: Empirical sizes (in percent) of nominal 5% level tests for different sample sizes. We have $n = 40$ throughout, a single factor, $\mu_1 = 0$. Factor innovations are i.i.d. standard normally distributed. The factor loadings are drawn from a normal distribution with mean 1/2 and unit variance. The initial values of factor and idiosyncratic innovations are zero.

T	50	100	200	500	1000
ADF Observed (Trend)	4.8	5.1	5.2	5.1	5.1
ADF Difference (Trend)	6.3	6.2	6.3	5.8	5.4
ADF Level (Trend)	5.6	5.8	6.1	5.7	5.4
ADF Average Mult. (Trend)	10.9	11.7	12.4	12.7	12.9
Rank Observed	3.9	4.3	4.6	4.8	4.9
Rank Level	3.5	4.1	4.2	4.5	4.5
Rank Average	2.5	2.9	3.0	3.2	3.2
Rank Average mult.	2.5	2.9	3.0	3.2	3.2
Trend UMP Observed	3.0	3.4	3.8	4.2	4.5
Trend UMP Level	6.0	6.3	6.3	6.6	6.7
Trend UMP Average	1.8	2.2	2.5	2.8	2.9
Trend UMP Average Mult.	3.3	3.8	3.9	4.2	4.5

Table 3.2: Empirical sizes (in percent) of nominal 5% level tests for different sample sizes. We have $n = 40$ throughout, a single factor with trend $\mu_1 = 1$. Factor innovations are i.i.d. standard normally distributed. The factor loadings are drawn from a normal distribution with mean 1/2 and unit variance. The initial values of factor and idiosyncratic innovations are zero.

are estimated using principal components, but using averages, as suggested for the UMP tests, does not work well (see 'ADF Average Mult. (Trend)' in Table 3.2). Concerning the optimal tests ('Trend UMP'), we see in Table 3.2 that the method from Section 3.4.2 (labeled 'Average Mult.') is clearly superior to that of Section 3.4.1 (labeled 'Average') in terms of size. Although the tests are a bit undersized (based on asymptotic critical values), the size of the UMP test based on Section 3.4.2 is very close to that of the UMP test with observed factors.

For reference, we have also included rank-based tests, as suggested in Hallin, Van den Akker, and Werker (2011). Proving that our factor estimates lead to adaptive rank-based tests is beyond the scope of this chapter, but they would be instrumental in getting valid tests for non-normal factor innovations and they perform well in our simulations even under normality. They have the added advantage, that the scaling has no influence, i.e., our factor estimators from Sections 3.4.1 and 3.4.2 yield to identical rank-based tests.

Figure 3.2 displays the size-corrected powers of those tests with reasonable sizes. The ADF tests have no power in excess of size. This again shows that what we actually tests is the presence of the deterministic trend: in our model it exists only under the hypothesis, allowing us to consider autoregressive parameters closer to one. In the Dickey-Fuller specification, on the other hand, the trend exists also under the alternative, so that we have no power at these alternatives.

In general, Figure 3.2 demonstrates that, for all tests, the convergence to the asymptotic power envelop is very slow; even for $T = 1000$ there is a considerable gap. This is for a fixed $n = 40$, however, since even the tests based on observed factors do not reach the power envelope quickly, a larger n would have no impact. Throughout sample sizes, the tests based on cross-section averages do considerable better than those based on principal components. It is unclear whether the principal component-based factor estimates will even lead to asymptotically adaptive tests. The Gaussian UMP tests do slightly better than their rank-based counterparts, although this difference closes quickly as T grows. Also, for small sample sizes, the scaling from Section 3.4.1 appears to do slightly better than that from Section 3.4.2, but the difference is even smaller.

3.6 Conclusion

We have shown how to conduct panel unit root tests for unobserved factors in a variety of settings. Throughout our specifications, not observing the factors does not preclude inference at the usual rate. However, the exact specification of the factor equation matters as much as it does in time-series case. Which specification is most realistic depends on the application as well as the exact hypothesis of interest. However, if the factor innovations can reasonably be augmented with deterministic trends this pays off not only in higher powers, but even in a faster convergence rate. In Section 3.4, we have demonstrated that cross-sectional averages can lead to optimal panel unit-root tests under cross-sectional cointegration when the factors have deterministic trends. A natural question is whether these findings generalize to the more commonly used setting, where the factor innovations have zero mean and but more than one factor is present. For illustration, reconsider the setting of Section 3.4.2, but with $\mu_{k,T} = 0, k = 1, \ldots, r$. When trying to estimate the central sequence S_T based on cross-sectional averages, the additional term $\bar{\lambda}_2 \frac{1}{T} \sum_{t=1}^{T} F_{2,t-1} \Delta F_{2,t}$ appears in the difference of the estimated and oracle central sequence. Note that if F_2 is integrated this converges to a Brownian functional while in the stationary case it converges in probability to a nonzero constant. Either way, additional factors do change the distribution of a central-sequence based test statistic, so we end this conclusion on a cautionary note: In the setting with zero-mean factor innovations, both sizes and powers of average based unit-root tests will be affected by the presence of additional stationary factors. When secondary factors are likely to be of importance and one does not want to specify deterministic trends under the hypothesis, cointegration-rank methods or Kalman-smoother based methods that directly identify the nonstationary factor are likely the way forward. For the testing problem with deterministic trends, however, cross-sectional averages provide a convenient way of conducting uniformly most powerful unit-root tests.

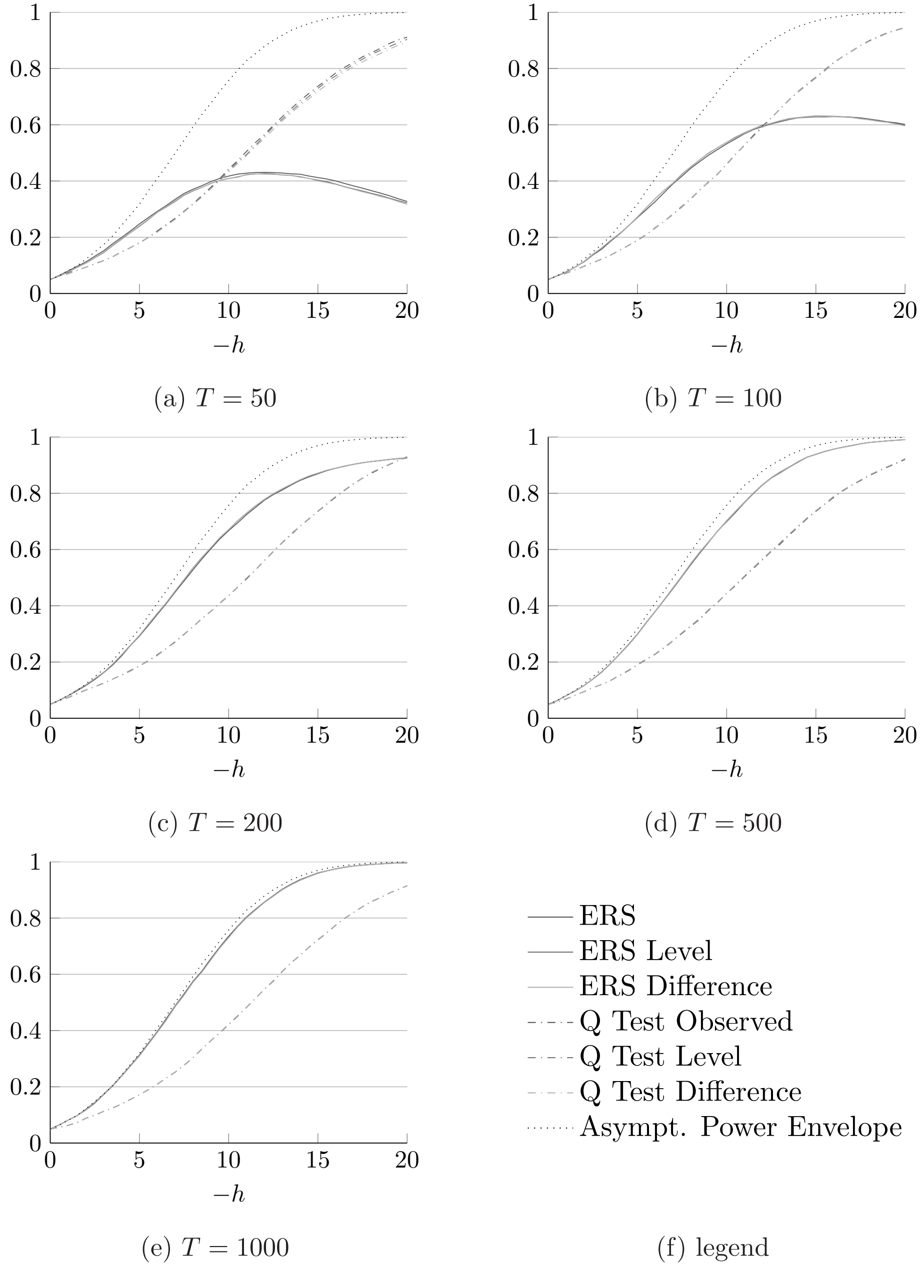
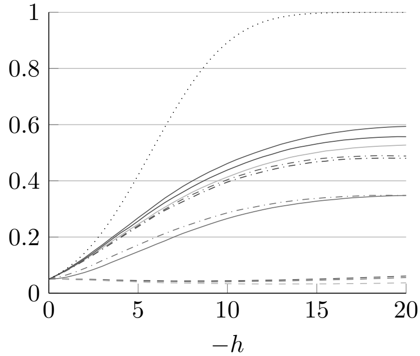
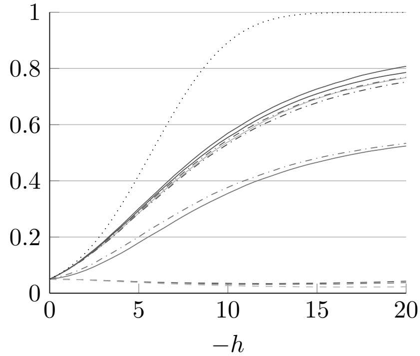


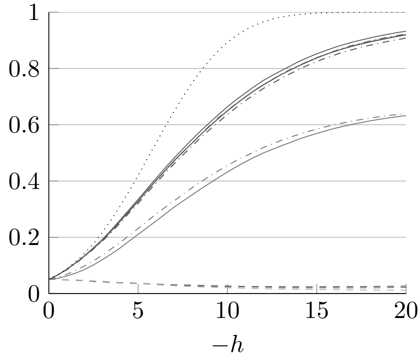
Figure 3.1: Size-corrected power of unit-root tests for $n = 40$ and different T. Single factor with trend $\mu_1 = 0$ and $\bar{\lambda} = 0.5$. The factor loadings are drawn from a normal distribution with mean 1/2 and unit variance. The initial values of factor and idiosyncratic innovations are zero.



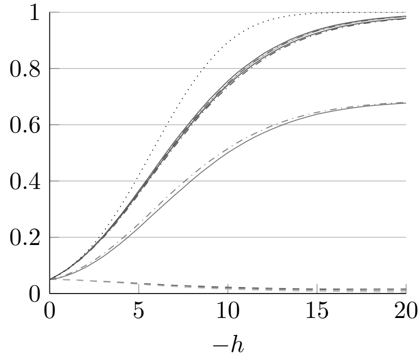
(a) $T = 50$



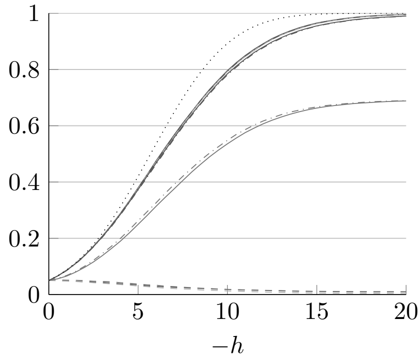
(b) $T = 100$



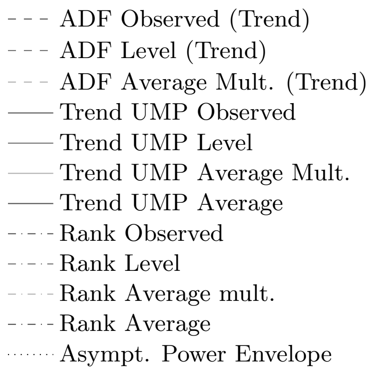
(c) $T = 200$



(d) $T = 500$



(e) $T = 1000$



(f) legend

Figure 3.2: Size-corrected power of unit-root tests for $n = 40$ and different T. Single factor with trend $\mu_1 = 1$ and $\bar{\lambda} = 0.5$. The factor loadings are drawn from a normal distribution with mean 1/2 and unit variance. The initial values of factor and idiosyncratic innovations are zero.

3.A Proofs

3.A.1 Factors without Trends

The following lemma establishes that negligibility under the hypothesis can be used interchangeably with negligibility under the alternatives.

Lemma 3.A.1 *Local alternatives as in* (3.4) *are mutually contiguous.*

Proof Using (3.5) and (3.7), we get that

$$\frac{\mathrm{d}\tilde{P}_{0,n,T}}{\mathrm{d}\tilde{P}_{0,n,T}} \overset{\tilde{P}_{0,n,T}}{\Rightarrow} \exp(-hS - \frac{1}{2}h^2 H). \tag{3.A.1}$$

We will show contiguity using Le Cam's First Lemma as, e.g., in Van der Vaart (2000) Lemma 6.4. To show that $\tilde{P}_{h,n,T}$ is contiguous to $\tilde{P}_{0,n,T}$, we use the equivalence of (i) and (iii) of that Lemma. Writing $M(t) = \int_0^t hW(r)\,\mathrm{d}W(r)$, we can rewrite the right-hand side of (3.A.1) as $\exp(-hS - \frac{1}{2}h^2 H) = \exp(-M(1) - \frac{1}{2}\langle M\rangle_1)$, i.e., a stochastic exponential which indeed has expectation one. To show that $\tilde{P}_{0,n,T}$ is contiguous to $\tilde{P}_{h,n,T}$ we can use the equivalence of part (i) and (ii) of the Lemma by again using (3.A.1) and noting that $\exp(-hS - \frac{1}{2}h^2 H) > 0$, since S and H are bounded in probability. ■

Proof of Proposition 3.3.1

Proof Thanks to Lemma 3.A.1, we only have to prove Proposition 3.3.1 under the hypothesis $\tilde{P}_{0,n,T}$, as probability convergence to zero under alternatives then follows from contiguity. Therefore, all following calculations proceed under the hypothesis $\tilde{P}_{0,n,T}$. Let A be a lagged cumulative sum operator, i.e., a $T \times T$ matrix with ones below the diagonal and zero on and above the diagonal. Rewrite S_T as

$$S_T = \frac{1}{T\omega_f^2}\sum_{t=2}^{T} \Delta F_{1,t}F_{1,t-1} + \frac{1}{T}f'(\Sigma_f^{-1} - \omega_f^{-2}I)Af$$

and note that

$$\mathrm{E}\frac{1}{T}f'(\Sigma_f^{-1} - \omega_f^{-2}I)Af = \mathrm{tr}((\Sigma_f^{-1} - \omega_f^{-2}I)A\Sigma_f) = -\omega_f^{-2}\,\mathrm{tr}(A\Sigma_f).$$

As $A + A' = \iota\iota' - I$,

$$\frac{1}{T}\,\mathrm{tr}(A\Sigma_f) = \frac{1}{T}\iota'\Sigma_f\iota - \frac{1}{T}\,\mathrm{tr}(\Sigma_f) = \sum_{m=-T+1}^{T-1}(1 - |m|/T)\gamma_f(m) - \gamma_f(0) \to \omega_f^2 - \gamma_f(0).$$

For the variance, we obtain, using again the formulas for quadratic forms of Gaussian random variables,

$$\begin{aligned}&\mathrm{Var}\,\frac{1}{T}f'(\Sigma_f^{-1} - \omega_f^{-2}I)Af \\ &\leq \frac{1}{T^2}\left\|(\Sigma_f^{-1} - \omega_f^{-2}I)A\Sigma_f\right\|_F^2 + \frac{1}{T^2}\left\|(I - \omega_f^{-2}\Sigma_f)A\right\|_F \left\|(\Sigma_f^{-1} - \omega_f^{-2}I)A\Sigma_f\right\|_F.\end{aligned}$$

We have

$$\left\|(\Sigma_f^{-1} - \omega_f^{-2} I) A \Sigma_f\right\|_F \leq \left\|\Sigma_f^{-1}\right\|_{\text{spec}} \left\|(I - \omega_f^{-2}\Sigma_f)A\right\|_F \|\Sigma_f\|_{\text{spec}} = O(1)o(T)O(1),$$

thanks to Assumption 3.1 and Lemma 2.A.1 Conclude that

$$S_T = \tilde{S}_T - \frac{1}{2} + \frac{\gamma_f(0)}{2\omega_f^2} + o_P(1) = \tilde{S}_T - \frac{\delta_f}{\omega_f^2} + o_P(1),$$

where $\delta_f = \frac{\omega_f^2 - \gamma_f(0)}{2}$ is the one-sided long-run variance of f and $\tilde{S}_T = \frac{1}{T\omega_f^2}\sum_{t=2}^T \Delta F_{1,t} F_{1,t-1}$. Similarly, we obtain, for $\tilde{H}_T = \frac{1}{T^2\omega_f^2}\sum_{t=1}^T F_{t-1}^2$,

$$H_T - \tilde{H}_T = \frac{1}{T^2} f' A' (\Sigma_f^{-1} - \omega_f^{-2} I) A f \to 0$$

in probability, as the (scaled) expectation of the difference is given by

$$\operatorname{tr}(A'(\Sigma_f^{-1} - \omega_f^{-2} I) A \Sigma_f) \leq \left\|A'(\Sigma_f^{-1} - \omega_f^{-2} I)\right\|_F \|A\|_F \|\Sigma_f\|_{\text{spec}} = o(T)O(T)O(1),$$

using the same arguments as above, and the (scaled) variance is bounded by

$$\left\|A'(\Sigma_f^{-1} - \omega_f^{-2} I) A \Sigma_f\right\|_F,$$

to which the same bound applies.[11]

Having significantly simplified the likelihood ratio, we now show that $|\hat{H}_T - H_T| = o_{\tilde{P}_{0,n,T}}(1)$. We have

$$\begin{aligned}|\hat{H}_T - \tilde{H}_T| &= \left|\frac{1}{T^2}\sum_{t=2}^T \left(\frac{\hat{F}_{1,t-1}^2}{\hat{\omega}_f^2} - \frac{F_{1,t-1}^2}{\omega_f^2}\right)\right| \\ &\leq \left|\frac{1}{T^2 R_{n,T}^2 \omega_f^2}\sum_{t=2}^T \left(\hat{F}_{1,t-1}^2 - R_{n,T}^2 F_{1,t-1}^2\right)\right| + \left|\frac{1}{\hat{\omega}_f^2} - \frac{1}{R_{n,T}^2\omega_f^2}\right| \frac{1}{T^2}\sum_{t=2}^T \hat{F}_{1,t-1}^2.\end{aligned}$$

Since $\frac{1}{T^2}\sum_{t=2}^T \hat{F}_{1,t-1}^2 = O_{\tilde{P}_{0,n,T}}(1)$ by (3.7) and (3.A.2), it is sufficient to show that the two absolute values converge to zero. Using the identity $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ and Cauchy-Schwarz, we obtain

$$\left|\frac{1}{T^2}\sum_{t=2}^T \left(\hat{F}_{1,t-1}^2 - R_{n,T}^2 F_{1,t-1}^2\right)\right| \leq \frac{1}{T} MSE_T + \frac{2}{T^2}\sum_{t=2}^T |R_{n,T} F_{1,t-1}(\hat{F}_{1,t-1} - R_{n,T}F_{1,t-1})| \tag{3.A.2}$$

$$\leq \frac{1}{T} MSE_T + 2\sqrt{R_{n,T}^2 H_T \frac{1}{T} MSE_T}. \tag{3.A.3}$$

To show that $\hat{\omega}_f^2$ converges to $R_{n,T}^2\omega_f^2$, we recall that by assumption, $\hat{\omega}_f^2$ converges to the long-run variance of $\hat{F}$, i.e., $\hat{\omega}_f^2 \to \lim_{T\to\infty} \operatorname{Var} \hat{F}_T/\sqrt{T}$. We have

$$\frac{1}{T}(\hat{F}_T - \hat{R}F_T)^2 \leq MSE_T \to 0$$

[11] Similar results on the simplification of the likelihood ratio could be obtained from Phillips (1987a).

in probability. As $\hat{F}_T^2/T$ is assumed to be uniformly integrable, and the same applies to F_T^2/T thanks to Gaussianity, we also obtain that $\frac{1}{T}(\hat{F}_T - RF_T)^2$ converges to zero in L_1, or put differently, $\frac{1}{T}\operatorname{Var}(\hat{F}_T - RF_T) \to 0$. Conclude that $\hat{\omega}_f^2 - R_{n,T}^2\omega_f^2 \to 0$. As $R_{n,T}^2\omega_f^2 > 0$, this implies convergence to zero of $\frac{1}{\hat{\omega}_f^2} - \frac{1}{R_{n,T}^2\omega_f^2}$.

Now consider $\hat{S}_T$. We first take care of the scaling as above, i.e.,

$$|\hat{S}_T + \frac{\hat{\delta}_f}{\hat{\omega}_f^2} - \tilde{S}_T| \leq \left| \frac{1}{TR_{n,T}^2\omega_f^2} \sum_{t=2}^{T} \left(\Delta\hat{F}_{1,t}\hat{F}_{1,t-1} - R_{n,T}^2\Delta F_{1,t}F_{1,t-1}\right)\right| + \left|\frac{1}{\hat{\omega}_f^2} - \frac{1}{R_{n,T}^2\omega_f^2}\right| \left|\frac{1}{T}\sum_{t=2}^{T}\Delta\hat{F}_{1,t}\hat{F}_{1,t-1}\right|.$$

From telescoping and triangle inequality, we have

$$\left|\frac{2}{T}\sum_{t=2}^{T}\left(\Delta\hat{F}_{1,t}\hat{F}_{1,t-1} - R_{n,T}^2\Delta F_{1,t}\hat{F}_{1,t-1}\right)\right| \leq \frac{|\hat{F}_{1,1}^2 - R_{n,T}^2F_{1,1}^2|}{T} + \frac{|\hat{F}_{1,T}^2 - R_{n,T}^2F_{1,T}^2|}{T} + \frac{1}{T}\sum_{t=1}^{T}|(\Delta\hat{F}_{1,t})^2 - (R_{n,T}\Delta F_{1,t})^2|. \quad (3.A.4)$$

For the first two terms, use $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ to note that

$$\frac{|\hat{F}_{1,t}^2 - R_{n,T}^2F_{1,t}^2|}{T} \leq \frac{(\hat{F}_{1,t} - R_{n,T}F_{1,t})^2}{T} + \frac{2|R_{n,T}F_{1,t}|}{\sqrt{T}}\frac{|\hat{F}_{1,t} - R_{n,T}F_{1,t}|}{\sqrt{T}}.$$

From Assumption 3.1 we get that $F_{1,t}/\sqrt{T} = O_{\tilde{P}_{0,n,T}}(1)$ even for a growing $t = T$, so that the condition on MSE_T implies both summands converging to zero.

For the term in (3.A.4), write

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}|(\Delta\hat{F}_{1,t})^2 - (R_{n,T}\Delta F_{1,t})^2| \leq & \frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_{1,t} - R_{n,T}\Delta F_{1,t})^2 & (3.A.5)\\ & + \frac{2}{T}\sum_{t=2}^{T}|R_{n,T}\Delta F_{1,t}(\Delta\hat{F}_{1,t} - R_{n,T}\Delta F_{1,t})| \\ \leq & \frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_{1,t} - R_{n,T}\Delta F_{1,t})^2 & (3.A.6)\\ & + \sqrt{\frac{R_{n,T}^2}{T}\sum_{t=2}^{T}(\Delta F_{1,t})^2\frac{1}{T}\sum_{t=2}^{T}(\Delta F_{1,t} - R_{n,T}\Delta\hat{F}_{1,t})^2}.\end{aligned}$$

By Assumption 3.1 and a Law of Large Numbers, $\frac{1}{T}\sum_{t=1}^{T}(\Delta F_{1,t})^2$ converges in probability to $\gamma_f(0)$. For the first summand and the second part under the root we use the identity $(a+b)^2 \leq 2a^2 + 2b^2$, to demonstrate that

$$\frac{1}{T}\sum_{t=2}^{T}(\Delta\hat{F}_{1,t} - R_{n,T}\Delta F_{1,t})^2 = \frac{1}{T}\sum_{t=2}^{T}(\hat{F}_{1,t} - R_{n,T}F_{1,t} + R_{n,T}F_{1,t-1} - \hat{F}_{1,t-1})^2 \leq 4MSE_T$$

converges to zero in probability.

Note that we impose conditions on the factor estimates only under the null hypothesis of an integrated factor, for which results are readily available. We have now shown that $(\hat{S}_T, \hat{H}_T) = (S_T, H_T) + o_{\tilde{P}_{0,n,T}}(1)$. However, thanks to Lemma 3.A.1, we can use $o_{P_0}(1)$ and $o_{P_h}(1)$ terms interchangeably. ■

Proof of Proposition 3.3.2

Proof In the following, all probabilities are evaluated under $\tilde{P}_{0,n,T}$. We first verify (3.12). The key insight is that $(\lambda' \otimes \tilde{A}\Sigma_f \tilde{A}')(\lambda\lambda' \otimes \tilde{A}\Sigma_f \tilde{A}' + \Omega_\eta \otimes \Sigma_\eta)^{-1}$ can be written as the Kronecker product $(\frac{\lambda'}{\lambda'\Omega_\eta^{-1}\lambda}\Omega_\eta^{-1}) \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1}$. To verify this, note

$$\begin{aligned}
&\left((\frac{\lambda'}{\lambda'\Omega_\eta^{-1}\lambda}\Omega_\eta^{-1}) \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1} \right) (\lambda\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}' + \Omega_\eta \otimes \Sigma_\eta) \\
=&\lambda' \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1} \tilde{A}\Sigma_f\tilde{A}' + \frac{\lambda'}{\lambda'\Omega_\eta^{-1}\lambda} \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1} \Sigma_\eta \\
=&\lambda' \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1} \left(\tilde{A}\Sigma_f\tilde{A}' + \frac{\Sigma_\eta}{\lambda'\Omega_\eta^{-1}\lambda}\right) \\
=&\lambda' \otimes \left(I_T + \Sigma_\eta \frac{(\tilde{A}\Sigma_f\tilde{A}')^{-1}}{\lambda'\Omega_\eta^{-1}\lambda}\right)^{-1} \left(I_T + \frac{\Sigma_\eta}{\lambda'\Omega_\eta^{-1}\lambda}(\tilde{A}\Sigma_f\tilde{A}')^{-1}\right)(\tilde{A}\Sigma_f\tilde{A}') \\
=&\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}' \\
=&(\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}')(\lambda\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}' + \Omega_\eta \otimes \Sigma_\eta)^{-1}(\lambda\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}' + \Omega_\eta \otimes \Sigma_\eta).
\end{aligned}$$

As $\lambda\lambda' \otimes \tilde{A}\Sigma_f\tilde{A}' + \Omega_\eta \otimes \Sigma_\eta$ has full rank, this shows the desired equality.

We now prove Proposition 3.3.2, using (3.12) to relate the Kalman smoother to the generalized principal component estimator. Note that $MSE_T = \frac{1}{T}\left\|\hat{F}^{KS} - F_1\right\|_F$, so from Section 3.3.2 we obtain $\left\|\hat{F}^{GPC} - F_1\right\|_F = o_P(T)$. Thus it is sufficient to show that $\left\|\hat{F}^{KS} - \hat{F}^{GPC}\right\|_F = o_P(T)$. As $\hat{F}^{GPC} = Y\hat{\Omega}_\eta^{-1}\frac{\hat{\lambda}}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}}$, we have

$$\begin{aligned}
\left\|\hat{F}^{KS} - \hat{F}^{GPC}\right\|_F =& \left\| \left(\left(I_T + \hat{\Sigma}_\eta \frac{(\tilde{A}\hat{\Sigma}_f\tilde{A}')^{-1}}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}}\right)^{-1} - I_T \right) \hat{F}^{GPC} \right\|_F \\
\leq& \left\| \left(I_T + \hat{\Sigma}_\eta \frac{(\tilde{A}\hat{\Sigma}_f\tilde{A}')^{-1}}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}}\right)^{-1} - I_T \right\|_{\text{spec}} \left\|\hat{F}^{GPC}\right\|_F .
\end{aligned}$$

Note that the matrix in the former norm equals $(\tilde{A}\hat{\Sigma}_f\tilde{A}'\hat{\Sigma}_\eta\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda} + I_T)^{-1}$. Thus, as $\left\|\hat{F}^{GPC}\right\|_F = O_P(T)$, we need $\left\|(\tilde{A}\hat{\Sigma}_f\tilde{A}'\hat{\Sigma}_\eta\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda} + I_T)^{-1}\right\|_{\text{spec}} = o_P(1)$. For this, write

$$\begin{aligned}
\left\|(\tilde{A}\hat{\Sigma}_f\tilde{A}'\hat{\Sigma}_\eta\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda} + I_T)^{-1}\right\|_{\text{spec}} =& \frac{1}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}} \left\| \left(\tilde{A}\hat{\Sigma}_f\tilde{A}'\hat{\Sigma}_\eta + \frac{I_T}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}}\right)^{-1} \right\|_{\text{spec}} \\
\leq& \frac{1}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}} \left\| \left(\tilde{A}\hat{\Sigma}_f\tilde{A}'\hat{\Sigma}_\eta\right)^{-1} \right\|_{\text{spec}}
\end{aligned}$$

$$\leq \frac{1}{\hat{\lambda}'\hat{\Omega}_\eta^{-1}\hat{\lambda}} \left\| \tilde{A}^{-1} \right\|_{\text{spec}}^2 \left\| \hat{\Sigma}_f^{-1} \right\|_{\text{spec}} \left\| \hat{\Sigma}_\eta \right\|_{\text{spec}}$$
$$= O_P(n^{-1}) O(1) O_P(1) O_P(1).$$

$\tilde{A}^{-1}$ is a simple tridiagonal Toeplitz matrix whose spectral norm converges to 2 and is thus $O(1)$. The rates on the covariance matrices follow from the bounds on the spectral densities.■

3.A.2 Factors with Trends

Proof of Lemma 3.4.1

Proof By contiguity, it once more suffices to do all calculations under the hypothesis of a unit-root, i.e. under $\tilde{P}_{0,h_2,n,T}$, where $\Delta Y_{i,t} = \lambda_i 1(\mu_{1,T} + f_{1,t}) + \Delta E_{i,t}$. We first show that $s(\overline{\Delta Y}_{\cdot,t}) \to \bar{\lambda}_1$ in probability. For this, write

$$\begin{aligned} s^2(\overline{\Delta Y}_{\cdot,t}) =& \frac{1}{T}\sum_{t=1}^T \left(\bar{\lambda}_{1,\cdot}(\mu_{1,T} + f_{1,t}) + \overline{\Delta E}_{\cdot,t} - \bar{\lambda}_{1,\cdot}\mu_{1,T} - \frac{\bar{\lambda}_{1,\cdot}}{T}\sum_{t=1}^T f_{1,t} - \frac{1}{T}\sum_{t=1}^T \overline{\Delta E}_{\cdot,t} \right)^2 \\ =& \frac{1}{T}\sum_{t=1}^T \left(\bar{\lambda}_{1,\cdot}(f_{1,t} - \frac{1}{T}\sum_{t=1}^T f_{1,t}) + \overline{\Delta E}_{\cdot,t} - \frac{1}{T}\sum_{t=1}^T \overline{\Delta E}_{\cdot,t} \right)^2 \\ =& \bar{\lambda}_{1,\cdot}^2 s^2(f_{1,t}) + s^2(\overline{\Delta E}_{\cdot,t}) + \frac{2}{T}\sum_{t=1}^T (f_{1,t} - \frac{1}{T}\sum_{t=1}^T f_{1,t})(\overline{\Delta E}_{\cdot,t} - \frac{1}{T}\sum_{t=1}^T \overline{\Delta E}_{\cdot,t}). \end{aligned}$$

Thanks to our second-moment assumptions on and u_t and ergodicity, $s^2(f_{1,t})$ converges to its population counterpart, i.e., one, as $T \to \infty$. Using Cauchy Schwarz, we get

$$\frac{1}{T}\sum_{t=1}^T |f_{1,t} - \frac{1}{T}\sum_{t=1}^T f_{1,t}||\overline{\Delta E}_{\cdot,t} - \frac{1}{T}\sum_{t=1}^T \overline{\Delta E}_{\cdot,t}| \leq \sqrt{s^2(f_{1,t}) s^2(\overline{\Delta E}_{\cdot,t})},$$

so that demonstrating that $s^2(\overline{\Delta E}_{\cdot,t}) \to 0$ in probability is enough to show that

$$s^2(\overline{\Delta Y}_{\cdot,t}) \to \bar{\lambda}_1^2 \text{ in probability.} \tag{3.A.7}$$

We will demonstrate this by showing $s^2(\overline{\Delta E}_{\cdot,t}) \to 0$ in L_1. We have, using cross-sectional independence,

$$\begin{aligned} \mathrm{E}s^2(\overline{\Delta E}_{\cdot,t}) =& \mathrm{E}\frac{1}{T}\sum_{t=1}^T \left(\frac{1}{n}\sum_{i=1}^n \Delta E_{i,t} \right)^2 - \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^T \frac{1}{n}\sum_{i=1}^n \Delta E_{i,t} \right)^2 \\ \leq& \mathrm{E}\frac{1}{T}\sum_{t=1}^T \frac{1}{n^2}\sum_{i=1}^n (\Delta E_{i,t})^2 = \frac{\sigma_e^2}{n} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Now we can show the approximation of our test statistic. For convenience write $w_t := \left(\frac{t}{T+1} - \frac{1}{2} \right)$ and note that $\sum_{t=1}^T w_t = 0$ and $|w_t| \leq 1$ for all t. We have

$$\mathcal{T}(\Delta \hat{F}_{1,T}) - \mathcal{T}(\Delta \hat{F}_{1,T}1) = \frac{1}{\sqrt{T}}\sum_{t=1}^T w_t(\Delta \hat{F}_{1,t} - \Delta F_{1,t})$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \left(\frac{\bar{\lambda}_{1,\cdot}}{s(\overline{\Delta Y}_{\cdot,t})} - 1\right)(\mu + f_{1,t}) + \frac{1}{s(\overline{\Delta Y}_{\cdot,t})}\frac{1}{n\sqrt{T}}\sum_{t=1}^{T} w_t \sum_{i=1}^{n} \Delta E_{i,t} \\
=&\left(\frac{\bar{\lambda}_{1,\cdot}}{s(\overline{\Delta Y}_{\cdot,t})} - 1\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t f_{1,t} + \frac{1}{s(\overline{\Delta Y}_{\cdot,t})}\left(\frac{T}{T+1} - \frac{1}{2}\right)\frac{1}{n}\sum_{i=1}^{n}\frac{E_{iT}}{\sqrt{T}} \\
&- \frac{1}{s(\overline{\Delta Y}_{\cdot,t})}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\frac{1}{n}\sum_{i=1}^{n}\frac{E_{it}}{T+1}.
\end{aligned}
$$

First, note that

$$
\operatorname{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t f_{1,t}\right) = \frac{\sum_{t=1}^{T} w_t^2}{T} \leq 1,
$$

so that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t f_{1,t}$ is bounded in probability. Combining this with (3.A.7), the first summand converges to zero. For the second summand, note that $\frac{1}{n}\sum_{i=1}^{n}\frac{E_{iT}}{\sqrt{T}}$ has mean zero and that

$$
\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{E_{iT}}{\sqrt{T}}\right)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\frac{\mathrm{E}E_{iT}^2}{T} = \frac{1}{nT}\mathrm{E}E_{iT}^2, \tag{3.A.8}
$$

by the cross-sectional dependence of the idiosyncratic errors. Now, if the $E_{i,\cdot}$ are covariance stationary, i.e., we impose Assumption 3.5(1), this converges to zero even for fixed n. Under Assumption 3.5(2), however, we have

$$
\frac{1}{T}\mathrm{E}E_{iT}^2 = \sum_{s=-T}^{T}\frac{T-|s|}{T}\gamma(s) \leq \gamma(0) + 2\sum_{s=1}^{\infty}\gamma(s),
$$

which is bounded by assumption. Therefore, under Assumption 3.5(2), the second summand goes to zero in probability as long as also $n \to \infty$. The last summand is treated similarly: $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\frac{E_{it}}{T+1}$ has mean zero and the expectation of its square is given by

$$
\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{E_{it}}{T+1}\right)^2\right] =&\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{T^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{E}E_{it}E_{is} \\
=&\frac{1}{nT^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{E}\sum_{u=1}^{t}\Delta E_{iu}\sum_{v=1}^{s}\Delta E_{iv} \\
=&\frac{1}{nT^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{t}\sum_{v=1}^{s}\mathrm{E}\Delta E_{iu}\Delta E_{iv} \\
=&\frac{1}{nT^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{t}\sum_{v=1}^{s}\gamma(u-v) \\
\leq&\frac{1}{nT^3}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{v=1}^{T}\gamma(u-v) = \frac{1}{nT}\sum_{u=1}^{T}\sum_{v=1}^{T}\gamma(u-v) \\
=&\sum_{s=-T}^{T}\frac{T-|s|}{nT}\gamma(s) \leq \gamma(0)/n + \frac{2}{n}\sum_{s=1}^{\infty}\gamma(s),
\end{aligned}
$$

so we can conclude that the last term converges to zero as well as $n, T \to \infty$. Finally, since $s^2(\overline{\Delta Y}_{\cdot,t}) \to \bar{\lambda}_1^2$, $\left(\frac{\bar{\lambda}_{1,\cdot}}{s(\overline{\Delta Y}_{\cdot,t})} - 1\right)$ converges in probability to zero. ■

Proof of Lemma 3.4.2

Proof The long run variance of $\overline{\Delta Y}_{\cdot,t}$ equals

$$\operatorname{Var}\left(\frac{\bar{Y}_{\cdot,T}}{\sqrt{T}}\right) = \bar{\lambda}_{1,\cdot}^2 \operatorname{Var}(F_{1,T}/\sqrt{T}) + \sum_{k=2}^{r} \bar{\lambda}_{k,\cdot}^2 \operatorname{Var}(F_{k,T})/T + \operatorname{Var}(\bar{E}_{\cdot,t}/\sqrt{T}).$$

Since all but the first factors are stationary, $\operatorname{Var}(F_{k,T})$ does not depend on T for $k \geq 2$, so the second summand converges to zero. Also the third part does, as shown in (3.A.8). By assumption, $\operatorname{Var}(F_{1,T}/\sqrt{T}) = 1$, so indeed the long run variance converges to $\bar{\lambda}_1^2$.

Now we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t(\Delta\hat{F}_{1,t} - \Delta F_{1,t})$$
$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \left(\frac{\bar{\lambda}_{1,\cdot}}{\hat{\bar{\lambda}}_1} - 1\right) \Delta F_{1,t} + \sum_{k=2}^{r} \frac{\bar{\lambda}_{k,\cdot}}{\hat{\bar{\lambda}}_1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \Delta F_{k,t} + \frac{1}{\hat{\bar{\lambda}}_1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \Delta \bar{E}_{\cdot,t}.$$

The last term was handled in the proof of Lemma 3.4.1; it converges to zero. The second term can be handled in a similar way: We have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w_t \Delta F_{k,t} = \left(\frac{T}{T+1} - \frac{1}{2}\right) \frac{F_{k,T}}{\sqrt{T}} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} \frac{F_{k,t}}{T+1}, \tag{3.A.9}$$

which, for $k \geq 2$, goes to zero in probability: Due to $F_{k,t}$ being stationary, $\mathrm{E}|F_{k,t}|$ is a constant, so that both summands converge to zero in L_1. ■

Proof of Proposition 3.4.1

Proof We will show that $\mathcal{T}(\Delta\hat{F}_{1,T}) \overset{P_{r_1}}{\to} N(0, 1/12)$ in distribution for the $\Delta\hat{F}_{1,t}$ introduced in Lemma 3.4.2. First, we introduce the process $G_t = \frac{1}{\sigma}\sum_{k=1}^{r_1} \bar{\lambda}_k F_{k,t}, t = 1, \ldots, T$, that our factor estimate will approximate. From Assumption 3.13, we have

$$\Delta G_t = \frac{1}{\sigma} \sum_{k=1}^{r_1} \bar{\lambda}_k \mu_k + \frac{1}{\sigma} \sum_{k=1}^{r_1} \bar{\lambda}_k f_{k,t} \sim N\left(\frac{1}{\sigma} \sum_{k=1}^{r_1} \bar{\lambda}_k \mu_k, 1\right).$$

Because of Assumption 3.13 and the unit variance, it follows from Theorem 2.1 in Hallin, Van den Akker, and Werker (2011) that $\mathcal{T}(\Delta G_T) \to N(0, 1/12)$ in distribution. We proceed to show that $\mathcal{T}(\Delta G_T) - \mathcal{T}(\Delta\hat{F}_{1,T}) = o_{P_{r_1}}(1)$. Analogous to the proof of Lemma 3.4.2, we first show that the long run variance of $\overline{\Delta Y}_{\cdot,t}$ converges to σ^2: We have

$$\operatorname{Var}\left(\frac{\bar{Y}_{\cdot,T}}{\sqrt{T}}\right) = \sum_{k=1}^{r_1} \bar{\lambda}_{k,\cdot}^2 \operatorname{Var}(F_{k,T}/\sqrt{T}) + \sum_{k=r_1+1}^{r} \bar{\lambda}_{k,\cdot}^2 \operatorname{Var}(F_{k,T})/T + \operatorname{Var}(\bar{E}_{\cdot,t}/\sqrt{T}),$$

where we can handle all terms completely analogous to the proof of Lemma 3.4.2. Using

$$\overline{\Delta Y}_{\cdot,t} = \sum_{k=1}^{r_1} \bar{\lambda}_{k,\cdot} \Delta F_{1,t} + \sum_{k=r_1+1}^{r} \bar{\lambda}_{k,\cdot} \Delta F_{1,t} + \Delta \bar{E}_{\cdot,t},$$

we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t(\Delta\hat{F}_{1,t} - \Delta G_t) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t(\overline{\Delta Y}_{\cdot,t}/\hat{\bar{\lambda}}_1 - \frac{1}{\sigma}\sum_{k=1}^{r_1}\bar{\lambda}_k \Delta F_{k,t})$$
$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \left(\sum_{k=1}^{r_1}\left(\frac{\bar{\lambda}_{k,\cdot}}{\hat{\bar{\lambda}}_1} - \frac{\bar{\lambda}_k}{\sigma}\right)\Delta F_{k,t} + \frac{1}{\hat{\bar{\lambda}}_1}\left(\sum_{k=r_1+1}^{r}\bar{\lambda}_{k,\cdot}\Delta F_{1,t} + \Delta\bar{E}_{\cdot,t}\right)\right)$$
$$= \sum_{k=1}^{r_1}\left(\frac{\bar{\lambda}_{k,\cdot}}{\hat{\bar{\lambda}}_1} - \frac{\bar{\lambda}_k}{\sigma}\right)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \Delta F_{k,t}$$
$$+ \frac{1}{\hat{\bar{\lambda}}_1}\left(\sum_{k=r_1+1}^{r}\bar{\lambda}_{k,\cdot}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \Delta F_{1,t} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \Delta\bar{E}_{\cdot,t}\right).$$

The two elements of the second term have been handled in the proof of Lemma 3.4.2. Thus, using that $\left(\frac{1}{\hat{\bar{\lambda}}_1} - \frac{1}{\sigma}\right) \to 0$ and recalling from the same proof that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \Delta F_{k,t}$ is bounded in probability, we are done. ∎

Chapter 4

Using Stationary Covariates to Test for Common Stochastic Trends in High-Dimensional Panels

Abstract

This chapter considers panel unit-root tests in the presence of stationary covariates and cross-sectional dependence. Our starting point is the popular PANIC framework and we analyze the potential power gains due to observing additional stationary covariates, focusing on unit-root tests that are robust to cross-sectional cointegration, i.e., tests for a unit root in the common unobserved factors. The stationary, observed covariates are assumed to be unit-specific but allowed to be cross-sectionally correlated. We differentiate two cases: one in which the contribution of the factor of interest to the covariance structure of the covariate can be perfectly identified and a more general one, where the contribution of the factor innovations in the covariate equation is perturbed by another unobserved common shock.

In the former case, the inclusion of stationary covariates leads to vastly more powerful tests, with a faster convergence rate. We first analyze the problem for an observed factor, and show that the statistical

experiment is locally asymptotically mixed normal (LAMN). This implies that no UMP test exists, but we obtain an asymptotically optimal invariant test. We also demonstrate how to conduct valid inference based on estimated factors. The improved rate allows us to compare different factor estimation schemes in terms of resulting asymptotic power. When implemented well, the asymptotic power of estimated-factor based tests is relatively close to the observed-factor power envelope.

In the second case, with the additional perturbation, the statistical problem is closely related to that of univariate unit-root tests with stationary factors that have been studied in Elliott and Jansson (2003) and Hansen (1995). We demonstrate that the original time-series experiment is locally, asymptotically Brownian Functional (LABF) but converges to the better understood LAMN case as the contribution of the covariate grows to one. Moreover, we show that the covariate-augmented Dickey Fuller (CADF) test of Hansen (1995) becomes optimal invariant as the share of the variation explained by the covariate converges to unity. This explains why the tests of Hansen (1995) are competitive in terms of power to those of Elliott and Jansson (2003), in particular when the covariate is more important. We show that both the CADF tests and the point-optimal tests can also be implemented in a panel setting with unobserved common factors and that their optimality properties carry over to this panel setup.

4.1 Introduction

Exploiting stationary covariates to improve the power of unit-root tests was first proposed by Hansen (1995) for univariate time series. It is assumed that, in addition to the series to be tested, one observes a stationary covariate that is correlated with the innovations for the series of interest. As the covariates are assumed to be stationary under both the null and alternative hypotheses, they can be used to remove part of the variation in the error term, leading to more powerful tests. The original proposal by Hansen (1995) achieves this by augmenting a Dickey-Fuller regression with a covariate. This is easy to implement and improves power significantly compared to the no-covariate setup. Elliott and Jansson (2003) reconsider this setup and propose point-optimal tests, that are uniformly close to the power envelope if the alternative is chosen in a smart way. Applications of covariate-augmented unit-root tests include Elliott and Pesavento (2006), who study purchasing power parity on a country-by country basis, using macroeconomic time-series like differenced nominal exchange rates, differences in money supply, and interest rate differentials as stationary covariates.

Testing the stationarity properties of unobserved common factors and idiosyncratic deviations separately when dealing with panel data was first proposed by Bai and Ng (2004). As part of their PANIC (Panel Analysis of Nonstationarity in Idiosyncratic and Common components) approach, they suggest to apply Dickey-Fuller tests to common factors estimated by principal component analysis. More recently, covariates have been used to further improve the power of panel unit-root tests. However, these papers have only dealt with testing the idiosyncratic deviations of each panel unit for a unit root, rather than testing the common shocks as is the focus of this chapter. We take the PANIC framework as a starting point and analyze the consequences of observing additional stationary covariates.

Unobserved latent factors are commonly used in the panel unit-root literature, introducing cross-sectional correlation either in the observations directly, as in the PANIC approach, or in the innovations, as in Moon and Perron (2004). Although these two setups are equivalent in the absence of observed covariates (see Wichert et al. (2019)), studies that augment these setups with stationary covariates obtain very different conclusions. Becheri, Drost, and Van den Akker (2015b), who consider the Moon and Perron (2004) setup with cross-sectionally correlated errors, show that covariates can improve the power of unit-root tests for the idiosyncratic parts. However, the statistical problem in that case becomes more complex, as the experiment is locally asymptotically mixed normal (LAMN) instead of locally asymptotically normal (LAN). Juodis and Westerlund (2018), on the other hand, consider observed covariates in the PANIC setting of Bai and Ng (2004) and conclude that the limit experiment is of the same type as when no covariates are considered.[1]

[1] This holds for the case of no idiosyncratic trends being present, however, the observed

As a further difference, Becheri, Drost, and Van den Akker (2015b) only consider covariates that do not differ per cross-sectional unit. Juodis and Westerlund (2018) consider the other extreme: their covariates are cross-sectionally independent. Moreover, both papers focus on the testing problem for the idiosyncratic parts and not the common factors. In this chapter, we propose an intermediate framework, where the covariates are unit-specific but correlated, with the correlation being driven by common factors. To realize power gains, the factor of interest must also drive (part of) the correlation between the covariates. As Hansen (1995), we assume that the covariates (and thus their common factor innovations) are known to be stationary.

Our approach can be considered a panel-analogue of Hansen (1995), where again the testing problem for the factors is isolated from that of the idiosyncratic parts. However, in the panel setting the potential gains from using stationary covariates are more pronounced: depending on the exact specification, rather than just a fixed increase in local power, the covariates here can even deliver a faster convergence rate. Nevertheless, the intuition for the additional power is the same: observing stationary covariates allows us to predict part of the innovations in addition to the observations of interest. If both the innovations and the levels were observed, checking for a unit-root would be trivial algebra. Thus, the covariate being known to be stationary and related to the innovations offers an additional way of testing the unit-root hypothesis.

Following this intuition, the amount of power that can be gained by including the covariate then depends on how well the covariate approximates the innovations of the factor of interest. We consider two specifications: In the first, the factor of interest's innovations determine the cross-sectional correlation of the covariate as an additively separable factor. In this case, intuition suggests that, as the cross-sectional dimension grows, the covariates' factor structure can be estimated more accurately and the power of the test will increase both with the number of cross-section units as well as the number of time periods. In fact, we formally show that the convergence rate will be equal

---

covariate can prevent the large loss of power due to estimating heterogeneous deterministic trends.

to that of typical unit-root tests that are not robust to cross-sectional cointegration, highlighting the large potential gains that covariates may deliver in this setting.

On the other hand, estimating this factor structure in the presence of multiple factors in the covariate equation is nontrivial.[2] More importantly, the assumption of the factor innovation entering without further perturbation does not always follow from economic theory and will be hard to verify empirically. Therefore, we also consider a more robust version of this problem, where not the innovation of interest, but only a correlated disturbance enters the factor structure of the covariate.[3] In this case, the original power envelope of Elliott and Jansson (2003) applies, so that the number of cross-sectional units will not enter the convergence rate. The challenge is now to find tests that get close to this power envelope, despite being applied to estimated factors.

We offer the following contributions. In the first setting, where the contribution of the factor of interest to the covariance structure of the covariate can be perfectly identified, we analyze how to optimally combine the information inherent in the series with that of the covariates by deriving the limit experiment. We find that the experiment is of the Locally Asymptotically Mixed Normal (LAMN) type .[4] The LAMN result suggests tests with certain optimality properties. In particular, we show that an easily implementable test corresponding to a t test in the limit experiment is optimal invariant, asymptotically normal and we derive its asymptotic power. Moreover, simulations demonstrate good finite sample performance.

---

[2] Principle components would not be sufficient here, as one would need to take into account the time-series structure of the innovations to prevent a lagged level of the factor to appear as an additional factor.

[3] When considering the first setup, this is equivalent to the presence of an additional factor with exactly the same loadings, so that they cannot be identified separately.

[4] This is the same limit experiment as Becheri, Drost, and Van den Akker (2015b) obtain for testing the idiosyncratic parts with observed common factors. In both cases, observing a common component leads to a term in the likelihood ratio that does not abide by a cross-sectional central limit theorem. However, the origins of the common and idiosyncratic parts in the likelihood ratios are completely different.

As the focus of this chapter is on panel unit root tests with unobserved factors, we consider the implications of applying these tests to unobserved factors. Contrary to other common setups, the limiting distribution of our test statistics is not unaffected by using estimates when the idiosyncratic parts are nonstationary. However, we develop the appropriate size-corrections for a general class of linear factor estimates. Moreover, we derive the local asymptotic powers of these tests based on estimated factors and show that power is maximized by using weights corresponding to principal component estimates.[5]

For the second setting, in which the covariate is related to the innovations less directly, we also start by considering the observed factor case. Contrary to the first setting, this observed factor case is now the same as that studied in Elliott and Jansson (2003) and Hansen (1995). However, the limit experiment has not been derived to the best of our knowledge. It is of the more general Local Asymptotic Brownian functional (LABF) type, making optimal inference even more complicated than in the first case. However, we observe that as the correlation between the innovation and the covariate approaches one, we again obtain a simpler LAMN experiment, motivating the use of a t-test in the original limit experiment. This test turns out to be identical to the Covariate Augmented Dickey Fuller (CADF) test of Hansen (1995) and is thus simple to implement. This sheds light on why the CADF test performs well compared to the point-optimal tests by Elliott and Jansson (2003), especially when the influence of the covariate is large.[6] As a final step, we again show that the asymptotic distribution of this test is unaffected when applied

[5] The efficiency of the factor estimation method appearing in the asymptotic powers is an additional benefit of the faster convergence of our first setting. While some authors have argued for using cross-sectional averages instead of principal component estimates to estimate the factor, in this setting we show that using principal components is preferable asymptotically.

[6] Elliott and Jansson (2003) liken the relation of their point-optimal tests to the Hansen (1995) tests to the relation between the Elliott, Rothenberg, and Stock (1996) and Dickey and Fuller (1979) tests. However, in the setting with an additional covariate, the case for using the point-optimal tests is weaker, as the choice of a reasonable alternative depends on the importance of the covariate and due to our optimality result when the correlation approaches one.

to the unobserved factor of interest.

4.2 Setup and Assumptions

We consider a panel analogue of Hansen (1995), featuring unit-specific, stationary covariates in addition to the observations of interest. Cross-sectional dependence is accounted for using unobserved common factors in the spirit of the PANIC approach of Bai and Ng (2004). This leaves us with, for $i = 1, \dots, n$ and $t = 1, \dots T$,

$$y_{it} = \lambda_i F_t + E_{it}, \tag{4.1}$$

$$F_t = \rho F_{t-1} + f_t, \tag{4.2}$$

$$E_{it} = \rho_E E_{i,t-1} + \eta_{it}, \tag{4.3}$$

$$x_{it} = \gamma_i f_t + u_{it}, \tag{4.4}$$

where y is observed, F is an unobserved univariate common factor, E are unobserved idiosyncratic shocks and x is a stationary covariate. We assume zero starting values, i.e., $F_0 = 0$, $E_{i0} = 0$ and $u_{i0} = 0$ for all $i = 1, \dots, n$.[7]

In a panel context, stationary covariates have only been used in the unit-root testing problem for the idiosyncratic components E. Becheri, Drost, and Van den Akker (2015b) consider the case of observed factors.[8] Thus, their observed covariates are cross-section common and not unit-specific. On the other hand, Juodis and Westerlund (2018), consider unit specific covariates that are cross-sectionally independent and for each unit exhibit correlation

[7] When starting values are large, these can have a significant influence on the power of panel unit-root tests, see Müller and Elliott (2003) for a general discussion and Aristidou, Harvey, and Leybourne (2017) and Westerlund (2013) for the special case of covariate-augmented unit-root tests. As the goal of our chapter is to study the implications of observing panel data, we focus on the most studied case of zero (or small) starting values.

[8] Becheri, Drost, and Van den Akker (2015b) consider the influence of observing the factor innovations f_t in an alternative setup where the factor model is in the innovations and the factors and idiosyncratic parts are of the same order on integration, on the power of unit root tests for the observations Y.

with the error term η_{it}. We assume an intermediate case, where the covariates as well as the observations of interest have cross sectional dependence based on a factor structure. Moreover, we focus on the testing problem for the factor F. Equation (4.4) assumes that x satisfies a factor model, with f, the innovation of the factor under investigation, being one of the factors.[9] Depending on the application, one might not be comfortable with the assumption that the factor innovation of interest enters the covariate equation directly. Therefore, we also consider an alternative specification

$$x_{it} = \gamma_i(f_t + g_t) + u_{it}, \tag{4.5}$$

where g can be considered an extra factor that happens to have the same loadings as f. Alternatively, g can be interpreted as the difference between the factor that drives the correlation in the covariate equation and that in the observations. Contrary to having separately identifiable additional factors, i.e., factors that do not share their loadings with f_t, having a factor with exactly the same loadings significantly changes the testing problems and entails an entirely new analysis (see Section 4.4).

Throughout, the null hypothesis is $\rho = 1$, i.e., F and (therefore) each panel unit has a unit root, with local alternatives

$$\rho = 1 + \frac{h}{Tn^{\nu}}. \tag{4.6}$$

The coefficient of the idiosyncratic parts, ρ_E, is assumed to be fixed. We choose ν depending on the specification to ensure alternatives contiguous to the null hypothesis. In the first specification, where the covariate satisfies (4.4), we obtain $\nu = 1/2$, i.e., a faster convergence rate thanks to the presence of the stationary covariates, see Theorem 4.3.1. However, the robustified, second specification (using (4.5)) means that even as $n \to \infty$, f cannot be perfectly estimated, so, as we formally show in Section 4.4, the time-series power envelope and a convergence rate with $\nu = 0$ apply.

---

9 In this chapter we consider the case of a single factor, but the results are expected to be robust to additional factors entering (4.4). However, estimating these is non-trivial and left for future work.

4.2.1 Notation

Collect the panel units in the $T \times n$ matrices Y, E, x, η, and u. Also, write Y_i, E_i, x_i, η_i, and u_i for their ith columns, respectively. Introduce the $n \times 1$ matrix Λ to contain the factor loadings λ_i and the $n \times 1$ matrix Γ to contain the covariate loadings γ_i. Finally, let $F = \{F_t\}_{t=1}^{T}$ and define f analogously. Also, denote $F_{-1} = (F_0, \ldots, F_{T-1})'$ and $\hat{F}_{-1} = (\hat{F}_0, \ldots, \hat{F}_{T-1})'$ and note that with this notation, we have $y = F\Lambda' + E$ and $x = f\Gamma' + u$. The $T \times T$ covariance matrices of f, η_i, and u_i are denoted by Σ_f, $\Sigma_{\eta,i}$ and $\Sigma_{u,i}$, respectively, with long-run variances ω_f^2, $\omega_{\eta,i}^2$, and $\omega_{u,i}^2$.

For a matrix C, let $\|C\|_F$ denote its Frobenius norm and $\|C\|_{\text{spec}}$ its spectral norm. Henceforth, A will denote a cumulative sum operator (a $T \times T$ matrix with ones below the diagonal and zeros on and above) so that $F_{-1} = A\Delta F$, while I denotes an identity matrix. Throughout, W_1 and W_2 are independent standard Brownian motions, (time) integrals of these are understood to be from zero to one and we abbreviate $\int_0^1 W_1(t)\mathrm{d}t$ by $\int W_1$. By $\rightarrow$ we denote convergence in $\mathbb{R}^d$, $\Rightarrow$ denotes convergence in distribution, while combining any convergence mode with the statement $(n, T \rightarrow \infty)$ refers to n and T going to infinity jointly as in Phillips and Moon (1999).

4.2.2 Basic Assumptions

In line with the existing optimality literature on panel unit-root tests, we restrict ourselves to Gaussian innovations.[10] In our baseline specification (4.1)–(4.4) we allow for very general serial correlation. This is nontrivial from a methodological point of view, as we deal with both a growing number of time series and cross-sectional observations, so that, despite the imposed Gaussianity, we are dealing with infinite-dimensional nuisance parameters.

Assumption 4.1 *For each i, the covariate innovations $\{u_{it}\}$, factor innovations $\{f_t\}$, $\{g_t\}$ and idiosyncratic innovations $\{\eta_{it}\}$ are stationary Gaussian*

[10] Without distributional assumptions we could not write out the likelihood ratios that form the basis of our asymptotic analysis. Thus, Gaussianity is required for our optimality results. Nevertheless, we expect the test statistics to have good sizes even with non-Gaussian observations.

time series independent of each other. Moreover, the spectral densities f_i of the u_{it} are uniformly bounded away form zero and twice continuously differentiable with uniformly bounded second derivative, i.e.,

$$\inf_{i,m} f_i(m) > 0, \quad \sup_{i,m} |f_i''(m)| < \infty. \tag{4.7}$$

Similarly, the spectral density of the factors is bounded away from zero and their autocorrelation functions γ_f satisfy

$$\sum_{m=0}^{\infty} (m+1)|\gamma_f(m)| < \infty.$$

Remark 4.2.1 *The condition on the second derivative of the spectral density ensures a uniform (across panel units) decay of the autocorrelation function, while the bound on the infimum is necessary to obtain bounds on the inverse of the covariance matrices. If the $\{u_{it}\}$ were causal ARMA processes, these conditions would be satisfied as long as the roots are uniformly (across cross-section units) bounded away from the unit circle.*

Assumption 4.2 *The factors are strong, i.e., there exist positive definite constants Ψ_Λ and Ψ_Γ such that $\lim_{n\to\infty} \frac{1}{n}\Lambda'\Lambda = \Psi_\Lambda$ and $\lim_{n\to\infty} \frac{1}{n}\Gamma'\Gamma = \Psi_\Gamma$. Moreover, we have, for some $\omega_{u,\Gamma}^{-2} > 0$, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\gamma_i^2}{\omega_{u,i}^2} \to \omega_{u,\Gamma}^{-2}.$$

4.3 Limit Experiments and Feasible Tests

We now analyze the statistical problems related to testing $h = 0$, i.e., the presence of unit roots in Y induced by a unit root in the common factor. This analysis yields power envelopes and suggests tests with certain optimality properties. Section 4.3.1 considers the experiment where the factor of interest is observed. Section 4.3.2 adapts the resulting procedure for use with an estimated factor and Section 4.3.3 relaxes the requirement that the idiosyncratic parts of the covariate are independent of those in the main covariate.

4.3.1 Observed Factors

In the presence of unit-specific covariates, optimal unit-root tests are not available in the literature, even if the factor is observed.[11] Thus, in this section, we treat the simplest case: a single *observed* factor F that drives both the cross-sectional correlation in the Y as well as that in the x. Moreover, we assume all nuisance parameters are known; this will be relaxed later.

If the factor of interest F is observed, the observations Y are not relevant for the testing problem, i.e., observing Y in addition to F and x yields no additional power.[12] They will, however, be used later to approximate unobserved factors and nuisance parameters.

Denote by P_h the joint law of F, x, and Y under (4.1)–(4.4) and (4.6). The following proposition characterizes the limit experiment associated with the testing problem $h = 0$ vs $h < 0$.

Proposition 4.3.1 *Under Assumptions 4.1 and 4.2, the likelihood ratio in the experiment with a single, observed factor and known nuisance parameters, satisfies, under* P_0*,*

$$\log \frac{\mathrm{d}P_h}{\mathrm{d}P_0} = h\Delta_{n,T} - \frac{1}{2}h^2 J_{n,T} + o_P(1),$$

with

$$\Delta_{n,T} = \frac{-1}{\sqrt{n}T} \sum_{i=1}^{n} \sum_{t=1}^{T} \frac{\gamma_i}{\omega_{u,i}^2} F_{t-1}(x_{it} - \gamma_i \Delta F_t)$$

and

$$J_{n,T} = \frac{1}{nT^2} \sum_{i=1}^{n} \frac{\gamma_i^2}{\omega_{u,i}^2} \sum_{t=1}^{T} F_{t-1}^2.$$

Moreover, under P_0*,* $(\Delta_{n,T}, J_{n,T}) \Rightarrow (\Delta, J)$ *as* $(n, T \to \infty)$*, where* $J = \omega_f^2 \omega_{u,\Gamma}^{-2} \int W_1^2$*, and* $\Delta | J \sim N(0, J)$*.*

[11] Hansen (1995) has studied the testing problem in a time-series context, i.e., with a finite number of covariates. Here, however, we assume that covariates are unit specific.

[12] This changes when we allow for correlation between the u and η, see Section 4.3.2.

Proposition 4.3.1, proved in Section 4.A.1, implies that the testing problem is of the LAMN type. Although this limit experiment is well understood, this implies that no asymptotically UMP test exist, see Jeganathan (1995). If all nuisance parameters were known, we could, however, conduct simple inference by using the asymptotic normality of $t_{n,T} := \Delta_{n,T}/\sqrt{J_{n,T}}$, leading to optimal invariant[13] tests. Note that in $\Delta_{n,T}$ and $J_{n,T}$ we have replaced covariance matrices by long-run matrices as even with an observed factor, estimating the entire variance-covariance matrix of the innovations u and f is impractical. Thanks to this replacement, as we show in Section 4.3.2, we can implement tests based on $t_{n,T}$ that do not require any knowledge about nuisance parameters.

Proposition 4.3.2 states the distributions under alternatives. As we are dealing with a mixed Gaussian shift limit experiment instead of the standard Gaussian shift limit experiment, the distribution contains stochastic integrals and no simple closed form solutions for powers exist. However, Figure 4.1, plots powers for varying values of $\omega_f \omega_{u,\Gamma}^{-1}$.

Proposition 4.3.2 *Suppose Assumptions 4.1 and 4.2 hold. Under P_h,*

$$t_{n,T} \Rightarrow Z - h\sqrt{\omega_f^2 \omega_{u,\Gamma}^{-2} \int W_1^2 \mathrm{d}t},$$

as $(n, T \to \infty)$, where Z is standard normally distributed and independent of W_1.

Our LAMN result facilitates a simple proof of Proposition 4.3.2 using only the properties of the limit experiment, see Section 4.A.2. In the simple case of i.i.d. time series without cross sectional heterogeneity and variances σ_u^2 and σ_f^2, the result can also be derived using direct calculations that might be more intuitive to some readers. As $\Delta F_t = f_t + \frac{h}{\sqrt{n}T} F_{t-1}$ we have

$$\Delta_{n,T} = \frac{1}{\sqrt{n}T\sigma_u^2} \sum_{t=1}^{T} F_{t-1} \sum_{i=1}^{n} \gamma_i u_{it} - \frac{h}{nT^2\sigma_u^2} \sum_{i=1}^{n} \gamma_i^2 \sum_{t=1}^{T} F_{t-1}^2. \tag{4.8}$$

[13] That is, tests that achieve nominal size even conditional on $J_{n,T}$.

Note that, from Lemma 4.A.1 in Section 4.A.6, under all alternatives, the limiting distribution of the first part equals that of $\Delta_{n,T}$ under the null hypothesis. Further, note that the second part equals $J_{n,T}$ and, again using Lemma 4.A.1 its distribution is the same irrespective of the alternative,[14] i.e.,

$$J_{n,T} \Rightarrow \Psi_\Gamma \frac{\sigma_f^2}{\sigma_u^2} \int W_1^2 \mathrm{d}t.$$

Combining yields

$$t_{n,T} = \frac{1}{\sqrt{n}T\sigma_u^2\sqrt{J_{n,T}}} \sum_{t=1}^{T} F_{t-1} \sum_{i=1}^{n} \gamma_i u_{it} - h\sqrt{J_{n,T}} \Rightarrow Z - h\sqrt{\Psi_\Gamma \frac{\sigma_f^2}{\sigma_u^2} \int W_1^2 \mathrm{d}t}.$$

4.3.2 Tests when the Factor is Unobserved

In the previous section we have established how to conduct optimal tests when the factor is observed and that we can exploit the covariates to obtain power in $\sqrt{n}T$ instead of T neighbourhoods of the unit root. We now show that even for testing an unobserved factor, we can achieve the improved rate. If the idiosyncratic parts are stationary, the power envelope and the power of our proposed tests are the same as if the factor was observed. If $\rho_E = 1$, however, we observe a modest loss in asymptotic power. Note that this loss in power is atypical: both in the no-covariate case as studied in Bai and Ng (2004) and in our robust specification in Section 4.4 using estimated factors in place of observed ones does not change the limiting distributions of the test statistics, even when the idiosyncratic parts are nonstationary. In our setting, however, not observing the factors necessitates a size correction, which leads to a loss of power. In this section we discuss how to implement a version of the test that can be applied to unobserved factors and compare the power loss relative to the observed factor case for different factor estimates.

Our method of choice estimates ΔF by applying principal components to Δy. The resulting estimates, $\Delta\hat{F}$, are then used in combinations with the covariates x to estimate Γ using OLS. Next, consider the limiting distribution of

[14] This is thanks to our alternatives being in $\sqrt{n}T$ neighbourhoods around the unit root. See Phillips (1987b) for the behaviour in T regions around the unit root. The present case is obtained by letting the local parameter converge to zero.

a version of $t_{n,T}$ based on those estimates. To estimate the factors sufficiently well we need to make additional assumptions on the relative size of n and T. This is common in the literature, as these unit-root tests are typically applied to panels of macroeconomic time-series.

Assumption 4.3 *We have $n/T \to 0$ as $(n, T \to \infty)$.*

To create tests that can be implemented in practice, we also need estimates of the long-run variances. These requirements are standard, and satisfied, for example, by kernel estimates, see Remark 2.9 in Moon, Perron, and Phillips (2014).

Assumption 4.4 *We have estimators $\hat{\omega}^2_{u,i}$ and $\hat{\omega}^2_{\eta,i}$ of the long-run variances of the $\{u_{it}\}$ and $\{\eta_{it}\}$, respectively, satisfying, under P_0, as $(n, T \to \infty)$,*

$$\max_{i=1,\ldots,n} \mathbb{E}(\hat{\omega}^2_{u,i} - \omega^2_{u,i})^2 = o(1/n), \quad \max_{i=1,\ldots,n} \mathbb{E}(\hat{\omega}^2_{\eta,i} - \omega^2_{\eta,i})^2 = o(1/n).$$

Theorem 4.3.1 *Let $(\Delta\hat{F}, \hat{\Lambda})$ be the principal components estimator based on Δy and let*

$$\hat{\Gamma}' = (\Delta\hat{F}'\Delta\hat{F})^{-1}\Delta\hat{F}'x$$

with its i-th element denoted $\hat{\gamma}_i$. Assume $\kappa^2_\eta := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \lambda_i^2\omega^2_{\eta,i}$ exists and define the estimators $\hat{\kappa_\eta}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\lambda}_i^2\hat{\omega}^2_{\eta,i}$, $\hat{\omega_{u,\Gamma}}^{-2} = \frac{1}{n}\sum_{i=1}^n \frac{\hat{\gamma}_i^2}{\hat{\omega}^2_{u,i}}$, and $\hat{\Psi}_\Lambda = \frac{1}{n}\hat{\Lambda}'\hat{\Lambda}$, using the long-run variance estimators of Assumption 4.4. Let

$$\tilde{t}_{n,T} = \hat{t}_{n,T}\Big/\sqrt{1 + \hat{\Psi}_\Lambda^{-2}\hat{\omega_{u,\Gamma}}^{-2}\hat{\kappa_\eta}^2},$$

where $\hat{t}_{n,T} = \frac{\hat{\Delta}_{n,T}}{\sqrt{\hat{J}_{n,T}}}$ with

$$\hat{\Delta}_{n,T} = \frac{-1}{\sqrt{n}T}\sum_{i=1}^n\sum_{t=1}^T \frac{\hat{\gamma}_i}{\hat{\omega}^2_{u,i}}\hat{F}_{t-1}(x_{it} - \Delta\hat{F}_t\hat{\gamma}_i) \text{ and}$$

$$\hat{J}_{n,T} = \frac{1}{nT^2}\sum_{i=1}^n \frac{\hat{\gamma}_i^2}{\hat{\omega}^2_{u,i}}\sum_{t=1}^T \hat{F}^2_{t-1}.$$

Then, under Assumptions 4.1–4.4, with $\rho_E = 1$ and under P_h, as $(n, T \to \infty)$,

$$\tilde{t}_{n,T} \Rightarrow Z - h\frac{\sqrt{\omega_f^2\omega_{u,\Gamma}^{-2}\int W_1^2\mathrm{d}t}}{\sqrt{1 + \Psi_\Lambda^{-2}\omega_{u,\Gamma}^{-2}\kappa^2_\eta}}.$$

The proof is provided in Section 4.A.3. Theorem 4.3.1 implies that we can again use standard normal inference, based on the adapted test statistic $\tilde{t}_{n,T}$.

The loss in power due to not observing the factors can be attributed to the term

$$F'_{-1}\Delta E\Lambda(\Lambda'\Lambda)^{-1}\Gamma'\Omega^{-1}\Gamma,$$

which constitutes part of the difference between the central sequence $\Delta_{n,T}$ and its estimated counterpart $\hat{\Delta}_{n,T}$. Under $\rho_E = 1$, $\frac{1}{\sqrt{n}}\Delta E\Lambda$ gives rise to κ_η^2 as its long-run variance. However, if E is stationary and ΔE is thus overdifferenced, the corresponding long-run variance is zero, making a size correction superfluous.[15] Thus we obtain the following corollary.

Corollary 4.3.1 *Under the assumptions of Theorem 4.3.1 but with $\rho_E = 0$, we have, under P_h, as $(n, T \to \infty)$,*

$$\hat{t}_{n,T} \Rightarrow Z - h\sqrt{\omega_f^2\omega_{u,\Gamma}^{-2}\int W_1^2 \mathrm{d}t}.$$

Therefore, whether $\hat{t}_{n,T}$ or $\tilde{t}_{n,T}$ is to be used depends on whether the idiosyncratic parts are stationary. If this is unclear, one can use $\tilde{t}_{n,T}$ for robust (but potentially conservative) inference.

Remark 4.3.1 *To avoid identification problems, in Theorem 4.3.1, we immediately replace both factor and their loadings by estimates. Its proof demonstrates that the power loss occurs due to having to estimate the factor in $(x_{it} - \Delta\hat{F}_t\hat{\gamma}_i)$ in the central sequence. The estimation of the loadings is then adaptive, i.e., we show that the terms due to estimating the loadings converge to zero. Note that, thanks to Le Cam's First Lemma and our LAMN result, we only have to show this adaptivity under the null hypothesis. Convergence to zero of these terms under the local alternatives is then implied by contiguity.*

Remark 4.3.2 *Instead of using principal components, one could also use cross-sectional averages as a simple way to show that unit-root tests are possible in $\sqrt{n}T$ neighbourhoods of the unit root even when the factor is not*

[15] More formally, note that if $E = \eta$, then $F'_{-1}\Delta E_i = \sum_{t=0}^{T-1}\Delta F_t(\eta_{iT} - \eta_{it}) = O_P(\sqrt{T})$.

observed. For panel unit-root tests in the PANIC framework this is a well-documented technique (see Reese and Westerlund (2015)) and it is often considered to be competitive with principal components. However, in this testing problem, we can clearly identify the increase in asymptotic power due to using principal components instead of simple averages when $\rho_E = 1$.

To gain intuition, again assume i.i.d. innovations and normalize the average of the λ_i *to one. Consider, in addition to* $t_{n,T}$ *and* $\tilde{t}_{n,T}$, *a version of* $\tilde{t}_{n,T}$ *that is based on cross-sectional averages. To ensure a standard normal limiting distribution under the null hypothesis, this requires a different normalization than when using principal components: the asymptotic distributions of all three test statistics under alternative* h *can be written as*

$$Z - h\sqrt{\frac{\sigma_f^2 \Psi_\Gamma}{\sigma_u^2 + p}} \int W_1^2 \mathrm{d}t,$$

where $p = 0$ *when the factor is observed,* $p = \sigma_\eta^2 \Psi_\Gamma \Psi_\Lambda^{-1}$ *in case of principle components, and* $p = \sigma_\eta^2 \Psi_\Gamma$ *when using averages. As expected, the efficiency loss due to using averages over principal components depends on the variability of the factor loadings in the observations* (Ψ_Λ)*, while the loss from not observing the factor depends mainly on the variance of the idiosyncratic terms* (σ_η^2)*. For general weights* w*, satisfying* $w'\lambda = 1$*, the variance of the weighted* η *is* $w'w$*. If* λ *were known, one could use the OLS weights* $\lambda/(\lambda'\lambda)$ *to obtain a variance of* $(\lambda'\lambda)^{-1}$ *instead of 1. This explains why the variance correction is divided by* Ψ_Λ *and it also shows that there is no asymptotic loss in efficiency from the* λ *being estimated by principal components rather than being observed. Figure 4.1 compares these asymptotic powers for a range of typical parameter values.*

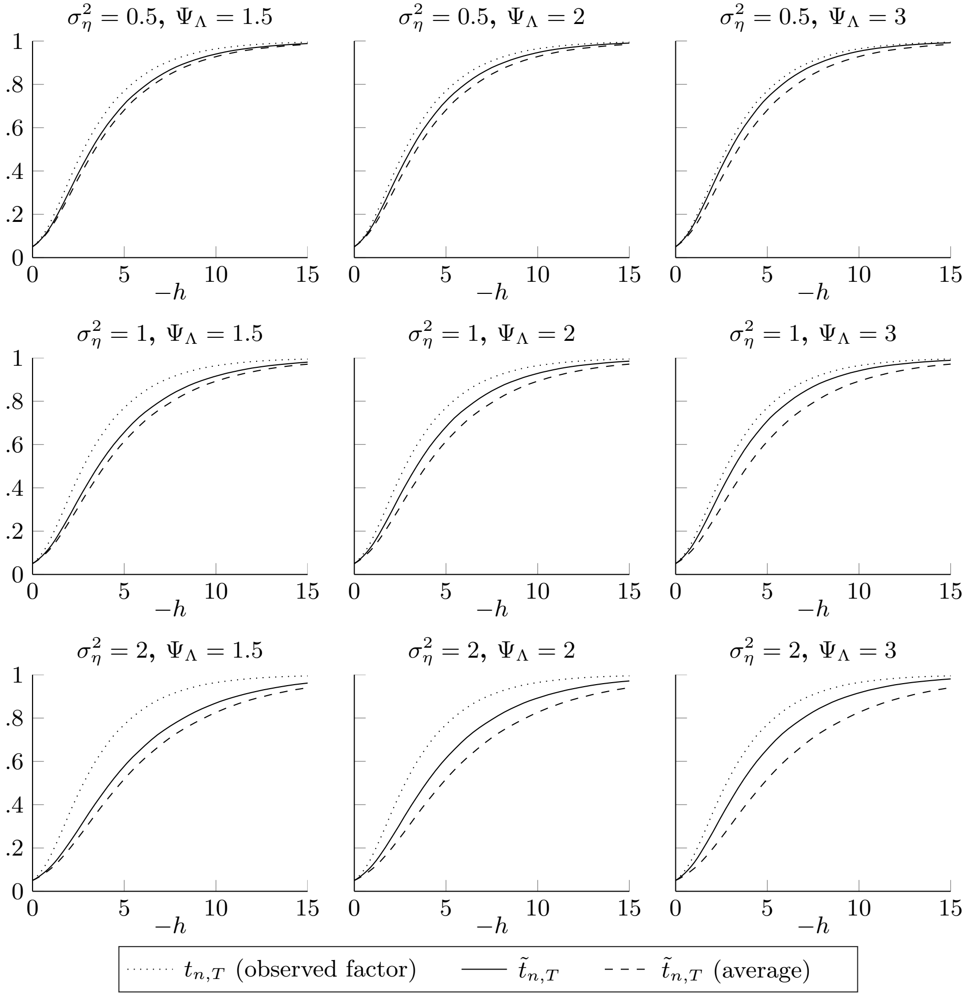


Figure 4.1: Asymptotic power of unit-root t-tests as a function of $-h$ for varying values of σ_η^2 and Ψ_Λ. Here $\sigma_f^2 = \sigma_u^2 = 1$. The dotted line represents the observed-factor benchmark, the solid line presents the asymptotic power of $\tilde{t}_{n,T}$ and the dashed line represents a version of $\tilde{t}_{n,T}$ where the factor is estimated using cross-sectional averages rather than principal components.

4.3.3 Correlated Innovations

We now consider the more complicated case where the idiosyncratic parts of the observations are correlated with those of the covariate. To allow for such correlation in a simple way we augment (4.3) as

$$E_{it} = \rho_E E_{i,t-1} + \eta_{it} + \beta u_{it}. \tag{4.9}$$

Note that now it becomes worthwhile to exploit the observations y to improve on the tests even when the factor of interest is observed. The aim of this section is to demonstrate that exploiting the correlation between the error terms leads only to minor increases in power, so we focus on the observed factor case. For this more complicated setup, we also consider the simplifying Assumption 4.5. The results in this section can be extended to allow for serial correlation along the same lines as in Section 4.3.1.

Assumption 4.5 *The covariate innovations* $\{u_{it}\}$*, factor innovations* $\{f_{kt}\}$ *and idiosyncratic innovations* $\{\eta_{it}\}$ *are i.i.d. normally distributed, with variances* σ_u^2*,* σ_f^2*, and* σ_η^2*, respectively. Moreover,* $\rho_E = 0$ *and the factors are independent of* $\{u_{it}\}$ *and* $\{\eta_{it}\}$*.*

Proposition 4.3.3 below demonstrates that, based on a similar t-test, one can again use standard normal critical values for testing the stationarity of the observed factor. The proof, found in Section 4.A.4, also demonstrates that the experiment is LAMN and the proposed t-test is thus conditionally optimal. This test is again motivated by the limit-experiment, which is of the same type as the one in Section 4.3.1. However, thanks to the correlation between the two equations, this test can also exploit the fact that, under the hypothesis $\eta = E_{it} - \beta(x_{it} - \gamma_i \Delta F_t)$, but that the right-hand side will be correlated with F_{t-1} under (local) alternatives. This leads to an increase in asymptotic power, depending on the magnitude of β, as shown in Figure 4.2. Some limited gains are possible when σ_η^2 is small relative to σ_u^2 and σ_f^2.

Proposition 4.3.3 *Suppose Assumptions 4.2 and 4.5 hold and let*

$$t_{n,T}^{\beta} := \Delta_{n,T}^{\beta} / \sqrt{J_{n,T}^{\beta}},$$

with

$$\Delta^{\beta}_{n,T} = \tilde{\Delta}_{n,T} - \beta \frac{1}{\sqrt{n}T\sigma^2_\eta} \sum_{i=1}^{n} \sum_{t=1}^{T} \gamma_i F_{t-1}(E_{it} - \beta(x_{it} - \gamma_i \Delta F_t))$$

and

$$J^{\beta}_{n,T} = \frac{1}{nT^2} \sum_{i=1}^{n} \left(\frac{1}{\sigma^2_u} + \frac{\beta^2}{\sigma^2_\eta} \right) \gamma_i^2 \sum_{t=1}^{T} F^2_{t-1}.$$

Under P_h, and (4.9)*, as* $(n, T \to \infty)$*,*

$$t^{\beta}_{n,T} \Rightarrow Z - h \sqrt{\Psi_\Gamma \left(\frac{\sigma^2_f}{\sigma^2_u} + \beta^2 \frac{\sigma^2_f}{\sigma^2_\eta} \right) \int W_1^2 \mathrm{d}t},$$

where Z is standard normally distributed and independent of W_1.

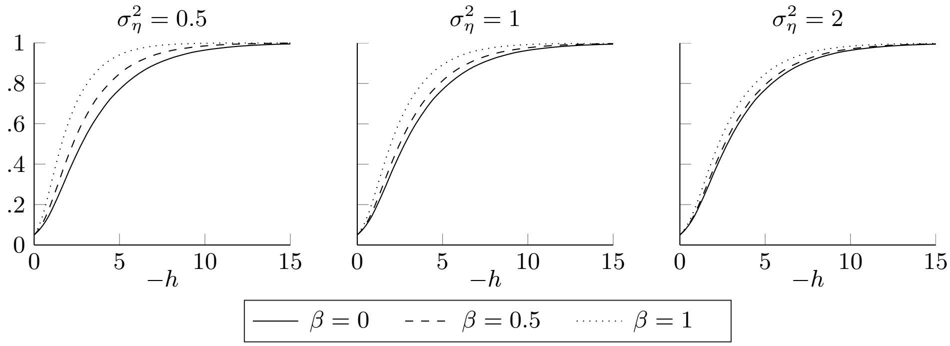


Figure 4.2: Asymptotic power of observed factor unit-root test $t^{\beta}_{n,T}$ for varying values of σ^2_η and β. Here $\sigma^2_f = \sigma^2_u = 1$.

Remark 4.3.3 *Here we allow for any fixed value of β. Note that while the gains in asymptotic power are small for the magnitudes of β that are likely to be encountered in practice, as $\beta \to \infty$, the power for any fixed h converges to 1.*

Remark 4.3.4 *Proposition 4.3.3 assumes that $\rho_E = 0$, i.e., the idiosyncratic parts are stationary. If the idiosyncratic parts E are known to have an autoregressive unit root, one can simply replace E_{it} by ΔE_{it} in the central sequence.*

The intermediate case, however, where the persistence parameter of the idiosyncratic parts is local-to-unity at the contiguous rate $\sqrt{n}T$ is left for further research. [16]

Remark 4.3.5 *Note that the correlation between E and u also implies a typical regression relation between x and Δy. In particular, under the null hypothesis and with $\rho_E = 1$, we obtain*

$$\Delta y_{it} = (\lambda_i - \beta\gamma_i)\Delta F_t + \beta x_{it} + \eta_{it}.$$

4.4 Robust Specification

In this section we consider the alternative covariate setup as introduced in (4.5). That is, we introduce a new factor

$$b_t := f_t + g_t, \tag{4.10}$$

that can be considered as the relevant factor in the covariate equation. b_t is allowed to differ from the factor that is to be tested for a unit root, with the perturbation g_t representing their difference. The assumptions on g_t are analogous to Assumption 4.5.

Assumption 4.6 *The covariate factor innovations b_t are i.i.d. normally distributed, with variance σ_g^2 and independent of all other innovations.*

Similarly to the previous section, we first consider an auxiliary model, where the factor is observed. In this setup it turns out to be beneficial to assume that b_t is observed as well. This will be relaxed later, but for now brings

16 When $\beta \neq 0$, we need to remove the influence of the x_{it} before removing the factors. However, to understand how x_{it} enters the observations y_{it} in this setup we need to to know ρ_E with sufficient accuracy. If, for example, we impose a unit root in the E equation whereas in fact $\rho_E = 0.5$, f_t directly enters the y equation. So when we estimate the factors based on y, we in fact do not only obtain F but also f, which leads to size distortions for the unit-root tests based on those estimates. Therefore, if ρ_E was localized at the contiguous rate it becomes a non-adaptive nuisance parameter. The joint likelihood ratio then involves both equations. The joint experiment appears to also be of the LAMN type, but the resulting test statistics are more complicated.

one back to the univariate covariate setup as studied by Elliott and Jansson (2003) and Hansen (1995). While Hansen (1995) has suggested adding the covariate b_t to ADF regressions, Elliott and Jansson (2003) derive point-optimal tests. However, in the spirit of the previous sections, we do not start from a specific test statistic, but first investigate the structure of the limit experiment. We show that the limit experiment in the time-series covariate setup is LABF and that it gets closer to the LAMN case considered in the previous section as the influence of the observed factor grows.[17] This observation motivates the use of a simple t-test (albeit with slightly adjusted critical values) that turns out to be competitive in power with the point optimal tests proposed by Elliott and Jansson (2003) but more robust to the choice of alternative. It turns out that this test is equivalent to the approach of Hansen (1995) and can thus be easily implemented using OLS regression. In Proposition 4.4.2 we exploit our expansion of the likelihood ratio to show that both tests can be applied to estimated common factors with no loss in asymptotic power.

In this more robust specification, we localize the autoregressive parameter of the factor as $\rho_1 = 1 + \frac{h}{T}$ and denote by P_h the joint law of Y, x, F, and b under (4.2) and (4.10). The following proposition characterizes the limit experiment associated with the testing problem $h = 0$ vs $h < 0$. The result is immediate from the Gaussian likelihood ratio and we omit the proof.

Proposition 4.4.1 *Under Assumptions 4.1, 4.2, 4.5 and 4.6, the likelihood ratio in the experiment with an observed factor, known nuisance parameters, and a known covariate b, satisfies*

$$\log \frac{\mathrm{d}P_h}{\mathrm{d}P_0} = h\tilde{\Delta}_T - \frac{1}{2}h^2 \tilde{J}_T, \tag{4.11}$$

with

$$\tilde{\Delta}_T = \frac{1}{T}\sum_{t=1}^{T} F_{t-1}\left(\frac{\Delta F_t}{\sigma_f^2} + \frac{\Delta F_t - b_t}{\sigma_g^2}\right)$$

[17] Indeed the DGP in Section 4.3 corresponds to $R^2 = 1$. However, note that the contiguity rate is different, and that the LAMN result here arises from an entirely different sequence of experiments. This also implies that the optimality properties of the tests developed in this section for $R^2 < 1$ do not carry over to the case of Section 4.3.

and

$$\tilde{J}_T = \frac{1}{T^2}\sum_{t=1}^{T} F_{t-1}^2\left(\frac{1}{\sigma_f^2}+\frac{1}{\sigma_g^2}\right).$$

Note that, under the null hypothesis, with W_1 and W_2 being independent Brownian motions,

$$\tilde{\Delta}_T \Rightarrow \int W_1 \mathrm{d}(W_1 + \frac{\sigma_f}{\sigma_g}W_2) \text{ and } \tilde{J}_T \Rightarrow \left(1+\frac{\sigma_f^2}{\sigma_g^2}\right)\int W_1^2 \mathrm{d}t,$$

as $T \to \infty$. Thus, as σ_f^2 gets large, relative to σ_g^2 (or, in the notation of Elliott and Jansson (2003), $R^2 = \frac{\sigma_f^2}{\sqrt{\sigma_f^2+\sigma_g^2}\sigma_f}$ goes to one), i.e., the influence of the covariate increases, the limit experiment mimics more and more the LAMN case of the previous sections.[18] This suggests to use again the simple t-statistic

$$t_T := \frac{\tilde{\Delta}_T}{\sqrt{\tilde{J}_T}},$$

as this will yield optimal invariant tests in the setting of the previous section where $R^2 \to 1$. Note, on the other hand, that t_T approaches the regression Dickey Fuller statistic as $R^2 \to 0$. Therefore, we expect this test statistic to behave reasonably well also for intermediate R. Figure 4.3 compares the powers of t_T to that of the point-optimal tests. As expected, the CADF test t_T performs particularly well when R^2 is large.

Remark 4.4.1 *For easier intuition we have again treated this case based on i.i.d. errors. However, due to the standard ingredients of our test, existing results for time series (as stated, for example, in Elliott, Rothenberg, and Stock (1996)) imply that our test could be applied to serially correlated data by replacing variances by long-run variances and subtracting the correction term δ_f/ω_f^2, where δ_f and ω_f^2 are estimates of the one-sided long-run variance and*

[18] To see the relation to Proposition 4.3.1, note that $\int W_1 \mathrm{d}W_2$ indeed has a normal distribution conditional on $\int W_1^2 \mathrm{d}t$. As σ_f^2 gets large relative to σ_g^2, the term $\int W_1 \mathrm{d}W_1$ becomes negligible. It is also evident that as $\sigma_f/\sigma_g \to \infty$, one needs a different standardization to obtain a non-degenerate limiting distribution. This corresponds to the analysis in Section 4.3, which uses a different contiguity rate.

long-run variance based on ΔF, respectively. Note that this is the same correction term as in the no-covariate case, as the contribution of $\int W_1 \mathrm{d}W_2$ requires no correction term, c.f., Proposition 4.3.1. Although the test statistic is identical to that of Hansen (1995) in the i.i.d. case, this suggests an alternative way of handling serial correlation.

Remark 4.4.2 *Contrary to the model studied in Section 4.3, a correlation between the idiosyncratic parts, as in* (4.9)*, would not change the likelihood ratio or the attainable power. Thus, the results in this section apply regardless of whether β is zero or not.*

Remark 4.4.3 *To allow for multiple factors in the covariate equation, one could generalize* (4.5) *to*

$$x_{it} = \sum_{k=1}^{K} \gamma_{i,k}(f_t + g_{k,t}) + u_{it}.$$

With $b_{k,t} = f_t + g_{k,t}$ observed for $k = 1, \ldots, K$, and σ_k^2 denoting the variance of $g_{k,t}$, the central sequence and Fisher information would read

$$\begin{aligned}\tilde{\Delta}_T^K =& \frac{1}{T}\sum_{t=1}^{T} F_{t-1}\left(\frac{\Delta F_t}{\sigma_f^2} + \sum_{k=1}^{K}\frac{\Delta F_t - b_{k,t}}{\sigma_k^2}\right) \quad \textit{and} \\ \tilde{J}_T^K =& \frac{1}{T^2}\sum_{t=1}^{T} F_{t-1}^2\left(\frac{1}{\sigma_f^2} + \sum_{k=1}^{K}\frac{1}{\sigma_k^2}\right).\end{aligned}$$

Note that we have chosen this formulation of the generalized setup over a formulation with $K + 1$ independent factors, as this allows us to treat the $b_{k,t}$ as observed without worrying about identifying the single relevant factor.[19]

4.4.1 Implementing the tests with unobserved factors

In this section we demonstrate that $\tilde{\Delta}_T$ and $\tilde{J}_T$ can be approximated without actually observing F or b. This not only enables adaptive testing based on

[19] As principal components only identifies factors up to a rotation, imposing a rotation that eliminates f_t from $K - 1$ factors would make it harder to implement the resulting test statistic in practice.

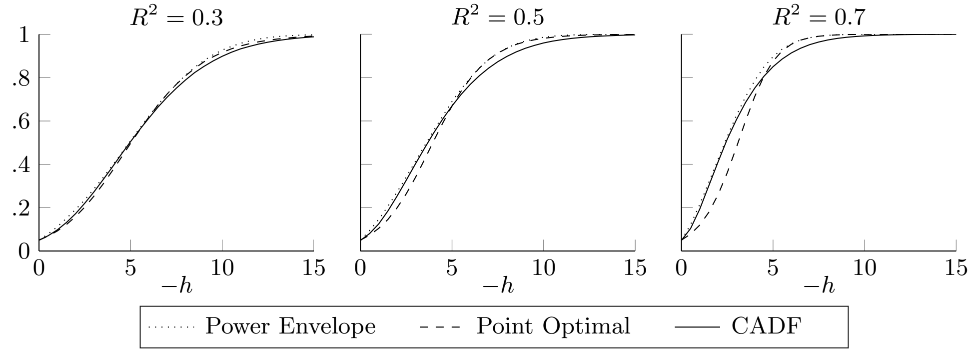


Figure 4.3: Local asymptotic powers of unit-root tests and pointwise power envelope. The 'Point Optimal' test of Elliott and Jansson (2003) is implemented for a fixed alternative $h = -7$ as recommended by the authors.

t_T, but also based on (4.11) for a fixed h, i.e., based on point-optimal tests. Contrary to the case in Section 4.3, here there is no change in the asymptotic distribution of the test statistics, and thus no loss in power due to not observing the factors.

Proposition 4.4.2 *Reconsider the principal components estimator $\hat{\Lambda}, \hat{F}$ based on ΔF from Theorem 4.3.1 and similarly denote the principal components estimator based on x by $\hat{\Gamma}$ and $\hat{b}$.*[20] *With $\hat{\sigma}_f^2$ and $\hat{\sigma}_g^2$ consistent estimates of σ_f^2 and σ_g^2, respectively, we have, under Assumptions 4.1–4.5 and 4.6, and under P_h, as $(n, T \to \infty)$,*

$$\hat{\Delta}_T - \tilde{\Delta}_T = o_P(1) \text{ and } \hat{J}_T - \tilde{J}_T = o_P(1),$$

with

$$\hat{\Delta}_T = \frac{1}{T} \sum_{t=1}^{T} \hat{F}_{t-1} \left(\frac{\Delta \hat{F}_t}{\hat{\sigma}_f^2} + \frac{\Delta \hat{F}_t - \hat{b}_t}{\hat{\sigma}_g^2} \right) \text{ and}$$

$$\hat{J}_T = \frac{1}{T^2} \sum_{t=1}^{T} \hat{F}_{t-1}^2 \left(\frac{1}{\hat{\sigma}_f^2} + \frac{1}{\hat{\sigma}_g^2} \right).$$

[20] Note that with our identification assumptions it is necessary to get a common scaling of the principal component estimates across the two equations. We achieve this by recovering the scaling of $\hat{\Gamma}$ from a regression on $\Delta \hat{F}$ on x.

Proposition 4.4.2 implies that, even in this robust specification, there are sizeable gains from using observed covariates when testing unobserved common factors, as the covariate-based tests maintain their asymptotic powers as if the factors were observed. The following corollaries formalize this notion.

Corollary 4.4.1 *Let*

$$\hat{t}_T := \frac{\hat{\Delta}_T}{\sqrt{\hat{J}_T}}.$$

Then, under P_h and as $(n, T \to \infty)$, $\hat{t}_T$ has the same limiting distribution as t_T. Consider a test that rejects for small values of $\hat{t}_T$. Using the critical values from Table 1 in Hansen (1995), this test has correct asymptotic size and, as $(n, T \to \infty)$ and subsequently $R^2 \to 1$, its power converges to that of the optimal invariant test.

Corollary 4.4.2 *Let*

$$p^{\bar{c}} = \bar{c}\hat{\Delta}_T - \frac{1}{2}\bar{c}^2 \hat{J}_T.$$

A test rejecting for large values of $p^{\bar{c}}$ is asymptotically point optimal for testing $P_{\bar{c}}$ against P_0 and critical values from Table 1 in Elliott and Jansson (2003) lead to correct asymptotic sizes as $(n, T \to \infty)$.

Remark 4.4.4 *Natural estimators for the variances are $\hat{\sigma}_f^2 = \Delta\hat{F}'\Delta\hat{F}/T$ and $\hat{\sigma}_g^2 = (\Delta\hat{F}' - \hat{b})'(\Delta\hat{F} - \hat{b})/T$ and these indeed lead to the asymptotically well-behaved tests. However, in small samples and with estimated factors, other estimates lead to significantly higher powers. In particular, take*

$$\hat{\sigma}_f^2 = \frac{\hat{b}'\hat{b}\Delta\hat{F}'\Delta\hat{F} - (\Delta\hat{F}'\hat{b})^2}{T(\hat{b}'\hat{b} - \Delta\hat{F}'\hat{b})} \quad \text{and} \quad \hat{\sigma}_g^2 = \frac{\hat{b}'\hat{b}\Delta\hat{F}'\Delta\hat{F} - (\Delta\hat{F}'\hat{b})^2}{T\Delta\hat{F}'\hat{b}}. \tag{4.12}$$

These are the estimators implicitly used in the point-optimal tests proposed in Elliott and Jansson (2003) and although unintuitive, simple algebra confirms that these are indeed consistent estimators of the variances. Figure 4.6 shows that using these estimates over the natural estimator significantly improves small-sample powers for both CADF and point-optimal tests.

4.5 Finite-Sample Performance

In this section, we demonstrate to what extent observing stationary covariates can improve inference in finite samples. We treat the specifications of Sections 4.3 and 4.4 separately as the contiguity rates are different and even in small samples a comparison is not insightful.

For Section 4.3, we simulate (4.1)–(4.4), drawing the factor loadings λ and γ from independent normal distributions with means and variances equal to one. For the innovations, we consider i.i.d., AR(1) and MA(1) specifications (both with parameter 0.4). The long-run variances of the η and u are drawn from a log-normal distribution with mean and variance one (except where a different ω_u^2 is indicated, this refers to the mean of the lognormal distribution), to create heterogeneous long-run variances. The long-run variances of the f are unity. All results are based on 20 000 replications.

Table 4.1 presents sizes in the observed-factor benchmark $t_{n,T}$, for nominal 5% level tests. We consider the test statistics based on kernel-estimated long-run variances, using a Bartlett kernel with Newey-West bandwidth. It is evident that the proposed corrections for serial correlations work well even in small samples. Table 4.2 presents the corresponding results for an unobserved factor and also here the sizes are reasonably closer to nominal (but slightly under-sized rather than over-sized for small sample sizes). Finally, Table 4.3 presents sizes in the robust framework of Section 4.4. Here we treat all parameters the same as in the previous framework, while adapting the variance of g to yield the desired level of R^2. We use the variance estimates from (4.12) for better finite-sample powers, c.f. Remark 4.4.4. We observe that sizes are close to nominal levels, but slightly lower when n is large relative to T.

We now study the powers of the aforementioned tests, with the same DGP as before (but focusing on i.i.d. innovations) and local alternatives generated by (4.6). Figure 4.4 considers the tests from Section 4.3. Here the emphasis is on the finite-sample implications of not observing the factors. Recall that not observing the factors does lead to an asymptotic power loss here. However, when the variance of the error terms across the two equations is comparable, the loss of not observing the factor is not large: even in small samples the convergence to the asymptotic powers appears quite fast. In particular, the distance to the asymptotic power envelope in small samples appears to mirror the case of observed factors. Combining the insights from Figure 4.4 with the asymptotic powers presented in Figure 4.1 gives a good idea of the expected finite sample powers for other parameter values.

Figure 4.5 considers powers in the robust framework of Section 4.4. We compare our proposed CADF tests based on estimated factors to the estimated Dickey-Fuller tests proposed in Bai and Ng (2004). For a moderately correlated covariate of $R^2 = 0.5$, the proposed CADF test outperforms Dickey-Fuller tests even in small samples. However, when n is small, its power is significantly lower than the asymptotic power envelope. Figure 4.6 revisits the phenomenon discussed in Remark 4.4.4, highlighting how a particular set of variance estimates outperforms the more intuitive ones. For small n, the finite sample gains from using the more complicated variance estimates are sizeable.

It is evident that the power gains from exploiting the stationary covariates are very large even in small samples. This applies in the robust specification of Section 4.4, where the tests that take into account the covariate significantly outperform Dickey-Fuller tests even in the smallest sample sizes and with a moderately correlated covariate. In the baseline specification, where the factor contribution can be perfectly identified, the gains are very significant even in small samples and small correlations.

		i.i.d., $\omega_u^2 =$			AR(1), $\omega_u^2 =$			MA(1), $\omega_u^2 =$		
n	T	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
50	25	6.8	6.9	7.1	6.0	6.0	5.9	5.9	6.0	5.7
50	50	5.2	5.2	5.3	5.2	5.1	5.0	5.3	5.1	5.2
50	100	5.0	4.9	4.7	4.6	5.1	4.8	4.7	4.8	4.7
100	50	5.5	5.4	5.5	5.3	5.1	5.1	5.4	5.5	5.1
100	100	4.8	4.8	4.8	5.0	4.9	4.9	4.6	4.7	4.8
100	200	5.0	4.9	4.8	4.7	4.9	4.9	5.1	4.9	4.9
200	100	4.8	4.9	4.9	5.0	4.7	4.9	4.6	4.5	4.8
200	200	5.0	4.7	4.8	4.9	4.9	4.8	4.9	4.9	4.8
200	400	5.0	4.8	4.9	4.9	5.0	5.1	5.0	4.9	4.9

Table 4.1: Sizes (in percent) of nominal 5% level test based on $t_{n,T}$ (observed factor).

		i.i.d., $\omega_u^2 =$			AR(1), $\omega_u^2 =$			MA(1), $\omega_u^2 =$		
n	T	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
50	25	2.9	3.1	3.4	2.0	2.2	2.4	2.4	2.5	2.8
50	50	3.4	3.5	3.6	3.1	2.9	3.3	3.1	3.1	3.2
50	100	4.2	4.1	4.0	4.0	4.1	4.1	3.9	3.9	3.8
100	50	3.2	3.1	3.3	2.7	2.8	3.1	2.8	2.9	2.9
100	100	4.1	4.0	3.9	3.8	3.9	3.9	3.6	3.7	3.7
100	200	4.5	4.3	4.2	4.4	4.3	4.3	4.2	4.3	4.4
200	100	4.0	4.0	4.0	3.6	3.6	3.8	3.5	3.6	3.7
200	200	4.3	4.2	4.3	4.5	4.3	4.3	4.2	4.3	4.4
200	400	4.7	4.6	4.6	4.6	4.7	4.6	4.6	4.5	4.6

Table 4.2: Sizes (in percent) of nominal 5% level test based on $\tilde{t}_{n,T}$ (unobserved factor).

		i.i.d., $R^2 =$			AR(1), $R^2 =$			MA(1), $R^2 =$		
n	T	0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
50	25	3.6	4.8	6.2	3.4	3.7	4.3	3.9	4.2	4.6
50	50	4.3	5.5	6.5	3.8	4.2	4.8	4.2	4.3	4.8
50	100	5.1	6.3	7.1	4.1	4.5	5.1	4.2	4.6	5.4
100	50	3.2	4.0	4.8	3.8	4.0	4.5	3.7	3.8	4.3
100	100	3.9	4.8	5.3	3.9	4.3	4.6	4.1	4.2	4.5
100	200	4.8	5.1	5.6	4.5	4.7	5.3	4.3	4.4	5.2
200	100	3.3	4.0	4.4	3.8	4.0	4.8	3.6	3.8	4.3
200	200	3.8	4.1	4.6	4.2	4.4	5.0	4.1	4.6	4.8
200	400	4.5	4.4	5.0	4.3	4.6	5.3	4.2	4.4	5.1

Table 4.3: Sizes (in percent) of nominal 5% level test based on $\hat{t}_T$ (unobserved factor).

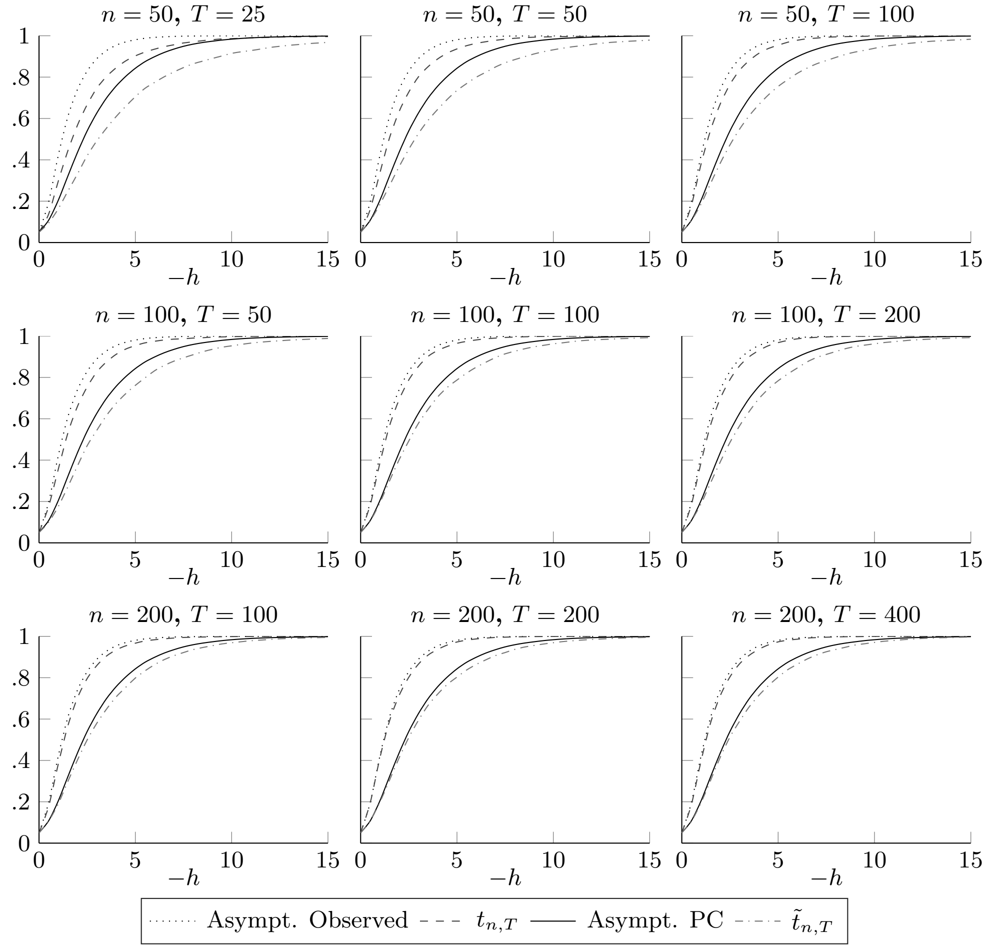


Figure 4.4: Size-corrected powers and asymptotic powers of unit-root tests as a function of $-h$ for varying sample sizes. All innovations are i.i.d. normally distributed.

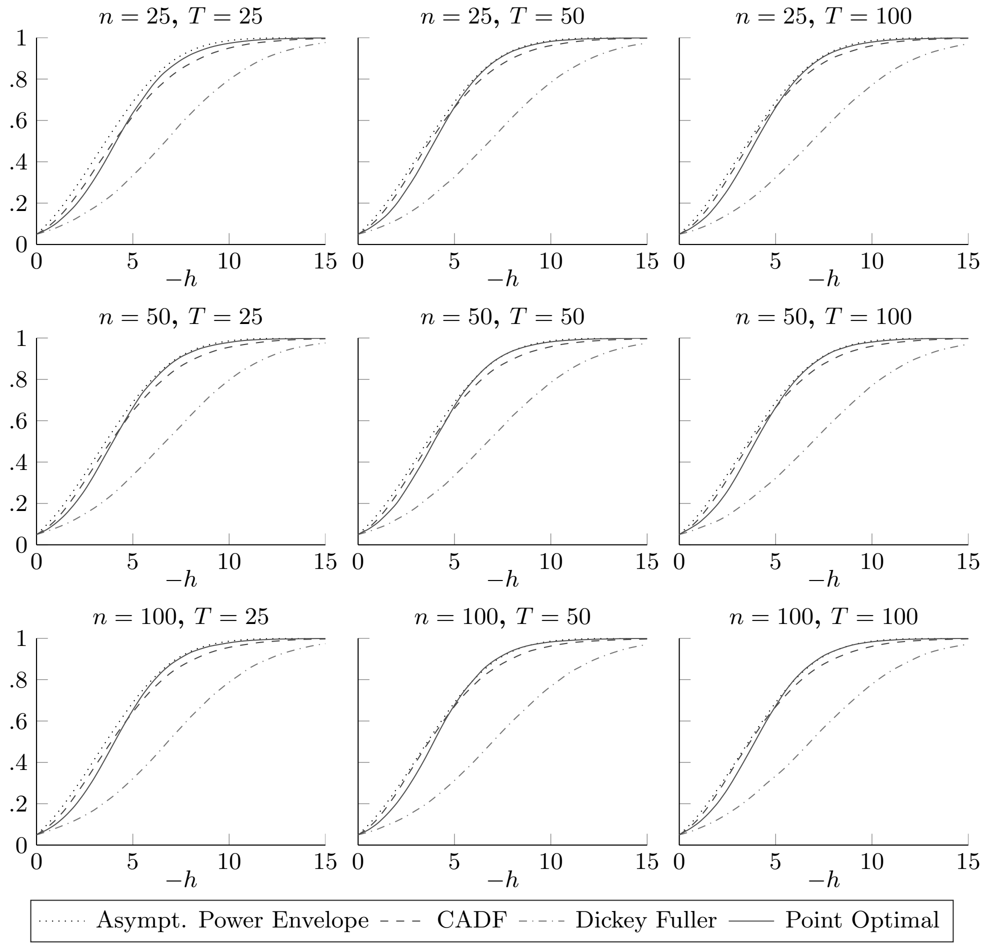


Figure 4.5: Size-corrected powers and asymptotic powers of unit-root tests as a function of $-h$ for varying sample sizes. All innovations are i.i.d. normally distributed and $R^2 = 0.5$.

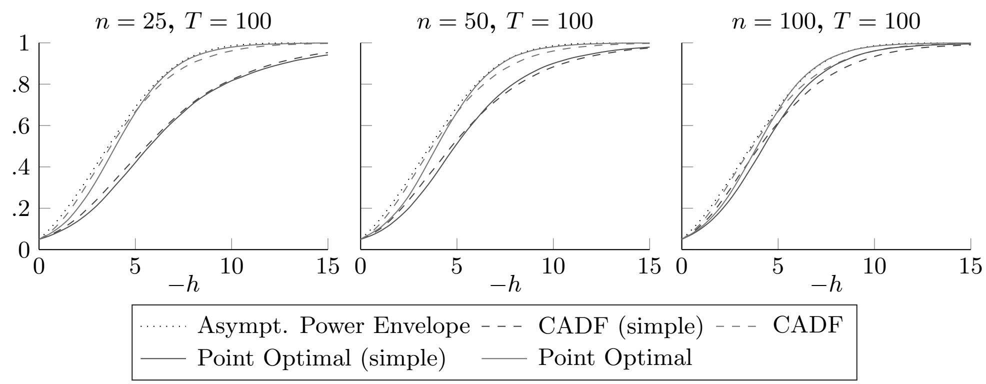


Figure 4.6: Size-corrected powers and asymptotic powers of unit-root tests as a function of $-h$ for varying sample sizes. All innovations are i.i.d. normally distributed and $R^2 = 0.5$.

4.6 Conclusion

We have demonstrated that, also for testing unobserved common factors, it can be highly beneficial to use observed stationary covariates. The gains over the classical PANIC procedure are particularly large when the contribution of the factor innovation to the covariate can be separately identified, in which case the covariate overcomes the rate-disadvantage due to allowing for cross-sectional cointegration in panel unit-root tests. Even when one is not comfortable assuming that the factor in the covariate equation can be perfectly identified, there are gains from taking into account the covariate. In this case the statistical problem is reduced to that of the well-studied time-series case, i.e., there is no improvement in the convergence rate, but the covariate nevertheless significantly improves local powers. Whether the rate improvement from having the covariates is attainable depends on the application at hand. Most applications would likely require multiple factors, requiring more advanced factor estimates than principal components. Although they do not suffer from these problems, the robust tests provide considerably higher finite-sample powers compared to tests without covariates, even if the convergence rate is the same. Finally, the analysis under the faster convergence rate also facilitates the asymptotic comparison of different factor estimation methods, highlighting the optimality properties of principal components even asymptotically.

4.A Proofs

4.A.1 Proof of Proposition 4.3.1

Proof Noting that, conditional on F, the observations Y and x are both cross-sectionally independent, we obtain the Gaussian likelihood ratio

$$
\begin{aligned}
\log \frac{\mathrm{d}P_h}{\mathrm{d}P_0} =& -\frac{1}{2}\sum_{i=1}^{n}(x_{i\cdot} - \gamma_i(\Delta F - \frac{h}{\sqrt{n}T}F_{-1}))'\Sigma_{u,i}^{-1}(x_{i\cdot} - \gamma_i(\Delta F - \frac{h}{\sqrt{n}T}F_{-1})) \\
& - \frac{1}{2}(\Delta F - \frac{h}{\sqrt{n}T}F_{-1})'\Sigma_f^{-1}(\Delta F - \frac{h}{\sqrt{n}T}F_{-1}) \\
& + \frac{1}{2}\sum_{i=1}^{n}(x_{i\cdot} - \gamma_i\Delta F)'\Sigma_{u,i}^{-1}(x_{i\cdot} - \gamma_i\Delta F) + \frac{1}{2}\Delta F'\Sigma_f^{-1}\Delta F \\
=& h\left(\frac{-1}{\sqrt{n}T}\sum_{i=1}^{n}\gamma_i F_{-1}'\Sigma_{u,i}^{-1}(x_{i\cdot} - \gamma_i\Delta F) + \frac{1}{\sqrt{n}T}F_{-1}'\Sigma_f^{-1}\Delta F\right) \\
& - \frac{1}{2}h^2\frac{1}{T^2}F_{-1}'\left(\sum_{i=1}^{n}\frac{\gamma_i^2}{n}\Sigma_{u,i}^{-1} + \frac{1}{n}\Sigma_f^{-1}\right)F_{-1} \\
=:& h\tilde{\Delta}_{n,T} - \frac{1}{2}h^2\tilde{J}_{n,T}.
\end{aligned}
$$

We start by characterizing the limiting distribution of $(\tilde{\Delta}_{n,T}, \tilde{J}_{n,T})$ under the null hypothesis. Then, we will show that $(\Delta_{n,T}, J_{n,T}) = (\tilde{\Delta}_{n,T}, \tilde{J}_{n,T}) + o_P(1)$. Throughout the proofs, we use freely the well-known results of weak convergence to stochastic integrals in the time-series case, namely, with $\tilde{u}_t$ being i.i.d standard normally distributed innovations independent of F, we have, under the null hypothesis of F being a random walk,[21]

$$
\frac{1}{T}\sum_{t=1}^{T}F_{t-1}f_t \Rightarrow \int W_1\mathrm{d}W_1, \quad \frac{1}{T}\sum_{t=1}^{T}F_{t-1}\tilde{u}_t \Rightarrow \int W_1\mathrm{d}W_2, \quad \frac{1}{T^2}\sum_{t=1}^{T}F_{t-1}^2 \Rightarrow \int W_1^2(t)\mathrm{d}t, \tag{4.A.1}
$$

with the convergences also holding jointly.[22]

[21] This also extends to stationary Gaussian processes with more general autocovariance functions, if one normalizes by the square-root of the long-run variance. One way to obtain this generalization is an application of Lemma 4.A.2.

[22] In this very simple case, this is a direct consequence of the probability convergence (formally on a different probability space) to the stochastic integral, which then implies joint convergence. For similar convergence under more general conditions, see, for example, Kurtz and Protter (1991) and Phillips (1987b).

First, we show that, under the null hypothesis, the direct contribution of the factor innovations to the likelihood ratio is negligible. For this, recall the cumulate sum operator A, a matrix with ones below the diagonal and zeros on and above, so that $F_{-1} = A\Delta F$. We have, under the null hypothesis,

$$\frac{1}{T}F'_{-1}\Sigma_f^{-1}\Delta F = \frac{1}{T}f'A'\Sigma_f^{-1}f = O_P(1),$$

since it has mean $\operatorname{tr} A' = 0$ and a variance given by

$$\begin{aligned}
&\frac{1}{T^2}\operatorname{tr}((A'\Sigma_f^{-1} + \Sigma_f^{-1}A)\Sigma_f(A'\Sigma_f^{-1} + \Sigma_f^{-1}A)\Sigma_f) \\
\leq &\frac{1}{T^2}\left\|(A'\Sigma_f^{-1} + \Sigma_f^{-1}A)\Sigma_f\right\|_F^2 \\
\leq &\left(\frac{1}{T}\|A\|_F + \frac{1}{T}\left\|\Sigma_f^{-1}\right\|_{\text{spec}}\|A\Sigma_f\|_F\right)^2 = O(1),
\end{aligned}$$

using that $\|A\|_F = O(T)$ and the eigenvalues of Σ_f are bounded and bounded away from zero. As this term is divided by $\sqrt{n}$ it indeed becomes negligible.

Under the null hypothesis, $x_{it} - \gamma_i\Delta F_t = u_{it}$, so under P_0, the remaining term of $\tilde{\Delta}_{n,T}$ is given by

$$\Delta^*_{n,T} := \frac{-1}{T\sqrt{n}}F'_{-1}\sum_{i=1}^n \gamma_i\Sigma_{u,i}^{-1}u_{i,\cdot}. \tag{4.A.2}$$

Note that, with $\Sigma^{-1} = \frac{1}{n}\sum_{i=1}^n \gamma_i^2\Sigma_{u,i}^{-1}$ and $\Sigma^{-1/2}$ the matrix square root of Σ^{-1},

$$\Sigma^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i\Sigma_{u,i}^{-1}u_{i\cdot} \sim N(0, I_T).$$

Therefore, $\Delta^*_{n,T}$ is equal in distribution to

$$\check{\Delta}_{n,T} := \frac{1}{T}F'_{-1}\Sigma^{-1/2}\tilde{u}_{\cdot}, \tag{4.A.3}$$

for a $\tilde{u}_{\cdot} \sim N(0, I_T)$ independent of F_{-1}. Moreover, for each n, T, $(\Delta^*_{n,T}, J_{n,T})$ has the same distribution as $(\check{\Delta}_{n,T}, J_{n,T})$. We proceed to derive the limiting distribution of $(\check{\Delta}_{n,T}, J_{n,T})$. A direct application of (4.A.1) is precluded by the presence of $\Sigma^{-1/2}$. However, it turns out that it is possible to replace $\Sigma^{-1/2}$ by a scalar in (4.A.3) without changing the limiting distribution. For this, let $\Sigma_T(f)$ denote a $T\times T$ Toeplitz matrix based on the spectral density function[23] f, i.e.

$$(\Sigma_T(f))_{k,j} = \frac{1}{2\pi}\int_0^{2\pi} f(\lambda)e^{-\mathbb{i}(k-j)\lambda}\mathrm{d}\lambda, \tag{4.A.4}$$

with $\mathbb{i}^2 = -1$. Lemma A.1 in Wichert et al. (2019) implies that, for f the spectral density function of a time series with m-summable autocovariances, $\left\|A'(\Sigma_T(f) - f(0)I)\right\|_F =$

[23] Conversely, $f(\lambda) = \sum_{k=-\infty}^{\infty}\gamma(k)e^{\mathbb{i}k\lambda}$. Often, the spectral density is defined as this divided by 2π, but we stick to (4.A.4) for convenience. This normalization implies that the long-run variance equals $f(0)$.

$O(\sqrt{T})$. To apply this result to (4.A.3), we make use of both the inverse of a Toeplitz matrix and its square root being asymptotically equivalent to certain Toeplitz matrices as well, see, for example, Section 5 in Gray (2005). In particular, let f_i be the spectral density of $\{u_{it}\}$. Then $\left\|\Sigma_{u,i}^{-1} - \Sigma_T(1/f_i)\right\|_{\text{spec}} = o(1)$. Thanks to the uniform bounds on the spectral densities, this result also holds uniformly across cross-section units, see Lemma 4.A.3. Therefore,

$$\left\|\Sigma^{-1} - \Sigma_T\Big(\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{f_i}\Big)\right\|_{\text{spec}} \leq \sup_{i\in\mathbb{N}} \left\|\Sigma_{u,i}^{-1} - \Sigma_T(1/f_i)\right\|_{\text{spec}} \frac{1}{n}\sum_{i=1}^n \gamma_i^2 \to 0$$

as $(n, T \to \infty)$. Thanks to a uniform continuity property of the matrix square-root (see Schmitt (1992), Lemma 2.2), and the fact that the eigenvalues of $\Sigma_T(\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{f_i})$ are bounded away from zero thanks to the bound on the spectral densities, this also implies that

$$\left\|\Sigma^{-1/2} - \Sigma_T\Big(\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{f_i}\Big)^{1/2}\right\|_{\text{spec}} \to 0.$$

Combined, this implies, together with the statement on the matrix square root above,

$$\left\|\Sigma^{-1/2} - \Sigma_T\Big(\Big(\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{f_i}\Big)^{1/2}\Big)\right\|_{\text{spec}} = o(1). \tag{4.A.5}$$

So define $f_n := \left(\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i^2}{f_i}\right)^{1/2}$, $\omega_{u,\Gamma,n}^{-1} := f_n(0)$ and split

$$\frac{1}{T}F_{-1}'\Sigma^{-1/2}\tilde{u}_{\cdot} = \frac{\omega_{u,\Gamma,n}^{-1}}{T}F_{-1}'\tilde{u}_{\cdot} + \frac{1}{T}f'A'(\Sigma^{-1/2} - \Sigma_T(f_n))\tilde{u}_{\cdot} + \frac{1}{T}f'A'(\Sigma_T(f_n) - \omega_{u,\Gamma,n}^{-1}I)\tilde{u}_{\cdot}.$$

We will show that only the first summand contributes to the limiting distribution. To see that the second summand is asymptotically negligible, note that it has zero mean conditional on f and its variance is given by

$$\begin{aligned}\mathbb{E}\operatorname{Var}\frac{1}{T}f'A'(\Sigma^{-1/2} - \Sigma_T(f_n))\tilde{u}_{\cdot}|f =& \frac{1}{T^2}\mathbb{E}f'A'(\Sigma^{-1/2} - \Sigma_T(f_n))(\Sigma^{-1/2} - \Sigma_T(f_n))Af \\ =& \frac{1}{T^2}\operatorname{tr} A'(\Sigma^{-1/2} - \Sigma_T(f_n))(\Sigma^{-1/2} - \Sigma_T(f_n))A\Sigma_f \\ \leq& \frac{1}{T^2}\left\|A\right\|_F \left\|\Sigma^{-1/2} - \Sigma_T(f_n)\right\|_{\text{spec}}^2 \left\|A\Sigma_f\right\|_F \\ =& \frac{1}{T^2}O(T)o(1)O(T) = o(1).\end{aligned}$$

Similarly, the variance of the third summand is given by

$$\begin{aligned}\frac{1}{T^2}\operatorname{tr} A'(\Sigma_T(f_n) - \omega_{u,\Gamma,n}^{-1}I)(\Sigma_T(f_n) - \omega_{u,\Gamma,n}^{-1}I)A\Sigma_f \leq& \frac{1}{T^2}\left\|A'(\Sigma_T(f_n) - \omega_{u,\Gamma,n}^{-1}I)\right\|_F^2 \left\|\Sigma_f\right\|_{\text{spec}} \\ =& \frac{1}{T^2}o(T^2)O(1).\end{aligned}$$

We obtain the $o(T^2)$ rate by an application of Lemma 4.A.2, which requires the f_n to be twice continuously differentiable with uniformly bounded second derivative. This follows from the analogous assumption on the f_i, as

$$f_n''(\lambda) = \frac{1}{n}\sum_{i=1}^n \frac{2}{f_i^3(\lambda)}f_i'(\lambda) - \frac{1}{n}\sum_{i=1}^n \frac{1}{f_i^2(\lambda)},$$

so as the f_i are uniformly bounded away from zero and their second derivatives are uniformly bounded, the same applies to f_n.

Thanks to these calculations, the dependence on n has disappeared, allowing us to use the existing convergence results recalled in (4.A.1), i.e.,

$$\frac{1}{T}\sum_{t=1}^{T} F_{t-1}\tilde{u}_{\cdot,t} \to \omega_f \int W_1 \mathrm{d}W_2,$$

as $T \to \infty$ and thus also as $(n, T \to \infty)$. Conclude that, under the null hypothesis

$$\tilde{\Delta}_{n,T} \Rightarrow \omega_f \omega_{u,\Gamma}^{-1} \int W_1 \mathrm{d}W_2 =: \Delta.$$

Moreover, also from (4.A.1), this convergence holds jointly with

$$\frac{1}{T^2}\sum_{t=1}^{T} F_{t-1}^2 \Rightarrow \omega_f^2 \int W_1^2(t)\mathrm{d}t.$$

Thus, the weak limit of $\tilde{J}_{n,T}$ equals that of

$$J_{n,T} := \frac{1}{nT^2}\sum_{i=1}^{n}\frac{\gamma_i^2}{\omega_{u,i}^2}\sum_{t=1}^{T} F_{t-1}^2 \Rightarrow \omega_f^2\omega_{u,\Gamma}^{-2}\int W_1^2(t)\mathrm{d}t =: J.$$

Thus we have demonstrated the joint convergence the central sequence $\tilde{\Delta}_{n,T}$ and empirical Fisher information $\tilde{J}_{n,T}$: $(\tilde{\Delta}_{n,T}, \tilde{J}_{n,T}) \Rightarrow (\Delta, J)$.

For completeness, we now demonstrate that $\Delta|J \sim N(0, J)$, by considering its moment generating function. Note that, conditional on W_1, $\int W_1 \mathrm{d}W_2 \sim N(0, \int W_1^2 \mathrm{d}t)$. Thus

$$\mathbb{E}[\exp(t\Delta)|J] = \mathbb{E}[\mathbb{E}[\exp(t\Delta)|W_1]|J] = \mathbb{E}[\exp(t^2 J/2)|J] = \exp(t^2 J/2),$$

as required.

Finally, we show that $(\Delta_{n,T}, J_{n,T}) = (\tilde{\Delta}_{n,T}, \tilde{J}_{n,T}) + o_P(1)$. We have, under the null hypothesis,

$$\Delta_{n,T} - \tilde{\Delta}_{n,T} = \frac{1}{\sqrt{n}T}\sum_{i=1}^{n}\gamma_i f' A'(\omega_{u,i}^2 I - \Sigma_{u,i})\Sigma_{u,i}^{-1}\frac{1}{\omega_{u,i}^2}u_{i,\cdot},$$

which has zero mean and a variance of

$$\begin{aligned}
&\frac{1}{nT^2}\sum_{i=1}^{n}\frac{\gamma_i^2}{\omega_{u,i}^4}\mathbb{E} f' A'(\omega_{u,i}^2 I - \Sigma_{u,i})\Sigma_{u,i}^{-1}(\omega_{u,i}^2 I - \Sigma_{u,i})Af \\
=&\frac{\gamma_i^2}{nT^2}\sum_{i=1}^{n}\frac{2}{\omega_{u,i}^4}\operatorname{tr} A'(\omega_{u,i}^2 I - \Sigma_{u,i})\Sigma_{u,i}^{-1}(\omega_{u,i}^2 I - \Sigma_{u,i})A\Sigma_f \\
\leq&\frac{\gamma_i^2}{nT^2}\sum_{i=1}^{n}\frac{2}{\omega_{u,i}^4}\left\|A'(\omega_{u,i}^2 I - \Sigma_{u,i})\right\|_F^2 \left\|\Sigma_{u,i}^{-1}\right\|_{\text{spec}}\left\|\Sigma_f\right\|_{\text{spec}} \\
=&\frac{2}{nT^2}\sum_{i=1}^{n} O(1)o(T)^2 O(1)O(1) = o(1),
\end{aligned}$$

as the orders are uniform across i, thanks to Lemma 4.A.2. The analogous difference for the Fisher information is given by

$$\begin{aligned}\tilde{J}_{n,T} - J_{n,T} =& \frac{1}{nT^2} \sum_{i=1}^{n} \gamma_i^2 F_{-1}' (\Sigma_{u,i}^{-1} - \frac{1}{\omega_{u,i}^2} I_T) F_{-1} \\ =& \frac{1}{nT^2} f' \sum_{i=1}^{n} \frac{\gamma_i^2}{\omega_{u,i}^2} A' (\omega_{u,i}^2 I_T - \Sigma_{u,i}) A f.\end{aligned}$$

Its mean is given by

$$\begin{aligned}\frac{1}{nT^2} \operatorname{tr} \sum_{i=1}^{n} \frac{\gamma_i^2}{\omega_{u,i}^2} A' (\omega_{u,i}^2 I_T - \Sigma_{u,i}) A \Sigma_f \leq& \frac{1}{nT^2} \sum_{i=1}^{n} \frac{\gamma_i^2}{\omega_{u,i}^2} \sup_i \left\| A' (\omega_{u,i}^2 I_T - \Sigma_{u,i}) \right\|_F \left\| A \Sigma_f \right\|_F \qquad (4.A.6) \\ =& \frac{1}{T^2} o(T) O(T) = o(1).\end{aligned}$$

As the variance is bounded by twice the square of the right-hand side of (4.A.6), the difference between the two Fisher informations converges to zero in L_2, as does the difference between the central sequences, implying joint convergence. ■

4.A.2 Proof of Proposition 4.3.2

Proof As the experiment is LAMN, we can apply the general version of Le Cam's Third Lemma (see, for example, Chapter 6 in Van der Vaart (2000)). Here, it states that the limiting distribution of $t_{n,T}$ under under P_h is given by the probability measure L_h, defined by

$$L_h(B) = \mathbb{E} \mathbb{1}_B (\Delta / \sqrt{J}) \exp(h\Delta - \frac{1}{2} h^2 J).$$

This implies that the distribution of $t_{n,T}$ under local alternatives matches that of its analogue in the limit experiment. For completeness, we show that this distribution indeed has the representation given in Proposition 4.3.2: its moment generating function is given by (using the conditional normality of Δ)

$$\begin{aligned}\int \exp(tx) \mathrm{d}L_h(x) =& \mathbb{E} \exp(t \frac{\Delta}{\sqrt{J}}) \exp(h\Delta - \frac{1}{2} h^2 J) \\ =& \mathbb{E}\mathbb{E}[\exp(t \frac{\Delta}{\sqrt{J}} + h\Delta - \frac{1}{2} h^2 J) | J] \\ =& \mathbb{E}[\exp(-\frac{1}{2} h^2 J) \mathbb{E}[\exp((\frac{t}{\sqrt{J}} + h)\Delta) | J]] \\ =& \mathbb{E}[\exp(-\frac{1}{2} h^2 J) \exp(\frac{J}{2} (\frac{t}{\sqrt{J}} + h)^2)] \\ =& \exp(t^2/2) \mathbb{E}[\exp(th\sqrt{J})],\end{aligned}$$

which is indeed the moment generating function of an independent normal distribution added to $h\sqrt{J}$. ■

4.A.3 Proof of Theorem 4.3.1

Proof Due to the complexity of this proof, we now switch to vector notation. To obtain estimates of F, Λ, and Γ, we first estimate Λ and F using principal components based one Δy. Note that it is crucial to obtain the estimates of F from the y equation and not estimating the x equation directly. We only use the x equation to obtain estimates of Γ, by applying OLS to estimated factors.

First, we establish some preliminary results. Lemma 3.1 in Wichert et al. (2019) implies that principal components applied to Δy yield estimates $\hat{\Lambda}$ satisfying

$$\left\|\hat{\Lambda} - \Lambda H_K\right\|_F = o_p(1),$$

with H_K a $K \times K$ matrix satisfying $\|H_K\|_F = O_P(1)$ and $\left\|H_K^{-1}\right\|_F = O_P(1)$. As H_K does nothing but complicate notation, we will suppose $H_K = I_K$ in the remainder of this proof.

Given loading estimates $\hat{\Lambda}$, we can estimate F using OLS, i.e.,

$$\begin{aligned}\hat{F} =& y\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} = y\Lambda(\Lambda'\Lambda)^{-1} + y(\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1}) \qquad (4.A.7)\\ =& F + E\Lambda(\Lambda'\Lambda)^{-1} + y(\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1}). \qquad (4.A.8)\end{aligned}$$

Note

$$\begin{aligned}\left\|(\hat{\Lambda}'\hat{\Lambda})^{-1} - (\Lambda'\Lambda)^{-1}\right\|_F \leq& \frac{1}{n^{3/2}}\left\|(\hat{\Lambda}'\hat{\Lambda}/n)^{-1}\right\|_F \left\|\Lambda'\Lambda/\sqrt{n} - \hat{\Lambda}'\hat{\Lambda}/\sqrt{n}\right\|_F \left\|(\Lambda'\Lambda/n)^{-1}\right\|_F\\ =& \frac{1}{n^{3/2}} O_P(1) o_P(1) O(1) = o_p(n^{-3/2}),\end{aligned}$$

so

$$\begin{aligned}\left\|\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1}\right\|_F \leq& \left\|\hat{\Lambda} - \Lambda\right\|_F \left\|(\Lambda'\Lambda)^{-1}\right\|_F + \left\|\hat{\Lambda}\right\|_F \left\|(\hat{\Lambda}'\hat{\Lambda})^{-1} - (\Lambda'\Lambda)^{-1}\right\|_F\\ =& o_P(1)o_P(n^{-1}) + o_P(\sqrt{n})o_p(n^{-3/2}) = o_P(n^{-1}). \qquad (4.A.9)\end{aligned}$$

Both $\|y\|_F$ and $\|E\|_F$ are $O_P(\sqrt{n}T)$, and the same holds for $\|E\Lambda\|_F$ as

$$\|E\Lambda\|_F^2 = \sum_{t=1}^{T}\sum_{i=1}^{n} \lambda_i^2 E_{it}^2 = O_P(nT^2).$$

Therefore, using (4.A.8), we obtain

$$\begin{aligned}\left\|F - \hat{F}\right\|_F =& \left\|E\Lambda(\Lambda'\Lambda)^{-1} + y(\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1})\right\|_F \qquad (4.A.10)\\ \leq& O_P(\sqrt{n}T)O(n^{-1}) + O_P(\sqrt{n}T)o_P(n^{-1}) = O_P(T/\sqrt{n}). \qquad (4.A.11)\end{aligned}$$

Using that $\|\Delta y\|_F$ and $\|\Delta E\Lambda\|_F$ are $O_P(\sqrt{nT})$, the same arguments yield

$$\left\|\Delta F - \Delta\hat{F}\right\|_F = O_P(\sqrt{T/n}).$$

Further, again using the ΔF version of (4.A.8), we have

$$\left\|\Delta F'(\Delta\hat{F} - \Delta F)\right\|_F \leq \left\|\Delta F'\Delta E\right\|_F O(n^{-1/2}) + \|\Delta F\|_F \|\Delta y\|_F o_P(n^{-1})$$

$$=O_P(\sqrt{nT})O(n^{-1/2}) + O_P(\sqrt{T})O_P(\sqrt{nT})o_P(n^{-1}) = o_P(T/\sqrt{n}).$$

Finally, for the estimated loadings in the x equation, we obtain, under the null hypothesis,

$$\begin{aligned}
\left\|\hat{\Gamma}' - \Gamma'\right\|_F &= \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\Delta\hat{F}'x - \Gamma'\right\|_F \\
&= \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\Delta\hat{F}'(\Delta\hat{F}\Gamma' + (\Delta F - \Delta\hat{F})\Gamma' + u) - \Gamma'\right\|_F \\
&= \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\Delta\hat{F}'((\Delta F - \Delta\hat{F})\Gamma' + u)\right\|_F \\
&\leq \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\right\|_F \left\|\Delta F'(\Delta F - \Delta\hat{F})\right\|_F \|\Gamma\|_F \\
&\quad + \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\right\|_F \left\|\Delta F - \Delta\hat{F}\right\|_F^2 \|\Gamma\|_F + \left\|(\Delta\hat{F}'\Delta\hat{F})^{-1}\right\|_F \left\|\Delta\hat{F}'u\right\|_F \\
&= O_P(T^{-1})o_P(T/\sqrt{n})O_P(\sqrt{n}) \\
&\quad + O_P(T^{-1})O_P(T/n)O_P(\sqrt{n}) + O_P(T^{-1})O_p(\sqrt{nT}) \\
&= o_P(1) + O_P(n^{-1/2}) + O_P(\sqrt{n/T}) = o_P(1).
\end{aligned}$$

We now show that the distribution of the empirical Fisher information $J_{n,T}$ is not affected by estimating the factors, loadings, and long-run variances. Denote by Ω be the $n \times n$ matrix with the long-run variances of the u_{it} on the diagonal, i.e., $\Omega_{i,i} = \omega_{u,i}^2$, and zeros off the diagonal. Let $\hat{\Omega}$ be analogously defined, using the estimates $\hat{\omega_{u,i}}$ from Assumption 4.4 in place of $\omega_{u,i}$. We have

$$\begin{aligned}
\hat{J}_{n,T} - J_{n,T} &= \frac{1}{nT^2}(\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma}\hat{F}_{-1}'\hat{F}_{-1} - \Gamma'\Omega^{-1}\Gamma F_{-1}'F_{-1}) \\
&\leq \frac{1}{nT^2}\left\|\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma} - \Gamma'\Omega^{-1}\Gamma\right\|_F \left\|\hat{F}_{-1}\right\|_F^2 + \left\|\Gamma'\Omega^{-1}\Gamma\right\|_F \left\|\hat{F}_{-1} - F_{-1}\right\|_F (\|F_{-1}\|_F + \left\|\hat{F}_{-1}\right\|_F) \\
&= \frac{1}{nT^2}\left(o_P(n)O_P(T) + O(n)O_P(T/\sqrt{n})O(T)\right) = o_P(T^{-1}) + O_P(n^{-1/2}) = o_P(1),
\end{aligned}$$

where the rate on the first norm follows from

$$\begin{aligned}
&\left\|\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma} - \Gamma'\Omega^{-1}\Gamma\right\|_F \\
&\leq \left\|\hat{\Gamma} - \Gamma\right\|_F \left\|\hat{\Omega}^{-1}\right\|_{\text{spec}} \left\|\hat{\Gamma}\right\|_F + \|\Gamma\|_F \left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_{\text{spec}} \left\|\hat{\Gamma}\right\|_F + \|\Gamma\|_F \left\|\Omega^{-1}\right\|_{\text{spec}} \left\|\hat{\Gamma} - \Gamma\right\|_F \\
&= o_P(1)O_P(1)O_P(\sqrt{n}) + O(\sqrt{n})o_P(1)O_P(\sqrt{n}) + O(\sqrt{n})O(1)o_P(1) = o_P(n),
\end{aligned}$$

as the $\omega_{u,i}^2$ are bounded away from zero. Conclude that the limiting distribution of $J_{n,T}$ is not affected by using factor and loading estimates instead of observed factors. Thanks to our LAMN result and Le Cam's First Lemma, this also holds under local alternatives.

Now consider $\hat{\Delta}_{n,T}$. We have

$$\begin{aligned}
\sqrt{n}T(\hat{\Delta}_{n,T} - \Delta_{n,T}) &= \hat{F}_{-1}'(x - \Delta\hat{F}\hat{\Gamma}')\hat{\Omega}^{-1}\hat{\Gamma} - F_{-1}'(x - \Delta F\Gamma')\Omega^{-1}\Gamma \\
&= F_{-1}'(\Delta F - \Delta\hat{F})\Gamma'\Omega^{-1}\Gamma + F_{-1}'\Delta\hat{F}(\Gamma' - \hat{\Gamma}')\Omega^{-1}\Gamma \\
&\quad + \left(\hat{F}_{-1}' - F_{-1}'\right)(x - \Delta\hat{F}\hat{\Gamma}')\Omega^{-1}\Gamma + \hat{F}_{-1}'(x - \Delta\hat{F}\hat{\Gamma}')(\hat{\Omega}^{-1} - \Omega^{-1})\Gamma \\
&\quad + \hat{F}_{-1}'(x - \Delta\hat{F}\hat{\Gamma}')\hat{\Omega}^{-1}(\hat{\Gamma} - \Gamma)
\end{aligned}$$

$$=: I + II + III + IV + V.$$

Term I will be responsible for the loss of power from not observing the factors. We first show that terms II, III, and IV are asymptotically negligible. We have, under H_0,

$$II \leq \left\| F'_{-1}\Delta\hat{F} \right\|_F \left\| \Gamma' - \hat{\Gamma}' \right\|_F \|\Omega^{-1}\|_{\text{spec}} \|\Gamma\|_F = O_P(T)o_P(1)O_P(1)O_P(\sqrt{n}) = o_P(\sqrt{n}T),$$

where we have used the difference version of (4.A.10) to bound the first norm. For III, note that

$$x - \Delta\hat{F}\hat{\Gamma}' = u + \Delta F(\Gamma' - \hat{\Gamma}') + (\Delta F - \Delta\hat{F})\hat{\Gamma}' \qquad (4.A.12)$$

Using (4.A.10) twice, noting that $\Lambda' E' u\Omega^{-1}\Lambda = O(nT)$, $y'u\Omega^{-1}\Lambda = O(nT)$, $\Lambda' E'\Delta F = O(\sqrt{n}T)$, $y'\Delta F = O(\sqrt{n}T)$, $\Lambda' E'\Delta E\Lambda = O(nT)$, $y'\Delta E\Lambda = O(nT)$, $\Lambda' E'\Delta y = O(nT)$, $y'\Delta y = O(nT)$, and recalling (4.A.9), we have

$$\begin{aligned}
III \leq & \left((\Lambda'\Lambda)^{-1}\Lambda' E' + ((\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\Lambda}' - (\Lambda'\Lambda)^{-1}\Lambda'))y'\right) \\
& \left(u + \Delta F o_P(1) + \Delta E\Lambda(\Lambda'\Lambda)^{-1} + \Delta y(\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1})\right)\Omega^{-1}\Gamma \\
\leq & (o_P(n^{-1}) + O_P(n^{-1}))O(nT) + (o_P(n^{-1}) + O_P(n^{-1}))O(\sqrt{n}T)o_P(1)O_P(1)O_P(\sqrt{n}) \\
& + (o_P(n^{-1}) + O_P(n^{-1}))O(nT)(o_P(n^{-1}) + O_P(n^{-1}))O(1)O(\sqrt{n}) \\
= & O_P(T) + o_P(T) + O_P(T/\sqrt{n}) = o_P(\sqrt{n}T).
\end{aligned}$$

For IV, we use Hoelder's inequality, then Cauchy-Schwarz and once more (4.A.12) to obtain

$$\begin{aligned}
IV &\leq \|\hat{F}'_{-1}(u + \Delta F(\Gamma' - \hat{\Gamma}') + (\Delta F - \Delta\hat{F})\hat{\Gamma}')(\hat{\Omega}^{-1} - \Omega^{-1})\|_1 \|\Gamma\|_\infty \\
&\leq \left\|\hat{F}'_{-1}(u + \Delta F(\Gamma' - \hat{\Gamma}') + (\Delta F - \Delta\hat{F})\hat{\Gamma}')\right\|_F \left\|\hat{\Omega}^{-1} - \Omega^{-1}\right\|_F \|\Gamma\|_\infty \\
&= O_P(\sqrt{n}T)o_P(1)O_P(1),
\end{aligned}$$

as

$$\begin{aligned}
&\left\|\hat{F}'_{-1}(u + \Delta F(\Gamma' - \hat{\Gamma}') + (\Delta F - \Delta\hat{F})\hat{\Gamma}')\right\|_F \\
&\quad \leq \left\|\hat{F}'_{-1}u\right\|_F + \left\|\hat{F}'_{-1}\Delta F\right\|_F \left\|\Gamma - \hat{\Gamma}\right\|_F + \left\|\hat{F}'_{-1}(\Delta F - \Delta\hat{F})\right\|_F \left\|\hat{\Gamma}\right\|_F \\
&\quad = O_P(\sqrt{n}T) + O_P(T)o_P(1) + O_P(Tn^{-1/2})O_P(\sqrt{n}) = O_P(\sqrt{n}T),
\end{aligned}$$

where we have used

$$\begin{aligned}
\left\|\hat{F}'_{-1}(\Delta F - \Delta\hat{F})\right\|_F &= \left\|\hat{F}'_{-1}\Delta E\Lambda(\Lambda'\Lambda)^{-1} + \hat{F}'_{-1}\Delta y(\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1})\right\|_F \\
&\leq \left\|\hat{F}'_{-1}\Delta E\Lambda\right\|_F \|(\Lambda'\Lambda)^{-1}\|_F + \left\|\hat{F}'_{-1}\Delta y\right\|_F \left\|\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1} - \Lambda(\Lambda'\Lambda)^{-1}\right\|_F \\
&= O_P(\sqrt{n}T)O(n^{-1}) + O_P(\sqrt{n}T)o_P((n^{-1}) = O_P(Tn^{-1/2}). \qquad (4.A.13)
\end{aligned}$$

Similarly,

$$V \leq \left\|\hat{F}'_{-1}u\right\|_F \left\|\hat{\Gamma} - \Gamma\right\|_F + \left\|\hat{F}'_{-1}\Delta F\right\|_F \left\|\Gamma - \hat{\Gamma}\right\|_F^2 + \left\|\hat{F}'_{-1}(\Delta F - \Delta\hat{F})\right\|_F \left\|\hat{\Gamma}\right\|_F$$

$$=O_P(\sqrt{n}T)o_P(1) + O_P(T)o_P(1) + O_P(Tn^{-1/2})O_P(\sqrt{n}) = o_P(\sqrt{n}T).$$

Finally, we consider the first, non-negligible term. From (4.A.13) it follows that

$$I = F'_{-1}\Delta E\Lambda(\Lambda'\Lambda)^{-1}\Gamma'\Omega^{-1}\Gamma + o_P(\sqrt{n}T).$$

The long-run variance of $\Delta E\Lambda/\sqrt{n}$ is given, under $\rho_E = 1$, by

$$\lim_{T\to\infty} \mathrm{Var}\, \frac{1}{\sqrt{nT}} \sum_{t=1}^{T}\sum_{i=1}^{n} \lambda_i \eta_{it} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 \,\mathrm{Var} \lim_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \eta_{it} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^2 \omega_{\eta,i}^2 =: \kappa_{\eta,n}^2$$

Thus, with $\kappa_\eta^2 := \lim_{n\to\infty} \kappa_{\eta,n}^2$,

$$\frac{1}{\sqrt{n}T} I \Rightarrow \Psi_\Lambda^{-1}\omega_{u,\Gamma}^{-2}\omega_f\kappa_\eta \int W_1 \mathrm{d}W_2.$$

Inserting this this into $\hat{t}_{n,T}$ yields

$$\hat{t}_{n,T} = \frac{1}{\sqrt{n}T\sqrt{J_{n,T}}} F'_{-1}\left(-u\Omega^{-1}\Gamma - \Delta E\Lambda(\Lambda'\Lambda)^{-1}\Gamma'\Omega^{-1}\Gamma\right) - h\sqrt{J_{n,T}} + o_P(1).$$

Thanks to independence of u and ΔE, the limit of the long-run variance of

$$\frac{1}{\sqrt{nT}}\left(u\Omega^{-1}\Gamma - \Delta E\Lambda(\Lambda'\Lambda)^{-1}\Gamma'\Omega^{-1}\Gamma\right)$$

is given by

$$\omega_{u,\Gamma}^{-2} + \Psi_\Lambda^{-2}\omega_{u,\Gamma}^{-4}\kappa_\eta^2,$$

so

$$\hat{t}_{n,T} \Rightarrow \sqrt{1 + \Psi_\Lambda^{-2}\omega_{u,\Gamma}^{-2}\kappa_\eta^2}\, Z - h\sqrt{\omega_f^2\omega_{u,\Gamma}^{-2}\int W_1^2 \mathrm{d}t}.$$ ■

4.A.4 Proof of Proposition 4.3.3

Proof Considering the same likelihood ratio as in Proposition 4.3.1 but under (4.9) instead of (4.3), the additional terms

$$-\frac{1}{2\sigma_\eta^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(E_{it} - \beta(x_{it} - \gamma_i(\Delta F_t - \frac{h}{\sqrt{n}T}F_{t-1}))\right)^2 + \frac{1}{2\sigma_\eta^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(E_{it} - \beta(x_{it} - \gamma_i\Delta F_t))^2$$

appear. This entails a new central sequence of

$$\Delta_{n,T}^{\beta} = \tilde{\Delta}_{n,T} - \beta\frac{1}{\sqrt{n}T\sigma_\eta^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_i F_{t-1}(E_{it} - \beta(x_{it} - \gamma_i\Delta F_t))$$

and a new empirical Fisher information of

$$J_{n,T}^{\beta} = \frac{1}{nT^2}\sum_{i=1}^{n}\left(\frac{1}{\sigma_u^2} + \frac{\beta^2}{\sigma_\eta^2}\right)\gamma_i^2\sum_{t=1}^{T}F_{t-1}^2 \Rightarrow \sigma_f^2\Psi_\Gamma\left(\frac{1}{\sigma_u^2} + \frac{\beta^2}{\sigma_\eta^2}\right)\int W_1^2 \mathrm{d}t.$$

Similarly, under the null hypothesis,

$$\begin{aligned}\Delta^{\beta}_{n,T} =&\tilde{\Delta}_{n,T} - \beta\frac{1}{\sqrt{nT}\sigma^2_\eta}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_i F_{t-1}\eta_{it} = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_i F_{t-1}\left(\frac{1}{\sigma^2_u}u_{it} - \frac{\beta}{\sigma^2_\eta}\eta_{it}\right)\\ \Rightarrow&\sigma_f\sqrt{\Psi_\Gamma\left(\frac{1}{\sigma^2_u}+\frac{\beta^2}{\sigma^2_\eta}\right)}\int W_1 \mathrm{d}W_2.\end{aligned}$$

To obtain the distribution under alternatives, again inserting $\Delta F_t = f_t + \frac{h}{\sqrt{nT}}F_{t-1}$ yields

$$\begin{aligned}\Delta^{\beta}_{n,T} =&\tilde{\Delta}_{n,T} - \beta\frac{1}{\sigma^2_\eta\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_i F_{t-1}(\eta_{it} + \beta\gamma_i\frac{h}{\sqrt{nT}}F_{t-1})\\ =&\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\gamma_i\sum_{t=1}^{T}F_{t-1}\left(\frac{u_{it}}{\sigma^2_u} - \frac{\beta}{\sigma^2_\eta}\eta_{it}\right) - h\left(\frac{1}{\sigma^2_u}+\frac{\beta^2}{\sigma^2_\eta}\right)\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\gamma_i F^2_{t-1},\end{aligned}$$

where we used (4.8) in the last step. Thus,

$$t_{n,T} = \frac{1}{\sqrt{nT}\sqrt{J^{\beta}_{n,T}}}\sum_{i=1}^{n}\gamma_i\sum_{t=1}^{T}F_{t-1}\left(\frac{u_{it}}{\sigma^2_u} - \frac{\beta}{\sigma^2_\eta}\eta_{it}\right) - h\sqrt{J^{\beta}_{n,T}}.$$

Recalling Lemma 4.A.1 and the discussion under the null hypothesis above, this indeed has the desired limiting distribution. ∎

4.A.5 Proof of Proposition 4.4.2

Proof Thanks to contiguity, we only have to show the desired approximation under the null hypothesis. From the proof of Theorem 4.3.1 recall that $\left\|F - \hat{F}\right\|_F = O_P(T/\sqrt{n})$ and $\left\|\Delta F - \Delta\hat{F}\right\|_F = O_P(\sqrt{T/n})$. By the same logic, we obtain $\left\|b - \hat{b}\right\|_F = O_P(\sqrt{T/n})$. Therefore, $\Delta\hat{F}'\Delta\hat{F} - \Delta\hat{F}'\Delta\hat{F} = O_P(n^{-1/2})$ and $\hat{\sigma}^2_f$ converges in probability to σ^2_f. Similarly, $\hat{\sigma}^2_g$ converges to σ^2_g, so that replacing the variances by its estimates does not change the limiting distribution. To show that $\hat{\Delta}_T - \tilde{\Delta}_T = o_P(1)$ it therefore suffices to demonstrate that $\hat{F}'\Delta\hat{F} - F'\Delta F = o_P(T)$ and $\hat{F}'\hat{b} - F'b = o_P(T)$. For the former, recall (4.A.9) to write

$$\begin{aligned}\hat{F}'\Delta\hat{F} - F'\Delta F =&F'(\Delta\hat{F} - \Delta F) + \Delta\hat{F}'(\hat{F} - F)\\ \leq&F'\Delta E\Lambda(\Lambda'\Lambda)^{-1} + F'\Delta y o_P(n^{-1}) + \Delta\hat{F}'E\Lambda(\Lambda'\Lambda)^{-1} + \Delta\hat{F}'y o_P(n^{-1})\\ \leq&\left\|F'\Delta E\Lambda\right\|_F\left\|(\Lambda'\Lambda)^{-1}\right\|_F + (\left\|F'\Delta F\right\|_F\left\|\Lambda\right\|_F + \left\|F'\Delta E\right\|_F)o_P(n^{-1})\\ &+\left\|\Delta\hat{F}'E\Lambda\right\|_F\left\|(\Lambda'\Lambda)^{-1}\right\|_F + (\left\|\Delta\hat{F}'F\Lambda'\right\|_F + \left\|\Delta\hat{F}'E\right\|_F)o_P(n^{-1})\\ =&O_P(\sqrt{n}T)O_P(n^{-1}) + (O_P(T)O_P(\sqrt{n}))o_P(n^{-1}) = O_P(n^{-1/2}T).\end{aligned}$$

For the latter part, the same rates apply. This is due to the fact, that, under the null hypothesis, the assumptions on b mimic those on ΔF. ∎

4.A.6 Additional Lemmas

Lemma 4.A.1 states that certain stochastic integral limits are not affected by considering local alternatives in $\sqrt{n}T$ neighbourhoods of the unit-root. The first two results follow directly as a limiting case of Lemma 1 in Phillips (1987b), who considers T neighbourhoods of unity.

Lemma 4.A.1 *Let $\{v_t\}_{t=1}^{\infty}$ be i.i.d. normally distributed and independent of F, denote its variance by σ_v^2, let V be its cumulative sums and let W_1 and W_2 be two independent Brownian motions. Then, for any alternative h, we have, as $T \to \infty$,*

1. $\frac{1}{T}\sum_{t=1}^{T} F_{t-1}v_t \Rightarrow \int W_1 \mathrm{d}W_2$,
2. $\frac{1}{T^2}\sum_{t=1}^{T} F_{t-1}^2 \Rightarrow \int W_1^2 \mathrm{d}t$,
3. $\frac{1}{T^2}\sum_{t=1}^{T} F_{t-1}V_{t-1} = O_P(1)$.

Proof For Item 3, note

$$\sum_{t=1}^{T}\sum_{s=1}^{t-1}\Delta F_s \sum_{s=1}^{t-1} v_s = \sum_{t=1}^{T} v_t \sum_{s=1}^{T}(T - \max(s,t))\Delta F_s.$$

Note that this term has mean zero (also conditionally on the f) while the variance is given by

$$\begin{aligned}\mathbb{E}\,\mathrm{Var}(\sum_{t=1}^{T} v_t \sum_{s=1}^{T}(T-\max(s,t))\Delta F_s|f) &= \sum_{t=1}^{T}\mathbb{E}(\sum_{s=1}^{T}(T-\max(s,t))\Delta F_s)^2\sigma_v^2 \\ &= \sum_{t=1}^{T}\mathrm{Var}(\sum_{s=1}^{T}(T-\max(s,t))\Delta F_s)\sigma_v^2.\end{aligned}$$

Again we split $\Delta F_s = f_s + \frac{h}{\sqrt{n}T}F_{s-1}$ and treat both variances separately. We have, for every t,

$$\mathrm{Var}(\sum_{s=1}^{T}(T-\max(s,t))f_s) = \sum_{s=1}^{T}(T-\max(s,t))^2\sigma_f^2 \leq T^3\sigma_f^2$$

and, for every t,

$$\begin{aligned}&\mathrm{Var}(\frac{h}{\sqrt{n}T}\sum_{s=1}^{T}(T-\max(s,t))F_{s-1}) \\ =&\frac{h^2}{nT^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}(T-\max(s_1,t))(T-\max(s_2,t))\,\mathrm{Cov}(F_{s_1-1},F_{s_2-1}) \\ \leq&\frac{h}{nT^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}(T^2\sqrt{T^2\sigma_f^4} = \frac{h}{n}T^3.\end{aligned}$$

Thus $\mathrm{Var}\sum_{t=1}^{T} F_{t-1}V_{t-1} = O(T^4)$ as required. ■

Lemma 4.A.2 *Consider a collection H of spectral densities f_h, $h \in H$, define $\omega_h^2 = 2\pi f_h(0)$ and denote by $\Sigma_T(f_h)$ the $T \times T$ Toeplitz matrix associated with f_h (i.e., $(\Sigma_T(f_h))_{k,l}$ is the $|k-l|$th Fourier coefficient of f_h). If the f_h are twice continuously differentiable with $\sup_{\lambda,h} |f_h''(\lambda)| < \infty$, then*

$$\sup_h \left\| A'(\Sigma_T(f_h) - \omega_h^2 I) \right\|_F = o(T).$$

Proof Calculation analogous to Lemma A.1 in Wichert et al. (2019) yield, with $\gamma_h(m)$ the mth Fourier coefficient of f_h,

$$\begin{aligned}\left\| A'(\Sigma_T(f_h) - \omega_h^2 I_T) \right\|_F^2 &= \sum_{s=1}^{T} \sum_{t=1}^{T} \left(\sum_{m=s-t+1}^{T-t} \gamma_h(m) - \omega_h^2 1_{s<t} \right)^2 \\ &\leq 5T \left(\sum_{m=-\infty}^{\infty} |\gamma_h(m)| \right) \sum_{m=1}^{\infty} \min(m,T) |\gamma_h(m)|\end{aligned}$$

Integrating by parts twice in (4.A.4) we obtain

$$\gamma_h(m) = \frac{1}{2\pi(\mathring{\imath} m)^2} \int_0^{2\pi} f_h''(\lambda) e^{-\mathring{\imath} m\lambda} \mathrm{d}\lambda \leq \frac{1}{|\mathring{\imath} m|^2} \sup_\lambda |f_h''(\lambda)| \sup_\lambda |e^{-\mathring{\imath} m\lambda}| = \frac{1}{m^2} \sup_\lambda |f_h''(\lambda)|. \tag{4.A.14}$$

Thus, combining yields

$$\frac{1}{T} \sup_h \left\| A'(\Sigma_T(f_h) - \omega_h^2 I) \right\|_F \leq \sup_{\lambda,h} |f_h''(\lambda)| \left(\sum_{m=-\infty}^{\infty} \frac{1}{m^2} \right)^{1/2} \left(\frac{1}{T} \sum_{m=1}^{\infty} \min(m,T) \frac{1}{m^2} \right)^{1/2},$$

which converges to zero. ■

Lemma 4.A.3 *Let H be a collection of time series with spectral density functions f_i and autocorrelation functions $(\gamma_i(m))_{m=0}^{\infty}$, $i \in H$. If the f_i are twice continuously differentiable with $\sup_{\lambda,i} |f_i''(\lambda)| < \infty$, and $\inf_{i,\lambda} f_i(\lambda) > 0$, then $\sup_{i \in H} \left\| \Sigma_T(f_i)^{-1} - \Sigma_T(1/f_i) \right\|_{spec} = o(1)$.*

Proof We proceed analogous to Gray (2005), see Chapters 4 and 5 for additional details on some of the inequalities used, who shows this result for a single time series. The proof proceeds by first approximating the inverse by an inverse circulant matrix $\mathcal{C}_T(f_i)$ and then approximating the inverse circulant by a Toeplitz matrix, i.e., we split

$$\begin{aligned}\left\| \Sigma_T(f_i)^{-1} - \Sigma_T(1/f_i) \right\|_{\mathrm{spec}} &\leq \left\| \Sigma_T(f_i)^{-1} - \mathcal{C}_T(f_i)^{-1} \right\|_{\mathrm{spec}} + \left\| \mathcal{C}_T(f_i)^{-1} - \Sigma_T(1/f_i) \right\|_{\mathrm{spec}} \\ &=: I + II.\end{aligned}$$

For I, it is sufficient to show that $\left\| \Sigma_T(f_i) - \mathcal{C}_T(f_i) \right\|_{\mathrm{spec}} = o(1)$, as the norm of both inverses is bounded by the inverse of the minimum of the spectral density. We have

$$\begin{aligned}\left\| \Sigma_T(f_i) - \mathcal{C}_T(f_i) \right\|_{\mathrm{spec}} &\leq \left\| \Sigma_T(f_i) - \mathcal{C}_T(\hat{f}_i) \right\|_{\mathrm{spec}} + \left\| \mathcal{C}_T(\hat{f}_i) - \mathcal{C}_T(f_i) \right\|_{\mathrm{spec}} \\ &= 2 \sum_{m=1}^{T-1} \frac{m}{T} \gamma_i(m)^2 + \frac{1}{T} \sum_{m=0}^{T-1} (f_i(2\pi m/T) - \hat{f}_i(2\pi m/T))^2.\end{aligned}$$

Note that the second summand is bounded by $2\sum_{m=T+1}^{\infty} \gamma_i(m)$, see p. 39 of Gray (2005). Therefore, the bound in (4.A.14) implies the desired result. ■

References

Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, Vol. 59, No. 3, p. 817.

Aristidou, C., D. I. Harvey, and S. J. Leybourne (2017). The Impact of the Initial Condition on Covariate Augmented Unit Root Tests. *Journal of Time Series Econometrics*, Vol. 9, No. 1, p. 20150013.

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics*, Vol. 122, pp. 137–183.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, Vol. 70, pp. 191–221.

— (2004). A PANIC Attack on Unit Roots and Cointegration. *Econometrica*, Vol. 72, pp. 1127–1177.

— (2010). Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation. *Econometric Theory*, Vol. 26, No. 04, pp. 1088–1114.

Baltagi, B. H. and C. Kao (2000). Nonstationary panels, cointegration in panels and dynamic panels: A survey. In: *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*. Ed. by B. H. Baltagi, T. B. Fomby, and R. C. Hill. Emerald, pp. 7–51.

Banerjee, A. (1999). Panel Data Unit Roots and Cointegration: An Overview. *Oxford Bulletin of Economics and Statistics*, Vol. 61, No. s1, pp. 607–629.

Banerjee, A., M. Marcellino, and I. Masten (2017). Structural FECM: Cointegration in large-scale structural FAVAR models. *Journal of Applied Econometrics*, Vol. 32, pp. 1069–1086.

Barigozzi, M. and L. Trapani (2018). Determining the dimension of factor structures in nonstationary large datasets. eprint: 1806.03647.

Becheri, I. G., F. C. Drost, and R. van den Akker (2015a). Asymptotically UMP Panel Unit Root Tests—the Effect of Heterogeneity in the Alternatives. *Econometric Theory*, Vol. 31, No. 3, pp. 539–559.

— (2015b). Unit root tests for cross-sectionally dependent panels: The influence of observed factors. *Journal of Statistical Planning and Inference*, Vol. 160, pp. 11–22.

Becheri, I. G., F. C. Drost, R. van den Akker, and O. Wichert (2016). The power envelope of panel unit root tests in case stationary alternatives offset explosive ones. *Statistics & Probability Letters*, Vol. 108, pp. 1–8.

Breitung, J. and S. Das (2007). Testing For Unit Roots in Panels with a Factor Structure. *Econometric Theory*, Vol. 24, pp. 88–108.

Breitung, J. and M. H. Pesaran (2008). Unit Roots and Cointegration in Panels. In: *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Ed. by L. Mátyás and P. Sevestre. Berlin, Heidelberg: Springer, pp. 279–322.

Brockwell, P. J. and R. A. Davis (1991). Time Series: Theory and Methods. Springer Series in Statistics. New York, NY: Springer.

Carvalho, M. d. and P. Júlio (2012). Digging out the PPP hypothesis: an integrated empirical coverage. *Empirical Economics*, Vol. 42, No. 3, pp. 713–744.

Choi, I. (2006). Nonstationary Panels. In: *Palgrave Handbook of Econometrics*. Ed. by H. Hassani, T. C. Mills, and K. Patterson.

— (2015). Almost All About Unit Roots. Foundations, Developments, and Applications. Cambridge: Cambridge University Press.

— (2017). Efficient estimation of nonstationary factor models. *Journal of Statistical Planning and Inference*, Vol. 183, pp. 18–43.

Davidson, J. (1994). Stochastic Limit Theory: An Introduction for Econometricians. Advanced Texts in Econometrics. Oxford University Press.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, Vol. 74, pp. 427–431.

Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Vol. 164, No. 1.

Elliott, G. and M. Jansson (2003). Testing for unit roots with stationary covariates. *Journal of Econometrics*, Vol. 115, No. 1, pp. 75–89.

Elliott, G. and E. Pesavento (2006). On the Failure of Purchasing Power Parity for Bilateral Exchange Rates after 1973. *Journal of Money, Credit, and Banking*, Vol. 38, No. 6, pp. 1405–1430.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica*, Vol. 64, No. 4, p. 813.

Gengenbach, C., F. C. Palm, and J. P. Urbain (2010). Panel Unit Root Tests in the Presence of Cross-Sectional Dependencies: Comparison and Implications for Modelling. *Econometric Reviews*, Vol. 29, No. 2, pp. 111–145.

Gray, R. M. (2005). Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory*, Vol. 2, pp. 155–239.

Gumbel, E. J. (1954). The Maxima of the Mean Largest Value and of the Range. *The Annals of Mathematical Statistics*, Vol. 25, No. 1, pp. 76–84.

Gutierrez, L. (2006). Panel Unit-root Tests for Cross-sectionally Correlated Panels: A Monte Carlo Comparison. *Oxford Bulletin of Economics and Statistics*, Vol. 68, No. 4, pp. 519–540.

Hallin, M., R. van den Akker, and B. J. M. Werker (2011). A class of simple distribution-free rank-based unit root tests. *Journal of Econometrics*, Vol. 163, pp. 200–214.

— (2016). Semiparametric error-correction models for cointegration with trends: Pseudo-Gaussian and optimal rank-based tests of the cointegration rank. *Journal of Econometrics*, Vol. 190, pp. 46–61.

Hansen, B. E. (1995). Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power. *Econometric Theory*, Vol. 11, pp. 1148–1171.

Jansson, M. (2008). Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis. *Econometrica*, Vol. 76, No. 5, pp. 1103–1142.

Jansson, M. and M. Nielsen (2012). Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis. *Econometrica*, Vol. 80, pp. 2321–2332.

Jeganathan, P. (1995). Some Aspects of Asymptotic Theory with Applications to Time Series Models. *Econometric Theory*, Vol. 11, No. 5, pp. 818–887.

Juodis, A. and J. Westerlund (2018). Optimal Panel Unit Root Testing with Covariates. *The Econometrics Journal*, Vol. 22, pp. 57–72.

Kurtz, T. G. and P. Protter (1991). Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *The Annals of Probability*, Vol. 19, pp. 1035–1070.

Lütkepohl, H. (1996). Handbook of Matrices. Wiley.

Maciejowska, K. (2010). Common Factors in Nonstationary Panel Data with a Deterministic Trend.

Magnus, J.R. and H. Neudecker (1999). Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley Series in Probability and Statistics: Texts and References Section. Wiley.

Moon, H.R. and B. Perron (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics*, Vol. 122, No. 1, pp. 81–126.

Moon, H. R., B. Perron, and P. C. B. Phillips (2007). Incidental trends and the power of panel unit root tests. *Journal of Econometrics*, Vol. 141, No. 2, pp. 416–459.

— (2014). Point-Optimal Panel Unit Root Tests with Serially Correlated Errors. *The Econometrics Journal*, Vol. 17, No. 3, pp. 338–372.

Müller, U. K. and G. Elliott (2003). Tests for Unit Roots and the Initial Condition. *Econometrica*, Vol. 71, No. 4.

Newey, W. K. and K. D. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies*, Vol. 61, No. 4, pp. 631–653.

O'Connell, P. G. (1998). The overvaluation of purchasing power parity. *Journal of International Economics*, Vol. 44, No. 1, pp. 1–19.

Onatski, A. and C. Wang (2019). Spurious Factor Analysis. *Working paper*.

Papell, D. H. (2006). The Panel Purchasing Power Parity Puzzle. *Journal of Money, Credit, and Banking*, Vol. 38, No. 2, pp. 447–467.

Patterson, K. (2011). Unit root tests in time series, Volume 1: key concepts and problems. Palgrave Macmillan.

— (2012). Unit root tests in time series, Volume 2: extensions and developments. Palgrave Macmillan.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, Vol. 22, pp. 265–312.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013). Panel unit root tests in the presence of a multifactor error structure. *Journal of Econometrics*, Vol. 175, No. 2, pp. 94–115.

Phillips, P. C. B. (1987a). Time Series Regression with a Unit Root. *Econometrica*, Vol. 55, No. 2, p. 277.

Phillips, P. C. B. (1987b). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika*, Vol. 74, p. 535.

Phillips, P. C. B. and H. R. Moon (1999). Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, Vol. 67, pp. 1057–1111.

Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *The Econometrics Journal*, Vol. 6, pp. 217–259.

Poncela, P., E. Ruiz, and K. Miranda (2020). Factor extraction using Kalman filter and smoothing: this is not just another survey. DES - Working Papers. Statistics and Econometrics. WS. Universidad Carlos III de Madrid. Departamento de Estadistica.

Reese, S. and J. Westerlund (2015). Panicca: PANIC on Cross-Section Averages. *Journal of Applied Econometrics*, Vol. 31, No. 6, pp. 961–981.

Saldías, M. (2013). A market-based approach to sector risk determinants and transmission in the euro area. *Journal of Banking & Finance*, Vol. 37, No. 11, pp. 4534–4555.

Schmitt, B. A. (1992). Perturbation bounds for matrix square roots and pythagorean sums. *Linear Algebra and its Applications*, Vol. 174, pp. 215–227.

Serfling, R. (1980). Approximation Theorems of Mathematical Statistics. Wiley Series in Probability and Statistics. Wiley.

Silva, S. D., K. Hadri, and A. R. Tremayne (2009). Panel unit root tests in the presence of cross-sectional dependence: finite sample performance and an application. *The Econometrics Journal*, Vol. 12, No. 2, pp. 340–366.

Stock, J. H. and M. W. Watson (1988). Testing for Common Trends. *Journal of the American Statistical Association*, Vol. 83, pp. 1097–1107.

Van der Vaart, A. W. (2000). Asymptotic Statistics. Cambridge University Press.

Westerlund, J. and J. Breitung (2013). Lessons From a Decade of IPS and LLC. *Econometric Reviews*, Vol. 32, No. 5-6, pp. 547–591.

Westerlund, J. (2013). A computationally convenient unit root test with covariates, conditional heteroskedasticity and efficient detrending. *Journal of Time Series Analysis*, Vol. 34, No. 4, pp. 477–495.

— (2015). The power of PANIC. *Journal of Econometrics*, Vol. 185, No. 2, pp. 495–509.

Wichert, O., I. G. Becheri, F. C. Drost, and R. van den Akker (2019). Local Asymptotic Equivalence of the Bai and Ng (2004) and Moon and Perron (2004) Frameworks for Panel Unit Root Testing. eprint: `1905.11184`.

Zhou, B., R. van den Akker, and B. J. M. Werker (2019). Semiparametrically point-optimal hybrid rank tests for unit roots. *The Annals of Statistics*, Vol. 47, No. 5, pp. 2601–2638.

CENTER DISSERTATION SERIES

CentER for Economic Research, Tilburg University, the Netherlands

No.	Author	Title	ISBN	Published
638	Pranav Desai	Essays in Corporate Finance and Innovation	978 90 5668 639 0	January 2021
639	Kristy Jansen	Essays on Institutional Investors, Asset Allocation Decisions, and Asset Prices	978 90 5668 640 6	January 2021
640	Riley Badenbroek	Interior Point Methods and Simulated Annealing for Nonsymmetric Conic Optimization	978 90 5668 641 3	February 2021
641	Stephanie Koornneef	It's about time: Essays on temporal anchoring devices	978 90 5668 642 0	February 2021
642	Vilma Chila	Knowledge Dynamics in Employee Entrepreneurship: Implications for parents and offspring	978 90 5668 643 7	March 2021
643	Minke Remmerswaal	Essays on Financial Incentives in the Dutch Healthcare System	978 90 5668 644 4	July 2021
644	Tse-Min Wang	Voluntary Contributions to Public Goods: A multi-disciplinary examination of prosocial behavior and its antecedents	978 90 5668 645 1	March 2021
645	Manwei Liu	Interdependent individuals: how aggregation, observation, and persuasion affect economic behavior and judgment	978 90 5668 646 8	March 2021
646	Nick Bombaij	Effectiveness of Loyalty Programs	978 90 5668 647 5	April 2021
647	Xiaoyu Wang	Essays in Microeconomics Theory	978 90 5668 648 2	April 2021
648	Thijs Brouwer	Essays on Behavioral Responses to Dishonest and Anti-Social Decision-Making	978 90 5668 649 9	May 2021
649	Yadi Yang	Experiments on hold-up problem and delegation	978 90 5668 650 5	May 2021
650	Tao Han	Imperfect information in firm growth strategy: Three essays on M&A and FDI activities	978 90 5668 651 2	June 2021

No.	Author	Title	ISBN	Published
651	Johan Bonekamp	Studies on labour supply, spending and saving before and after retirement	978 90 5668 652 9	June 2021
652	Hugo van Buggenum	Banks and Financial Markets in Microfounded Models of Money	978 90 5668 653 6	August 2021
653	Arthur Beddock	Asset Pricing with Heterogeneous Agents and Non-normal Return Distributions	978 90 5668 654 3	September 2021
654	Mirron Adriana Boomsma	On the transition to a sustainable economy: Field experimental evidence on behavioral interventions	978 90 5668 655 0	September 2021
655	Roweno Heijmans	On Environmental Externalities and Global Games	978 90 5668 656 7	August 2021
656	Lenka Fiala	Essays in the economics of education	978 90 5668 657 4	September 2021
657	Yuexin Li	Pricing Art: Returns, Trust, and Crises	978 90 5668 658 1	September 2021
658	Ernst Roos	Robust Approaches for Optimization Problems with Convex Uncertainty	978 90 5668 659 8	September 2021
659	Joren Koëter	Essays on asset pricing, investor preferences and derivative markets	978 90 5668 660 4	September 2021
660	Ricardo Barahona	Investor Behavior and Financial Markets	978 90 5668 661 1	October 2021
660	Stefan ten Eikelder	Biologically-based radiation therapy planning and adjustable robust optimization	978 90 5668 662 8	October 2021
661	Maciej Husiatyński	Three essays on Individual Behavior and New Technologies	978 90 5668 663 5	October 2021
662	Hasan Apakan	Essays on Two-Dimensional Signaling Games	978 90 5668 664 2	October 2021
663	Ana Moura	Essays in Health Economics	978 90 5668 665 9	November 2021
664	Frederik Verplancke	Essays on Corporate Finance: Insights on Aspects of the General Business Environment	978 90 5668 666 6	October 2021

No.	Author	Title	ISBN	Published
665	Zhaneta Tancheva	Essays on Macro-Finance and Market Anomalies	978 90 5668 667 3	November 2021
666	Claudio Baccianti	Essays in Economic Growth and Climate Policy	978 90 5668 668 0	November 2021
667	Hongwei Zhang	Empirical Asset Pricing and Ensemble Machine Learning	978 90 5668 669 7	November 2021
668	Bart van der Burgt	Splitsing in de Wet op de vennootschapsbelasting 1969 Een evaluatie van de Nederlandse winstbelastingregels voor splitsingen ten aanzien van lichamen	978 90 5668 670 3	December 2021
669	Martin Kapons	Essays on Capital Markets Research in Accounting	978 90 5668 671 0	December 2021
670	Xolani Nghona	From one dominant growth mode to another: Switching between strategic expansion modes	978 90 5668 672 7	December 2021
671	Yang Ding	Antecedents and Implications of Legacy Divestitures	978 90 5668 673 4	December 2021
672	Joobin Ordoobody	The Interplay of Structural and Individual Characteristics	978 90 5668 674 1	February 2022
673	Lucas Avezum	Essays on Bank Regulation and Supervision	978 90 5668 675 8	March 2022
674	Oliver Wichert	Unit-Root Tests in High-Dimensional Panels	978 90 5668 676 5	April 2022

The four chapters of this PhD thesis all concern panel unit-root tests, i.e., tests for the stationarity properties of a large number of time-series. The first chapter analyzes the testing problem in case stationary alternatives offset explosives ones. While the panel units are assumed to be independent in the first chapter, the subsequent chapters consider 'second-generation' panel unit-root tests which allow the different time series to be correlated through a factor structure. Chapter 2 considers two common approaches of modeling this dependence and shows that the associated unit-root testing problems are asymptotically equivalent. Using Le Cam's theory of statistical experiments, an optimal test is derived jointly in both setups. Chapter 3 studies unit-root tests for the underlying common factors rather than the idiosyncratic parts. It is demonstrated that unit root tests can be applied to a number of different factor estimates as if the factor was observed. A similar result is obtained for the case in which the factors have non-zero mean innovations. The final Chapter 4 revisits the testing problem for the unobserved common factors but exploits additional observed covariates that are known to be stationary to obtain higher powers.

Oliver Wichert (Hamburg, Germany, 1992) obtained his Research Master in Economics in 2016 from Tilburg University. He carried out his doctoral studies under the supervision of Bas Werker, Feico Drost, and Ramon van den Akker, focusing unit-root tests for high-dimensional panel data. He currently works as an economist at the European Competition Practice of Charles River Associates, based in Brussels.

ISBN: 978 90 5668 676 5
DOI: 10.26116/center-lis-2203

TILBURG SCHOOL OF ECONOMICS
AND MANAGEMENT

CentER Dissertation Series