Cross-Sectional Dependence Robust Block Bootstrap Panel Unit Root Tests*

Franz C. Palm Stephan Smeekes Jean-Pierre Urbain

Department of Quantitative Economics
Maastricht University
November 28, 2008

Abstract

In this paper we consider the issue of unit root testing in cross-sectionally dependent panels. We consider panels that may be characterized by various forms of cross-sectional dependence including (but not exclusive to) the popular common factor framework. We consider block bootstrap versions of the group-mean Im, Pesaran, and Shin (2003) and the pooled Levin, Lin, and Chu (2002) unit root coefficient DF-tests for panel data, originally proposed for a setting of no cross-sectional dependence beyond a common time effect. The tests, suited for testing for unit roots in the observed data, can be easily implemented as no specification or estimation of the dependence structure is required. Asymptotic properties of the tests are derived for T going to infinity and N finite. Asymptotic validity of the bootstrap tests is established in very general settings, including the presence of common factors and even cointegration across units. Properties under the alternative hypothesis are also considered. In a Monte Carlo simulation, the bootstrap tests are found to have rejection frequencies that are much closer to nominal size than the rejection frequencies for the corresponding asymptotic tests. The power properties of the bootstrap tests appear to be similar to those of the asymptotic tests.

1 Introduction

The use of panel data to test for unit roots and cointegration has become very popular recently. A major problem with tests for unit roots (and cointegration) in univariate time series is that they lack power for small sample sizes. Therefore one of the reasons people have

*Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail addresses: F.Palm@ke.unimaas.nl, S.Smeekes@ke.unimaas.nl, J.Urbain@ke.unimaas.nl.
This paper was presented at the conference on Factor Structures for Panel and Multivariate Time Series Data in Maastricht, September 2008 and at an Econometric Institute seminar at Erasmus University Rotterdam, November 2008. We thank the conference and seminar participants for helpful comments and suggestions. We thank NWO for financial support. The usual disclaimer applies.

turned to panel data, is to utilize the cross-sectional dimension to increase power. Another reason to use panel data is that one might be interested in testing a joint unit root hypothesis for N entities. The so-called first-generation panel unit root tests such as the tests proposed by Levin et al. (2002) and Im et al. (2003) are examples where the cross-sectional dimension is used to construct tests that have higher power than individual unit root tests. However, all the first-generation tests rely on independence along the cross-sectional dimension.

It was soon realized that cross-sectional independence is a highly unrealistic assumption for most settings encountered in practice, and it has been shown that the first-generation tests exhibit large size distortions in the presence of cross-sectional dependence (e.g. O'Connell, 1998). Therefore, so called second-generation panel unit root tests have been constructed to take the cross-sectional dependence into account in some way. These second-generation tests assume specific forms of the cross-sectional dependence as their application depends on modelling the structure of the dependence. Most tests model the cross-sectional dependence in the form of common factors, although the way the common factors are dealt with differs for each test. Examples of second-generation panel unit root tests are Bai and Ng (2004), Moon and Perron (2004), and Pesaran (2007). An extensive Monte Carlo comparison of these tests can be found in Gengenbach, Palm, and Urbain (2008). Breitung and Das (2008) provide an analytical comparison of several first- and second-generation tests in the presence of factor structures.

While the second-generation panel unit root tests can deal with common factor structures and contemporaneous dependence, they cannot deal with other forms of cross-sectional dependence, with the exception of Pedroni, Vogelsang, Wagner, and Westerlund (2008). Of particular interest for practical applications are dynamic interrelationships (an example of which is Granger causality). Our goal in this paper is to present panel unit root tests that can deal not only with common factors, but also with a wide range of other plausible dynamic dependencies.

The tool we use to achieve this is the bootstrap, and in particular the block bootstrap method. Two very useful features of the block bootstrap are that one does not have to model the dependence (both temporal and cross-sectional) in order to apply it, and that it is valid to use under a wide range of possible data generating processes (DGPs). This makes it an appropriate tool to use in this setting with N fixed, possibly large, and large T asymptotics.

Of course, the idea to use the bootstrap in cross-sectionally dependent panels is not new and has already been proposed by Maddala and Wu (1999),[1] but so far no one has considered the theoretical properties of the block bootstrap in this setup. There are theoretical results available for other bootstrap and related resampling methods. Chang (2004) considers sieve bootstrap unit root tests, but the sieve bootstrap can only be applied in panels under restrictive assumptions on the cross-sectional dependence. Kapetanios (2008) proposes a

[1] Also see Fachin (2007) and Di Iorio and Fachin (2008) for some successful applications of the block bootstrap in testing for cointegration in panels.

bootstrap resampling scheme which resamples in the cross-sectional dimension instead of the usual time dimension, but this is based on cross-sectional independence. Choi and Chue (2007) consider subsampling, which does allow for more general dependence, but as the authors themselves state (p. 235) "Notwithstanding these nice features of the subsampling approach, depending on the nature of the problem at hand, other methods like bootstrapping may work better in finite samples."

Hence, the properties of the block bootstrap are still largely unknown in this setting, while in fact the block bootstrap is quite popular among practitioners. We try to fill this gap by providing theoretical results, mainly about asymptotic validity, of block bootstrap panel unit root tests. The block bootstrap method we consider here is the moving-blocks bootstrap (Künsch, 1989), and is an extension of the univariate bootstrap unit root test proposed by Paparoditis and Politis (2003). We will consider a very general DGP that can capture many different interesting and relevant forms of cross-sectional and time dependence.

Our results provide the theoretical justification, supported by Monte Carlo evidence, for the use of the proposed panel unit root tests in applications where one is interested in testing for a unit root in the observed data, and where cross-sectional dependence of possibly unknown form might be present in the data. The tests can be easily implemented, as they do not require the specification and estimation of the cross-sectional dependence structure. For example, it is not necessary to know the number of common factors, nor to estimate these factors. It is not even necessary to know whether common factors are present in the data at all.

The structure of the paper is as follows. Section 2 explains the model and assumptions. The test statistics and the construction of the bootstrap versions are discussed in Section 3. We establish the asymptotic validity of the bootstrap tests (for $T \to \infty$ and N fixed) for various settings in Section 4. Finite sample performance, including block length selection, is investigated in Section 5. Section 6 concludes. All proofs and preliminary results are contained in the Appendix.

Finally, a word on notation. We use $|\cdot|$ to denote the Euclidean norm for vectors and matrices, i.e. $|v| = (v'v)^{1/2}$ for a vector v and $|M| = (\text{tr}\, M'M)^{1/2}$ for a matrix M. $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Convergence in distribution (probability) is denoted by $\xrightarrow{d}$ ($\xrightarrow{p}$). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript $*$ to the standard notation.

2 Cross-sectionally dependent panels

Let us first describe the model that we use for panels with possible unit roots and that allows for various types of cross-sectional and temporal dependence.

Let $y_t = (y_{1,t}, \ldots, y_{N,t})'$ $(t = 1, \ldots, T)$ be generated as

$$y_t = \Lambda F_t + w_t, \tag{1}$$

where $\Lambda = (\lambda_1, \dots, \lambda_N)'$, $F_t = (F_{1,t}, \dots, F_{d,t})'$ and $w_t = (w_{1,t}, \dots, w_{N,t})'$. Hence, f_t are common factors (d in total), Λ are the factor loadings, and v_t are the idiosyncratic components. Let $y_0 = 0$.

We let the factors and the idiosyncratic components be generated by

$$\begin{aligned} F_t &= \Phi F_{t-1} + f_t, \\ w_t &= \Theta w_{t-1} + v_t, \end{aligned} \tag{2}$$

where $\Phi = diag(\phi_1, \dots, \phi_d)$ and $\Theta = diag(\theta_1, \dots, \theta_N)$.

Furthermore we let f_t and v_t be constructed as

$$\begin{bmatrix} v_t \\ f_t \end{bmatrix} = \Psi(L)\varepsilon_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{v,t} \\ \varepsilon_{f,t} \end{bmatrix}, \tag{3}$$

where $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$ ($\Psi_0 = I$). We also divide $\Psi(z)$ as $\Psi(z) = (\Psi_1(z)', \Psi_2(z)')'$ where $\Psi_i(z) = (\Psi_{i1}(z), \Psi_{i2}(z))$, $i = 1, 2$.

We only need some mild conditions on $\Psi(z)$ and ε_t.

Assumption 1.

(i) $\det(\Psi(z)) \neq 0$ for all $\{z \in \mathbb{C} : |z| = 1\}$ and $\sum_{j=0}^{\infty} j|\Psi_j| < \infty$.

(ii) ε_t is i.i.d. with $\mathrm{E}\,\varepsilon_t = 0$, $\mathrm{E}\,\varepsilon_t\varepsilon_t' = \Sigma$ and $\mathrm{E}\,|\varepsilon_t|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

Our null hypothesis is H_0: $y_{i,t}$ has a unit root for all $i = 1, \dots, N$. As in Bai and Ng (2004) and Breitung and Das (2008), we can discern three different settings under which this can occur.

(A) $\theta_i = \phi_j = 1$ for all $i = 1, \dots, N$ and $j = 1, \dots, d$: both the common factors and the idiosyncratic components have a unit root. This is our first main setting.

(B) $|\theta_i| < 1$ for all $i = 1, \dots, N$, $\phi_j = 1$ for all $j = 1, \dots, d$: the common factors have a unit root while the idiosyncratic components are stationary. This is the setting where the units are cross-sectionally cointegrated. In accordance with most of the literature we shall call this *cross-unit cointegration*. We also discuss this case in detail.[2]

(C) $\theta_i = 1$ for all $i = 1, \dots, N$, $|\phi_j| < 1$ for all $j = 1, \dots, d$: the common factors are stationary while the idiosyncratic components have a unit root. We shall not discuss this case in detail in Section 4 but its properties can easily be derived from the previous two cases.

[2] We could also easily think of a setting in between setting A and B, i.e. one where $|\theta_i| < 1$ for all $i \in \mathcal{I}_1$ and $\theta_i = 1$ for all $i \in \mathcal{I}_2$ (with $\mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \dots, N\}$). In other words, where part of the units are cointegrated and others are not. We will not analyze this setting in detail as it is basically contained in the analysis of settings (A) and (B).

Note that we are not interested in which of the three settings occur, instead we simply want to test if $y_{i,t}$ has a unit root for all i.

We can discern different alternative hypotheses.[3] The following two are of interest to us.

- H_1^a: $y_{i,t}$ is stationary for all $i = 1, \dots, N$. This implies that $|\theta_i| < 1$ for all $i = 1, \dots, N$ and $|\phi_j| < 1$ for all $j = 1, \dots, d$.
- H_1^b: $y_{i,t}$ is stationary for a (significant) portion of the units. This implies that $|\phi_j| < 1$ for all $j = 1, \dots, d$; while $|\theta_i| < 1$ for all $i \in \mathcal{I}_1$ and $\theta_i = 1$ for all $i \in \mathcal{I}_2$, with $\mathcal{I}_1 \cup \mathcal{I}_2 = \{1, \dots, N\}$ and $n_1/N = \kappa > 0$, where n_1 is the number of elements of $\mathcal{I}_1$.[4]

Remark 1. Note that while the setting we adopt is fairly comparable to factor models such as those considered in Bai and Ng (2004) and Breitung and Das (2008), it is more general in several ways. First, it is very common to assume $\Psi_{12}(z) = \Psi_{21}(z) = 0$ and $\Sigma_{12} = \Sigma_{21} = 0$ such that the factors are independent of the idiosyncratic components. There is however no need to do so in order to obtain our theoretical results, and therefore we will not make this assumption in general. Whenever this assumption is made, this will be explicitly mentioned.

Moreover, and more importantly, in most common factor models only weak dependence between the idiosyncratic components is allowed. We do not make this assumption; instead we allow for a wide array of possible dependencies between the idiosyncratic components, both through Σ and $\Psi(z)$. Especially the lag polynomial allows for a wide range of dependencies, including all sorts of dynamic dependencies.

It is therefore that setting (A) is our main setting of interest, as simply setting $\lambda_i = 0$ for all $i = 1, \dots, N$ results in a model without common factors, where the cross-sectional dependence is completely generated by Σ and $\Psi(z)$. This setting is therefore the most general. We also analyze setting (B) as it has generated a lot of attention in the literature (mainly due to Bai and Ng, 2004), but it is in fact a very specialized setting that lacks the generality of setting (A).

Remark 2. One might wonder if we can actually call w_t idiosyncratic components given the degree of interdependence that we allow for, as Σ_{11} and $\Psi_{11}(z)$ might be non-diagonal without restrictions beyond full rank and $\Psi_{12}(z)$ might be non-zero. The reason why we keep doing so however is that we would like our setup to encompass two types of models. The first is the traditional approximate factor model, for which one would place additional conditions on the DGP to ensure that the idiosyncratic components would only be weakly dependent. The second is the multivariate time series model where we allow for common components

[3] Di Iorio and Fachin (2008) discuss several alternative hypotheses that are relevant when testing for the null of no panel cointegration. They also argue that the choice of the test statistic should depend on the alternative hypothesis. Their arguments are valid for the unit root setting as well.

[4] In principle we could also let some of the factors be $I(1)$ provided they have zero loadings on the units in $\mathcal{I}_1$. We do not consider this however.

as well as for dependence through a VARMA structure (and where the term idiosyncratic components is rather meaningless).

Hence, while we formulate our setup as a multivariate time series model, we retain the terminology belonging to the factor model to emphasize that such a model is covered by our setup as well. Note that in our simulations in Section 5 we will restrict the dependence between the idiosyncratic components to be weak.

3 Bootstrap unit root tests in panels

3.1 Test statistics

We will consider bootstrapping simplified versions of the Levin, Lin, and Chu (2002) [LLC] and Im, Pesaran, and Shin (2003) [IPS] test statistics. The first simplification is that we take the test statistics before corrections for mean and variance. The reason is that adding or multiplying the original test statistic and the bootstrap test statistic with the same number will obviously not have an effect on the performance of the tests. This is therefore a completely harmless simplification.

The second simplification is that we consider DF instead of ADF tests. Usually, the main reason to use ADF type of tests is to obtain asymptotically pivotal statistics. However, in the presence of complicated cross-sectional dependence it is often not possible to obtain asymptotically pivotal statistics anyway. There is therefore little reason (at least asymptotically) to use ADF instead of DF tests.

The third simplification is that we look at the DF coefficient test rather than the t-test. The main reason for this is that block bootstrapping naively studentized statistics leads to serious problems in terms of accuracy of the tests as discussed for example in Section 3.1.2 of Härdle, Horowitz, and Kreiss (2003). As this is a second order problem, it does not lead to invalidity of the bootstrap, but it does cause the bootstrap to converge at a slower rate than the standard asymptotic approximation.[5]

Given all these modifications, we prefer to call our test statistics "pooled" and "group-mean" instead of LLC and IPS, respectively. Note though that the essence of the LLC and IPS tests remains in our tests and that our methods can be trivially extended to the original LLC and IPS statistics if one so desires.[6]

Consider the pooled regression

$$\Delta y_{i,t} = \beta y_{i,t-1} + u_{i,t}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T, \tag{4}$$

[5] While this is a well known result in the statistics literature, it seems to have been widely ignored in the (applied) econometrics literature.

[6] Note that these tests could also be implemented when we have an unbalanced panel with different numbers of observations T_i over time, provided of course the number of observations increase. The implementation of the block bootstrap in such a setting would, while possible, become considerably more complicated.

for which we define the pooled statistic as

$$\tau_p = T\hat{\beta} \quad \text{where } \hat{\beta} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t}}{\sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1}^2}. \tag{5}$$

Also consider the individual regressions for $i = 1, \ldots, N$

$$\Delta y_{i,t} = \beta_i y_{i,t-1} + u_{i,t}, \qquad t = 1, \ldots, T. \tag{6}$$

Then we define our group-mean statistic as

$$\tau_{gm} = \frac{1}{N}\sum_{i=1}^{N} T\hat{\beta}_i \quad \text{where } \hat{\beta}_i = \frac{\sum_{t=2}^{T} y_{i,t-1}\Delta y_{i,t}}{\sum_{t=2}^{T} y_{i,t-1}^2}. \tag{7}$$

3.2 Bootstrap algorithm

We employ the following block bootstrap algorithm, which is a multivariate extension of the algorithm proposed by Paparoditis and Politis (2003) to test for unit roots in univariate time series.

Bootstrap Algorithm.

1. For $i = 1, \ldots, N$ estimate

$$y_{i,t} = \rho_i y_{i,t-1} + u_{i,t} \tag{8}$$

consistently by OLS and calculate

$$\hat{u}_{i,t} = y_{i,t} - \hat{\rho}_i y_{i,t-1} - \frac{1}{T-1}\sum_{t=2}^{T}(y_{i,t} - \hat{\rho}_i y_{i,t-1}). \tag{9}$$

Let $\hat{u}_t = (\hat{u}_{1,t}, \ldots, \hat{u}_{N,t})'$.

2. Choose a block length b (smaller than T). Draw $i_0, \ldots, i_{k-1}$ i.i.d. from the uniform distribution on $\{1, 2, \ldots, T-b\}$, where $k = \lfloor (T-2)/b \rfloor + 1$ is the number of blocks.

3. Construct the bootstrap errors $u_1^*, \ldots, u_T^*$ as follows. Let $u_1^* = y_1$. For $t > 1$, let

$$u_t^* = \hat{u}_{i_m+s}, \tag{10}$$

where $m = \lfloor (t-2)/b \rfloor$ and $s = t - mb - 1$.

4. Construct y_t^* recursively as

$$y_t^* = y_{t-1}^* + u_t^*. \tag{11}$$

5. Calculate the bootstrap versions of the group-mean and pooled statistics. Using the regression

$$\Delta y_{i,t}^* = \beta y_{i,t-1}^* + u_{i,t}^*, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{12}$$

calculate

$$\tau_p^* = T\hat{\beta}^*, \quad \text{where} \quad \hat{\beta}^* = \frac{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{i=1}^N \sum_{t=2}^T y_{i,t-1}^{*2}}, \tag{13}$$

and using the regressions for $i = 1, \ldots, N$,

$$\Delta y_{i,t}^* = \beta_i^* y_{i,t-1}^* + u_{i,t}^*, \qquad t = 1, \ldots, T \tag{14}$$

calculate

$$\tau_{gm}^* = \frac{1}{N} \sum_{i=1}^N T\hat{\beta}_i^*, \quad \text{where} \quad \hat{\beta}_i^* = \frac{\sum_{t=2}^T y_{i,t-1}^* \Delta y_{i,t}^*}{\sum_{t=2}^T y_{i,t-1}^{*2}}. \tag{15}$$

6. Repeat Steps 2 to 5 B times, obtaining bootstrap test statistics $\tau_\kappa^{*b}, b = 1, \ldots, B, \kappa = p, gm$, and select the bootstrap critical value c_α^* as $c_\alpha^* = \max\{c : \sum_{b=1}^B I(\tau_\kappa^{*b} < c) \leq \alpha\}$, or equivalently as the α-quantile of the ordered τ_κ^{*b} statistics. Reject the null of a unit root if τ_κ, calculated from equation (5) if $\kappa = p$ or equation (7) if $\kappa = gm$, is smaller than c_α^*, where α is the nominal level of the test.

Note that a crucial role in the analysis of our block bootstrap method will be played by the series

$$u_{i,t} = y_{i,t} - \rho_i y_{i,t-1}. \tag{16}$$

As in Paparoditis and Politis (2003), $\rho_i = 1$ should correspond to a unit root in $y_{i,t}$, while $\rho_i < 1$ should correspond to $y_{i,t}$ being stationary. Given our estimation of ρ_i in step 1, ρ_i is implicitly defined as

$$\rho_i = \lim_{t\to\infty} \frac{\mathrm{E}(y_{i,t-1}y_{i,t})}{\mathrm{E}(y_{i,t-1}^2)}, \tag{17}$$

which fulfills these correspondences (Paparoditis and Politis, 2003, Example 2.1).[7] Note that under H_0 we simply have that $u_{i,t} = y_{i,t} - y_{i,t-1}$ for all $i = 1, \ldots, N$ or in vector notation $u_t = \Delta y_t$.

[7] Given our definition of ρ_i it is clear that under stationarity we will always have $|\rho_i| < 1$. Paparoditis and Politis (2003, Example 2.2) show that if one estimates and hence implicitly defines ρ_i differently, for example through an ADF regression, it is not always the case that $\rho_i > -1$.

We need that the estimator in step 1 satisfies the properties $\hat{\rho}_i - \rho_i = O_p(T^{-1})$ if $\rho_i = 1$ and $\hat{\rho}_i - \rho_i = o_p(1)$ if $\rho_i < 1$. Our OLS estimator satisfies these properties (Paparoditis and Politis, 2003, Remark 2.3).

We also need the following assumption on the block length.

Assumption 2. Let $b \to \infty$ and $b = o(T^{1/2})$ as $T \to \infty$.

Remark 3. While we do not consider deterministic components, our tests can be modified to account for them in the same way as discussed by Levin et al. (2002) and Im et al. (2003). The crucial issue regarding the bootstrap tests is to implement exactly the same deterministic specification in the calculation of the test statistic on the bootstrap sample as in the calculation of the test statistic on the original sample. The only further modification of the bootstrap algorithm would be to include the appropriate deterministic components in step 1 as well.

We will not discuss deterministic components in detail in this paper as it would detract from our main objective to deal with cross-sectional dependence. There is a large literature on deterministic components and their impact. Part of the literature, for example on the local power of panel unit root tests in the case of incidental trends (Moon, Perron, and Phillips, 2007), depends on $N \to \infty$ and will therefore not apply here, although in finite samples these results will most likely have an impact on our tests as well. We would like to stress that the bootstrap will not solve any problems that arise due to the implementation of deterministic components. In order to avoid shifting the focus from dealing with the cross-sectional dependence to dealing with deterministic components, we do not consider them in this paper and refer to the existing literature instead (cf. Breitung and Das, 2005; Moon et al., 2007).

Remark 4. Unlike the methods considered by Moon and Perron (2004) and Pesaran (2007), which are essentially tests on the presence of a unit root in the idiosyncratic components as pointed out by Bai and Ng (2007), our methods are tests on the presence of a unit root in the observed data. Therefore in our setup there is no need to consider the properties of the common factors separately.

4 Asymptotic properties

In this section we will investigate the asymptotic properties of our (bootstrap) test statistics by letting T go to infinity while keeping N fixed. We study only T asymptotics for two reasons. First, it is standard practice in studies on resampling methods; see for example Chang (2004) and Choi and Chue (2007). Second, it is very difficult to obtain meaningful results for infinite N with our general model without making several stringent additional assumptions. However, as neither our bootstrap method nor our proofs of asymptotic validity depend on the finiteness of N, there is no reason to expect that asymptotic validity breaks down with joint T and N asymptotics.

4.1 Asymptotic properties under the main null hypothesis

In this section we investigate the validity of the bootstrap procedure proposed above in setting (A), i.e. where $\phi_i = 1$ for all $j = 1, \ldots, d$ and $\theta_i = 1$ for all $i = 1, \ldots, N$ or equivalently $\Phi = I_d$ and $\Theta = I_N$.

Note that under this null hypothesis we can write

$$u_t = \Delta y_t = \Gamma' x_t, \tag{18}$$

where $\Gamma = (I_N, \Lambda)'$, and

$$x_t = (v_t', f_t')' = \Psi(L)\varepsilon_t. \tag{19}$$

4.1.1 Asymptotic properties of the test statistics

We start by presenting the asymptotic distributions for the original series. After all, the bootstrap test statistics should mimic these distributions. The first step is the invariance principle, or functional central limit theorem.

Lemma 1. *Let y_t be generated under H_0 setting* (A) *and let Assumption 1 hold. Then, as $T \to \infty$,*

$$S_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \xrightarrow{d} B(r),$$

where $B(r) = \Gamma' \Psi(1) \Sigma^{1/2} W(r)$ and $W(r)$ denotes a $(N+d)$-dimensional standard Brownian motion.

Next define

$$\Omega = \lim_{T\to\infty} T^{-1} \operatorname{E} \left(\sum_{t=1}^{T} u_t \right) \left(\sum_{t=1}^{T} u_t \right)' \quad \text{and} \quad \Omega_0 = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \operatorname{E}(u_t u_t').$$

The limiting distributions now follow straightforwardly.

Theorem 1. *Let y_t be generated under H_0 setting* (A) *and let Assumption 1 hold. Then, as $T \to \infty$,*

$$\tau_p \xrightarrow{d} \frac{\sum_{i=1}^{N} \int_0^1 B_i(r) dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})}{\sum_{i=1}^{N} \int_0^1 B_i(r)^2 dr}$$

and

$$\tau_{gm} \xrightarrow{d} \frac{1}{N}\sum_{i=1}^{N} \frac{\int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})}{\int_0^1 B_i(r)^2 dr},$$

where $B_i(r)$ is the i-th element of $B(r) = \Gamma'\Psi(1)\Sigma^{1/2}W(r)$ and ω_i ($\omega_{0,i}$) is the (i,i)-th element of Ω (Ω_0).

Remark 5. To see how the Brownian motion $B(r)$ depends on the idiosyncratic components and on the factors, consider the following. Let $B_v(r) = \Psi_1(1)\Sigma^{1/2}W(r)$ be the Brownian motion generated by the idiosyncratic components and $B_f(r) = \Psi_2(1)\Sigma^{1/2}W(r)$ the Brownian motion generated by the common factors. With this definition $B(r) = B_v(r) + \Lambda B_f(r)$. Note that if $\Psi_{12}(L) = \Psi_{21}(L) = 0$ and $\Sigma_{12} = \Sigma_{21} = 0$ we can write $B_v(r) = \Psi_{11}(1)\Sigma_{11}^{1/2}W_1(r)$ and $B_f(r) = \Psi_{22}(1)\Sigma_{22}^{1/2}W_2(r)$ where $W_1(r)$ is of dimension N and $W_2(r)$ is of dimension d. For the i-th element of $B(r)$, $B_i(r)$, we can then write $B_i(r) = B_{v,i}(r) + \lambda_i' B_f(r)$.

4.1.2 Asymptotic properties of the bootstrap test statistics

Next we turn to the bootstrap test statistics. The first step is the bootstrap invariance principle.

Lemma 2. *Let y_t be generated under H_0 setting* (A)*. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$S_T^*(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} u_t^* \xrightarrow{d^*} B(r) \quad \textit{in probability.}$$

Lemma 2 shows that the bootstrap partial sum process correctly mimics the original partial sum process. The limiting distributions of the bootstrap test statistics now follow as given below.

Theorem 2. *Let y_t be generated under H_0 setting* (A)*. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$\tau_p^* \xrightarrow{d^*} \frac{\sum_{i=1}^{N}\int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})}{\sum_{i=1}^{N}\int_0^1 B_i(r)^2 dr} \quad \textit{in probability}$$

and

$$\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N}\sum_{i=1}^{N} \frac{\int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})}{\int_0^1 B_i(r)^2 dr} \quad \textit{in probability.}$$

Theorem 2 establishes the asymptotic validity of the proposed tests.

4.2 Asymptotic properties of the tests under cross-unit cointegration

In this section we look at setting (B), i.e. where $\Phi = I_N$ and $\theta_i < 1$ for all $i = 1, \ldots, N$ in 2. Note that in this case we may write

$$u_t = \Delta y_t = \Lambda f_t + \Delta w_t = \Lambda f_t + (1 - L)(1 - \theta L)^{-1} v_t. \tag{20}$$

Now let

$$\bar{\Psi}(z) = \begin{bmatrix} (1 - \theta L)^{-1} \Psi_1(z) \\ \Psi_2(z) \end{bmatrix}, \tag{21}$$

such that

$$\begin{bmatrix} w_t \\ f_t \end{bmatrix} = \bar{\Psi}(L) \varepsilon_t. \tag{22}$$

Note that $\bar{\Psi}(z)$ satisfies Assumption 1 just as $\Psi(z)$.

4.2.1 Asymptotic properties of the test statistics

We start again by presenting the invariance principle for the original series.

Lemma 3. *Let y_t be generated under H_0 setting* (B)*. Let Assumption 1 hold. Then, as $T \to \infty$,*

$$S_T(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t \xrightarrow{d} \bar{B}(r),$$

where $\bar{B}(r) = \Lambda B_f(r)$ and $B_f(r) = \Psi_2(1) \Sigma^{1/2} W(r)$.

Note that the resulting Brownian motion $\bar{B}(r)$ has reduced rank as it is only generated by the factors and not the idiosyncratic components.

Define

$$\bar{\Omega} = \lim_{T \to \infty} T^{-1} \, \mathrm{E} \left(\sum_{t=1}^{T} u_t \right) \left(\sum_{t=1}^{T} u_t \right)' \quad \text{and} \quad \bar{\Omega}_0 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathrm{E}(u_t u_t').$$

Now we can derive the asymptotic distributions.

Theorem 3. *Let y_t be generated under H_0 setting* (B)*. Let Assumption 1 hold. Then, as $T \to \infty$,*

$$\tau_p \xrightarrow{d} \frac{\sum_{i=1}^{N} \int_0^1 \bar{B}_i(r) d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\sum_{i=1}^{N} \int_0^1 \bar{B}_i(r)^2 dr}$$

and

$$\tau_{gm} \xrightarrow{d} \frac{1}{N}\sum_{i=1}^{N} \frac{\int_0^1 \bar{B}_i(r)d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\int_0^1 \bar{B}_i(r)^2 dr},$$

where $\bar{B}_i(r)$ is the i-th element of $\bar{B}(r)$ and $\bar{\omega}_i$ ($\bar{\omega}_{0,i}$) is the (i,i)-th element of $\bar{\Omega}$ ($\bar{\Omega}_0$).

4.2.2 Asymptotic properties of the bootstrap test statistics

Next we turn to the bootstrap series. Before presenting the bootstrap invariance principle, some discussion is in order.

As can be seen in Lemma 3, the Brownian motion generated by the partial sum process has reduced rank as it is only driven by the factors. In order to properly replicate the structure of the original series, the same should be true for the bootstrap partial sum process.

In the proof of Lemma 2 it is shown that the bootstrap series u_t^* behaves approximately like u_{i_m+s}, ignoring centering for the moment. Summing over the variables within one block, we obtain

$$\sum_{s=1}^{b} u_{i_m+s} = \sum_{s=1}^{b}(\Lambda f_{i_m+s} + \Delta w_{i_m+s}) = \sum_{s=1}^{b} \Lambda f_{i_m+s} + w_{i_m+b} - w_{i_m},$$

as all intermediate terms cancel against each other. This is also what happens in the partial sum of the original series and what causes that only the factors contribute to the Brownian motion.

However, summing both over the blocks and within the blocks, we obtain

$$\begin{aligned}\sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} u_{i_m+s} &= \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(\sum_{s=1}^{b} \Lambda f_{i_m+s} + w_{i_m+b} - w_{i_m} \right) \\ &= \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \Lambda f_{i_m+s} + \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left(w_{i_m+b} - w_{i_m} \right),\end{aligned}$$

where now the endpoints of the blocks do not cancel against each other as the blocks are randomly selected. The first term in this sum is the partial sum process of the factors, which generates the Brownian motion in Lemma 3 if we divide by $T^{1/2}$.

The second part is the partial sum process of the idiosyncratic components which generates an (unwanted) Brownian motion by dividing by $k^{1/2}$. As this rate is slower than $T^{1/2}$ by Assumption 2, the second part will vanish at rate $T^{1/2}/k^{1/2}$, so at rate $b^{1/2}$. Therefore, an increasing block length is crucial to make the second part vanish. In finite samples however one will always have a non-zero partial sum of the idiosyncratic components, although the magnitude will depend on both the sample size and the actual block length. Due to this, the covariance matrix of the resulting Brownian motion will always be of full rank in finite samples instead of reduced rank as in Lemma 3. It might therefore be expected that in this setting the block bootstrap might not work optimally in finite samples, although it is also clear that large block lengths should improve the performance of the tests in this setting.

Remark 6. This result is closely related to the result obtained by Paparoditis and Politis (2003, Lemma 8.5) in their discussion about the difference-based block bootstrap (DBB), in which one also bootstraps an over-differenced series. However, where the different bootstrap stochastic order leads to serious (power) problems for the DBB, it is what preserves the validity of the bootstrap tests in the case of cross-unit cointegration. The result described above is formalized in Lemma A.9 in the Appendix.

Given the discussion above, it is clear that the bootstrap validity is preserved in this setting, giving rise to the following bootstrap invariance principle.

Lemma 4. *Let y_t be generated under H_0 setting* (B)*. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$S_T^*(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} u_t^* \xrightarrow{d^*} \bar{B}(r) \quad \text{in probability.}$$

Finally we derive the limiting distributions of the test statistics, again establishing asymptotic validity of the bootstrap tests.

Theorem 4. *Let y_t be generated under H_0 setting* (B)*. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$\tau_p^* \xrightarrow{d^*} \frac{\sum_{i=1}^N \int_0^1 \bar{B}_i(r) d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\sum_{i=1}^N \int_0^1 \bar{B}_i(r)^2 dr} \quad \text{in probability}$$

and

$$\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N} \sum_{i=1}^N \frac{\int_0^1 \bar{B}_i(r) d\bar{B}_i(r) + \frac{1}{2}(\bar{\omega}_i - \bar{\omega}_{0,i})}{\int_0^1 \bar{B}_i(r)^2 dr} \quad \text{in probability.}$$

4.3 Asymptotic properties under the alternative hypothesis

Let us start by considering the alternative H_1^a (stationarity for all $y_{i,t}$).

Let us define

$$y_t = \Lambda(I_d - \Phi L)^{-1} f_t + (I_N - \Theta L)^{-1} v_t = \Gamma' \Psi^+(L) \varepsilon_t, \tag{23}$$

where

$$\Psi^+(L) = \begin{bmatrix} (I_N - \Theta L)^{-1} \Psi_1(L) \\ (I_d - \Phi L)^{-1} \Psi_2(L) \end{bmatrix}. \tag{24}$$

Note that the lag polynomial $\Psi^+(z)$ meets the conditions in Assumption 1.

We start by describing the asymptotic properties of our test statistics.

Lemma 5. *Let y_t be generated under H_1^a. Let Assumption 1 hold. Then, as $T \to \infty$,*

$$T^{-1}\tau_p \xrightarrow{p} \frac{\sum_{i=1}^N (\gamma_i(1) - \gamma_i(0))}{\sum_{i=1}^N \gamma_i(0)}$$

and

$$T^{-1}\tau_{gm} \xrightarrow{p} N^{-1} \sum_{i=1}^N \frac{\gamma_i(1) - \gamma_i(0)}{\gamma_i(0)},$$

where $\gamma_i(j) = \mathrm{E}(y_{i,t-j}y_{i,t})$.

Lemma 5 shows that both test statistics diverge to $-\infty$ under H_1^a as $\gamma_i(1) < \gamma_i(0)$ for all $i = 1, \ldots, N$. This is a necessary, but for bootstrap tests not sufficient step in showing consistency of the tests. The second step that is needed is to show that the bootstrap tests, and correspondingly the bootstrap critical values, do not diverge under H_1^a.

To that end, let $P = diag(\rho_1, \ldots, \rho_N)$ and consequently $u_t = (I_N - PL)y_t$. Then

$$u_t = (I_N - PL)\Gamma'\Psi^+(L)\varepsilon_t = \Psi^{++}(L)\varepsilon_t \tag{25}$$

where $\Psi^{++}(L) = (I_N - PL)\Gamma'\Psi^+$. Note that the summability condition from Assumption 1 still holds for this lag polynomial. Therefore we can give the following theorem.

Theorem 5. *Let y_t be generated under H_1^a. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$\tau_p^* \xrightarrow{d^*} \frac{\sum_{i=1}^N \int_0^1 B_i^+(r)dB_i^+(r) + \frac{1}{2}(\omega_i^+ - \omega_{0,i}^+)}{\sum_{i=1}^N \int_0^1 B_i^+(r)^2 dr} \quad \textit{in probability,}$$

and

$$\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N} \sum_{i=1}^N \frac{\int_0^1 B_i^+(r)dB_i^+(r) + \frac{1}{2}(\omega_i^+ - \omega_{0,i}^+)}{\int_0^1 B_i^+(r)^2 dr} \quad \textit{in probability,}$$

where $B_i^+(r)$ is the i-th element of $B^+(r) = \Psi^{++}(1)\Sigma^{1/2}W(r)$ and ω_i^+ and $\omega_{0,i}^+$ are the (i,i)-th elements of $\Omega^+ = \Psi^{++}(1)\Sigma\Psi^{++}(1)'$ and $\Omega_0^+ = \sum_{j=0}^\infty \Psi_j^{++}\Sigma\Psi_j^{++\prime}$, respectively.

Note that Lemma 5 and Theorem 5 jointly establish the consistency of our tests.

Let us now consider H_1^b. Again we first look at the properties of the test statistics. Let us first without loss of generality assume that the first n_1 units are I(0), while the rest is I(1). Hence, $\rho_i < 1$ for $i = 1, \ldots, n_1$ and $\rho_i = 1$ for $i = n_1 + 1, \ldots, N$.

Lemma 6. *Let y_t be generated under H_1^b. Let Assumption 1 hold. Then, as $T \to \infty$,*

$$\tau_p \xrightarrow{d} \frac{\sum_{i=1}^{n_1}(\gamma_i(1)-\gamma_i(0)) + \sum_{i=n_1+1}^{N}\left(\int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})\right)}{\sum_{i=n_1+1}^{N}\int_0^1 B_i(r)^2 dr}$$

and

$$T^{-1}\tau_{gm} \xrightarrow{p} N^{-1}\sum_{i=1}^{n_1}\frac{\gamma_i(1)-\gamma_i(0)}{\gamma_i(0)},$$

where $\gamma_i(j) = \mathrm{E}(y_{i,t-j}y_{i,t})$.

We see that the group-mean statistic diverges to $-\infty$ as it should. The pooled statistic does not diverge however, which means it is not consistent against this alternative. This is in fact not surprising, given that the pooled test is designed as a large N-test for homogeneous alternatives (also see Remark 8).

Let us turn to the bootstrap series and define $u_t = y_t - Py_{t-1}$, where now part of the ρ_i are equal to one and the rest is smaller than one. We may then write that

$$u_t = \Psi^{\#}(L)\varepsilon_t, \tag{26}$$

where the values for $\Psi^{\#}(L)$ for the I(1) components are determined as in the analysis under the null, and for the I(0) components as in the analysis above. The summability condition will obviously still hold and therefore we can directly state the limiting distributions as a corollary.

Corollary 1. *Let y_t be generated under H_1^b. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$\tau_p^* \xrightarrow{d^*} \frac{\sum_{i=1}^{N}\int_0^1 B_i^{\#}(r)dB_i^{\#}(r) + \frac{1}{2}(\omega_i^{\#} - \omega_{0,i}^{\#})}{\sum_{i=1}^{N}\int_0^1 B_i^{\#}(r)^2 dr} \quad \textit{in probability},$$

and

$$\tau_{gm}^* \xrightarrow{d^*} \frac{1}{N}\sum_{i=1}^{N}\frac{\int_0^1 B_i^{\#}(r)dB_i^{\#}(r) + \frac{1}{2}(\omega_i^{\#} - \omega_{0,i}^{\#})}{\int_0^1 B_i^{\#}(r)^2 dr} \quad \textit{in probability},$$

where $B_i^{\#}(r)$ is the i-th element of $B^{\#}(r) = \Psi^{\#}(1)\Sigma^{1/2}W(r)$ and $\omega_i^{\#}$ and $\omega_{0,i}^{\#}$ are the (i,i)-th elements of $\Omega^{\#} = \Psi^{\#}(1)\Sigma\Psi^{\#}(1)'$ and $\Omega_0^{\#} = \sum_{j=0}^{\infty}\Psi_j^{\#}\Sigma\Psi_j^{\#\prime}$, respectively.

Note that Lemma 6 and Corollary 1 jointly establish the consistency of the bootstrap group-mean test. Also note that the inconsistency of the pooled test does not depend on the bootstrap distribution, but purely on the original test statistic.

Remark 7. It might seem that our bootstrap method does not correctly reproduce the asymptotic null distribution if the alternative is true as the nuisance parameters are different than for example in Theorem 2, but this is not so straightforward. It all depends on how exactly the alternative is formulated related to the null. Had we formulated our alternative as $y_t = Py_{t-1} + u_t$ where $u_t = \Gamma'\Psi(L)\varepsilon_t$, the nuisance parameters would have been the same. The key to understanding this is that the process under the null corresponding to the process in (1) and (2) with Φ and Θ implying stationarity is not necessarily the same process with $\Phi = I_d$ and $\Theta = I_N$.

Remark 8. A few qualifications are in order regarding the inconsistency of the pooled test. First, the actual location of the pooled test can be seen to depend on both the proportion of stationary units (through n_1 in the sums) and the distance from the null (through the quantity $\gamma_i(1) - \gamma_i(0)$). If either becomes larger, the statistic will become more negative. Second, if T increases, the denominator will become smaller as the sum over the stationary units disappears (the b_{iT} part in the proof). Hence the test statistic will grow larger with increasing T, but the denominator will not go to zero as the nonstationary part does not vanish. Both factors imply that the actual power of the test can still be non-trivial and even reach 1.

5 Small sample performance

In this section we will investigate the small sample properties of our tests using Monte Carlo simulations. First we perform a simulation study to investigate the properties of our tests while fixing the block length to be a function of T only. Next we will perform a separate and smaller simulation study to investigate the selection of block lengths.

5.1 Monte Carlo design

We consider the following DGP for the simulation study.

$$y_t = \Lambda F_t + w_t, \tag{27}$$

where

$$\begin{aligned} F_t &= \phi F_{t-1} + f_t, \\ w_{i,t} &= \theta_i w_{i,t-1} + v_{i,t}. \end{aligned} \tag{28}$$

Furthermore,

$$\begin{aligned} v_t &= A_1 v_{t-1} + \varepsilon_{1,t} + B_1 \varepsilon_{1,t-1}, \\ f_t &= \alpha_2 f_{t-1} + \varepsilon_{2,t} + \beta_2 \varepsilon_{2,t-1}, \end{aligned} \tag{29}$$

where $\varepsilon_{2,t} \sim N(0,1)$ and

$$\varepsilon_{1,t} \sim N\left(0, \Sigma\right),$$

where Σ is generated as in Chang (2004):

1. Generate an $N \times N$ matrix $U \sim U[0,1]$. Construct $H = U(U'U)^{-1/2}$.

2. Generate N eigenvalues $\lambda_1, \dots, \lambda_N$ with $\lambda_1 = r$, $\lambda_N = 1$ and $\lambda_i \sim U[r,1]$ for $i = 2, \dots, N-1$.

3. Let $\Lambda = diag(\lambda_1, \dots, \lambda_N)$. Then let $\Sigma = H\Lambda H'$.

We consider both $r = 1$ (no cross-sectional dependence) and $r = 0.1$.

We consider five settings regarding the parameters in equations (27) and (29) in accordance with Gengenbach et al. (2008).

I No common factor, unit root for all idiosyncratic components: $\lambda_i = 0$, $\theta_i = 1$ for all $i = 1, \dots, N$.

II Unit root in common factor and idiosyncratic components: $\phi = 1$, $\theta_i = 1$ for all $i = 1, \dots, N$ and $\lambda_i \sim U[-1,3]$.

III Unit root in common factor, stationary idiosyncratic components: $\phi = 1$, $\theta_i \sim U[0.8,1]$ and $\lambda_i \sim U[-1,3]$. This is the setting of cross-unit cointegration.

IV No common factor, stationary idiosyncratic component: $\theta_i \sim U[0.8,1]$ and $\lambda_i = 0$ for all $i = 1, \dots, N$. This is under the alternative hypothesis.[8]

V Stationary common factor and idiosyncratic component: $\phi = 0.95$, $\theta_i \sim U[0.8,1]$ and $\lambda_i \sim U[-1,3]$. This is also under the alternative hypothesis.[8]

We consider two different options for the parameters A_1 and B_1:

1. No dynamic dependence: $A_1 = B_1 = 0$.

2. Dynamic autoregressive moving-average cross-sectional dependence: A_1 and B_1 are non-diagonal.

   We let $A_1 = \Xi$, where

$$\Xi = \begin{bmatrix} \xi_1 & \xi_1\eta_1 & \xi_1\eta_1^2 & \cdots & \xi_1\eta_1^{N-1} \\ \xi_2\eta_2 & \xi_2 & \xi_2\eta_2^2 & \cdots & \xi_2\eta_2^{N-2} \\ \vdots & & \ddots & & \vdots \\ \xi_N\eta_N^{N-1} & \cdots & \xi_N\eta_N^2 & \xi_N\eta_N & \xi_N \end{bmatrix}, \tag{30}$$

[8]The reported power estimates are not size adjusted.

where $\xi_i, \eta_i \sim U[-0.5, 0.5]$. To ensure stationarity and invertibility we impose that $\det(I_N - A_1 z) \neq 0$ for $\{z \in \mathbb{C} : |z| \leq 1.2\}$.

Furthermore we let $B_1 = \Omega$. We construct Ω in much the same way as Σ. Let $M = HLH'$ where $H = U(U'U)^{-1/2}$, with U a $N \times 1$-vector of $U[0,1]$-variables, and L is a diagonal matrix with on the diagonal $L_1, \ldots, L_N$ where $L_1 = 0.1$, $L_N = 1$ and $L_2, \ldots L_{N-1} \sim U[0.1, 1]$. We then let $\Omega = 2 * M - I_N$. By generating Ω this way we assure that $I_N + \Omega$ is of full rank. Note that invertibility is not guaranteed (on purpose).

The parameters of the common factor in (29), α_2 and β_2, are taken in accordance with the setting for the idiosyncratic components, so if the dependence for the idiosyncratic components is of the ARMA type, then the same will hold for the common factor. Note that for both Σ and the $\Psi(1)$ matrix derived from A and B the eigenvalues are bounded if $N \to \infty$; as such these parameters can be regarded as weak dependence parameters.

For all combinations of the parameters described above we consider all combinations of $T = 25, 50, 100$ and $N = 5, 25, 50$. As several parameters in our DGP are chosen randomly, we repeat the simulations for each setting ten times, and store the mean, median, minimum and maximum. We only report results for the mean here. The mean is representative as in general there is little dispersion between the simulation results. The other results are available upon request. The results are based on 2000 simulations and the Warp-Speed bootstrap (Giacomini, Politis, and White, 2007) is used to obtain estimates for the rejection frequencies of the bootstrap tests.

In our simulation study we consider the LLC and IPS tests (with lag lengths selected by BIC), denoted by τ_{llc} and τ_{ips} respectively, and the bootstrap pooled and group-mean tests, denoted by τ_p and τ_{gm}. We also consider a bootstrap test based on the median of the individual test statistics, denoted by τ_{med}. This test might be more robust to outlying units than the test based on the mean (also see the discussion in Di Iorio and Fachin, 2008). While we do not consider this test explicitly in our theoretical analysis as the median presents difficulties for asymptotic analysis, it is clear that a median based test will be valid as well as we can show that the joint bootstrap distribution of the individual DF statistics is asymptotically valid. Block lengths of the bootstrap tests were taken as $b = 1.75T^{1/3}$, which amounts to blocks of length 6, 7 and 9 for sample size 25, 50 and 100 respectively, which is within the range usually considered in the literature. We return to the issue of block length selection in Section 5.3.

5.2 Monte Carlo results

Table 1 presents results for the setting without common factors. It can be noted in general that the asymptotic tests have poor size for $T = 25$, which is mainly caused by the performance of the BIC, as this tends to select too large lag lengths for $T = 25$.[9] From $T = 50$ on this does not happen anymore. The first part of the table presents results for the setting without any dependence (both temporal and cross-sectional). It can be seen that the asymptotic tests have good size properties for $T = 50$ and $T = 100$, while the bootstrap tests are undersized increasing in N. The second part lists results for the setting where there is only contemporaneous correlation. The asymptotic tests have slight positive size distortions here, while the bootstrap tests are somewhat undersized. The third and fourth part of the table give results for the model with autoregressive moving-average errors. It is clear here that the asymptotic tests are quite oversized, while the bootstrap tests perform well although there is some undersize for large N. There is little difference between the three bootstrap tests.

Table 2 present the results for the model with a nonstationary common factor and nonstationary idiosyncratic components. For all three settings considered the table shows that the bootstrap tests have good size properties, while the asymptotic tests have large size distortions increasing with N. The bootstrap tests again perform very similarly.

Table 3 gives the results for the model with cross-unit cointegration, i.e. with a nonstationary common factor and stationary idiosyncratic components. The asymptotic tests have very large size distortions, and while the size distortions of the bootstrap tests are significantly less, they are still large. As expected it indeed seems that the bootstrap tests do not perform very well in this setting. The problem partly arises, especially for the group-mean test, because for some units the loadings will be very close to zero, thereby making that unit effectively stationary and hence inflating the test statistic. In such a situation we may expect the median-based test to be more robust, and it indeed seems to perform somewhat better than the group-mean test although it still suffers from considerable size distortions.

Table 4 presents results for the model under the alternative without a common factor. The power of the bootstrap tests is satisfactory, and as expected, increases with both T and N. The only setting in which we can directly compare the power of the asymptotic and the bootstrap tests is the setting of no dependence, and here power results are very similar. Given that the bootstrap tests are somewhat undersized, this shows that the power of the bootstrap tests is good. In the other settings the power of the bootstrap test is somewhat less than the power of the asymptotic tests, which can be explained by the size distortions of the asymptotic tests. Note that the bootstrap tests perform similarly.

Table 5 gives results for power with a common factor. It can be seen that the power of the bootstrap tests still increases with T and N, although power is less than in Table 4 and especially the increase in power with N is less. This is not surprising as the common factor which is present in every unit ensures that the information on the order of integration is not increased by much by the addition of units in the panel. The fact that the power of the asymptotic tests is higher than the power of the bootstrap tests can be explained by the

[9] A similar result was obtained by Hlouskova and Wagner (2006).

Table 1: Size properties without common factors (setting I)

A_1, B_1	Σ	T	N	τ_{llc}	τ_p	τ_{ips}	τ_{gm}	τ_{med}
$A_1 = 0$,	$r = 1$	25	5	0.140	0.024	0.141	0.020	0.025
$B_1 = 0$		25	25	0.211	0.001	0.183	0.005	0.009
		25	50	0.260	0.000	0.207	0.001	0.002
		50	5	0.076	0.031	0.051	0.024	0.033
		50	25	0.063	0.004	0.060	0.011	0.014
		50	50	0.055	0.000	0.056	0.003	0.004
		100	5	0.077	0.032	0.049	0.032	0.035
		100	25	0.062	0.009	0.051	0.014	0.020
		100	50	0.056	0.001	0.051	0.005	0.010
	$r = 0.1$	25	5	0.151	0.026	0.159	0.022	0.031
		25	25	0.200	0.003	0.197	0.006	0.010
		25	50	0.236	0.000	0.215	0.001	0.002
		50	5	0.111	0.033	0.070	0.028	0.037
		50	25	0.077	0.006	0.072	0.009	0.017
		50	50	0.067	0.001	0.069	0.004	0.005
		100	5	0.113	0.040	0.066	0.031	0.039
		100	25	0.084	0.013	0.064	0.015	0.021
		100	50	0.073	0.003	0.067	0.007	0.012
$A_1 = \Xi$,	$r = 1$	25	5	0.215	0.054	0.207	0.082	0.055
$B_1 = \Omega$		25	25	0.235	0.004	0.198	0.032	0.016
		25	50	0.280	0.000	0.237	0.010	0.004
		50	5	0.154	0.052	0.123	0.097	0.054
		50	25	0.113	0.008	0.097	0.032	0.020
		50	50	0.109	0.001	0.106	0.023	0.010
		100	5	0.152	0.067	0.110	0.099	0.064
		100	25	0.130	0.013	0.108	0.028	0.023
		100	50	0.117	0.004	0.096	0.026	0.015
	$r = 0.1$	25	5	0.222	0.056	0.212	0.063	0.051
		25	25	0.252	0.003	0.238	0.020	0.012
		25	50	0.265	0.000	0.214	0.007	0.003
		50	5	0.197	0.049	0.146	0.059	0.046
		50	25	0.144	0.011	0.131	0.052	0.030
		50	50	0.127	0.001	0.119	0.018	0.008
		100	5	0.187	0.055	0.129	0.094	0.054
		100	25	0.158	0.012	0.129	0.029	0.023
		100	50	0.143	0.004	0.119	0.027	0.015

Table 2: Size properties with common factors (setting II)

A_1, B_1	Σ	T	N	τ_{llc}	τ_p	τ_{ips}	τ_{gm}	τ_{med}
$A_1 = 0$,	$r = 1$	25	5	0.212	0.029	0.188	0.024	0.030
$B_1 = 0$		25	25	0.288	0.014	0.337	0.015	0.023
		25	50	0.352	0.009	0.425	0.013	0.014
		50	5	0.165	0.036	0.104	0.030	0.037
		50	25	0.218	0.021	0.280	0.022	0.030
		50	50	0.263	0.018	0.362	0.020	0.023
		100	5	0.160	0.039	0.095	0.030	0.038
		100	25	0.213	0.030	0.258	0.025	0.035
		100	50	0.253	0.020	0.342	0.022	0.024
	$r = 0.1$	25	5	0.229	0.030	0.204	0.029	0.034
		25	25	0.314	0.025	0.389	0.021	0.027
		25	50	0.359	0.023	0.467	0.021	0.024
		50	5	0.193	0.036	0.122	0.031	0.036
		50	25	0.283	0.032	0.332	0.026	0.031
		50	50	0.296	0.027	0.393	0.025	0.028
		100	5	0.182	0.040	0.114	0.030	0.036
		100	25	0.277	0.034	0.316	0.030	0.038
		100	50	0.315	0.031	0.390	0.031	0.035
$A_1 = \Xi$,	$r = 0.1$	25	5	0.269	0.022	0.243	0.023	0.022
$B_1 = \Omega$		25	25	0.351	0.012	0.381	0.013	0.016
		25	50	0.406	0.007	0.448	0.010	0.011
		50	5	0.253	0.025	0.177	0.050	0.024
		50	25	0.348	0.015	0.358	0.018	0.020
		50	50	0.378	0.013	0.411	0.016	0.021
		100	5	0.252	0.032	0.172	0.032	0.031
		100	25	0.373	0.023	0.362	0.025	0.028
		100	50	0.420	0.023	0.425	0.022	0.028

Table 3: Size properties with cross-unit cointegration (setting III)

A_1, B_1	Σ	T	N	τ_{llc}	τ_p	τ_{ips}	τ_{gm}	τ_{med}
$A_1 = 0$,	$r = 1$	25	5	0.410	0.129	0.332	0.103	0.102
$B_1 = 0$		25	25	0.629	0.198	0.585	0.173	0.198
		25	50	0.698	0.224	0.643	0.173	0.216
		50	5	0.612	0.200	0.463	0.169	0.170
		50	25	0.782	0.267	0.620	0.259	0.251
		50	50	0.816	0.281	0.671	0.282	0.283
		100	5	0.674	0.240	0.550	0.333	0.257
		100	25	0.798	0.303	0.642	0.282	0.275
		100	50	0.845	0.336	0.688	0.399	0.336
	$r = 0.1$	25	5	0.461	0.096	0.356	0.086	0.084
		25	25	0.612	0.128	0.561	0.111	0.126
		25	50	0.667	0.166	0.598	0.148	0.177
		50	5	0.554	0.141	0.362	0.136	0.137
		50	25	0.742	0.179	0.552	0.183	0.184
		50	50	0.764	0.184	0.591	0.175	0.183
		100	5	0.651	0.171	0.431	0.187	0.149
		100	25	0.806	0.177	0.622	0.198	0.176
		100	50	0.819	0.211	0.633	0.253	0.210
$A_1 = \Xi$,	$r = 0.1$	25	5	0.427	0.047	0.357	0.049	0.041
$B_1 = \Omega$		25	25	0.549	0.062	0.492	0.071	0.064
		25	50	0.597	0.067	0.522	0.065	0.068
		50	5	0.466	0.085	0.309	0.125	0.089
		50	25	0.670	0.098	0.516	0.122	0.092
		50	50	0.698	0.089	0.539	0.107	0.087
		100	5	0.464	0.100	0.305	0.136	0.115
		100	25	0.701	0.117	0.527	0.165	0.104
		100	50	0.738	0.109	0.575	0.148	0.101

Table 4: Power properties without common factors (setting IV)

A_1, B_1	Σ	T	N	τ_{llc}	τ_p	τ_{ips}	τ_{gm}	τ_{med}
$A_1 = 0$,	$r = 1$	25	5	0.607	0.507	0.651	0.354	0.337
$B_1 = 0$		25	25	0.829	0.866	0.980	0.894	0.892
		25	50	0.875	0.958	0.999	0.996	0.996
		50	5	0.754	0.757	0.829	0.810	0.773
		50	25	0.995	0.999	1.000	1.000	1.000
		50	50	1.000	1.000	1.000	1.000	1.000
		100	5	0.905	0.929	0.989	0.974	0.946
		100	25	1.000	1.000	1.000	1.000	1.000
		100	50	1.000	1.000	1.000	1.000	1.000
	$r = 0.1$	25	5	0.553	0.508	0.608	0.357	0.361
		25	25	0.832	0.827	0.985	0.850	0.854
		25	50	0.878	0.919	1.000	0.981	0.980
		50	5	0.856	0.630	0.887	0.648	0.633
		50	25	0.998	0.989	1.000	1.000	0.999
		50	50	1.000	1.000	1.000	1.000	1.000
		100	5	0.928	0.943	0.996	0.985	0.950
		100	25	1.000	1.000	1.000	1.000	1.000
		100	50	1.000	1.000	1.000	1.000	1.000
$A_1 = \Xi$,	$r = 1$	25	5	0.573	0.491	0.625	0.498	0.391
$B_1 = \Omega$		25	25	0.798	0.618	0.972	0.862	0.768
		25	50	0.866	0.711	0.998	0.984	0.934
		50	5	0.816	0.700	0.865	0.779	0.675
		50	25	0.991	0.955	1.000	1.000	0.997
		50	50	1.000	0.983	1.000	1.000	1.000
		100	5	0.896	0.891	0.975	0.950	0.955
		100	25	1.000	0.996	1.000	1.000	1.000
		100	50	1.000	0.999	1.000	1.000	1.000
	$r = 0.1$	25	5	0.575	0.354	0.609	0.413	0.361
		25	25	0.767	0.754	0.949	0.860	0.820
		25	50	0.841	0.729	0.997	0.981	0.941
		50	5	0.755	0.693	0.814	0.729	0.607
		50	25	0.989	0.934	1.000	0.999	0.996
		50	50	0.999	0.989	1.000	1.000	1.000
		100	5	0.985	0.762	0.994	0.884	0.830
		100	25	0.999	0.995	1.000	1.000	1.000
		100	50	1.000	1.000	1.000	1.000	1.000

Table 5: Power properties with common factors (setting V)

A_1, B_1	Σ	T	N	τ_{llc}	τ_p	τ_{ips}	τ_{gm}	τ_{med}
$A_1 = 0$,	$r = 1$	25	5	0.597	0.213	0.512	0.173	0.177
$B_1 = 0$		25	25	0.784	0.345	0.795	0.306	0.344
		25	50	0.840	0.377	0.858	0.326	0.365
		50	5	0.883	0.505	0.783	0.504	0.474
		50	25	0.989	0.677	0.957	0.672	0.658
		50	50	0.997	0.723	0.974	0.750	0.722
		100	5	0.978	0.822	0.970	0.830	0.802
		100	25	1.000	0.944	1.000	0.964	0.933
		100	50	1.000	0.956	0.999	0.981	0.961
	$r = 0.1$	25	5	0.590	0.189	0.506	0.140	0.148
		25	25	0.767	0.265	0.772	0.236	0.263
		25	50	0.822	0.271	0.812	0.234	0.266
		50	5	0.828	0.415	0.682	0.423	0.399
		50	25	0.981	0.503	0.935	0.487	0.497
		50	50	0.994	0.497	0.958	0.509	0.508
		100	5	0.963	0.695	0.928	0.792	0.731
		100	25	1.000	0.840	0.996	0.883	0.834
		100	50	1.000	0.853	0.998	0.912	0.850
$A_1 = \Xi$,	$r = 0.1$	25	5	0.534	0.078	0.480	0.087	0.074
$B_1 = \Omega$		25	25	0.707	0.096	0.674	0.085	0.100
		25	50	0.753	0.124	0.720	0.126	0.124
		50	5	0.801	0.243	0.669	0.244	0.190
		50	25	0.953	0.270	0.876	0.329	0.253
		50	50	0.972	0.272	0.903	0.328	0.276
		100	5	0.967	0.482	0.904	0.619	0.454
		100	25	0.998	0.527	0.987	0.712	0.494
		100	50	0.999	0.596	0.989	0.756	0.564

large size distortions of the asymptotic tests in this case. The bootstrap tests all have similar power properties, although the median-based test seems to be somewhat less powerful than the group-mean test.

5.3 Block length selection

The Monte Carlo experiment in the previous section was done with fixed block lengths. It is well known from the literature on block bootstrap that the block length selected can have an a large effect on the performance of any kind of application of the block bootstrap. That is of course valid here as well. Added to the usual issues relating to the structure of the temporal dependence, block length selection is also important in our setting in the case of cross-unit cointegration, where one can expect that large blocks are needed based on the discussion in Section 4.2.2. Our discussion here mirrors the discussion in Paparoditis and Politis (2003, Section 6.1), who discuss the selection of block lengths for univariate unit root tests.

Quite some research has been done on optimal block length selection in the framework of stationary time series. As noted in Paparoditis and Politis (2003) in order to talk about optimality one needs to set a criterion that is to be optimized. This criterion will depend on the type of application of the bootstrap (variance estimation, confidence intervals, hypothesis tests, etc.). Using higher order asymptotics, it has been found for stationary series that an optimal block length b_{opt} is of the form

$$b_{\text{opt}} = CT^{1/\kappa}, \tag{31}$$

where κ is a known integer depending on the type of application and C is usually unknown and depends on the data. Härdle et al. (2003) and Lahiri (2003) give an overview on optimal block lengths in stationary time series.

Several methods have been proposed in the setting where one can describe b_{opt} as in (31). Some are based on estimation of C by exploiting the dependence of C on certain quantities that can be estimated. Bühlmann and Künsch (1999) and Politis and White (2004) are examples of such methods that are applicable for variance estimation. Lahiri, Furukawa, and Lee (2007) propose a *plug-in* method, based on the jackknife-after-bootstrap, that is also applicable for confidence intervals and hypothesis test.

A different method is the *subsampling* approach by Hall, Horowitz, and Jing (1995). The attractive feature of this method is that it avoids estimation of C. This feature, as well as the ease of its implementation, has made this method a popular choice among practitioners. It does however require knowledge of κ to implement it.

The problem with nonstationary time series is that κ is unknown here, as the required asymptotic expansions have not been developed yet. This makes it very difficult to implement any of the methods discussed above using a well funded choice of κ. Paparoditis and Politis (2003) discuss this issue and propose some heuristic ideas to determine κ.

An alternative strategy to the methods discussed above is provided by the *minimum volatility* method and *calibration* method proposed by Politis, Romano, and Wolf (1999). These methods do not require knowledge of κ. The minimum volatility method involves calculating critical values using a range of block lengths and selecting the optimal one in the region where the critical values have the lowest volatility.

We will focus here on the calibration method,[10] which we will describe below. In particular, we will consider the Warp-Speed calibration method, which was considered as a modification of the original calibration method by Giacomini et al. (2007) for the purpose of constructing confidence intervals. We present the procedure for hypothesis tests below for completeness.

[10] We also considered the minimum volatility method, the subsampling method by Hall et al. (1995) and the plug-in method by Lahiri et al. (2007), the latter two with the value for κ based on the results for stationary time series, but all these methods were inferior to the calibration method; see Remark 9.

Block length selection by Warp-Speed calibration.

1. Choose a starting value b_0 for the block length. Using this value, generate K bootstrap samples: $(\{y_t^1\}, \ldots, \{y_t^K\})$. Calculate the statistic of interest for each bootstrap sample, say $\hat{\theta}^k(b_0)$ for $k = 1, \ldots, K$. Using the empirical distribution of the statistics, calculate the bootstrap critical value $c(b_0)$.

2. Let $(b_1, \ldots, b_M)$ be the candidate block lengths. For each $i = 1, \ldots, M$ and $k = 1, \ldots, K$, construct one bootstrap resample from the bootstrap sample $\{y_t^k\}$ using block length b_i, call this $\{y_t^k(i)\}$. Using each resample calculate the statistic of interest, say $\hat{\theta}^{*k}(b_i)$.

3. Using the distribution of $\hat{\theta}^{*k}(b_i)$ for $k = 1, \ldots, K$, calculate the bootstrap resample critical value $c^*(b_i)$ for all $i = 1, \ldots, M$.

4. Select the optimal block length b_{opt} such that

$$b_{\text{opt}} = \underset{b_i, i=1\ldots,M}{\arg\min} |c^*(b_i) - c(b_0)|. \tag{32}$$

To reduce the dependence on b_0 one can apply this algorithm iteratively, by using b_{opt} as the starting block length in the next iteration and continuing until convergence.

To analyze the performance of the method, we performed a small Monte Carlo experiment using the same DGP as in Section 5.1 applying the tests τ_p and τ_{gm}. Based on 500 simulations, we let the block length be selected by the Warp-Speed calibration method, and using the same seed, we run the tests for a wide range of fixed block lengths (up to 0.75 times the sample size) to determine the optimal block length. As starting block lengths we take the fixed block lengths from the previous section, while we take $K = 199$. Due to computational costs we do not iterate the algorithm.

Results for size are given in Table 6. Optimal block lengths are determined as that block length which gives an empirical rejection frequency the closest to the nominal level (5%). It can be seen that while the optimal rejection frequencies are not obtained using the block length selection method, the rejection frequencies for setting I and II are reasonably close. However, while the selected block lengths do increase for setting III, they do not increase sufficiently compared to the optimal block lengths and size distortions persist.

Results for power are presented in Table 7. Optimal block lengths here are selected as the block lengths that give the highest power possible. One should regard this with caution, as optimal block lengths under the alternative hypothesis are difficult to define, as higher power could come at the expense of good size properties under the null. It is therefore not clear that high power is the criterion that should be optimized.[11] What is clear though, is that choosing an unnecessarily large block length will decrease power. The results show that the calibration method performs reasonably satisfactorily.

[11] Note that even when using size-adjusted power this problem would still be present.

Table 6: Size properties with block length selection

					τ_p				τ_{gm}			
Set.	A_1, B_1	Σ	T	N	RF	AvB	OpB	OpRF	RF	AvB	OpB	OpRF
I	$A_1 = 0$,	$r = 1$	25	5	0.022	3.992	1	0.038	0.018	4.172	1	0.038
	$B_1 = 0$		25	25	0.002	2.778	1	0.012	0.012	2.762	2	0.038
			50	5	0.028	6.472	5	0.050	0.020	6.468	2	0.050
			50	25	0.012	3.610	1	0.020	0.014	3.760	1	0.038
		$r = 0.1$	25	5	0.044	4.414	4	0.050	0.014	4.378	2	0.050
			25	25	0.010	2.944	1	0.014	0.012	2.912	1	0.062
			50	5	0.034	6.990	1	0.052	0.034	6.482	5	0.054
			50	25	0.016	3.658	1	0.024	0.014	3.672	3	0.038
	$A_1 = \Xi$,	$r = 1$	25	5	0.032	4.056	2	0.038	0.012	4.504	2	0.044
	$B_1 = \Omega$		25	25	0.002	2.304	1	0.008	0.008	2.332	2	0.032
			50	5	0.042	7.968	6	0.052	0.024	9.906	13	0.050
			50	25	0.006	2.918	3	0.010	0.046	4.144	5	0.052
		$r = 0.1$	25	5	0.024	4.502	4	0.048	0.042	5.066	10	0.042
			25	25	0.008	2.578	1	0.024	0.028	2.968	3	0.032
			50	5	0.042	7.212	6	0.054	0.036	7.092	4	0.050
			50	25	0.008	3.314	5	0.012	0.010	3.434	2	0.020
II	$A_1 = 0$,	$r = 1$	25	5	0.022	4.878	2	0.046	0.020	4.950	2	0.052
	$B_1 = 0$		25	25	0.018	3.510	2	0.042	0.008	3.286	1	0.060
			50	5	0.018	8.850	12	0.050	0.034	7.086	5	0.050
			50	25	0.018	5.566	2	0.046	0.024	5.064	1	0.050
		$r = 0.1$	25	5	0.028	5.494	6	0.046	0.020	4.938	1	0.048
			25	25	0.016	3.978	1	0.038	0.004	4.154	2	0.048
			50	5	0.020	10.292	5	0.048	0.032	8.892	2	0.056
			50	25	0.044	7.154	4	0.050	0.038	6.098	4	0.048
	$A_1 = \Xi$,	$r = 0.1$	25	5	0.006	4.946	3	0.020	0.004	5.128	5	0.014
	$B_1 = \Omega$		25	25	0.000	4.572	3	0.038	0.006	4.558	6	0.026
			50	5	0.010	8.606	2	0.036	0.012	8.802	7	0.028
			50	25	0.014	6.270	3	0.018	0.016	5.532	6	0.024
III	$A_1 = 0$,	$r = 1$	25	5	0.098	5.442	16	0.054	0.128	5.038	16	0.076
	$B_1 = 0$		25	25	0.192	4.208	17	0.078	0.144	3.976	19	0.070
			50	5	0.126	9.266	30	0.050	0.108	8.056	27	0.052
			50	25	0.230	5.984	37	0.090	0.266	5.140	33	0.114
		$r = 0.1$	25	5	0.024	6.626	5	0.046	0.018	6.394	1	0.052
			25	25	0.152	4.296	15	0.056	0.102	4.122	15	0.052
			50	5	0.044	11.222	10	0.048	0.038	9.448	26	0.052
			50	25	0.180	7.512	35	0.062	0.132	6.572	24	0.076
	$A_1 = \Xi$,	$r = 0.1$	25	5	0.046	5.140	5	0.050	0.044	4.756	8	0.052
	$B_1 = \Omega$		25	25	0.082	4.044	10	0.046	0.114	3.734	14	0.050
			50	5	0.032	9.536	6	0.048	0.018	8.686	3	0.040
			50	25	0.070	7.192	1	0.052	0.086	6.266	20	0.052

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 0.05); OpRF = rejection frequency corresponding to the optimal block length.

Table 7: Power properties with block length selection

					τ_p				τ_{gm}			
Set.	A_1, B_1	Σ	T	N	RF	AvB	OpB	OpRF	RF	AvB	OpB	OpRF
IV	$A_1 = 0,$	$r = 1$	25	5	0.218	4.088	2	0.436	0.186	4.060	2	0.244
	$B_1 = 0$		25	25	0.916	2.928	1	0.964	0.986	2.596	1	0.990
			50	5	0.740	5.974	5	0.856	0.884	5.438	4	0.948
			50	25	1.000	3.732	1	1.000	1.000	3.168	1	1.000
		$r = 0.1$	25	5	0.686	4.530	4	0.816	0.340	4.194	1	0.636
			25	25	0.984	2.868	1	0.992	0.966	2.606	2	0.976
			50	5	0.652	6.592	5	0.792	0.748	5.730	3	0.850
			50	25	1.000	3.766	1	1.000	1.000	3.118	1	1.000
	$A_1 = \Xi,$	$r = 1$	25	5	0.102	3.788	6	0.132	0.192	3.908	4	0.270
	$B_1 = \Omega$		25	25	0.476	2.458	4	0.428	0.756	2.390	3	0.860
			50	5	0.582	7.586	4	0.682	0.570	7.412	1	0.798
			50	25	0.988	3.192	2	0.992	1.000	2.850	1	1.000
		$r = 0.1$	25	5	0.234	4.076	4	0.330	0.250	3.966	3	0.400
			25	25	0.642	2.414	2	0.704	0.784	2.390	3	0.880
			50	5	0.264	6.976	3	0.410	0.198	6.702	4	0.354
			50	25	0.998	3.362	2	1.000	1.000	2.994	1	1.000
V	$A_1 = 0,$	$r = 1$	25	5	0.118	4.712	1	0.300	0.138	4.386	1	0.254
	$B_1 = 0$		25	25	0.328	3.802	1	0.550	0.242	3.608	2	0.402
			50	5	0.340	8.188	3	0.486	0.372	6.868	3	0.438
			50	25	0.538	6.338	1	0.682	0.484	5.638	2	0.684
		$r = 0.1$	25	5	0.226	4.652	2	0.406	0.160	4.484	1	0.268
			25	25	0.182	4.342	1	0.250	0.100	4.276	2	0.240
			50	5	0.182	9.964	3	0.292	0.214	8.812	1	0.364
			50	25	0.414	7.096	1	0.508	0.580	5.858	1	0.752
	$A_1 = \Xi,$	$r = 0.1$	25	5	0.064	5.120	8	0.120	0.058	4.574	2	0.136
	$B_1 = \Omega$		25	25	0.058	4.638	3	0.092	0.042	4.108	5	0.096
			50	5	0.110	8.774	2	0.184	0.098	8.166	3	0.158
			50	25	0.200	6.424	7	0.286	0.258	5.508	8	0.288

RF = rejection frequency with block length selection; AvB = average block length selected; OpB = optimal block length (such that the corresponding rejection frequency is as close as possible to 1); OpRF = rejection frequency corresponding to the optimal block length.

To conclude, using the calibration method improves on using a fixed block length, but it is not optimal. It is clear that a lot of work still needs to be done on this topic, especially from a theoretical perspective.

Remark 9. As mentioned before, we compared the calibration method to the subsampling approach of Hall et al. (1995), the plug-in method of Lahiri et al. (2007) and the minimum volatility method. The subsampling method tends to select block lengths in a somewhat unpredictable way, although the obtained rejection frequencies are reasonably close (but somewhat inferior) to those obtained with the calibration method. The plug-in method generally favors too small block lengths, regardless of the underlying DGP. The minimum volatility method selects block lengths almost uniformly over the range of allowed lengths, thereby selecting too large block lengths in general. The results are available on request.

6 Conclusion

We have established the asymptotic validity of two block bootstrap panel unit root tests for a model that includes various kinds of cross-sectional and temporal dependence. This includes a common factor structure and possibly cross-unit cointegration. The tests are very simple pooled and group-mean tests based on the popular LLC and IPS tests. The finite sample properties of our test statistics have also been investigated and shown to be satisfactory in general. There also seems to be little difference between the bootstrap tests considered.

While for most specific settings (in particular cross-unit cointegration) some tests can be found that perform better for that particular setting, it is a lot more difficult to find a test that is valid for all the settings for which our bootstrap tests are valid. Moreover, there are currently very few tests that are valid in the empirically relevant case of dynamic cross-sectional dependence, while our tests are valid even in that setting. Our tests are very easy to implement as no specification and estimation of the dependence structure is necessary, and will therefore be very useful for practice when the true form of the cross-sectional (and temporal) dependence is not known and robustness to the unknown cross-sectional dependence matters. In fact, quite a lot of practitioners already use the bootstrap to account for cross-sectional dependence for the reasons listed above. Hence, this work provides the necessary theoretical justification.

On the basis of the theoretical and simulation results in this paper, we conclude that it is legitimate to use the proposed tests in practice when testing for unit roots in the observed data of a panel of fixed N entities, in the presence of various forms of cross-sectional dependence. The block bootstrap algorithm described in Section 3 can be straightforwardly implemented whereby block lengths can be selected using the Warp-Speed calibration method.

This study still leaves several ends open. First, while we touched upon the subject of block length selection, much still needs to be done as at the moment there does not exist a fully satisfactory method to select block lengths. Second, while our derivations do not depend on small N in any way, it will be interesting to see what happens if $N \to \infty$. As explained, such a theoretical analysis is very difficult in our setting but it is certainly worth further research. Third, the specification of deterministic components remains an open issue. While a "naive" implementation of deterministic components is quite straightforward, and can even be seen to be valid without too much difficulty, experience has shown that including "naive" deterministic terms in panels is hardly ever a good solution. Thus, further investigation of this issue is also merited.

References

Bai, J. and S. Ng (2004). A PANIC attack on unit roots and cointegration. *Econometrica 72*, 1127–1177.

Bai, J. and S. Ng (2007). Panel unit root tests with cross-section dependence: A further investigation. Working paper.

Breitung, J. and S. Das (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica 59*, 414–433.

Breitung, J. and S. Das (2008). Testing for unit roots in panels with a factor structure. *Econometric Theory 24*, 88–108.

Bühlmann, P. and H. R. Künsch (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis 31*, 295–310.

Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency. *Journal of Econometrics 120*, 263–293.

Choi, I. and T. K. Chue (2007). Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices. *Journal of Applied Econometrics 22*, 401–428.

Di Iorio, F. and S. Fachin (2008). Testing for cointegration in dependent panels via residual-based bootstrap methods. Working paper.

Fachin, S. (2007). Long-run trends in internal migrations in Italy: a study in panel cointegration with dependent units. *Journal of Applied Econometrics 22*, 401–428.

Gengenbach, C., F. C. Palm, and J.-P. Urbain (2008). Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling. *Econometric Reviews*, forthcoming.

Giacomini, R., D. N. Politis, and H. White (2007). A Warp-Speed method for conducting Monte Carlo experiments involving bootstrap estimators. Working paper.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*, 561–574.

Härdle, W., J. L. Horowitz, and J.-P. Kreiss (2003). Bootstrap methods for time series. *International Statistical Review 71*, 435–459.

Hlouskova, J. and M. Wagner (2006). The performance of panel unit root and stationarity tests: results from a large scale simulation study. *Econometric Reviews 25*, 85–116.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*, 53–74.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *Econometrics Journal 11*, 377–395.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*, 1217–1241.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag.

Lahiri, S. N., K. Furukawa, and Y.-D. Lee (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Statistical Methodology 4*, 292–321.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics 108*, 1–24.

Maddala, G. S. and S. Wu (1999). A comparative study for unit root tests with panel data and a simple new test. *Oxford Bulletin of Economics and Statistics 61*, 631–652.

Moon, H. R. and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics 122*, 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 141*, 416–459.

O'Connell, P. G. J. (1998). The overvaluation of purchasing power parity. *Journal of International Economics 44*, 1–19.

Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica 71*, 813–855.

Pedroni, P., T. J. Vogelsang, M. Wagner, and J. Westerlund (2008). Robust unit root and cointegration rank tests for time series panels. Working paper.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-sectional dependence. *Journal of Applied Econometrics 22*, 265–312.

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

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

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. New York: Springer-Verlag.

Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews 23*, 53–70.

A Appendix

Proof of Lemma 1. Note that by Assumption 1 $W_T(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \xrightarrow{d} \Sigma^{1/2}W(r)$. Then it follows from standard asymptotic theory for linear processes (see for example Phillips and Solo, 1992) that, uniformly in r,

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} x_t = \Psi(1)W_T(r) + o_p(1),$$

and consequently $T^{-1/2}\sum_{t=1}^{\lfloor Tr \rfloor} x_t \xrightarrow{d} \Psi(1)\Sigma^{1/2}W(r)$. The result then follows straightforwardly by the continuous mapping theorem. □

To prove Theorem 1 we need some moments that appear in the asymptotic distributions.

Lemma A.1. *Let y_t be generated under H_0 setting* (A)*. Let Assumption 1 hold. Then*

(i) $\Omega = \lim_{T\to\infty} T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} u_t\right)\left(\sum_{t=1}^{T} u_t\right)' = \Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma,$

(ii) $\Omega_0 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{E}(u_t u_t') = \sum_{j=0}^{\infty}\Gamma'\Psi_j\Sigma\Psi_j'\Gamma.$

Proof of Lemma A.1. For part (i), note that

$$\Omega = \lim_{T\to\infty} T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} u_t\right)\left(\sum_{t=1}^{T} u_t\right)' = \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\,u_s u_t'$$
$$= \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T} E(\sum_{j=0}^{\infty}\Gamma'\Psi_j\varepsilon_{s-j})(\sum_{j=0}^{\infty}\Gamma'\Psi_j\varepsilon_{t-j})'$$
$$= \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Gamma'\Psi_i\,\mathrm{E}\,\varepsilon_{s-i}\varepsilon_{t-j}'\Psi_j'\Gamma$$
$$= \lim_{T\to\infty} T^{-1}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Gamma'\Psi_i\left(\sum_{s=1}^{T}\sum_{t=1,t\neq s-i+j}^{T}\mathrm{E}\,\varepsilon_{s-i}\,\mathrm{E}\,\varepsilon_{t-j}'\right)\Psi_j'\Gamma$$
$$+ \lim_{T\to\infty} T^{-1}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Gamma'\Psi_i\left(\sum_{s=1}^{T}\mathrm{E}\,\varepsilon_{s-i}\varepsilon_{s-i}'\right)\Psi_j'\Gamma$$
$$= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Gamma'\Psi_i\Sigma\Psi_j'\Gamma = \Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma.$$

For part (ii) we have

$$\Omega_0 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{E}\left(u_t u_t'\right) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E(\sum_{j=0}^{\infty}\Gamma'\Psi_j\varepsilon_{t-j})(\sum_{j=0}^{\infty}\Gamma'\Psi_j\varepsilon_{t-j})'$$
$$= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\Gamma'\Psi_i\,\mathrm{E}\,\varepsilon_{t-i}\varepsilon_{t-j}'\Psi_j'\Gamma = \sum_{i=0}^{\infty}\Gamma'\Psi_i\Sigma\Psi_i'\Gamma.$$

This completes the proof. □

Lemma A.2. *Let y_t be generated under H_0 setting* (A). *Let Assumption 1 hold. Then, as $T\to\infty$, we have for $i = 1,\dots,N$,*

(i) $T^{-1}\sum_{t=1}^{T} y_{i,t-1}\Delta y_{i,t} \xrightarrow{d} \int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})$,

(ii) $T^{-2}\sum_{t=1}^{T} y_{i,t-1}^2 \xrightarrow{d} \int_0^1 B_i(r)^2 dr$.

Proof of Lemma A.2. The proof follows straightforwardly from Lemma 1, Lemma A.1 and the continuous mapping theorem. □

Proof of Theorem 1. The proof follows directly from Lemma A.2. □

In order to derive the bootstrap invariance principle we need three preliminary lemmas that build on each other. We use the fact that we have linear processes in our derivation. As for the original series, we first derive the properties for the bootstrap equivalent of ε_t which we then extend to u_t^*.

Lemma A.3 establishes some moments for this series, while Lemma A.4 establishes the corresponding invariance principle. Lemma A.5 then extends this to u_t^*.

Lemma A.3. *Define* $H_m^* = b^{-1/2}\sum_{s=1}^{b}(\varepsilon_{i_m+s} - \mathrm{E}^*\,\varepsilon_{i_m+s})$. *If Assumptions 1 and 2 hold, we have*

(i) $\mathrm{E}^*\,H_m^* = 0$,

(ii) $\mathrm{E}^*\,H_m^* H_m^{*\prime} = \Sigma + o_p(1)$.

Proof of Lemma A.3. Statement (i) is trivial. To prove statement (ii), write

$$
\begin{aligned}
\mathrm{E}^*\,H_m^* H_m^{*\prime} &= \mathrm{E}^*\left[b^{-1}\left(\sum_{s=1}^{b}(\varepsilon_{i_m+s} - \mathrm{E}^*\,\varepsilon_{i_m+s})\right)\left(\sum_{s=1}^{b}(\varepsilon_{i_m+s} - \mathrm{E}^*\,\varepsilon_{i_m+s})\right)'\right] \\
&= b^{-1}\sum_{s_1=1}^{b}\sum_{s_2=1}^{b}\left(\mathrm{E}^*\,\varepsilon_{i_m+s_1}\varepsilon_{i_m+s_2}' - \mathrm{E}^*\,\varepsilon_{i_m+s_1}\,\mathrm{E}^*\,\varepsilon_{i_m+s_2}'\right) \\
&= b^{-1}\sum_{s_1=1}^{b}\sum_{s_2=1}^{b}\left[\frac{1}{T-b}\sum_{t=1}^{T-b}\varepsilon_{t+s_1}\varepsilon_{t+s_2}' - \frac{1}{(T-b)^2}\left(\sum_{t=1}^{T-b}\varepsilon_{t+s_1}\right)\left(\sum_{t=1}^{T-b}\varepsilon_{t+s_2}'\right)\right] \\
&= \frac{1}{b(T-b)}\sum_{s_1=1}^{b}\sum_{s_2=1}^{b}\sum_{t=1}^{T-b}\varepsilon_{t+s_1}\varepsilon_{t+s_2}' - b^{-1}\left(\frac{1}{T-b}\sum_{s=1}^{b}\sum_{t=1}^{T-b}\varepsilon_{t+s}\right)\left(\frac{1}{T-b}\sum_{s=1}^{b}\sum_{t=1}^{T-b}\varepsilon_{t+s}\right)' \\
&= A_T + B_T.
\end{aligned}
$$

Let us first look at B_T. Note that

$$
\begin{aligned}
\frac{1}{T-b}\sum_{s=1}^{b}\sum_{t=1}^{T-b}\varepsilon_{t+s} &= \frac{b}{T}\sum_{t=1}^{T}\varepsilon_t + \frac{b}{T(T-b)}\sum_{t=1}^{T}\varepsilon_t - \frac{1}{T-b}\sum_{s=1}^{b}\sum_{t=1}^{s-1}\varepsilon_t \\
&\quad - \frac{1}{T-b}\sum_{s=1}^{b}\sum_{t=T-b+s+1}^{T}\varepsilon_t \\
&= O_p(bT^{-1/2}) + O_p(bT^{-3/2}) + O_p(b^{3/2}T^{-1}) + O_p(b^{3/2}T^{-1}),
\end{aligned}
$$

from which we can conclude that $B_T = O_p(bT^{-1})$.

Next we look at the first term. We have

$$
\begin{aligned}
A_T &= \frac{1}{b(T-b)} \sum_{s=1}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s}\varepsilon_{t+s}' + \frac{1}{b(T-b)} \sum_{s_1=1}^{b} \sum_{s_2=1, s_1\neq s_2}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s_1}\varepsilon_{t+s_2}' \\
&= \frac{b}{b(T-b)} \sum_{t=1}^{T} \varepsilon_t\varepsilon_t' - \frac{1}{b(T-b)} \sum_{s=1}^{b} \left(\sum_{t=1}^{s-1} \varepsilon_t\varepsilon_t' + \sum_{t=T-b+s+1}^{T} \varepsilon_t\varepsilon_t' \right) \\
&\quad + \frac{1}{b(T-b)} \sum_{s_1=1}^{b} \sum_{s_2=1, s_1\neq s_2}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s_1}\varepsilon_{t+s_2}' \\
&= T^{-1} \sum_{t=1}^{T} \varepsilon_t\varepsilon_t' + \frac{b}{T(T-b)} \sum_{t=1}^{T} \varepsilon_t\varepsilon_t' - \frac{1}{b(T-b)} \sum_{s=1}^{b} \left(\sum_{t=1}^{s-1} \varepsilon_t\varepsilon_t' + \sum_{t=T-b+s+1}^{T} \varepsilon_t\varepsilon_t' \right) \\
&\quad + \frac{1}{b(T-b)} \sum_{s_1=1}^{b} \sum_{s_2=1, s_1\neq s_2}^{b} \sum_{t=1}^{T-b} \varepsilon_{t+s_1}\varepsilon_{t+s_2}' \\
&= T^{-1} \sum_{t=1}^{T} \varepsilon_t\varepsilon_t' + O_p(bT^{-1}) + O_p(bT^{-1}) + O_p(bT^{-1/2}) = \Sigma + o_p(1).
\end{aligned}
$$

This concludes the proof of part (ii). □

Lemma A.4. *Let Assumption 1 and 2 hold. Then, as $T \to \infty$,*

$$
W_T^*(r) = T^{-1/2} \sum_{k=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\varepsilon_{i_m+s} - \mathrm{E}^* \varepsilon_{i_m+s}) \xrightarrow{d^*} \Sigma^{1/2} W(r) \quad \textit{in probability.}
$$

Proof of Lemma A.4. First note that

$$
\begin{aligned}
T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\varepsilon_{i_m+s} - \mathrm{E}^* \varepsilon_{i_m+s}) &= k^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} b^{-1/2} \sum_{s=1}^{b} (\varepsilon_{i_m+s} - \mathrm{E}^* \varepsilon_{i_m+s}) \\
&= k^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} H_m^*.
\end{aligned}
$$

We check the conditions of Corollary 2.2 of Phillips and Durlauf (1986) for the H_m^* terms. Weak stationarity follows straightforwardly by the definition of the block bootstrap. The moment condition (a), that $\mathrm{E}^* |H_{i,m}|^\beta = O_p(1)$ for some $2 \leq \beta < \infty$ is fulfilled with $\beta = 2$ by Lemma A.3.

By construction, each H_m^* is independent, thus fulfilling the mixing condition (b). Then the result follows from Corollary 2.2. □

Lemma A.5. *Let y_t be generated under H_0 setting* (A). *Let Assumptions 1 and 2 hold.*

Then, as $T \to \infty$,

$$T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) \xrightarrow{d^*} \Gamma' \Psi(1) \Sigma^{1/2} W(r).$$

Proof of Lemma A.5. As

$$\begin{aligned} T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) &= T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\Gamma' x_{i_m+s} - \mathrm{E}^* \Gamma' x_{i_m+s}) \\ &= \Gamma' \left(T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (x_{i_m+s} - \mathrm{E}^* x_{i_m+s}) \right), \end{aligned} \tag{33}$$

we focus on $T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (x_{i_m+s} - \mathrm{E}^* x_{i_m+s})$.

Using the Beveridge-Nelson decomposition we can write

$$x_{i_m+s} = \Psi(L) \varepsilon_{i_m+s} = \Psi(1) \varepsilon_{i_m+s} - \tilde{\Psi}(L)(\varepsilon_{i_m+s} - \varepsilon_{i_m+s-1}),$$

where $\tilde{\Psi}(z) = \sum_{j=0}^{\infty} \tilde{\Psi}_j z^j$, $\tilde{\Psi}_j = \sum_{i=j+1}^{\infty} \Psi_j$. Then

$$\begin{aligned} T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (x_{i_m+s} - \mathrm{E}^* x_{i_m+s}) &= T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} \Psi(1)(\varepsilon_{i_m+s} - \mathrm{E}^* \varepsilon_{i_m+s}) \\ &\quad - T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \Big((\tilde{\Psi}(L) \varepsilon_{i_m+b} - \mathrm{E}^* \tilde{\Psi}(L) \varepsilon_{i_m+b}) \\ &\quad - (\tilde{\Psi}(L) \varepsilon_{i_m} - \mathrm{E}^* \tilde{\Psi}(L) \varepsilon_{i_m}) \Big). \end{aligned}$$

We will show that $T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (\tilde{\Psi}(L) \varepsilon_{i_m+b} - \mathrm{E}^* \tilde{\Psi}(L) \varepsilon_{i_m+b}) = o_p^*(1)$. First note that

$$\begin{aligned} &\mathrm{P}^* \left[\left| T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \Big(\tilde{\Psi}(L) \varepsilon_{i_m+s} - \mathrm{E}^* \tilde{\Psi}(L) \varepsilon_{i_m+s} \Big) \right| > \epsilon \right] \\ &\quad \leq \frac{1}{\epsilon^2} \mathrm{E}^* \left| T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \Big(\tilde{\Psi}(L) \varepsilon_{i_m+s} - \mathrm{E}^* \tilde{\Psi}(L) \varepsilon_{i_m+s} \Big) \right|^2 = \frac{1}{\epsilon^2} E^* \left| G_{T,s}^* \right|^2 \end{aligned}$$

for $s = 0, b$ by the Markov inequality. Then, letting $\xi_t^* = \varepsilon_t - \mathrm{E}^* \varepsilon_t$,

$$\begin{aligned}
\mathrm{E}^* \left|G_{T,s}^*\right|^2 &= T^{-1}\,\mathrm{E}^* \left(\sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right)' \left(\sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right) \\
&= T^{-1} \sum_{m_1=0}^{\lfloor (k-1)r \rfloor} \sum_{m_2=0}^{\lfloor (k-1)r \rfloor} \mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_{m_1}+s-j}^*\right)' \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_{m_2}+s-j}^*\right) \\
&= T^{-1} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right)' \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right) \\
&\quad + T^{-1} \sum_{m_1=0}^{\lfloor (k-1)r \rfloor} \sum_{m_2=0, m_1 \neq m_2}^{\lfloor (k-1)r \rfloor} \mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_{m_1}+s-j}^*\right)' \mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_{m_2}+s-j}^*\right) \\
&= T^{-1} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right)' \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \xi_{i_m+s-j}^*\right) \\
&= T^{-1} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \mathrm{E}^* \left|\sum_{j=0}^{\infty} \tilde{\Psi}_j \left(\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j}\right)\right|^2
\end{aligned}$$

as

$$\mathrm{E}^* \left(\sum_{j=0}^{\infty} \tilde{\Psi}_j \left(\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j}\right)\right) = 0.$$

Now, by Minkowski's inequality, we have uniformly in r,

$$\begin{aligned}
\mathrm{E}^* \left|G_{T,s}^*\right|^2 &\leq T^{-1} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \left[\sum_{j=0}^{\infty} \left|\tilde{\Psi}_j\right| \left(\mathrm{E}^* \left|\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j}\right|^2\right)^{1/2}\right]^2 \\
&\leq 4kT^{-1} \left[\sum_{j=0}^{\infty} \left|\tilde{\Psi}_j\right| \left(\frac{1}{T-b} \sum_{t=1}^{T-b} \left|\varepsilon_{t+s-j}\right|^2\right)^{1/2}\right]^2 \\
&\leq 4kT^{-1} \left(\sum_{j=0}^{\infty} \left|\tilde{\Psi}_j\right|\right)^2 \max_j \frac{1}{T-b} \sum_{t=1}^{T-b} \left|\varepsilon_{t+s-j}\right|^2 .
\end{aligned}$$

A sufficient condition for

$$\sum_{j=0}^{\infty} \left|\tilde{\Psi}_j\right| < \infty$$

is that

$$\sum_{j=0}^{\infty} j\left|\Psi_j\right| < \infty,$$

see Phillips and Solo (1992, Lemma 2.1). This holds by Assumption 1. We also have that

$$\frac{1}{T-b}\sum_{t=1}^{T-b}\left|\varepsilon_{t+s-j}\right| = O_p(1)$$

by the moment conditions in Assumption 1. Therefore $\mathrm{E}^*\left|G_{T,s}^*\right|^2 = O_p(b^{-1})$ for $s = 0, b$ from which it follows that, uniformly in r,

$$T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(x_{i_m+s} - \mathrm{E}^*\, x_{i_m+s}) = \Psi(1)W_T^*(r) + o_p^*(1) \tag{34}$$

and therefore

$$T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(x_{i_m+s} - \mathrm{E}^*\, x_{i_m+s}) \xrightarrow{d^*} \Psi(1)\Sigma^{1/2}W(r) \quad \text{in probability} \tag{35}$$

by Lemma A.4. The proof is concluded by referring to (33) and applying the continuous mapping theorem. □

Proof of Lemma 2. Our proof is similar to Paparoditis and Politis (2003, Proof of Theorem 3.1). Note that

$$S_T^*(r) = T^{-1/2}y_1 + T^{-1/2}\sum_{m=0}^{M_r-1}\sum_{s=1}^{b}\hat{u}_{i_m+s} + T^{-1/2}\sum_{s=1}^{N_r}\hat{u}_{i_{M_r}+s},$$

where $M_r = \lfloor(\lfloor Tr\rfloor - 2)/b$ and $N_r = \lfloor Tr\rfloor - M_r b - 1$. As $T^{-1/2}y_1 = O_p(T^{-1/2})$, we write

$$S_T^*(r) = T^{-1/2}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\hat{u}_{i_m+s} - T^{-1/2}\sum_{s=N_r+1}^{b}\hat{u}_{i_{M_r}+s} + O_p(T^{-1/2}).$$

Now, for the i-th component of $S_T^*(r)$, $S_{T,i}^*(r)$ $(i = 1, \ldots, N)$, we can write

$$\begin{aligned}
S_{T,i}^*(r) &= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(u_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) \\
&\quad - T^{-1/2}(\hat{\rho}_i - \rho_i) \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(y_{i,i_m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \\
&\quad - T^{-1/2} \sum_{s=N_r+1}^{b} \left(u_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) \\
&\quad + T^{-1/2}(\hat{\rho}_i - \rho_i) \sum_{s=N_r+1}^{b} \left(y_{i,i_m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) + O_p(T^{-1/2}) \\
&= A_T^* - B_T^* - C_T^* + D_T^*,
\end{aligned}$$

which follows from the fact that

$$\begin{aligned}
\hat{u}_{i,i_m+s} &= y_{i,i_m+s} - \hat{\rho}_i y_{i,i_m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \hat{\rho}_i y_{i,i_m+s-1}) \\
&= y_{i,i_m+s} - \rho_i y_{i,i_m+s-1} + \rho_i y_{i,i_m+s-1} - \hat{\rho}_i y_{i,i_m+s-1} \\
&\quad - \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \rho_i y_{i,i_m+s-1} + \rho_i y_{i,i_m+s-1} - \hat{\rho}_i y_{i,i_m+s-1}) \\
&= u_{i,i_m+s} - (\hat{\rho}_i - \rho_i) y_{i,i_m+s-1} - \frac{1}{T-1} \sum_{t=2}^{T} (u_{i,t} - (\hat{\rho}_i - \rho_i) y_{i,t-1}).
\end{aligned} \tag{36}$$

We first look at C_T^*. We have uniformly in r

$$\begin{aligned}
C_T^* &= T^{-1/2} \sum_{s=N_r+1}^{b} u_{i,i_m+s} - \frac{b - N_r}{T^{1/2}(T-1)} \sum_{t=2}^{T} u_{i,t} \\
&= O_p^*(b^{1/2} T^{-1/2}) + O_p^*(b T^{-1}) = O_p^*(k^{-1/2}),
\end{aligned}$$

by the stationarity of u_t.

Next we turn to D_T^*. We have that

$$D_T^* = T^{-1/2}(\hat{\rho}_i - \rho_i) \sum_{s=N_r+1}^{b} y_{i,i_m+s-1} - \frac{1}{T^{1/2}(T-1)} (\hat{\rho}_i - \rho_i)(b - N_r) \sum_{t=2}^{T} y_{i,t-1},$$

which is of order $O_p^*(bT^{-1}) = O_p^*(k^{-1})$ uniformly in r as $\hat{\rho}_i - \rho_i = O_p(T^{-1})$.

Next we turn to B_T^*. Consider

$$
\begin{aligned}
&\mathrm{E}^* \left[\sum_{s=1}^{b} \left(y_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \\
&\quad = \frac{1}{T-b} \sum_{t=1}^{T-b} \left[\sum_{s=1}^{b} \left(y_{i,t+s-1} - \frac{1}{T-1} \sum_{\tau=2}^{T} y_{i,\tau-1} \right) \right] \\
&\quad = \frac{1}{T-b} \sum_{s=1}^{b} \sum_{t=1}^{T-b} y_{i,t+s-1} - \frac{b}{T-1} \sum_{t=2}^{T} y_{i,t-1} \\
&\quad = \frac{1}{T-b} \sum_{s=1}^{b} \left[\sum_{t=2}^{T} y_{i,t-1} - \sum_{t=1}^{s-1} y_{i,t} - \sum_{t=T-b+s}^{T-1} y_{i,t} \right] - \frac{b}{T-1} \sum_{t=2}^{T} y_{i,t-1} \\
&\quad = \frac{b}{T-b} \sum_{t=2}^{T} y_{i,t-1} - \frac{b}{T-1} \sum_{t=2}^{T} y_{i,t-1} - \frac{1}{T-b} \sum_{s=1}^{b} \left[\sum_{t=1}^{s-1} y_{i,t} + \sum_{t=T-b+s}^{T-1} y_{i,t} \right] \\
&\quad = \frac{b(b-1)}{(T-b)(T-1)} \sum_{t=2}^{T} y_{i,t-1} - \frac{1}{T-b} \sum_{s=1}^{b} \left[\sum_{t=1}^{s-1} y_{i,t} + \sum_{t=T-b+s}^{T-1} y_{i,t} \right] \\
&\quad = O_p(b^2 T^{-1/2}) + O_p(b^2 T^{-1/2}).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathrm{E}^* \left[\sum_{s=1}^{b} \left(y_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right]^2 \\
&\quad = \frac{1}{T-b} \sum_{t=1}^{T-b} \left[\sum_{s=1}^{b} \left(y_{i,t+s-1} - \frac{1}{T-1} \sum_{\tau=2}^{T} y_{i,\tau-1} \right) \right]^2 \\
&\quad = \frac{1}{T-b} \sum_{t=1}^{T-b} \left[\sum_{s=1}^{b} y_{i,t+s-1} - \frac{b}{T-1} \sum_{\tau=2}^{T} y_{i,\tau-1} \right]^2 \\
&\quad = \frac{1}{T-b} \sum_{t=1}^{T-b} \left[\left(\sum_{s=1}^{b} y_{i,t+s-1} \right)^2 - \frac{2b}{T-1} \left(\sum_{s=1}^{b} y_{i,t+s-1} \right) \left(\sum_{\tau=2}^{T} y_{i,\tau-1} \right) \right. \\
&\qquad\qquad \left. + \frac{b^2}{(T-1)^2} \left(\sum_{\tau=2}^{T} y_{i,\tau-1} \right)^2 \right] \\
&\quad = O_p(b^2 T).
\end{aligned}
$$

Therefore we have, uniformly in r,

$$
\begin{aligned}
&\mathrm{E}^* \left[T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(y_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right]^2 \\
&\quad = T^{-1}\, \mathrm{E}^* \left\{ \sum_{m_1=0}^{M_r} \sum_{m_2=0}^{M_r} \left[\sum_{s=1}^{b} \left(y_{i,i_{m_1}+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \right. \\
&\qquad\qquad \left. \times \left[\sum_{s=1}^{b} \left(y_{i,i_{m_2}+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \right\} \\
&\quad = T^{-1} \sum_{m=0}^{M_r} \mathrm{E}^* \left[\sum_{s=1}^{b} \left(y_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right]^2 \\
&\qquad + T^{-1} \sum_{m_1=0}^{M_r} \sum_{m_2=0, m_1 \neq m_2}^{M_r} \mathrm{E}^* \left[\sum_{s=1}^{b} \left(y_{i,i_{m_1}+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \\
&\qquad\qquad \times \mathrm{E}^* \left[\sum_{s=1}^{b} \left(y_{i,i_{m_2}+s} - \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1} \right) \right] \\
&\quad = T^{-1} O(k) O_p(b^2 T) + T^{-1} O(k^2) O_p(b^4 T^{-1}) = O_p(bT) + O_p(b^2),
\end{aligned}
$$

where we use that the blocks are independent. From this and the fact that $\hat{\rho}_i - \rho_i = O_p(T^{-1})$ as we are under the null hypothesis, it follows that

$$
B_T^* = O_p^*(b^{1/2} T^{-1/2}) = O_p^*(k^{-1/2}).
$$

Finally we look at A_T^*. We have that

$$
\begin{aligned}
A_T^* &= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(u_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) \\
&= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(u_{i,i_m+s} - \mathrm{E}^*\, u_{i,i_m+s} + \mathrm{E}^*\, u_{i,i_m+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right) \\
&= T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} (u_{i,i_m+s} - \mathrm{E}^*\, u_{i,i_m+s}) \\
&\quad + T^{-1/2} \sum_{m=0}^{M_r} \sum_{s=1}^{b} \left(\frac{1}{T-b} \sum_{t=1}^{T-b} u_{i,t+s} - \frac{1}{T-1} \sum_{t=2}^{T} u_{i,t} \right).
\end{aligned}
$$

Note that, uniformly in r,

$$
\begin{aligned}
&T^{-1/2}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(\frac{1}{T-b}\sum_{t=1}^{T-b}u_{i,t+s}-\frac{1}{T-1}\sum_{t=2}^{T}u_{i,t}\right)\\
&\quad=T^{-1/2}\sum_{m=0}^{M_r}\left[\frac{1}{T-b}\sum_{s=1}^{b}\left(\sum_{t=2}^{T}u_{i,t}-\sum_{t=2}^{s}u_{i,t}-\sum_{t=T-b+s+1}^{T}u_{i,t}\right)-\frac{b}{T-1}\sum_{t=2}^{T}u_{i,t}\right]\\
&\quad=T^{-1/2}M_r\left[\frac{b(b-1)}{(T-b)(T-1)}\sum_{t=2}^{T}u_{i,t}-\frac{1}{T-b}\sum_{s=1}^{b}\left(\sum_{t=2}^{s}u_{i,t}+\sum_{t=T-b+s+1}^{T}u_{i,t}\right)\right]\\
&\quad=T^{-1/2}O(k)O_p^*(b^2T^{-3/2})+T^{-1/2}O(k)O_p^*(b^{3/2}T^{-1})=O_p^*(k^{-1/2}).
\end{aligned}
$$

Combining all the previous results, and realizing they hold for all $i=1,\dots,N$, we have that, uniformly in r,

$$
S_T^*(r)=T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(u_{i,i_m+s}-\mathrm{E}^*\,u_{i,i_m+s})+o_p^*(1), \tag{37}
$$

where we also take the sum up to $\lfloor (k-1)r\rfloor$ instead of M_r which is the same asymptotically. The proof is then concluded by applying Lemma A.5. □

The next step is to determine the moments of the bootstrap series corresponding to the moments in Lemma A.1.

Lemma A.6. *Let y_t be generated under H_0 setting* (A). *Let Assumptions 1 and 2 hold. Then, as $T\to\infty$,*

(i) $\Omega^*=T^{-1}\left[\mathrm{E}^*\left(\sum_{t=1}^T u_t^*\right)\left(\sum_{t=1}^T u_t^*\right)'-\mathrm{E}^*\left(\sum_{t=1}^T u_t^*\right)\mathrm{E}^*\left(\sum_{t=1}^T u_t^*\right)'\right]=\Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma+o_p(1)$,

(ii) $\Omega_0^*=T^{-1}\sum_{t=1}^T\left[\mathrm{E}^*(u_t^*u_t^{*\prime})-\mathrm{E}^*\,u_t^*\,\mathrm{E}^*\,u_t^{*\prime}\right]=\sum_{i=0}^\infty\Gamma'\Psi_i\Sigma\Psi_i'\Gamma+o_p(1)$.

Proof of Lemma A.6. We start with part (i). Using the arguments in the proof of Lemma 2 (take $r=1$) we can show that

$$
T^{-1/2}\sum_{t=1}^{T}u_t^*=T^{-1/2}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s})+o_p^*(1). \tag{38}
$$

Therefore

$$
\mathrm{E}^*\left[T^{-1/2}\sum_{t=1}^{T}u_t^*\right]=o_p(1)
$$

and

$$\Omega^* = T^{-1}\,\mathrm{E}^*\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s})\right)\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s})\right)' + o_p(1).$$

Using the Beveridge-Nelson decomposition, we can show, as in the proof of Lemma A.5, that

$$T^{-1/2}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s}) = T^{-1/2}\Gamma'\Psi(1)\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^*\,\varepsilon_{i_m+s}) + o_p^*(1).$$

Consequently

$$\begin{aligned}
\Omega^* &= T^{-1}\Gamma'\Psi(1)\,\mathrm{E}^*\left[\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^*\,\varepsilon_{i_m+s})\right)\right.\\
&\quad\times\left.\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^*\,\varepsilon_{i_m+s})\right)'\right]\Psi(1)'\Gamma + o_p(1)\\
&= k^{-1}\Gamma'\Psi(1)\sum_{m=0}^{k-1}\mathrm{E}^*\left[\left(b^{-1/2}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^*\,\varepsilon_{i_m+s})\right)\right.\\
&\quad\times\left.\left(b^{-1/2}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^*\,\varepsilon_{i_m+s})\right)'\right]\Psi(1)'\Gamma + o_p(1)\\
&= \Gamma'\Psi(1)\Sigma\Psi(1)'\Gamma + o_p(1),
\end{aligned}$$

where we use the independence of the blocks and the last line follows from Lemma A.3. This concludes the proof of part (i).

The proof of part (ii) is similar to part (i). By (36) and the arguments used in the proof of Lemma 2, we can straightforwardly show that

$$\Omega_0 = T^{-1}\sum_{m=0}^{k-1}\sum_{s=1}^{b}\mathrm{E}^*\left(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s}\right)\left(u_{i_m+s}-\mathrm{E}^*\,u_{i_m+s}\right)' + o_p(1).$$

Then, using that $u_{i_m+s} = \sum_{j=0}^{\infty} \Gamma' \Psi_j \varepsilon_{i_m+s-j}$, we can write

$$\begin{aligned}
\Omega_0^* &= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(\sum_{j=0}^{\infty} \Gamma' \Psi_j (\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j}) \right) \\
&\quad \times \left(\sum_{j=0}^{\infty} \Gamma' \Psi_j (\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j}) \right)' + o_p(1) \\
&= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i \, \mathrm{E}^* \left[(\varepsilon_{i_m+s-i} - \mathrm{E}^* \varepsilon_{i_m+s-i}) \right. \\
&\quad \left. \times \, (\varepsilon_{i_m+s-j} - \mathrm{E}^* \varepsilon_{i_m+s-j})' \right] \Psi_j' \Gamma + o_p(1) \\
&= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i \frac{1}{T-b} \sum_{t=1}^{T-b} \left(\varepsilon_{t+s-i} \varepsilon_{t+s-j}' \right) \Psi_j' \Gamma \\
&\quad - T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s-i} \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s-j}' \Psi_j' \Gamma + o_p(1) \\
&= A_T + B_T + o_p(1).
\end{aligned}$$

Note that

$$|B_T| \leq b \max_{1 \leq s \leq b} \left| \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+s} \right|^2 |\Gamma|^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\Psi_i| |\Psi_j| = b O_p(T^{-1}) O(1) = O_p(k^{-1}).$$

Then

$$\begin{aligned}
A_T &= \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \Gamma' \Psi_i \left(\varepsilon_{t+s-i} \varepsilon_{t+s-j}' \right) \Psi_j' \Gamma \\
&= \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \Gamma' \Psi_i \left(\varepsilon_{t+s-i} \varepsilon_{t+s-i}' \right) \Psi_i' \Gamma \\
&\quad + \frac{1}{(T-b)b} \sum_{t=1}^{T-b} \sum_{s=1}^{b} \sum_{i=0}^{\infty} \sum_{j=0, j \neq i}^{\infty} \Gamma' \Psi_i \left(\varepsilon_{t+s-i} \varepsilon_{t+s-j}' \right) \Psi_j' \Gamma \\
&= A_{1,T} + A_{2,T}
\end{aligned}$$

and furthermore

$$|A_{2,T}| \leq \max_{1 \leq r,s \leq b, r \neq s} \left| \frac{1}{T-b} \sum_{t=1}^{T-b} \varepsilon_{t+r} \varepsilon_{t+s}' \right| |\Gamma|^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\Psi_i| |\Psi_j| = O_p(T^{-1/2}).$$

For $A_{1,T}$ note that it is easy to see that (see for example the proof of Lemma A.3)

$$\frac{1}{(T-b)b}\sum_{t=1}^{T-b}\sum_{s=1}^{b}\varepsilon_{t+s-i}\varepsilon'_{t+s-i} = \Sigma + o_p(1),$$

by which we can conclude that $\Omega_0^* = \sum_{i=0}^{\infty}\Gamma'\Psi_i\Sigma\Psi_i'\Gamma + o_p(1)$. □

Lemma A.7. *Let y_t be generated under H_0 setting* (A). *Let Assumptions 1 and 2 hold. Then, as $T\to\infty$, we have for $i=1,\dots,N$,*

(i) $T^{-1}\sum_{t=1}^{T} y^*_{i,t-1}\Delta y^*_{i,t} \xrightarrow{d} \int_0^1 B_i(r)dB_i(r) + \frac{1}{2}(\omega_i - \omega_{0,i})$,

(ii) $T^{-2}\sum_{t=1}^{T} y^{*2}_{i,t-1} \xrightarrow{d} \int_0^1 B_i(r)^2 dr$.

Proof of Lemma A.7. The result follows straightforwardly from Lemma 2, Lemma A.6 and the continuous mapping theorem. □

Proof of Theorem 2. The result follows directly from Lemma A.7. □

Proof of Lemma 3. As in Lemma 1, we have that by Assumption 1 $W_T(r) = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\varepsilon_t \xrightarrow{d} \Sigma^{1/2}W(r)$. Then it follows that

$$T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} f_t = \Psi_2(1)W_T(r) + o_p(1),$$

uniformly in r, and consequently $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} f_t \xrightarrow{d} \Psi_2(1)\Sigma^{1/2}W(r)$.

Now

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor} u_t &= T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Lambda f_t + T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Delta v_t \\ &= T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Lambda f_t + v_{\lfloor Tr\rfloor} - v_0 = T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Lambda f_t + O_p(T^{-1/2}) \end{aligned}$$

uniformly in r and $T^{-1/2}\sum_{t=1}^{\lfloor Tr\rfloor}\Lambda f_t \xrightarrow{d} \Lambda\Psi_2(1)\Sigma^{1/2}W(r)$. □

The next lemma is the counterpart of Lemma A.1.

Lemma A.8. *Let y_t be generated under H_0 setting* (B). *Let Assumption 1 hold. Then*

(i) $\bar{\Omega} = \lim_{T\to\infty} T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} u_t\right)\left(\sum_{t=1}^{T} u_t\right)' = \Lambda\Psi_2(1)\Sigma\Psi_2(1)'\Lambda'$,

(ii) $\bar{\Omega}_0 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{E}(u_t u_t') = \sum_{j=0}^{\infty}(\Lambda\Psi_{2,j}\Sigma\Psi'_{2,j}\Lambda' + (\bar{\Psi}_{1,j} - \bar{\Psi}_{1,j+1})\Sigma\Psi'_{2,j}\Lambda' + \Lambda\Psi_{2,j}\Sigma(\bar{\Psi}_{1,j} - \bar{\Psi}_{1,j+1})' + 2\bar{\Psi}_{1,j}\Sigma\bar{\Psi}'_{1,j} - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}'_{1,j+1} - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}'_{1,j})$.

Proof of Lemma A.8. For part (i), note that

$$\begin{aligned}
\bar{\Omega} &= \lim_{T\to\infty} T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T}(f_t+\Delta v_t)\right)\left(\sum_{t=1}^{T}(f_t+\Delta v_t)\right)' \\
&= \lim_{T\to\infty} T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} f_t + v_T - v_0\right)\left(\sum_{t=1}^{T} f_t + v_T - v_0\right)' \\
&= \lim_{T\to\infty}\left\{T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\, f_s f_t' + T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} f_t(v_T-v_0)'\right)\right. \\
&\quad \left. + T^{-1}\,\mathrm{E}\left((v_T-v_0)\sum_{t=1}^{T} f_t'\right) + T^{-1}\,\mathrm{E}(v_T-v_0)'(v_T-v_0)\right\} \\
&= \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathrm{E}\, f_s f_t' = \Lambda\Psi_2(1)\Sigma\Psi_2(1)'\Lambda',
\end{aligned}$$

where the last step follows straightforwardly as in the proof of Lemma A.1 part (i).

For part (ii) we have

$$\begin{aligned}
\bar{\Omega}_0 &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathrm{E}\left((f_t+\Delta v_t)(f_t+\Delta v_t)'\right) \\
&= \lim_{T\to\infty}\left\{T^{-1}\sum_{t=1}^{T}\mathrm{E}\, f_t f_t' + T^{-1}\sum_{t=1}^{T}\mathrm{E}\,\Delta v_t f_t' + T^{-1}\sum_{t=1}^{T}\mathrm{E}\, f_t\Delta v_t' + T^{-1}\sum_{t=1}^{T}\mathrm{E}\,\Delta v_t\Delta v_t'\right\} \\
&= A_T + B_T + B_T' + C_T.
\end{aligned}$$

Now

$$A_T = \sum_{j=0}^{\infty}\Lambda\Psi_{2,j}\Sigma\Psi_{2,j}'\Lambda',$$

analogous to the proof of Lemma A.1 part (ii). Similarly

$$B_T = \sum_{j=0}^{\infty}(\bar{\Psi}_{1,j}-\bar{\Psi}_{1,j+1})\Sigma\Psi_{2,j}'\Lambda'$$

and

$$C_T = \sum_{j=0}^{\infty}\left(2\bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j}' - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j+1}' - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j}'\right).$$

This completes the proof. □

Proof of Theorem 3. Using Lemma 3, Lemma A.8 and the continuous mapping theorem we can construct the counterpart of Lemma A.2. The result then follows. □

Lemma A.9. *Let y_t be generated under H_0 setting* (A). *Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) \xrightarrow{d^*} \Lambda \Psi_2(1) \Sigma^{1/2} W(r).$$

Proof of Lemma A.9. Note that

$$\begin{aligned}
T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) &= T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (\Lambda f_{i_m+s} + \Delta w_{i_m+s}) \\
&\quad - \mathrm{E}^*(\Lambda f_{i_m+s} + \Delta w_{i_m+s}) \\
&= \Lambda \left(T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (f_{i_m+s} - \mathrm{E}^* f_{i_m+s}) \right) \\
&\quad + T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (w_{i_m+b} - \mathrm{E}^* w_{i_m+b}) \\
&\quad - T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (w_{i_m} - \mathrm{E}^* w_{i_m}) \\
&= A_T^* + B_{T,0}^* + B_{T,b}^*.
\end{aligned} \tag{39}$$

We want to show that $B_{T,s}^* = O_p^*(b^{-1/2})$ uniformly in r for $s = 0, b$. First note that by equation (34)

$$B_{T,s}^* = T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (\bar{\Psi}_1(L)\varepsilon_{i_m+s} - \mathrm{E}^* \bar{\Psi}_1(L)\varepsilon_{i_m+s}) + o_p^*(1).$$

As

$$k^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} (\bar{\Psi}_1(L)\varepsilon_{i_m+s} - \mathrm{E}^* \bar{\Psi}_1(L)\varepsilon_{i_m+s}) = O_p^*(1),$$

it follows that $B_{T,s}^* = O_p^*(b^{-1/2})$ uniformly in r for $s = 0, b$.

Now we can show in exactly the same way as in the proof of Lemma A.5 that

$$T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (f_{i_m+s} - \mathrm{E}^* f_{i_m+s}) = \Psi_2(1) W_T^*(r) + o_p^*(1)$$

uniformly in r and consequently that

$$T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^* u_{i_m+s}) \xrightarrow{d^*} \Lambda\Psi_2(1)W(r) \quad \text{in probability.} \qquad \square$$

Proof of Lemma 4. As the order of the f_t determines the order of u_t, we can show in exactly the same way as in the proof of Lemma 2 that, uniformly in r,

$$S_T^*(r) = T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(u_{i,i_m+s}-\mathrm{E}^* u_{i,i_m+s}) + o_p^*(1). \tag{40}$$

The proof is then concluded by applying Lemma A.9. $\square$

We consider the bootstrap moments in the following lemma.

Lemma A.10. *Let y_t be generated under H_0 setting* (B). *Let Assumptions 1 and 2 hold. Then, as $T\to\infty$,*

(i) $\bar{\Omega}^* = T^{-1}\,\mathrm{E}\left(\sum_{t=1}^{T} u_t^*\right)\left(\sum_{t=1}^{T} u_t^*\right)' = \Lambda\Psi_2(1)\Sigma\Psi_2(1)'\Lambda' + o_p(1)$,

(ii) $\bar{\Omega}_0^* = T^{-1}\sum_{t=1}^{T}\mathrm{E}(u_t^* u_t^{*\prime}) = \sum_{j=0}^{\infty}(\Lambda\Psi_{2,j}\Sigma\Psi_{2,j}'\Lambda' + (\bar{\Psi}_{1,j}-\bar{\Psi}_{1,j+1})\Sigma\Psi_{2,j}'\Lambda' + \Lambda\Psi_{2,j}\Sigma(\bar{\Psi}_{1,j}-\bar{\Psi}_{1,j+1})' + 2\bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j}' - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j+1}' - \bar{\Psi}_{1,j}\Sigma\bar{\Psi}_{1,j}') + o_p(1)$.

Proof of Lemma A.10. We start with part (i). Using the arguments from the proofs of Lemma 2 and Lemma 4 (take $r=1$) we can again show that

$$T^{-1/2}\sum_{t=1}^{T} u_t^* = T^{-1/2}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^* u_{i_m+s}) + o_p^*(1). \tag{41}$$

from which it follows that

$$\bar{\Omega}^* = T^{-1}\,\mathrm{E}^*\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^* u_{i_m+s})\right)\left(\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^* u_{i_m+s})\right)' + o_p(1).$$

Combining the proof of Lemma A.5 and A.9 we can show that

$$T^{-1/2}\sum_{m=0}^{k-1}\sum_{s=1}^{b}(u_{i_m+s}-\mathrm{E}^* u_{i_m+s}) = T^{-1/2}\Lambda\Psi_2(1)\sum_{m=0}^{k-1}\sum_{s=1}^{b}(\varepsilon_{i_m+s}-\mathrm{E}^* \varepsilon_{i_m+s}) + o_p^*(1).$$

Consequently

$$\bar{\Omega}^* = \Lambda\Psi_2(1)\Sigma\Psi_2(1)'\Lambda' + o_p(1),$$

which follows in exactly the same way as in the proof of Lemma A.6. This concludes the proof of part (i).

Next we consider part (ii). As in the proof of Lemma A.6 we can show that

$$\bar{\Omega}_0^* = T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(u_{i_m+s} - \mathrm{E}^* u_{i_m+s}\right) \left(u_{i_m+s} - \mathrm{E}^* u_{i_m+s}\right)' + o_p(1).$$

Then we can write

$$\begin{aligned}
\bar{\Omega}_0^* &= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left((f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}) + (\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}) \right) \\
&\quad \times \left((f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}) + (\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}) \right)' + o_p(1) \\
&= T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}\right) \left(f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}\right)' \\
&\quad + T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}\right) \left(\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}\right)' \\
&\quad + T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}\right) \left(f_{i_m+s-j} - \mathrm{E}^* f_{i_m+s-j}\right)' \\
&\quad + T^{-1} \sum_{m=0}^{k-1} \sum_{s=1}^{b} \mathrm{E}^* \left(\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}\right) \left(\Delta v_{i_m+s-j} - \mathrm{E}^* \Delta v_{i_m+s-j}\right)' \\
&= A_T + B_T + B_T' + C_T.
\end{aligned}$$

Then we can easily show, in the same way as in the proof of Lemma A.6 part (ii), that

$$A_T^* = \sum_{j=0}^{\infty} \Lambda \Psi_{2,j} \Sigma \Psi_{2,j}' \Lambda' + o_p(1),$$

as well as

$$B_T = \sum_{j=0}^{\infty} (\bar{\Psi}_{1,j} - \bar{\Psi}_{1,j+1}) \Sigma \Psi_{2,j}' \Lambda' + o_p(1)$$

and

$$C_T = \sum_{j=0}^{\infty} \left(2\bar{\Psi}_{1,j} \Sigma \bar{\Psi}_{1,j}' - \bar{\Psi}_{1,j} \Sigma \bar{\Psi}_{1,j+1}' - \bar{\Psi}_{1,j} \Sigma \bar{\Psi}_{1,j}'\right) + o_p(1).$$

This completes the proof. □

Proof of Theorem 4. As for Theorem 3 we can construct the counterpart of Lemma A.7 using Lemma 4, Lemma A.10 and the continuous mapping theorem. The result then follows. □

Proof of Lemma 5. We can write

$$T^{-1}\tau_p = \frac{\sum_{i=1}^N T^{-1}\sum_{t=2}^T y_{i,t-1}\Delta y_{i,t}}{\sum_{i=1}^N T^{-1}\sum_{t=2}^T y_{i,t-1}^2} = \frac{\sum_{i=1}^N a_{iT}}{\sum_{i=1}^N b_{iT}}.$$

Now as $y_{i,t}$ is a stationary process for all $i = 1,\dots,N$, we have that

$$a_{iT} = T^{-1}\sum_{t=2}^T y_{i,t-1}y_{i,t} - T^{-1}\sum_{t=2}^T y_{i,t-1}^2 \xrightarrow{p} \gamma_i(1) - \gamma_i(0)$$

and

$$b_{iT} = T^{-1}\sum_{t=2}^T y_{i,t-1}^2 \xrightarrow{p} \gamma_i(0).$$

Similarly,

$$T^{-1}\tau_{gm} = N^{-1}\sum_{i=1}^N \frac{a_{iT}}{b_{iT}},$$

from which the result follows. □

Lemma A.11. *Let y_t be generated under H_1^a. Let Assumptions 1 and 2 hold. Then, as $T \to \infty$,*

$$T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) \xrightarrow{d^*} \Psi^{++}(1)\Sigma^{1/2}W(r).$$

Proof of Lemma A.11. Using the Beveridge-Nelson decomposition we can write

$$u_{i_m+s} = \Psi^{++}(L)\varepsilon_{i_m+s} = \Psi(1)^{++}\varepsilon_{i_m+s} - \tilde{\Psi}^{++}(L)(\varepsilon_{i_m+s} - \varepsilon_{i_m+s-1}),$$

where $\tilde{\Psi}^{++}(z) = \sum_{j=0}^{\infty}\tilde{\Psi}_j^{++}z^j$, $\tilde{\Psi}_j^{++} = \sum_{i=j+1}^{\infty}\Psi_j^{++}$. Then

$$\begin{aligned} T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}(u_{i_m+s} - \mathrm{E}^* u_{i_m+s}) &= T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\sum_{s=1}^{b}\Psi^{++}(1)(\varepsilon_{i_m+s} - \mathrm{E}^*\varepsilon_{i_m+s}) \\ &\quad - T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}\Big((\tilde{\Psi}^{++}(L)\varepsilon_{i_m+b} - \mathrm{E}^*\tilde{\Psi}^{++}(L)\varepsilon_{i_m+b}) \\ &\quad - (\tilde{\Psi}^{++}(L)\varepsilon_{i_m} - \mathrm{E}^*\tilde{\Psi}(L)\varepsilon_{i_m})\Big). \end{aligned}$$

We need to show that $T^{-1/2}\sum_{m=0}^{\lfloor (k-1)r\rfloor}(\tilde{\Psi}^{++}(L)\varepsilon_{i_m+b} - \mathrm{E}^*\tilde{\Psi}(L)^{++}\varepsilon_{i_m+b}) = o_p^*(1)$, uniformly in r. Completely analogous to the proof of Lemma A.5 this means showing that

$$\sum_{j=0}^{\infty}\left|\tilde{\Psi}_j^{++}\right|<\infty$$

or equivalently

$$\sum_{j=0}^{\infty}j\left|\Psi_j^{++}\right|<\infty.$$

This holds as we remarked that the summability condition continues to hold. □

Proof of Theorem 5. We start by showing that the invariance principle holds. As in the proof of Lemma 2 we have that

$$\begin{aligned}S_{T,i}^{*}(r)=&T^{-1/2}\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(u_{i,i_m+s}-\frac{1}{T-1}\sum_{t=2}^{T}u_{i,t}\right)\\&-T^{-1/2}(\hat{\rho}_i-\rho_i)\sum_{m=0}^{M_r}\sum_{s=1}^{b}\left(y_{i,i_m+s-1}-\frac{1}{T-1}\sum_{t=2}^{T}y_{i,t-1}\right)\\&-T^{-1/2}\sum_{s=N_r+1}^{b}\left(u_{i,i_m+s}-\frac{1}{T-1}\sum_{t=2}^{T}u_{i,t}\right)\\&+T^{-1/2}(\hat{\rho}_i-\rho_i)\sum_{s=N_r+1}^{b}\left(y_{i,i_m+s-1}-\frac{1}{T-1}\sum_{t=2}^{T}y_{i,t-1}\right)+O_p^{*}(T^{-1/2})\\=&A_T^{*}-B_T^{*}-C_T^{*}+D_T^{*}.\end{aligned}$$

As before, we have that, uniformly in r,

$$\begin{aligned}C_T^{*}&=T^{-1/2}\sum_{s=N_r+1}^{b}u_{i,i_m+s}-\frac{1}{T^{1/2}(T-1)}(b-N_r)\sum_{t=2}^{T}u_{i,t}\\&=O_p^{*}(b^{1/2}T^{-1/2})+O_p^{*}(bT^{-1})=O_p^{*}(k^{-1/2}),\end{aligned}$$

by the stationarity of u_t.

Turning to D_T^{*} we have that

$$D_T^{*}=T^{-1/2}(\hat{\rho}_i-\rho_i)\sum_{s=N_r+1}^{b}y_{i,i_m+s-1}-\frac{1}{T^{1/2}(T-1)}(\hat{\rho}_i-\rho_i)(b-N_r)\sum_{t=2}^{T}y_{i,t-1},$$

which is of order $o_p^{*}(b^{1/2}T^{-1/2})$, uniformly in r, by the stationarity of $y_{i,t}$.

Next we turn to B_T^*. We can write

$$B_T^* = T^{-1/2}(\hat{\rho}_i - \rho_i) \sum_{m=0}^{M_r} \sum_{s=1}^{b} y_{i,i_m+s} - \frac{1}{T^{1/2}(T-1)}(\hat{\rho}_i - \rho_i) M_r b \sum_{t=2}^{T} y_{i,t-1},$$

from which we can conclude that, uniformly in r, $B_T^* = o_p^*(1)$ as $\sum_{m=0}^{M_r} \sum_{s=1}^{b} y_{i,i_m+s} = O_p^*(T^{1/2})$ by the stationarity of $y_{i,t}$.

The results for A_T^* remain the same as in the proof of Lemma 2 from which we can conclude that, uniformly in r,

$$S_T^*(r) = T^{-1/2} \sum_{m=0}^{\lfloor (k-1)r \rfloor} \sum_{s=1}^{b} (u_{i,i_m+s} - \mathrm{E}^* u_{i,i_m+s}) + o_p^*(1). \tag{42}$$

The result now follows trivially by applying Lemma A.11. □

Proof of Lemma 6. We write

$$\begin{aligned}
\tau_p &= \frac{\sum_{i=1}^{N} T^{-1} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{i,t}}{\sum_{i=1}^{N} T^{-2} \sum_{t=2}^{T} y_{i,t-1}^2} \\
&= \frac{\sum_{i=1}^{n_1} T^{-1} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{i,t} + \sum_{i=n_1+1}^{N} T^{-1} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{i,t}}{T^{-1} \sum_{i=1}^{n_1} T^{-1} \sum_{t=2}^{T} y_{i,t-1}^2 + \sum_{i=n_1+1}^{N} T^{-2} \sum_{t=2}^{T} y_{i,t-1}^2} \\
&= \frac{\sum_{i=1}^{n_1} a_{iT} + \sum_{i=n_1+1}^{N} c_{iT}}{\sum_{i=1}^{n_1} T^{-1} b_{iT} + \sum_{i=n_1+1}^{N} d_{iT}}.
\end{aligned}$$

The convergence of a_{iT} and b_{iT} follow from the proof of Lemma 5. Furthermore, as in Lemma A.2, we have that

$$\begin{aligned}
c_{iT} &\xrightarrow{d} \int_0^1 B_i(r) dB_i(r), \\
d_{iT} &\xrightarrow{d} \int_0^1 B_i(r)^2 dr,
\end{aligned}$$

from which the result for τ_p follows.

For τ_{gm} we can write

$$T^{-1}\tau_{gm} = N^{-1} \sum_{i=1}^{n_1} \frac{a_{iT}}{b_{iT}} + N^{-1} \sum_{i=n_1+1}^{N} T^{-1} \frac{c_{iT}}{d_{iT}} = N^{-1} \sum_{i=1}^{n_1} \frac{a_{iT}}{b_{iT}} + O_p(T^{-1}).$$ □

Proof of Corollary 1. The proof is immediate by combining the proofs of Theorems 2 and 5. □