

# Bootstrap sequential tests to determine the stationary units in a panel

## Citation for published version (APA):

Smeeke, S. (2011). Bootstrap sequential tests to determine the stationary units in a panel. (METEOR Research Memorandum; No. 003). Maastricht: METEOR, Maastricht University School of Business and Economics.

## Document status and date:

Published: 01/01/2011

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Stephan Smeekes

**Bootstrap Sequential Tests to  
Determine the Stationary Units in  
a Panel**

RM/11/003

**METEOR**

Maastricht University School of Business and Economics  
Maastricht Research School of Economics  
of Technology and Organization

P.O. Box 616  
NL - 6200 MD Maastricht  
The Netherlands

# BOOTSTRAP SEQUENTIAL TESTS TO DETERMINE THE STATIONARY UNITS IN A PANEL

Stephan Smeekes\*

Department of Quantitative Economics  
Maastricht University  
December 6, 2010

## Abstract

We propose an approach to investigate the stationarity properties of individual units in a panel based on testing user-defined increasing proportions of hypothesized stationary units sequentially. Asymptotically valid critical values are obtained using the block bootstrap. This sequential approach has an advantage over multiple testing approaches, in particular if  $N$  is large and  $T$  is small, as it can exploit the cross-sectional dimension, which the multiple testing approaches cannot do effectively. A simulation study is conducted to analyze the relative performance of the approach in comparison with multiple testing approaches. The method is also illustrated by two empirical applications, in testing for unit roots in real exchange rates and log earnings data of households. The simulation study and applications demonstrate the usefulness of our method, in particular in panels with large  $N$  and small  $T$ .

*Keywords:* Sequential testing; unit roots; panel data; block bootstrap.

*JEL Classification:* C15, C23.

## 1 Introduction

Over the last decade a large number of unit root tests have been designed that can be applied in panel data. Most of these tests have as a null hypothesis that all units in the panel have a unit root. The alternative hypothesis differs; some tests have the alternative hypothesis that at least some series are stationary.<sup>1</sup> A rejection for such a test is hard to interpret; it could

---

\*Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: [S.Smeekes@maastrichtuniversity.nl](mailto:S.Smeekes@maastrichtuniversity.nl). This research was supported financially by the Netherlands Organisation for Scientific Research (NWO). I would like to thank Christoph Hanck for sharing his data. I also thank Eric Beutner, Stefano Fachin, Franz Palm and Jean-Pierre Urbain as well as participants at the Panel Data Conference 2010 in Amsterdam for helpful comments and suggestions.

<sup>1</sup>Throughout this paper, “stationary” should be read to mean integrated of order zero ( $I(0)$ ). In the same way, “nonstationary” implies  $I(1)$ .

be that just a few units are stationary, or that all units are stationary. These two opposites will usually have very different consequences for the economic interpretation of the tests, yet there is no way to distinguish. Other tests have as alternative hypothesis that all units are stationary. While such an alternative hypothesis might seem to help interpreting a rejection, this is often not so; many of these tests also have power if not all units are stationary (cf. Westerlund and Breitung, 2009, Myth 1); and hence a rejection is not convincing evidence that all series are indeed stationary.

For this reason, it is important to consider methods that can provide more information than just a rejection or no rejection for the whole panel. Methods that can give an estimate of the proportion of stationary units, or, even better, methods that can test which units are stationary, are therefore very valuable.

Recently several papers have investigated such methods. Ng (2008) proposes an estimator of the proportion of stationary units based on the value of the trend in cross-sectional averages of variances. This method delivers an estimate of the fraction of (non)stationary units, but cannot test which units are stationary. Hanck (2009) and Moon and Perron (2010) apply methods from the literature on multiple testing to determine which units are stationary. Hanck (2009) employs the bootstrap approach of Romano and Wolf (2005) to control the family-wise error rate (*FWE*) in testing for which countries PPP holds, an approach that is mainly suited for relatively small cross-sectional dimension  $N$ . Moon and Perron (2010) on the other hand aim to control the false discovery rate (*FDR*), an approach that is better suited to panels with larger  $N$ . Moon and Perron (2010) consider both asymptotic methods and the bootstrap method of Romano, Shaikh, and Wolf (2008a) to control the *FDR*, and find that the bootstrap method performs best in general.

In this paper, we propose an approach to determine the stationary units based on sequential testing, thereby avoiding the difficulties of controlling size in multiple testing. Our approach is similar in spirit to that of Kapetanios (2003), who was the first to consider sequential testing for the number of stationary units. Our method proceeds by testing user-specified fractions, or quantiles, of the units sequentially. The quantiles can be selected in such a way that the method tests individual units sequentially, in which case the method is mainly suited to panels with small  $N$ . We will show that this version of the method is closely related to the approach of Romano and Wolf (2005) to control for *FWE*; as a side-product of this analysis we propose a modification of their method that can be more powerful. By selecting “wider” quantiles the method can be made suitable for panels with large  $N$ . It is demonstrated that this method has several advantages over other large  $N$  methods; this holds in particular in large  $N$ , small  $T$  models, as unlike the existing methods, our approach is able to exploit the cross-sectional dimension to increase power.

The structure of the paper is as follows. The DGP is introduced in Section 2. In Section 3 the sequential tests are constructed and their asymptotic properties are derived. Section

4 presents two simulation studies comparing the properties for small and large  $N$  models to those of comparable multiple testing approaches. In Section 5 we present two applications. Section 6 concludes. Proofs are given in the Appendix.

A word on notation.  $\lfloor x \rfloor$  is the largest integer smaller than or equal to  $x$ . We denote  $x$  rounded to the nearest integer by  $\lceil x \rceil$ . 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. Convergence in distribution (probability) of bootstrap statistics is denoted  $\xrightarrow{d^*}$  ( $\xrightarrow{p^*}$ ), where this convergence is taken to take place in probability.  $|x|$  applied to a complex number  $x$  denotes its absolute value, while  $|\mathcal{G}|$  applied to a set  $\mathcal{G}$  denotes the cardinality of the set.  $\mathcal{G}^c$  denotes the compliment of the set  $\mathcal{G}$  taken with respect to the set  $\mathcal{N}_N = \{i \in \mathbb{N} : i \leq N\}$ , i.e.  $\mathcal{G}^c \cup \mathcal{G} = \mathcal{N}_N$  while  $\mathcal{G}^c \cap \mathcal{G} = \emptyset$ .

## 2 The model

Suppose we have a panel of observations  $y_{i,t}$ ,  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . Let  $y_t = (y_{1,t}, \dots, y_{N,t})'$  be generated as

$$y_t = \beta z_t + \Lambda F_t + w_t, \quad (1)$$

where  $\beta = (\beta_1, \dots, \beta_N)'$ ,  $\Lambda = (\lambda_1, \dots, \lambda_N)'$ ,  $F_t = (F_{1,t}, \dots, F_{d,t})'$  and  $w_t = (w_{1,t}, \dots, w_{N,t})'$ .  $F_t$  are common factors ( $d$  in total),  $\Lambda$  are the (non-random) factor loadings,  $w_t$  are idiosyncratic components and  $z_t$  are deterministic components, where  $z_t = 1$  or  $z_t = (1, t)'$ .

We let the factors and the idiosyncratic components be generated by

$$\begin{aligned} F_t &= \Phi F_{t-1} + f_t, & F_0 &= 0, \\ w_t &= \Gamma w_{t-1} + v_t, & w_0 &= 0, \end{aligned} \quad (2)$$

where  $\Phi = \text{diag}(\phi_1, \dots, \phi_d)$  and  $\Gamma = \text{diag}(\gamma_1, \dots, \gamma_N)$ .

Furthermore we let  $f_t$  and  $v_t$  be constructed as

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

where  $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$  ( $\Psi_0 = I$ ). We impose the following conditions on  $\Psi(z)$  and  $\varepsilon_t$ .

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

This is the same DGP as used by Palm, Smeekes, and Urbain (2010). Now define  $\rho_i$  as

$$\rho_i = \lim_{t \rightarrow \infty} \frac{\mathbb{E}(y_{i,t-1}y_{i,t})}{\mathbb{E}(y_{i,t-1}^2)}.$$

If  $\rho_i = 1$ , unit  $i$  has a unit root, if  $|\rho_i| < 1$ , unit  $i$  is stationary. Unit  $i$  has a unit root if there is a unit root in one of its two components  $\lambda_i'F_t$  or  $w_{i,t}$ . That is,  $\rho_i = 1$  if  $\gamma_i = 1$  or if  $\phi_j = 1$  and  $\lambda_{i,j} \neq 0$  for some  $j = 1, \dots, d$ , where  $\lambda_{i,j}$  is the  $j$ -th element of  $\lambda_i$ .

Let  $k_0 = 0, 1, \dots, N$  be the number of stationary units. Formally, define  $\mathcal{S} = \{i \in \mathcal{N}_N : |\rho_i| < 1\}$  as the set of stationary units, such that  $k_0 = |\mathcal{S}|$ , and  $\mathcal{U} = \{i \in \mathcal{N}_N : \rho_i = 1\}$  as the set of nonstationary units. Furthermore, let  $q_0 = k_0/N$  denote the proportion of stationary units.

**Remark 1.** Note that our DGP is very general as we allow for a wide range of temporal and cross-sectional dependencies. It is therefore not entirely appropriate to call  $w_t$  “idiosyncratic components”. For an extensive discussion of the DGP and the appropriateness of the terminology, see Remark 1 and 2 of Palm et al. (2010). We will not pay any more attention to this here, as our focus here is whether or not  $y_{i,t}$  has a unit root, and we are not interested in the cause of the unit root. Moreover, the whole sequential testing setup that we propose is independent of the structure of the DGP, and so this is a minor issue in this paper.

### 3 Sequential testing for unit roots

#### 3.1 Bootstrap Sequential Quantile Test

Let  $0 = q_1 < \dots < q_r < 1$  denote a set of  $r$  user-defined numbers, representing the proportions of stationary units to be tested sequentially. Let  $k_j = [q_j N]$  be the number of stationary units corresponding to  $q_j$ . We will go into detail later on how to select the numbers  $q_j$ ; we first focus on the method. Let  $H_0(q_j)$  denote the null hypothesis that a proportion of  $q_j$  units is stationary, or equivalently that  $k_j$  out of  $N$  units are stationary. In other words  $H_0(q_j) : |\mathcal{S}| = k_j = [q_j N]$ . Let  $H_1(q_{j+1})$  denote the alternative hypothesis that at least a proportion of  $q_{j+1}$  units, (or, equivalently,  $k_{j+1}$  out of  $N$  units) is stationary, i.e.  $|\mathcal{S}| \geq k_{j+1}$ . Let  $\tau(q_j, q_{j+1})$  be a test statistic to test  $H_0(q_j)$  vs.  $H_1(q_{j+1})$ , which rejects  $H_0(q_j)$  if  $\tau(q_j, q_{j+1}) < c_\alpha(q_j, q_{j+1})$ , where  $c_\alpha(q_j, q_{j+1})$  is an appropriate level  $\alpha$  critical value. The sequential testing procedure can now be described as below.

**Algorithm 1** (Sequential Quantile Test).

1. Test  $H_0(q_1)$  against  $H_1(q_2)$  using test statistic  $\tau(q_1, q_2)$ . Reject  $H_0(q_1)$  if  $\tau(q_1, q_2) < c_\alpha(q_1, q_2)$ .
2. If  $H_0(q_1)$  is not rejected, set  $\hat{q} = q_1$ . If  $H_0(q_1)$  is rejected, test  $H_0(q_2)$  against  $H_1(q_3)$ .

3. Keep testing until the null hypothesis  $H_0(q_j)$  cannot be rejected. In that case, set  $\hat{q} = q_j$ . If all null hypotheses up to  $H_0(q_r)$  are rejected, set  $\hat{q} = 1$ .

We now focus on how to construct the test statistic  $\tau(q_j, q_{j+1})$ . To do so we first need individual unit root tests statistics on each unit. Let  $\theta_i$  be any unit root test statistic applied to unit  $i$  that satisfies the following assumption.

**Assumption 2.** Let  $\theta_i$  be a statistic applied to unit  $i = 1, \dots, N$  that rejects the null hypothesis of a unit root for small values of  $\theta_i$ . Furthermore

- (i) Under the null hypothesis  $\theta_i$  is asymptotically distributed as  $\xi_i$ , i.e. if  $\rho_i = 1$ , then

$$\theta_i \xrightarrow{d} \xi_i, \quad \text{as } T \rightarrow \infty.$$

- (ii)  $\theta_i$  is consistent; in particular, if  $\rho_i < 1$ , then

$$\theta_i \xrightarrow{p} -\infty, \quad \text{as } T \rightarrow \infty.$$

Obviously many unit root tests statistics satisfy these conditions,<sup>2</sup> including the popular Dickey-Fuller tests (in this case  $\xi_i$  is the Dickey-Fuller distribution, or one of its detrended variants).<sup>3</sup> Note that in general  $\xi_i$  and  $\xi_j$  will be correlated if there is cross-sectional dependence in the data.

Now let  $\theta_{(1)}, \dots, \theta_{(N)}$  denote the order statistics of  $\theta_1, \dots, \theta_N$ , defined such that

$$\theta_{(1)} \leq \dots \leq \theta_{(N)}.$$

Then we take the test statistic as the order statistic corresponding to the alternative hypothesis to be tested, that is

$$\tau(q_j, q_{j+1}) = \theta_{(k_{j+1})} = \theta_{([q_{j+1}N])}. \quad (4)$$

While this choice of test statistic is a natural choice, it is not used often as asymptotic theory for the order statistics is notoriously difficult, in particular as  $\theta_1, \dots, \theta_N$  will not be

---

<sup>2</sup>Note that the assumption that the tests reject to the left is made for expositional simplicity (and because it is true for many unit root tests), it is in no way crucial for the construction of the methods as all arguments just turn around for tests that reject to the right.

<sup>3</sup>As is usual, we will use  $\xi_i$  interchangeably to denote the asymptotic distribution of  $\theta_i$  and the random variable with this distribution. For instance for the Dickey-Fuller distribution (for the coefficient test) we have

$$\xi_i = \frac{\int_0^1 B_i(r) dB_i(r)}{\int_0^1 B_i(r)^2 dr},$$

where  $B_i$  is a standard Brownian motion, which, just as is usual, can be interpreted both as the distribution itself or the random variable with that distribution.

independent due to the cross-sectional dependence in the panel. For this reason we propose to use the bootstrap to obtain critical values.

We now present the bootstrap implementation of our sequential test, which we call the Bootstrap Sequential Quantile Test (*BSQT*). Note that we can not only use this method to estimate  $q_0$ , but also to determine the members of  $\mathcal{S}$ . Let us first introduce the notation  $z_{(k:\mathcal{G})}$  as the  $k$ -th smallest element of the set  $\{z_i : i \in \mathcal{G}\}$ , or formally

$$z_{(k:\mathcal{G})} = \{z_i : i \in \mathcal{G} \text{ and } |\{j \in \mathcal{G} : z_j < z_i\}| = k - 1\}.$$

**Algorithm 2** (*BSQT*). To test  $H_0(q_j)$  vs.  $H_1(q_{j+1})$  in Algorithm 1, let  $\tau(q_j, q_{j+1}) = \theta_{([q_{j+1}N])} = \theta_{(k_{j+1})}$  and let

$$\mathbb{S}_{k_j} = \{i : \theta_i \leq \theta_{(k_j)}\}. \quad (5)$$

1. For each unit obtain the detrended series

$$\tilde{y}_{i,t} = y_{i,t} - \hat{\beta}_i' z_t,$$

where  $\hat{\beta}_i$  is an estimator of  $\beta_i$  obtained for example through OLS or GLS.

2. For each unit estimate

$$\hat{\rho}_i = \frac{\sum_{t=1}^T \tilde{y}_{i,t-1} \tilde{y}_{i,t}}{\sum_{t=1}^T \tilde{y}_{i,t-1}^2},$$

and calculate

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

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

3. Choose a block length  $b$ . Draw  $i_0, \dots, i_{k-1}$  i.i.d. from the uniform distribution on  $\{1, 2, \dots, T - b\}$ , where  $k = \lfloor (T - 2)/b \rfloor + 1$  is the number of blocks.
4. Construct the bootstrap errors  $u_2^*, \dots, u_T^*$  as follows.

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

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

5. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*,$$



where

$$\rho_i^* = 1 \quad \text{for all } i = 1, \dots, N.$$

6. Obtain  $\theta_i^*$  for all  $i \in \mathbb{S}_{k_j}^c$ , and obtain the bootstrap test statistic as

$$\tau^*(q_j, q_{j+1}) = \theta_{(k_{j+1}-k_j; \mathbb{S}_{k_j}^c)}^*.$$

7. Repeat Steps 3 to 6  $B$  times, obtaining bootstrap test statistics  $\tau^{*m}(q_j, q_{j+1})$ , for  $m = 1, \dots, B$ , and select the bootstrap critical value  $c_\alpha^*(q_j, q_{j+1})$  as  $c_\alpha^*(q_j, q_{j+1}) = \max\{c : B^{-1} \sum_{m=1}^B I(\tau^{*m}(q_j, q_{j+1}) < c) \leq \alpha\}$ , or equivalently as the  $\alpha$ -quantile of the ordered  $\tau^{*m}(q_j, q_{j+1})$  statistics.

It now follows directly that the set of units deemed stationary is simply equal to  $\mathbb{S}_{\hat{k}} = \mathbb{S}_{[\hat{q}N]}$  as defined in (5).

Before we develop the asymptotic properties of the procedure, let us briefly digress on the choice of the quantiles  $q_1, \dots, q_r$ . We argue that the choice of the quantiles should be determined by both economic and statistical arguments.

First of all note that, unless the quantiles are selected such that each unit is tested sequentially, the method will leave “holes” in between the units. For example, suppose that the quantiles to be tested are taken as  $q_j = (j - 1)/4$  (that is 0, 0.25, 0.5 and 0.75). Then the method will obviously not be able to detect if 35% or 40% of the units are stationary. Moreover, it is from the outset not clear how the method will behave if the true proportion of stationary units is in between selected quantiles.<sup>4</sup> In that sense, if the outcome of the method is that  $\hat{q} = q_j$ , this should not be interpreted to say that  $q_j$  is the true stationary proportion, but rather that the true stationary proportion should be found in the interval  $(q_{j-1}, q_{j+1})$  (also see Section 3.3). Hence, for a very precise result the quantiles should be taken close to one another.

On the other hand, the “holes” that are created are also the method’s strength. By skipping units, all information in these units is cross-sectionally bundled together in one test, which will increase power. Basically, in each step the method uses the cross-sectional information to increase power in the same way that a “regular” panel unit root test does. As such, for power purposes, it is best to take the quantiles as far away as possible.

Therefore we suggest to take the quantiles such that they are as far spread out as possible, while still being able to draw economically relevant conclusions from the outcomes. It is pointless to give a general rule of thumb, as for each specific application a different amount of precision is needed. The optimal quantiles will also depend on the sample sizes; the larger  $N$

---

<sup>4</sup>The intuition that the method is more likely to reject if the true proportion is closer to the alternative hypothesis will be confirmed in the asymptotic analysis.

is, the more units are in between two quantiles, and so they could be taken closer while still maintaining power. Power is also heavily influenced by  $T$ , and larger  $T$  also means quantiles can be taken closer to each other. Also, note that the quantiles do not have to be equally spaced; one could for instance imagine that the outcome of the question whether 0% or 10% of the units are stationary has a very different impact on the interpretation of the results than the outcome of the question whether 10% or 20% of the units are stationary.

Finally, we have already alluded to setting the quantiles such that each unit is tested sequentially. Specifically, this amounts to setting  $q_j = (j - 1)/N$ ,  $j = 1, \dots, N$ . This deserves special attention, as it has a fundamentally different interpretation than when setting “fixed” quantiles. First, note that all problems mentioned above related to the holes do not apply here. Second, while the fixed quantile method has a nice interpretation for large and even increasing  $N$ , this is not so for the unit based method (the quantiles collapse). As such, this method is really designed for panels with small or moderate  $N$ . Because of the special significance of this option, we refer to this method by a different name: *BSUT*, for Bootstrap Sequential Unit Test.

**Remark 2.** Strictly speaking the computation of the test statistic does not depend on  $q_j$ , and so we would not have to let it depend on it. However the distribution of the test statistic under the null obviously does depend on it, and consequently the bootstrap variant depends on it as well.

**Remark 3.** While the statistics  $\theta_i$  could be any unit root test statistics, we will from now mainly focus on the case where  $\theta_i$  is chosen as a Dickey-Fuller (DF) type statistic. The tests could contain an initial demeaning or detrending procedure that could either be standard OLS or GLS as in Elliott, Rothenberg, and Stock (1996). As DF statistic one could either take the coefficient test or the t-test, and the statistic can be augmented with lags (ADF) or not (also see Remark 5 below). For the applications in Section 5 we will also consider combinations of DF tests based on Harvey, Leybourne, and Taylor (2010) and Smeekes and Taylor (2010).

### 3.2 Asymptotic properties

We now proceed with the asymptotic analysis of the method. The asymptotic theory requires the following assumption on the block length (also see Remark 8).

**Assumption 3.** Let  $b \rightarrow \infty$  and  $b = o(T^{1/2})$  as  $T \rightarrow \infty$ .

The asymptotic distributions of the test statistics and bootstrap test statistics in a single step of the sequential procedure are given in Theorem 1. These distributions are not of much interest by themselves, but they are necessary to make statements about the asymptotic properties of the whole sequential method.

**Theorem 1.**

(i) Let  $y_{i,t}$  be generated by (1)-(3) and let Assumption 1 hold. Let  $\theta_1, \dots, \theta_N$  satisfy Assumption 2 and let  $\tau(q_j, q_{j+1})$  be defined as in (4). Then, as  $T \rightarrow \infty$ , we have that

$$\tau(q_j, q_{j+1}) \xrightarrow{p} -\infty \quad \text{if } q_{j+1} \leq q_0, \quad (6a)$$

$$\tau(q_j, q_{j+1}) \xrightarrow{d} \xi_{(k_{j+1}-k_0:\mathcal{W})} \quad \text{if } q_{j+1} > q_0. \quad (6b)$$

(ii) In addition to the conditions used in (i), let  $\theta_i^*$  and  $\tau^*(q_j, q_{j+1})$  be defined as in Algorithm 2 and let Assumption 3 hold. Then, as  $T \rightarrow \infty$ , for any  $q_0$ ,

$$\theta_i^* \xrightarrow{d^*} \xi_i \text{ in probability} \quad \text{for all } i \in \mathbb{S}_{k_j}^c \text{ (jointly)}, \quad (7)$$

and

$$\tau^*(q_j, q_{j+1}) \xrightarrow{d^*} \xi_{(k_{j+1}-k_j:\mathbb{S}_{k_j}^c)} \text{ in probability.} \quad (8)$$

This theorem shows the asymptotic validity of the bootstrap order statistics. Note that the bootstrap also correctly reproduces the cross-sectional dependence between the individual unit root test statistics, as shown in detail by Palm et al. (2010).

As mentioned above, we can now consider the properties of the sequential approach as a whole. The implications of Theorem 1 for the sequential approach are given in Corollary 1.

**Corollary 1.** *Under the assumptions of Theorem 1, we have that*

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) = 0 \quad \text{if } q_{j+1} \leq q_0, \quad (9a)$$

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) = 1 - a_{j,k_0}^\alpha \quad \text{if } q_j < q_0 < q_{j+1}, \quad (9b)$$

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) = 1 - \alpha \quad \text{if } q_j = q_0, \quad (9c)$$

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) \leq a_{j-1,k_0}^\alpha \quad \text{if } q_{j-1} < q_0 < q_j, \quad (9d)$$

$$\limsup_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) \leq \alpha \quad \text{if } q_{j-1} \geq q_0, \quad (9e)$$

where

$$a_{j,m}^\alpha = \mathbb{P} \left\{ \xi_{(k_{j+1}-m:\mathcal{W}_m)} < c_{\alpha,\infty}(q_j, q_{j+1}) \right\}, \quad (10)$$

with  $\mathcal{W}_m$  the set of  $N - m$   $I(1)$  units, and  $c_{\alpha,\infty}(q_j, q_{j+1})$  is such that

$$\mathbb{P} \left\{ \xi_{(k_{j+1}-k_j:\mathbb{S}_{k_j}^c)} < c_{\alpha,\infty}(q_j, q_{j+1}) \right\} = \alpha. \quad (11)$$

Furthermore, for any  $k_j \leq m < k_{j+1}$ ,

$$\alpha \leq a_{j,m}^\alpha \leq a_{j,m+1}^\alpha \leq 1. \quad (12)$$

Several conclusions follow from the results above. First, note that if we were to restrict ourselves to situations where we are able to pick the true proportion exactly, the method works perfectly asymptotically; that is, if  $q_j = q_0$ , we choose a too low quantile with probability 0 and a too high quantile with at most probability  $\alpha$ ; compare for instance Swensen (2006, Corollary 1) for a similar result in the context of sequential testing for cointegration rank. This confirms the asymptotic validity of the bootstrap critical values. Also note that, at least asymptotically, there is a simple way to avoid selecting quantiles that do not equal the true proportion, and that is by setting  $q_j = (j - 1)/N$  and using the *BSUT* method. However, as has been discussed before, this method will not always work well in finite samples, and is unsuited for applications when  $N$  is large in particular.

Corollary 1 also gives an insight into what happens asymptotically if the true proportion is in between the selected quantiles. Combining equations (9b) and (9d) with (12) shows that the closer  $q_j$  is to  $q_0$ , the higher the asymptotic probability is that  $q_j$  is selected by the sequential procedure. This is an intuitively logical and comforting result, which provides validation for the quantile-based approach. Obviously this does not eliminate the issues arising in this situation, but at least the method behaves in an appropriate way in these cases, such that one can reliably apply it when the quantiles are selected appropriately.<sup>5</sup>

**Remark 4.** Note that results (9b) and (12) are not only useful in our sequential context, but can also be used to describe the asymptotic power of “conventional” bootstrap panel unit root tests based on order statistics, such as the median, minimum and maximum.<sup>6</sup> In short, if the true number of stationary units is larger than the hypothesized number of stationary units under the null hypothesis, but smaller than the hypothesized number under the alternative hypothesis, asymptotic power is increasing with the true number of stationary units.

**Remark 5.** Although any unit root test can be used for  $\theta_i$ , in practice it is usually the best option to use tests for which their marginal distribution is “as nuisance-parameter free as possible”. If the marginal distributions of the individual tests depend on nuisance parameters, they may not live on the same scale. As a consequence any ranking of them becomes unreliable.<sup>7</sup> It is therefore important to make all the individual tests live on the same scale, which is the case if their marginal distributions are the same; also see the discussion in Romano and Wolf (2005, p. 1255).

---

<sup>5</sup>In Section 3.3 we discuss an extension which aims at eliminating this issue.

<sup>6</sup>See for example Di Iorio and Fachin (2008) for an application of such statistics to cointegration testing.

<sup>7</sup>This occurs for example if the variances differ. Suppose that the variance of  $\theta_i$  is greater than that of  $\theta_j$ . If in that case  $\theta_i$  is smaller than  $\theta_j$ ,  $\theta_i$  does not necessarily indicate more evidence against the unit root null than  $\theta_j$ , but it does get ranked before it.

**Remark 6.** Kapetanios (2003) used the IPS group-mean statistic of Im, Pesaran, and Shin (2003) as test statistic, deleting the “most stationary” unit every round. The major problem with this approach is that the IPS test, being an average of the individual DF tests, is strongly influenced by the nonstationary units. As such it may lack power compared to the test based on order statistics, in particular if only a few units are stationary. Moreover, the IPS test does not allow for cross-sectional dependence, although it can be augmented to deal with cross-sectional dependence as in the CIPS test of Pesaran (2007).

**Remark 7.** The properties of the *BSQT* method are only analyzed for fixed  $N$ , as currently all relevant bootstrap theory (in particular by Palm et al., 2010) only provides results for fixed  $N$ .<sup>8</sup> The extension to infinite  $N$  is very difficult with a general DGP as ours. However, neither in Palm et al. (2010) nor in our case is there any restriction on  $N$ ;  $N$  may be very large, just not increasing to infinity. The same caveat holds for the multiple testing approaches that we discuss later. Note though that the formulation of the *BSQT* approach with the quantiles directly lends itself to be extended to a setting with  $N$  increasing to infinity.

**Remark 8.** Block length selection in practice remains an open issue, in particular in the context of nonstationary time series. It is discussed in Palm et al. (2010, Section 5.3), where it found that the Warp-speed calibration method of Giacomini, Politis, and White (2007) performs reasonably well. However, that method depends on the test statistic employed, and that makes it more difficult to apply in a sequential context where a sequence of test statistics is used. An alternative method could be to set it equal to choice of bandwidth for long-variance estimation in Andrews (1991), as done by Gonçalves (2010) and Moon and Perron (2010). However, this method faces similar issues in a sequential context and its optimality properties in a nonstationary setting are unknown. In practice it seems advisable for empirical work to run the method at different fixed block lengths and investigate the sensitivity to these differences.

### 3.3 Extensions

We now consider two possible extensions of the proposed method. The first aims at increasing power, while the second deals with the case where the true proportion of stationary units lies between the selected quantiles.

#### 3.3.1 *BSQT*<sub>2</sub>

To motivate the first extension, note that the properties of the sequential approach and the limit distributions are for a large part based on the consistency of the individual unit root test

---

<sup>8</sup>Gonçalves (2010) considers the block bootstrap for panel linear regression models and proves asymptotic validity for increasing  $N$ . These results however are only valid for stationary data, and can therefore not be directly extended to the unit root model.

statistics  $\theta_i$ ; that is, if unit  $i$  is stationary,  $\theta_i \xrightarrow{p} -\infty$  and as such it will have no influence on the distribution of the order statistics corresponding to the nonstationary units. While this is true asymptotically, this may not be so in finite samples. While the unit root test statistics should take smaller values for stationary units, their finite sample distributions may overlap the distributions of the statistics corresponding to the nonstationary units and thus influence the finite sample distributions of the order statistics.

The *BSQT* method ignores this effect by deleting the units found stationary from the sample used in the next round. Therefore we now propose a modification, that we label *BSQT*<sub>2</sub>,<sup>9</sup> that tries to mimic the effect of the stationary units on the distribution of the order statistics by including the units as being stationary in the bootstrap. Unlike *BSQT* it does not delete the stationary units but instead includes these units as stationary in the bootstrap. The advantage of this second method is that the information in the stationary units is not discarded. While asymptotically negligible, it may increase the finite sample power of the *BSQT*<sub>2</sub> method.

**Algorithm 3** (*BSQT*<sub>2</sub>). To test  $H_0(q_j)$  vs.  $H_1(q_{j+1})$  in Sequential Test 1, let  $\tau(q_j, q_{j+1}) = \theta_{([q_{j+1}N])} = \theta_{(k_{j+1})}$  and let  $\mathbb{S}_{k_j}$  be as defined in (5). Perform bootstrap steps 1 to 4 of Algorithm 2 to obtain bootstrap errors  $u_i^*$ .

4. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*,$$

where

$$\rho_i^* = \begin{cases} \hat{\rho}_i & \text{if } i \in \mathbb{S}_{k_j} \\ 1 & \text{if } i \in \mathbb{S}_{k_j}^c \end{cases}.$$

5. Obtain  $\theta_i^*$  for all units, and let

$$\tau_2^*(q_j, q_{j+1}) = \theta_{(k_{j+1})}^*.$$

6. Repeat Steps 3 to 6  $B$  times, obtaining bootstrap test statistics  $\tau_2^{*m}(q_j, q_{j+1})$ , for  $m = 1, \dots, B$ , and select the bootstrap critical value  $c_\alpha^*(q_j, q_{j+1})$  as  $c_\alpha^*(q_j, q_{j+1}) = \max\{c : B^{-1} \sum_{m=1}^B I(\tau_2^{*m}(q_j, q_{j+1}) < c) \leq \alpha\}$ , or equivalently as the  $\alpha$ -quantile of the ordered  $\tau_2^{*m}(q_j, q_{j+1})$  statistics.

**Remark 9.** The asymptotic properties of *BSQT*<sub>2</sub> are very similar to those of *BSQT*. Given the results that are available regarding the validity of autoregressive bootstrapping in stationary time series (cf. Bose, 1988), combined with results for the block bootstrap (Künsch, 1989),

---

<sup>9</sup>When we apply the modification to the *BSUT* approach, we label it *BSUT*<sub>2</sub>.

we have that  $\hat{\rho}_i^* - \rho_i = o_p^*(1)$ , from which it then easily follows that, for  $i \in \mathcal{S}$ ,  $\theta_i^* \xrightarrow{P^*} -\infty$  as  $T \rightarrow \infty$ . Hence, whenever  $|\mathcal{S}| \geq |\mathbb{S}_{k_j}|$ , the properties of both methods are identical. The only difference between the two methods occurs in the setting of (9e), as there  $|\mathcal{S}| < |\mathbb{S}_{k_j}|$ , which means that some units that are in fact I(1) are generated in the bootstrap algorithm with  $\hat{\rho}_i^*$ . It is known that in that case the unit root distribution will not be replicated correctly (Basawa, Mallik, McCormick, Reeves, and Taylor, 1991) but as we only reach this situation at the end of the algorithm with a small probability, it is not a problem for the validity of the approach.

### 3.3.2 Iterative *BSQT*

As has been discussed extensively before, the major issue with the *BSQT* approach is that the true proportion of stationary units may lie in between the selected quantiles. While we argue that this problem can be overcome by careful consideration of the quantiles to be selected, it may still be desirable in some applications to have a more precise estimate of  $q_0$ .<sup>10</sup>

To this end we propose a modification of the *BSQT* method (which can be applied to *BSQT*<sub>2</sub> as well) that can provide more precise estimates. If  $\hat{q}$  is the estimate found from the sequential procedure, define  $\hat{q}_{-1}$  and  $\hat{q}_{+1}$  as one selected quantile smaller and larger than the estimate, respectively. In other words, if  $\hat{q} = q_m$ , then  $\hat{q}_{-1} = q_{m-1}$  and  $\hat{q}_{+1} = q_{m+1}$ . Now note that

$$\mathrm{P} \{q_0 \in [\hat{q}_{-1}, \hat{q}_{+1}]\} = 1 - \mathrm{P} \{q_0 < \hat{q}_{-1}\} - \mathrm{P} \{q_0 > \hat{q}_{+1}\}.$$

By (9a) and (9e) it follows that

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathrm{P} \{q_0 < \hat{q}_{-1}\} &= \lim_{T \rightarrow \infty} \mathrm{P} \{\hat{q} = q_j | q_{j-1} > q_0\} \leq \alpha, \\ \lim_{T \rightarrow \infty} \mathrm{P} \{q_0 > \hat{q}_{+1}\} &= \lim_{T \rightarrow \infty} \mathrm{P} \{\hat{q} = q_j | q_{j+1} < q_0\} = 0, \end{aligned}$$

from which we can conclude that

$$\lim_{T \rightarrow \infty} \mathrm{P} \{q_0 \in [\hat{q}_{-1}, \hat{q}_{+1}]\} \geq 1 - \alpha.$$

This confirms that we should interpret finding a certain  $\hat{q}$  as evidence that  $q_0$  is between  $\hat{q}_{-1}$  and  $\hat{q}_{+1}$ . The *BSQT* method stops here, but we may instead apply the sequential method again only on the interval  $[\hat{q}_{-1}, \hat{q}_{+1}]$ . Of course, after the second step the interval can further be shortened and the method applied again and so on until convergence to one unit takes place. Below we describe this iterative application of the *BSQT* method, denoted by *IBSQT*, more formally.

---

<sup>10</sup>As discussed before the problem can be avoided by using the *BSUT* method, but this is not always appropriate.

**Algorithm 4** (*IBSQT*).

1. Select a set of quantiles to be tested  $q_1^{(1)} < \dots < q_{r_1}^{(1)}$ , where  $q_1^{(1)} = 0$  and  $q_{r_1}^{(1)} < 1$  and apply *BSQT* or *BSQT*<sub>2</sub> to obtain an estimate  $\hat{q}^{(1)}$ .
2. Select a new set of quantiles to be tested  $q_1^{(2)} < \dots < q_{r_2}^{(1)}$ , where  $q_1^{(2)} = \hat{q}_{-1}^{(1)}$  and  $q_{r_2}^{(2)} < \hat{q}_{+1}^{(1)}$  and obtain a second round estimate  $\hat{q}^{(2)}$ .
3. Select a third round set of quantiles restricted to the set  $[\hat{q}_{-1}^{(2)}, \hat{q}_{+1}^{(2)}]$  to obtain an estimate  $\hat{q}^{(3)}$ . Continue in this fashion until convergence to one unit takes place in the  $I$ -th iteration and obtain a final estimate  $\hat{q}^{(I)}$ .

If at any stage of the algorithm  $\hat{q}^{(s)}$  is equal to the boundaries  $\hat{q}_{-1}^{(s-1)}$  or  $\hat{q}_{+1}^{(s-1)}$ , only consider the interval  $[\hat{q}^{(s)}, \hat{q}_{+1}^{(s)}]$  or  $[\hat{q}_{-1}^{(s)}, \hat{q}^{(s)}]$ , respectively, in the next stage.

In this way the *IBSQT* method can yield a precise estimate of the number of stationary units, while still being able to exploit the cross-sectional dimension. However, there is an asymptotic price to pay for these benefits. As the method is applied iteratively, with every iteration the probability of making an error is accumulated. If  $I$  iterations are needed, then it can be shown that

$$\lim_{T \rightarrow \infty} \text{P} \left\{ \hat{q}^{(I)} = q_0 \right\} \geq (1 - \alpha)^I.$$

It should be added though that usually the number of iterations will be fairly small (also see Remark 10), and that the probability of an error is generally smaller than  $\alpha$  in most iterations.<sup>11</sup>

**Remark 10.** In general the number of quantiles in the *BSQT*\* procedures should be relatively small to properly utilize the cross-sectional dimension. In every iteration the cross-sectional dimension grows smaller, and thus it becomes progressively harder to exploit it effectively. Choosing a relatively fine grid for the quantiles in the first stage will leave only few units within the second stage interval and will affect the power in the second stage negatively, in particular if a fine grid is chosen again. The same holds for every further iteration. Of course, even if only few quantiles are selected in each stage, finite sample power will drop fairly quickly with every iteration as the cross-sectional dimension is effectively reduced, often terminating the method quickly as from some point on it will keep ending up on the left boundary of the interval.

### 3.4 Sequential approaches in a multiple testing framework

It is interesting to investigate the similarities between our approach and approaches based on size control in multiple testing. An overview of multiple testing techniques is given by Ro-

---

<sup>11</sup>If the selected quantile at any stage is on the left boundary the probability is even equal to 0.



mano, Shaikh, and Wolf (2008b). In a panel context these methods have been used by Hanck (2009), Deckers and Hanck (2009) and Moon and Perron (2010) among others. Hanck (2009) tests for which countries PPP holds. To control for size in this multiple testing framework he employs the bootstrap method by Romano and Wolf (2005), which controls the family-wise error rate (*FWE*), which is defined as the probability of at least one false rejection.

Our *BSUT* approach, even though originating from a sequential perspective, is very similar to the method of Romano and Wolf (2005). Their method, which we call *RW*, can be described as follows using our terminology.

1. Test  $H_0(0)$  against  $H_1(1/N)$ . That is, obtain  $\tau(0, 1/N)$ . Reject  $H_0(0)$  if  $\tau(0, 1/N) < c_\alpha(0, 1/N)$ .
2. If  $H_0(0)$  is not rejected, set  $\hat{q} = 0$ . If  $H_0(0)$  is rejected, test  $H_0(1/N)$  against  $H_1(2/N)$ , still using the critical value  $c_\alpha(0, 1/N)$ . That is, reject  $H_0(1/N)$  if  $\tau(1/N, 2/N) < c_\alpha(0, 1/N)$ .
3. Keep testing using critical value  $c_\alpha(0, 1/N)$  until the null hypothesis  $H_0(j/N)$  cannot be rejected. If  $0 < j/N < 1$ , continue to the second step: Test  $H_0(j/N)$  against  $H_1((j+1)/N)$  by rejecting  $H_0(j/N)$  if  $\tau(j/N, (j+1)/N) < c_\alpha(j/N, (j+1)/N)$ . If  $H_0(j/N)$  cannot be rejected, set  $\hat{q} = j/N$ . If  $H_0(j/N)$  is rejected, test  $H_0((j+1)/N)$  against  $H_1((j+2)/N)$  using critical value  $c_\alpha(j/N, (j+1)/N)$ .
4. Keep testing using critical value  $c_\alpha(j/N, (j+1)/N)$  until the null hypothesis  $H_0(m/N)$  cannot be rejected. If  $j < m < N$ , continue to the next step, and so on. If all null hypotheses up to  $H_0((N-1)/N)$  can be rejected, set  $\hat{q} = 1$ .

The critical values  $c_\alpha((k-1)/N, k/N)$  are obtained in the same way as in *BSUT*. Very informally one could see the *RW* method as a “shortcut” to our *BSUT* method: instead of calculating critical values at every step, first the critical value from the previous step is used, given that  $c_\alpha((k-1)/N, k/N) \leq c_\alpha(k/N, (k+1)/N)$  for all  $k = 1, \dots, N-1$ .

Given the similarity between the two methods, the modification of *BSUT* immediately leads to an extension of *RW*: in analogy to the difference between *BSUT* and *BSUT*<sub>2</sub>, we could change how the critical values  $c_\alpha(k/N, (k+1)/N)$  in the *RW* procedure are found; instead of deleting the units deemed stationary, they can be incorporated into the bootstrap as stationary units. We will denote this modified *RW* procedure by *RW*<sub>2</sub>. In analogy with the sequential approaches we might expect the *RW*<sub>2</sub> method to be more powerful.

A different approach is to control the false discovery rate (*FDR*), as done by Deckers and Hanck (2009) and Moon and Perron (2010) in a panel setup. To define the *FDR* we must first define the false discovery proportion (*FDP*). The *FDP* is equal to the proportion of rejections that are false. The *FDR* is then defined as the expectation of the *FDP*. This generalized error rate is more “liberal” than the *FWE*, and therefore more suitable in large

$N$  panels. Controlling the *FWE*, the probability of just one false rejection, becomes very difficult for a large  $N$ , and does not even make sense if  $N \rightarrow \infty$ . This makes the *FWE* unsuitable in large panels, and controlling *FDR* is then an attractive alternative. In small panels however some researchers might be uncomfortable with the error that the *FDR* allows for. Moreover, controlling for *FDR* has the disadvantage that the expectation of *FDP* is controlled for, and not *FDP* itself; the *realized FDP* may be very different from its expected value (cf. Romano et al., 2008b, p. 423).

Moon and Perron (2010) develop unit root tests that can be applied in panel data to determine which units are stationary, based on controlling the *FDR*. They use several methods to control *FDR* but find that the bootstrap approach of Romano et al. (2008a) works best in controlling *FDR*. Deckers and Hanck (2009) also find favorable results for this bootstrap method.

Note though that this method cannot be directly compared to the *BSQT* method, as both are based on different principles. We can however compare the performance of both approaches, as they are designed for the same goal, although we have to keep the different philosophies (and therefore sometimes different results) in mind.

## 4 Simulations

### 4.1 Simulation DGP

We now perform two simulation studies; one in panels with small  $N$ , the second in panels with large  $N$ . While the methods that we analyze depend on this setting (details below), we use the same DGP for both, although with different  $T$  and  $N$ . Our DGP is very similar to the DGP used by Moon and Perron (2010) in their simulation study.<sup>12</sup>

$$\begin{aligned} y_{i,t} &= \mu_i + x_{i,t} \quad i = 1, \dots, N \quad t = 1, \dots, T, \\ x_{i,t} &= \rho_{i,T} x_{i,t-1} + u_{i,t}, \quad x_0 = 0, \end{aligned}$$

where  $u_{i,t}$  is a sum of a common and an idiosyncratic component

$$u_{i,t} = \lambda_i f_t + w_{i,t},$$

and the individual effects  $\mu_i$  are  $N(0, 1)$ .

---

<sup>12</sup>Our DGP is fairly simple and could be extended to allow for dynamic cross-dependencies, cointegration, etc, as Palm et al. (2010) showed that in such models the block bootstrap procedure is asymptotically valid. These issues however are not the focus of the present work, and our current DGP suffices to analyze the finite sample performance of the methods discussed here given the focus on the (non)stationarity properties of individual units.

For the common factor we let

$$f_t = 0.5f_{t-1} + \nu_{i,t},$$

where  $\nu_{i,t} \sim i.i.d.N(0,1)$ . For the factor loadings,  $\lambda_i$ , we either take  $\lambda_i = 0$ , in which case there is no cross-sectional dependence, or take  $\lambda_i \sim U[-1,3]$ , in which the cross-sectional dependence is generated by a factor structure.

The idiosyncratic components  $w_{i,t}$  are modeled as an  $ARMA(1,1)$  process,

$$w_{i,t} = \phi_i w_{i,t-1} + \varepsilon_{i,t} + \theta_i \varepsilon_{i,t-1}.$$

Here we take  $\phi_i = 0$  or  $\phi_i = U[-0.5, 0.5]$  for all  $i$ , and  $\psi_i = 0$  or  $\psi_i = U[-0.5, 0.5]$  for all  $i$ . Finally we take  $\varepsilon_{i,t} \sim i.i.d.N(0,1)$ .

We take  $N = 10$  combined with  $T = 50$  and  $T = 100$  for the small  $N$  panel, while we take  $N = 50$  with  $T = 100$  and  $N = 200$  with  $T = 25$  for the large  $N$  panel. Furthermore, for the stationary units in the panel,  $i \leq k_0$ , we take  $\rho_{i,T} \sim U[0, 1 - cT^{-1}]$  with  $c = 10$ , and for  $i > k_0$  we take  $\rho_{i,T} = 1$ . Table 1 gives all the combinations of parameters we use. All computations were performed in GAUSS 10.

INSERT TABLE 1 ABOUT HERE

## 4.2 Simulation study in small $N$ panels

In the small  $N$  panel we compare the  $BSUT$ ,  $BSUT_2$ ,  $RW$  and  $RW_2$  procedures, while we also add the bootstrap approach of Moon and Perron (2010) which is denoted by  $MP$ . For all methods we apply the block bootstrap based on residuals as described above. Note that Moon and Perron (2010) propose their method with the block bootstrap based on first differences. We slightly modified their method to make sure that any differences found are not caused by differences in bootstrap method. Also, for all methods we use the ADF t-test with OLS demeaning. As in Moon and Perron (2010), lag lengths were selected by MAIC (Ng and Perron, 2001) with a maximum of 4 lags. Block lengths are taken as  $b = 1.75T^{1/3}$  as in Palm et al. (2010). For  $k_0$ , the number of stationary units in the panel, we take  $k_0 = 0, 2, 5, 9$  for all parameter combinations. The level of all tests (or  $FWE/FDR$  when appropriate) is taken to be 5%.

The simulation results are given in Tables 2 and 3. We report the average  $\hat{k}$ , denoted by  $M(\hat{k})$ , the standard deviation of  $\hat{k}$ , denoted by  $S(\hat{k})$ , the average proportion of correctly found stationary units, denoted by  $CP$ , the  $FWE$  and the  $FDR$ . All results are based on 1000 Monte Carlo simulations and 499 bootstrap replications.

INSERT TABLE 2 ABOUT HERE

INSERT TABLE 3 ABOUT HERE

The four tests that asymptotically control  $FWE$  can be seen to control  $FWE$  in finite samples as well;  $FWE$  is always close to  $\alpha$  across all models. While the  $FWE$  control is not sensitive to the dynamic parameters, the ability to (correctly) reject is, as can be seen from the  $CP$  results in particular. The increase from  $T = 50$  to  $T = 100$  greatly improves the methods' ability to correctly pick up the stationary units;  $CP$  increases from roughly 0.5 to roughly 0.8. The  $FWE$  on the other hand is not very much affected. Cross-sectional dependence appears to have a small negative effect on the methods' ability to pick up the stationary units.

Comparing the  $MP$  test with the other tests, we see that it is usually able to identify more stationary units than the other methods, in particular if  $k_0$  is high. This is not strange, as if  $k_0$  is large, and correspondingly the number of rejections is high, more false rejections are allowed by the  $FDR$ . This is also reflected in the  $FWE$ , which can rise up to about 0.5 for large  $k_0$ . Note however that the  $MP$  does control  $FDR$  as it is supposed to do. It is therefore not very fair to compare the  $MP$  method directly with the  $FWE$  controlling methods; both have a different goal, and it is up to the applied researcher to determine if he is comfortable with controlling the  $FDR$  in such a small panel or wishes to control the  $FWE$ .

The  $BSUT$ ,  $BSUT_2$ ,  $RW$  and  $RW_2$  procedures perform very similarly. In general the  $BSUT_2$  and  $RW_2$  methods are slightly more powerful, indicating that there may indeed be a gain to including the stationary units as stationary in the bootstrap as opposed to deleting them. The gain is fairly small however. The  $BSUT$  and  $RW$  methods perform, as predicted, almost identically, and it would be hard to argue that the difference between them is more than simulation randomness.

Concluding, the sequential  $BSUT$  methods perform well in small  $N$  panels, but do not improve on the  $RW$  method. One might therefore argue what their added value is, beside offering the small modification that can be applied to  $RW$  as well. The answer, as expected from much of the foregoing, lies not in small  $N$  panels, but in large  $N$  panels.

### 4.3 Simulation study in large $N$ panels

We now consider the case of large  $N$  panels. We now take  $T = 100$  and  $N = 50$ , which represents a large macro panel, and  $T = 25$  and  $N = 200$ , which represents a micro panel. As true proportions of stationary units we take  $q_0 = 0, 0.2, 0.5, 0.9$ . The tests considered are  $BSQT$ ,  $BSQT_2$ ,  $IBSQT$ ,  $IBSQT_2$  and  $MP$ . Unit root tests used and lag selection are the same as in the previous section. We do not consider the  $BSUT$  and  $RW$  tests here, as these are mainly designed for small panels. For the sequential methods we take four equally spaced quantiles to be tested for the model with  $N = 50$ , and eight for  $N = 200$ . For the iterative approaches the same quantiles are used in the first step, with four quantiles in the following steps.

Results are given in Tables 4 and 5. As in the previous simulation study, we report  $M(\hat{q})$ ,  $CP$  and  $FDR$ . We do not report  $S(\hat{q})$ , as this is not a fair comparison, given that not all methods can select the same numbers. We also do not report the  $FWE$ , for reasons given before. The  $FDR$  of course remains a sensible criterion, but it is not entirely fair in this situation. As the  $BSQT$  and  $BSQT_2$  methods can only select the user-defined quantiles, they are bound to have a number of false rejections whenever the true number of stationary units lies in between two quantiles, and the higher number is selected. We know from Corollary 1 that this occurs with a probability larger than  $\alpha$ . This will obviously increase the  $FDR$ ; this behavior of the procedure however is known and accepted if we choose to apply the test. Therefore we construct a new criterion. Suppose that  $q_j < q_0 < q_{j+1}$ . Then, given the properties of the  $BSQT$  procedures, we would want the method to either select  $q_j$  or  $q_{j+1}$ . To have a comparable criterion for the  $IBSQT$  and  $MP$  methods, we construct an interval around  $q_0$  comparable to the interval  $[q_j, q_{j+1}]$ . We do not take this interval exactly however, as it could be that  $q_0$  is very close to  $q_j$ ; in such a case the  $MP$  or  $IBSQT$  methods could select a value slightly below  $q_j$ , which would be good, but be outside this interval. Therefore we select the interval as  $I_0 = [q_0 - q^*, q_0 + q^*]$ , where  $q^* = \max(q_0 - q_j, q_{j+1} - q_0)$  if  $q_j < q_0 < q_{j+1}$ , and  $q^* = (q_{j+1} - q_j)/2$  if  $q_0 = q_j$ .<sup>13</sup> The proportion of selected  $\hat{q}$  that is within  $I_0$  is reported as  $WI$ , and could be interpreted as a “power” measure. We also report the proportion selected higher than this interval in  $HI$ , which could be interpreted as a “size” measure.

INSERT TABLE 4 ABOUT HERE

INSERT TABLE 5 ABOUT HERE

For the macro panel, the  $MP$  method works very well; it is able to identify most of the stationary units, and controls  $FDR$  well. For the micro panel with  $N = 200$ , while  $FDR$  is still controlled, the proportion of stationary units picked up decreases quite strongly. This is not so unexpected, as the  $MP$  test cannot exploit the cross-sectional dimension effectively, and there are only 25 time series observations in this model. It therefore seems that the  $MP$  method performs well as long as  $T$  is not too small, and  $N$  is not too large.

While the pattern described above also applies to some extent to the  $BSQT$  methods, there are still some significant differences with the  $MP$  method. In the macro panel the  $BSQT$  methods are able to pick up slightly fewer stationary units than  $MP$  if  $q_0$  is not equal to a selected quantile, but if it is equal, the  $BSQT$  methods perform quite similarly. In the micro panel though the  $BSQT$  methods pick up much more stationary units than the  $MP$  method (although still not nearly all). Here the ability of the  $BSQT$  methods to exploit the

---

<sup>13</sup>Given our choices for the quantiles described above, this results in intervals around  $q_0 = 0, 0.2, 0.5, 0.9$  of  $[0, 0.125]$ ,  $[0, 0.4]$ ,  $[0.375, 0.525]$  and  $[0.75, 1]$  for  $N = 50$  and  $[0, 0.063]$ ,  $[0.125, 0.275]$ ,  $[0.437, 0.563]$  and  $[0.8, 1]$  for  $N = 200$

cross-sectional dimension really shows. Furthermore, from the reported  $HI$ , we can see that the estimates of stationary units are rarely too high. As expected,  $FDR$  is not controlled by  $BSQT$  if  $q_0$  is not equal to a tested quantile; if they are equal the methods do seem to control  $FDR$ .

The  $IBSQT$  approaches appear to perform quite well; even though they are not especially designed to do so, they appear to be able to control  $FDR$ . They do however tend to pick up somewhat less of the stationary units than the  $BSQT$  tests, although still more than the  $MP$  tests in the micro panel. This is caused by the fact that they lose power in higher iterations, as the cross-sectional dimension then effectively decreases. The  $BSQT_2$  and  $IBSQT_2$  methods are on average somewhat more powerful than the  $BSQT$  and  $IBSQT$  methods, although the difference is again fairly small.

All methods are less able to pick up the stationary units if there is a common factor; this is not surprising for the sequential tests, as their power basically comes from “pooling” the cross-section, and it is well known in the panel unit root literature that pooling is less effective if there is strong dependence across the units.

Also, in the micro panel serial correlation seems to have quite a negative effect on power. This is most likely caused by the fact that larger lag lengths are needed in those models, and the power of unit root tests are known to be affected quite a lot by the selected lag length. It might thus be a good idea in such panels to restrict the allowed lag lengths, and try to find other ways to make the marginal distributions live on the same scales, such as by scaling with bootstrap critical values (we will go into more detail in Section 5).

Concluding, the sequential methods perform well in panels with large  $N$ , even if  $T$  is small. In panels where  $T$  is large and  $N$  relatively large, the sequential methods perform nearly as well as the  $MP$  method. In panels with small  $T$  but very large  $N$ , the sequential methods keep performing reasonably well whereas the performance of the  $MP$  method deteriorates significantly.

## 5 Applications

In this section we consider two applications of the tests proposed in this paper. We first test if Purchasing Power Parity (PPP) holds, using a panel of real exchange rates from a group of countries. The second application is based on income data from the Panel Study of Income Dynamics (PSID).

### 5.1 Tests for PPP

One of the most popular applications of panel unit root tests has been in the analysis of PPP (see for example Wagner, 2008, for an overview). In such applications, evidence for PPP is

found if the (log) real exchange rate  $r_{i,t}$  is found stationary, where  $r_{i,t}$  is defined as

$$r_{i,t} = p_{i,t} - p^* - s_{i,t}.$$

$p_{i,t}$  is the log price index (CPI) of country  $i$ ,  $p^*$  is the log price index of the reference country, and  $s_{i,t}$  is the log nominal exchange rate between country  $i$  and the reference country.

Many studies apply a panel unit root test to the panel of real exchange rates and conclude that a rejection of the unit root null hypothesis is evidence for PPP in the panel. As discussed before, such a rejection does not automatically mean that PPP holds for *all* countries in the panel. It is well conceivable that PPP only holds for a subset of the countries, but that the evidence within this subset is strong enough to make the panel unit root test reject. The alternative approach is apply a unit root test such as the ADF test to each country individually (see for example Taylor, 2002). However, this approach does not control size and is likely to lead to false rejections of the null hypothesis.

For these reasons, Hanck (2009) proposes to test for PPP using the multiple testing approach of Romano and Wolf (2005), where he applies sieve bootstrap ADF tests as the unit root tests. Moon and Perron (2010) consider a similar application to PPP of their multiple testing approaches.

We now revisit the dataset of Hanck (2009), which is based on the long annual exchange rate data of Taylor (2002) and consists of data for 19 countries (with the US as reference country) for 105 years to illustrate the performance of the *BSQT* approach. Next to the *BSQT* approach, we consider the *BSQT*<sub>2</sub> approach, the *RW* and *RW*<sub>2</sub> approaches, and the bootstrap approach of Romano et al. (2008a) to control for *FDR* employed in Moon and Perron (2010) (denoted by *MP*). To make all results comparable, we use the same block bootstrap approach and same individual unit root statistics for all methods. In all methods, the “level” (size, *FWE* or *FDR*) is set to 5%.

As the individual unit root test statistics  $\theta_i$ , we consider the union of rejections (UR) statistic proposed by Smeeke and Taylor (2010),<sup>14</sup> which has the form of a scaled minimum of the four statistics.

$$\theta_i(\alpha) = \min \left( \left( \frac{x_i}{c_{i,GLS}^{\mu*}(\alpha)} \right) DF - GLS_i^\mu, \left( \frac{x_i}{c_{i,GLS}^{\tau*}(\alpha)} \right) DF - GLS_i^\tau, \right. \\ \left. \left( \frac{x_i}{c_{i,OLS}^{\mu*}(\pi)} \right) DF - OLS_i^\mu, \left( \frac{x_i}{c_{i,OLS}^{\tau*}(\alpha)} \right) DF - OLS_i^\tau \right).$$

Here  $DF - OLS_i$  and  $DF - GLS_i$  are the ADF test performed on unit  $i$  with OLS and GLS detrending respectively, while superscript  $\mu$  and  $\tau$  indicate whether demeaning or detrending is used respectively. The bootstrap critical values such as  $c_{i,GLS}^{\mu*}(\alpha)$  used in the scaling factors

---

<sup>14</sup>We use the “A” version of their statistic, which is asymptotically valid.

are determined in a preliminary bootstrap step as the individual level  $\alpha$  critical values of the four tests. This test statistic is the bootstrap adaptation of the method proposed by Harvey et al. (2010) of rejecting the null if either one of the four tests rejects (corrected to have the correct size). This way the test can deal with uncertainty regarding the initial condition (which affects the relative power of the OLS and GLS tests) and the presence of deterministic trends,<sup>15</sup> thus obviating the need to report multiple tests which is again subject to problems with controlling size.

In a panel context this becomes even more important, as for instance some exchange rates may exhibit deterministic trends while others do not.<sup>16</sup> In such a case it would be most beneficial to the power of the methods to only include the linear trend for those countries that actually contain the trend, which is clearly infeasible in a panel. However, because the union test statistic is applied to each individual unit, this statistic does automatically take this trend uncertainty into account.

A second benefit in a sequential/multiple testing framework is the use of the scaling factor  $x_i$ . Smeeke and Taylor (2010) take  $x_i = c_{i,GLS}^{\mu*}$ , which implies that the test statistics of unit  $i$  are scaled towards the critical values of the  $DF - GLS_i^\mu$  statistic. However, they argue in Remark 6 that any  $x_i < 0$  suffices. We therefore take  $x_i = -1$  for all  $i = 1, \dots, N$ , which means that all statistics are scaled towards -1. This step further ensures that the marginal distributions of the  $\theta_i$  are “as free of nuisance parameters as possible” as discussed in Remark 5, as the statistics are not scaled towards individual critical values (which may vary over  $i$ ) but to a fixed constant.

The lag lengths of the ADF statistics are selected by MAIC (with maximum lag length  $12(T/100)^{1/4}$ ), with lag length selection also done within the bootstrap. For the DF-GLS tests we apply the modification to MAIC proposed by Perron and Qu (2007) of selecting lags from OLS instead of GLS detrended data. We take a fairly large block length (20) to be able to deal with possible cross-unit cointegration (cf. Palm et al., 2010, Section 4.2.2).

In a (relatively) small  $N$  panel as this it would be natural to apply the *BSUT* method as well. As for this panel *BSUT* and *BSUT*<sub>2</sub> give exactly the same results as *RW* and *RW*<sub>2</sub> respectively, we do not report these separately. Instead, we apply the *BSQT* and *BSQT*<sub>2</sub> procedures with the selected quantiles  $q_j = (j - 1)/5$ ,  $j = 1, \dots, 5$ , to clearly see the differences with the unit-by-unit methods. Results are based on 4999 bootstrap replications, and all calculations were performed in GAUSS 10.<sup>17</sup>

INSERT TABLE 6 ABOUT HERE

---

<sup>15</sup>Harvey, Leybourne, and Taylor (2009) show in detail how uncertainty about the initial condition and/or the presence of deterministic trend affects ADF unit root tests.

<sup>16</sup>While it might seem natural to model exchange rates without a linear trend, Taylor (2002) provides evidence that such trends are present in the real exchange rates of some countries.

<sup>17</sup>GAUSS code to conduct the *BSQT* methods in combination with (possibly unions of) ADF statistics is available on [www.personeel.unimaas.nl/s.smeekes/research.htm](http://www.personeel.unimaas.nl/s.smeekes/research.htm).



The results are presented in Table 6. The second column,  $\theta_i$ , gives the values of the *UR* statistics for each country. It follows from the third column, which reports the result of the individual bootstrap *UR* tests, that for 9 countries the unit root null would have been rejected if each country was tested separately. The *BSQT*, *BSQT*<sub>2</sub> and *MP* test each reject for 8 countries, thus seeming to indicate that one *UR* rejection is false. Notably, the *RW* and *RW*<sub>2</sub> methods only reject for one or two countries, which seems rather low.<sup>18</sup> It might therefore be that these methods already suffer from low power, even in this panel with moderate (but still not very large) *T*. The higher rejections of *BSQT* and *BSQT*<sub>2</sub> on the one hand and *MP* on the other hand are caused by different factors though; *BSQT* and *BSQT*<sub>2</sub> have higher power because they exploit the panel dimension, while *MP* controls size in a more liberal way.

## 5.2 Unit root tests for PSID income data

Pesaran (2007) assesses the validity of the claim of Meghir and Pistaferri (2004) that the log of real earnings of households in the Panel Study of Income Dynamics (PSID) have a unit root. He applies his *CIPS* panel unit root test to the whole sample consisting of  $N = 181$  units, as well as to three subsamples consisting of college graduates (CLG,  $N = 58$ ), high school graduates (HSG,  $N = 87$ ) and high school drop outs (HSD,  $N = 36$ ). As  $T = 22$  this is a typical example of large  $N$ , small  $T$  panel, or what in the previous section was labeled a micro panel.

Pesaran (2007) finds mixed evidence regarding the unit root; the *CIPS* rejects for the full sample, but not for all subsamples.<sup>19</sup> This might lead to the conclusion that there is a relation between education level and stationarity properties, such that certain education groups have a unit root while others do not, and that the rejection for the subgroup drives the full sample rejection. An alternative explanation is that in all three subsamples there are both stationary and nonstationary units, without a pattern related to the education level, and that whether a rejection is observed or not is more of a “coincidence”, depending on a variety of factors.

There is however no way to find out which of the assertions is true with standard panel unit root tests. Therefore we now apply our *BSQT* and *IBSQT* methods (as well as the *MP* method) to the data used in Pesaran (2007).<sup>20</sup> As quantiles to be tested we take  $q_j = \frac{j-1}{6}$ ,  $j = 1, \dots, 6$ , while for the iterative approaches we divide the intervals in the later iterations into three quantiles. We again consider the *UR* test statistics already considered in the

---

<sup>18</sup>The results for *RW* and *RW*<sub>2</sub>, unlike the other methods, are somewhat sensitive to the selected block length. However, the number of stationary units found using these methods does remain significantly lower than the other methods irrespective of the block length, and so varying the block length does not change the overall conclusions.

<sup>19</sup>Which subgroups are rejected depends on the specification of the *CIPS* test, see Pesaran (2007) for details.

<sup>20</sup>The dataset used in Pesaran (2007) is available from the Journal of Applied Econometrics Data Archive ([www.econ.queensu.ca/jae](http://www.econ.queensu.ca/jae))

previous application. We take a block size of 5 (corresponding to  $1.75T^{1/3}$ ), while we select lags again using MAIC with the Perron and Qu (2007) modification with a maximum lag length of 3 to avoid losing too many observations.<sup>21</sup>

INSERT TABLE 7 ABOUT HERE

Results are listed in Table 7. Our results from  $BSQT$  and  $BSQT_2$  indicate that about one third to one half of the units are stationary.  $IBSQT$  and  $IBSQT_2$  result in fewer rejections, which is a typical example of the iterative method stopping at the left boundary because the power decreases in each subsequent iteration. Also note that  $BSQT_2$  is more powerful here than  $BSQT$ .

Note how the  $MP$  method leads to a very different result for these data; according to the  $MP$  method only very few units are stationary, which is not consistent with the results of Pesaran (2007), nor with the sequential methods. It therefore seems likely that the  $MP$  method suffers from low power because of the small  $T$ , as found in the simulation study. As discussed before, the  $BSQT$  methods overcome the small  $T$  by exploiting the cross-sectional dimension, which the  $MP$  method is unable to do.

The results for the subsamples are not obtained by applying the tests to the subsamples directly, but instead by ordering the units for which rejections were found in the complete sample into the three subgroups. It seems that the proportion of stationary units in each of the subsamples is fairly similar, from which we may conclude that there appears to be no relation between education level and unit root properties.

## 6 Conclusion

We have proposed new methods based on sequential tests to investigate the stationarity properties of individual units in a panel. The approach is based on testing user-defined quantiles sequentially, representing the proportions of stationary data the researcher wants to test for. By being based on quantiles, the method is applicable in panels with a large  $N$ , while it can also be made suitable for small  $N$  panels by testing individual units sequentially. The critical values are based on the block bootstrap and shown to be asymptotically valid.

We also consider two modifications. The first is a modification of the bootstrap procedure that leads to (somewhat) higher power in finite sample. This modification can also be applied to existing bootstrap methods based on multiple testing procedures. The second modification is an iterative application of the sequential approach designed to give more precise estimates of the number of stationary units, not restricted to the user-defined quantiles.

---

<sup>21</sup>We do not have to worry about the lag length being too small to cancel out nuisance parameters, as the scaling inherent in the  $UR$  approach (by setting  $x_i = -1$  for  $i = 1, \dots, N$ ) will take care of that automatically.

Through a simulation study we demonstrated the good performance of these methods in finite samples, where the sequential (and iterative) methods have a significant advantage over multiple testing approaches if  $N$  is rather large but  $T$  is small.

We also illustrated the tests by two empirical applications, in testing for unit roots in real exchange rates and log earnings data of households. These applications, and in particular the earnings application with large  $N$  and small  $T$ , again demonstrate the usefulness of these methods.

The methods developed in this paper are not restricted to unit root testing in panels; these sequential approaches can be used in many settings where multiple testing methods are used, but may not be optimal, especially in applications where  $N$  is very large but  $T$  may not be.

## References

- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Basawa, I. V., A. K. Mallik, W. P. McCormick, J. H. Reeves, and R. L. Taylor (1991). Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Communications in Statistics - Theory and Methods* 20, 1015–1026.
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* 16, 1709–1722.
- Deckers, T. and C. Hanck (2009). Multiple testing techniques in growth econometrics. MRPA Paper No. 17843.
- Di Iorio, F. and S. Fachin (2008). Testing for cointegration in dependent panels via residual-based bootstrap methods. Working paper.
- Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
- Giacomini, R., D. N. Politis, and H. White (2007). A Warp-Speed method for conducting Monte Carlo experiments involving bootstrap estimators. Working paper.
- Gonçalves, S. (2010). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory*, forthcoming.
- Hanck, C. (2009). For which countries did PPP hold? A multiple testing approach. *Empirical Economics* 37, 93–103.

- Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009). Unit root testing in practice: dealing with uncertainty over the trend and initial condition. *Econometric Theory* 25, 587–636.
- Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2010). Testing for unit roots in the presence of uncertainty over both the trend and initial condition. *Journal of Econometrics*, forthcoming.
- 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. (2003). Determining the stationarity properties of individual series in panel datasets. Department of Economics Working Paper No. 495, Queen Mary, University of London.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica* 72, 1–32.
- Moon, H. R. and B. Perron (2010). Beyond panel unit root tests: Using multiple testing to determine the non stationarity properties of individual series in a panel. Department of Economics working paper 2010-04, University of Montreal.
- Ng, S. (2008). A simple test for nonstationarity in mixed panels. *Journal of Business and Economic Statistics* 26, 113–127.
- Ng, S. and P. Perron (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
- Palm, F. C., S. Smeekes, and J.-P. Urbain (2010). Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics*, forthcoming.
- Paparoditis, E. and D. N. Politis (2003). Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–855.
- Perron, P. and Z. Qu (2007). A simple modification to improve the finite sample properties of Ng and Perron’s unit root tests. *Economics Letters* 94, 12–19.
- Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-sectional dependence. *Journal of Applied Econometrics* 22, 265–312.
- Romano, J. P., A. M. Shaikh, and M. Wolf (2008a). Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test* 17, 417–442.

- Romano, J. P., A. M. Shaikh, and M. Wolf (2008b). Formalized data snooping based on generalized error rates. *Econometric Theory* 24, 404–447.
- Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237–1282.
- Shaked, M. and J. G. Shanthikumar (2007). *Stochastic Orders*. New York: Springer.
- Smeekes, S. and A. M. R. Taylor (2010). Bootstrap union tests for unit roots. Granger Centre Discussion Paper 10/03, University of Nottingham.
- Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* 74, 1699–1714.
- Taylor, A. M. (2002). A century of purchasing-power parity. *Review of Economics and Statistics* 84, 139–150.
- Wagner, M. (2008). On PPP, unit roots and panels. *Empirical Economics* 35, 229–249.
- Westerlund, J. and J. Breitung (2009). Myths and facts about panel unit root tests. Working Papers in Economics No 380, University of Gothenburg.
- White, H. (2000). A reality check for data snooping. *Econometrica* 68, 1097–1126.

## A Appendix: Proofs

PROOF OF THEOREM 1. We first show (6). Note that  $\theta_i \xrightarrow{p} -\infty$  for  $i \in \mathcal{S}$  as  $T \rightarrow \infty$  by Assumption 2. Therefore, if  $k_0 = |\mathcal{S}|$ , we have that

$$\theta_{(1)}, \dots, \theta_{(k_0)} \xrightarrow{p} -\infty. \quad (13)$$

If  $q_{j+1} \leq q_0$ ,  $k_{j+1} \leq k_0$  and therefore  $\theta_{(k_{j+1})} \xrightarrow{p} -\infty$ , which proves (6a).

If  $k_{j+1} > k_0$ , then by (13) all order statistics below  $k_0$  have no effect on the distribution of the remaining order statistics. Therefore we only have to consider those  $\theta_i$  for which  $i \in \mathcal{U}$ , and hence

$$\theta_{(k_{j+1})} \xrightarrow{p} \theta_{(k_{j+1}-k_0:\mathcal{U})}.$$

The result in (6b) then follows directly from the continuous mapping theorem (cf. White, 2000, Lemma 2).

The result in (7) follows directly from Paparoditis and Politis (2003) and Palm et al. (2010), as their proofs of asymptotic validity can straightforwardly be extended to general

unit root statistics. While Paparoditis and Politis (2003) and Palm et al. (2010) focus on the DF coefficient test (without detrending), their results can easily be extended to more general test statistics. The crucial result that lies at the heart of any proof of validity is the invariance principle of the bootstrap partial sum process, which is derived in Lemmas 2 and 4 of Palm et al. (2010). As this result still applies here, it can be used to establish the asymptotic validity of more general unit root test statistics.

Finally, the result in (8) follows from the fact that in the bootstrap all units are generated under the null of a unit root, and the application of the continuous mapping theorem.  $\square$

PROOF OF COROLLARY 1. (9a) directly follows from (6a) coupled with (8). To prove (9b) note that for  $q_j < q_0 < q_{j+1}$ ,

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) &= \lim_{T \rightarrow \infty} \mathbb{P}(H_0(q_j) \text{ not rejected}) = \lim_{T \rightarrow \infty} \mathbb{P}\left\{\theta_{(k_{j+1})} \geq c_\alpha^*(q_j, q_{j+1})\right\} \\ &= \lim_{T \rightarrow \infty} \mathbb{P}\left\{\theta_{(k_{j+1}-k_0:\mathcal{U})} \geq c_\alpha^*(q_j, q_{j+1})\right\} = \mathbb{P}\left\{\xi_{(k_{j+1}-k_0:\mathcal{U})} \geq c_{\alpha,\infty}(q_j, q_{j+1})\right\} \\ &= 1 - a_{j,k_0}^\alpha, \end{aligned}$$

where the third equality follows from (13). (9c) is the probability of not rejecting the true null hypothesis  $H_0(q_j)$ , which is equal to  $1 - \alpha$  as  $\mathbb{S}_{k_j}^c$  is equal to  $\mathcal{U}$  in the limit with probability 1 by (13). To show (9d), note that if  $q_{j-1} < q_0 < q_j$ ,

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{q} = q_j) \leq \lim_{T \rightarrow \infty} \mathbb{P}(H_0(q_j - 1) \text{ rejected}) = \lim_{T \rightarrow \infty} \mathbb{P}\left\{\theta_{(k_j)} < c_\alpha^*(q_{j-1}, q_j)\right\} = a_{j-1,k_0}^\alpha,$$

which follows in the same way as (9b). To show (9e) assume without loss of generality that  $q_{j-1} = q_0$ . The result then follows from the fact that

$$\begin{aligned} \sum_{m=1}^r \mathbb{P}(\hat{q} = q_m) &= \sum_{m=1}^{j-2} \mathbb{P}(\hat{q} = q_m) + \mathbb{P}(\hat{q} = q_{j-1}) + \sum_{m=j}^r \mathbb{P}(\hat{q} = q_m) \\ &= 1 - \alpha + \sum_{m=j}^r \mathbb{P}(\hat{q} = q_m) = 1. \end{aligned}$$

Finally, the middle inequality in (12) follows if

$$\mathbb{P}\left\{\xi_{(k_{j+1}-m:\mathcal{U}_m)} < x\right\} \leq \mathbb{P}\left\{\xi_{(k_{j+1}-m-1:\mathcal{U}_{m+1})} < x\right\}$$

for any  $x \in \mathbb{R}$ , which is equivalent to

$$\xi_{(k_{j+1}-m-1:\mathcal{U}_{m+1})} \leq_{st} \xi_{(k_{j+1}-m:\mathcal{U}_m)}, \quad (14)$$

where “ $X \leq_{st} Y$ ” denotes that  $X$  is smaller in the usual stochastic order than  $Y$  (cf. Shaked and Shanthikumar, 2007, p. 3). (14) then follows directly from Corollary 6.B.24 in Shaked

and Shanthikumar (2007), as  $|\mathcal{U}_m| = N - m$  and therefore  $|\mathcal{U}_{m+1}| - (k_{j+1} - m - 1) = N - k_{j+1} = |\mathcal{U}_m| - (k_{j+1} - m)$ . The first inequality in (12) then follows by setting  $m = k_j$  along with (11), while the last inequality follows by definition. This completes the proof.  $\square$

DGP	$T$	$N$	$\lambda_i$	$\phi_i$	$\theta_i$
<b>Panel A: Small <math>N</math> simulations</b>					
1	50	10	0	0	0
2	50	10	0	U[-0.5,0.5]	U[-0.5,0.5]
3	50	10	U[-1,3]	0	0
4	50	10	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
5	100	10	0	0	0
6	100	10	0	U[-0.5,0.5]	U[-0.5,0.5]
7	100	10	U[-1,3]	0	0
8	100	10	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
<b>Panel B: Large <math>N</math> simulations</b>					
9	100	50	0	0	0
10	100	50	0	U[-0.5,0.5]	U[-0.5,0.5]
11	100	50	U[-1,3]	0	0
12	100	50	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
13	25	200	0	0	0
14	25	200	0	U[-0.5,0.5]	U[-0.5,0.5]
15	25	200	U[-1,3]	0	0
16	25	200	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]

Table 1: Parameter combinations simulation DGPs



DGP		$k_0 = 0$					$k_0 = 2$				
		<i>BSUT</i>	<i>BSUT</i> <sub>2</sub>	<i>RW</i>	<i>RW</i> <sub>2</sub>	<i>MP</i>	<i>BSUT</i>	<i>BSUT</i> <sub>2</sub>	<i>RW</i>	<i>RW</i> <sub>2</sub>	<i>MP</i>
1	$M(\hat{k})$	0.038	0.035	0.037	0.040	0.039	0.291	0.307	0.291	0.294	0.322
	$SD(\hat{k})$	0.201	0.199	0.199	0.215	0.213	0.508	0.537	0.502	0.527	0.600
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.121	0.126	0.121	0.122	0.130
	<i>FWE</i>	0.036	0.032	0.035	0.036	0.035	0.049	0.054	0.048	0.049	0.058
	<i>FDR</i>	0.036	0.032	0.035	0.036	0.035	0.043	0.045	0.043	0.040	0.043
2	$M(\hat{k})$	0.055	0.059	0.053	0.056	0.061	0.392	0.395	0.389	0.395	0.414
	$SD(\hat{k})$	0.253	0.264	0.249	0.259	0.278	0.528	0.547	0.534	0.547	0.599
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.165	0.165	0.164	0.166	0.167
	<i>FWE</i>	0.049	0.052	0.047	0.049	0.053	0.060	0.062	0.058	0.060	0.074
	<i>FDR</i>	0.049	0.052	0.047	0.049	0.053	0.052	0.050	0.049	0.049	0.053
3	$M(\hat{k})$	0.026	0.030	0.027	0.029	0.033	0.342	0.353	0.343	0.362	0.381
	$SD(\hat{k})$	0.171	0.208	0.174	0.205	0.210	0.517	0.592	0.523	0.606	0.659
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.158	0.160	0.158	0.163	0.166
	<i>FWE</i>	0.024	0.024	0.025	0.023	0.027	0.023	0.022	0.023	0.024	0.033
	<i>FDR</i>	0.024	0.024	0.025	0.023	0.027	0.019	0.017	0.019	0.018	0.021
4	$M(\hat{k})$	0.030	0.042	0.032	0.045	0.048	0.244	0.278	0.243	0.288	0.296
	$SD(\hat{k})$	0.212	0.326	0.217	0.356	0.371	0.508	0.686	0.500	0.694	0.775
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.101	0.102	0.100	0.109	0.109
	<i>FWE</i>	0.025	0.026	0.025	0.028	0.027	0.037	0.041	0.037	0.041	0.043
	<i>FDR</i>	0.025	0.026	0.025	0.028	0.027	0.033	0.034	0.034	0.033	0.033
5	$M(\hat{k})$	0.041	0.039	0.037	0.044	0.042	1.732	1.738	1.732	1.743	1.931
	$SD(\hat{k})$	0.198	0.194	0.189	0.210	0.206	0.573	0.569	0.566	0.567	0.631
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.843	0.846	0.845	0.849	0.900
	<i>FWE</i>	0.041	0.039	0.037	0.043	0.041	0.046	0.046	0.042	0.045	0.113
	<i>FDR</i>	0.041	0.039	0.037	0.043	0.041	0.018	0.018	0.016	0.017	0.042
6	$M(\hat{k})$	0.047	0.048	0.042	0.048	0.047	0.986	1.000	0.995	0.989	1.123
	$SD(\hat{k})$	0.216	0.232	0.201	0.232	0.221	0.596	0.615	0.584	0.614	0.737
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.466	0.470	0.469	0.465	0.507
	<i>FWE</i>	0.046	0.045	0.042	0.045	0.045	0.051	0.056	0.054	0.055	0.098
	<i>FDR</i>	0.046	0.045	0.042	0.045	0.045	0.031	0.031	0.033	0.030	0.047
7	$M(\hat{k})$	0.032	0.044	0.036	0.045	0.052	1.356	1.402	1.348	1.387	1.569
	$SD(\hat{k})$	0.217	0.400	0.262	0.404	0.444	0.752	0.831	0.742	0.813	1.015
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.657	0.670	0.655	0.663	0.723
	<i>FWE</i>	0.025	0.023	0.026	0.023	0.026	0.030	0.036	0.028	0.036	0.066
	<i>FDR</i>	0.025	0.023	0.026	0.023	0.026	0.015	0.018	0.015	0.018	0.030
8	$M(\hat{k})$	0.026	0.024	0.027	0.026	0.025	1.068	1.120	1.074	1.119	1.295
	$SD(\hat{k})$	0.159	0.153	0.168	0.165	0.162	0.622	0.705	0.641	0.689	0.819
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.519	0.537	0.521	0.540	0.606
	<i>FWE</i>	0.026	0.024	0.026	0.025	0.024	0.028	0.039	0.030	0.031	0.066
	<i>FDR</i>	0.026	0.024	0.026	0.025	0.024	0.015	0.018	0.016	0.015	0.030

Table 2: Simulation results for small  $N$  panels, part 1

DGP		$k_0 = 5$					$k_0 = 9$				
		<i>BSUT</i>	<i>BSUT</i> <sub>2</sub>	<i>RW</i>	<i>RW</i> <sub>2</sub>	<i>MP</i>	<i>BSUT</i>	<i>BSUT</i> <sub>2</sub>	<i>RW</i>	<i>RW</i> <sub>2</sub>	<i>MP</i>
1	$M(\hat{k})$	2.823	2.849	2.796	2.856	3.503	4.104	4.532	4.135	4.499	6.757
	$SD(\hat{k})$	1.169	1.197	1.168	1.187	1.405	1.775	1.889	1.748	1.919	2.435
	<i>CP</i>	0.556	0.561	0.551	0.562	0.667	0.455	0.501	0.459	0.498	0.728
	<i>FWE</i>	0.042	0.044	0.038	0.043	0.151	0.007	0.019	0.006	0.018	0.201
	<i>FDR</i>	0.013	0.013	0.012	0.013	0.036	0.001	0.003	0.001	0.002	0.022
2	$M(\hat{k})$	1.992	2.019	1.976	2.038	2.562	4.828	5.108	4.830	5.086	7.670
	$SD(\hat{k})$	1.018	1.061	1.022	1.060	1.414	1.572	1.719	1.616	1.745	2.116
	<i>CP</i>	0.393	0.399	0.390	0.402	0.490	0.534	0.564	0.534	0.562	0.821
	<i>FWE</i>	0.026	0.025	0.025	0.029	0.101	0.019	0.030	0.022	0.029	0.280
	<i>FDR</i>	0.009	0.008	0.009	0.009	0.026	0.003	0.004	0.004	0.004	0.029
3	$M(\hat{k})$	0.933	0.969	0.912	0.976	1.351	4.040	4.299	4.076	4.316	6.382
	$SD(\hat{k})$	0.912	0.998	0.901	0.998	1.505	2.267	2.387	2.267	2.412	2.995
	<i>CP</i>	0.183	0.190	0.179	0.191	0.258	0.448	0.477	0.452	0.479	0.690
	<i>FWE</i>	0.016	0.019	0.015	0.019	0.052	0.005	0.006	0.006	0.009	0.173
	<i>FDR</i>	0.011	0.010	0.011	0.010	0.018	0.001	0.001	0.001	0.001	0.018
4	$M(\hat{k})$	1.766	1.835	1.756	1.810	2.287	3.555	3.830	3.561	3.858	5.558
	$SD(\hat{k})$	1.204	1.297	1.205	1.272	1.687	2.232	2.393	2.210	2.390	3.175
	<i>CP</i>	0.347	0.358	0.345	0.353	0.433	0.394	0.425	0.395	0.427	0.601
	<i>FWE</i>	0.028	0.030	0.025	0.031	0.066	0.005	0.008	0.004	0.012	0.145
	<i>FDR</i>	0.011	0.011	0.009	0.013	0.022	0.001	0.001	0.001	0.002	0.015
5	$M(\hat{k})$	3.795	3.844	3.815	3.845	4.635	7.880	7.990	7.885	7.996	9.388
	$SD(\hat{k})$	0.785	0.772	0.754	0.771	0.833	1.132	1.096	1.140	1.087	0.667
	<i>CP</i>	0.752	0.761	0.755	0.761	0.880	0.871	0.883	0.872	0.884	0.990
	<i>FWE</i>	0.037	0.040	0.040	0.041	0.191	0.037	0.045	0.040	0.044	0.482
	<i>FDR</i>	0.008	0.008	0.009	0.009	0.039	0.004	0.005	0.004	0.005	0.048
6	$M(\hat{k})$	3.329	3.443	3.319	3.450	4.443	7.430	7.543	7.427	7.528	9.191
	$SD(\hat{k})$	1.070	1.120	1.077	1.136	1.252	1.080	1.058	1.084	1.080	0.780
	<i>CP</i>	0.658	0.677	0.656	0.679	0.836	0.823	0.836	0.823	0.834	0.977
	<i>FWE</i>	0.040	0.056	0.040	0.053	0.207	0.023	0.023	0.022	0.023	0.401
	<i>FDR</i>	0.009	0.013	0.010	0.011	0.044	0.003	0.002	0.002	0.002	0.040
7	$M(\hat{k})$	2.151	2.227	2.147	2.227	3.060	4.809	4.936	4.793	4.937	6.684
	$SD(\hat{k})$	0.895	1.045	0.897	1.018	1.493	1.269	1.351	1.269	1.364	1.898
	<i>CP</i>	0.428	0.440	0.427	0.442	0.588	0.534	0.547	0.532	0.547	0.724
	<i>FWE</i>	0.010	0.014	0.009	0.011	0.064	0.007	0.013	0.007	0.016	0.170
	<i>FDR</i>	0.004	0.005	0.003	0.004	0.019	0.001	0.002	0.001	0.002	0.018
8	$M(\hat{k})$	3.296	3.447	3.283	3.455	4.372	6.897	7.111	6.915	7.160	9.263
	$SD(\hat{k})$	1.124	1.174	1.120	1.178	1.322	1.729	1.769	1.725	1.743	1.094
	<i>CP</i>	0.652	0.677	0.649	0.679	0.822	0.764	0.787	0.766	0.792	0.978
	<i>FWE</i>	0.033	0.049	0.034	0.050	0.194	0.021	0.026	0.018	0.029	0.460
	<i>FDR</i>	0.008	0.012	0.009	0.012	0.043	0.002	0.003	0.002	0.003	0.046

Table 3: Simulation results for small  $N$  panels, part 2

DGP		$q_0 = 0$					$q_0 = 0.2$				
		<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>IBSQT</i>	<i>IBSQT</i> <sub>2</sub>	<i>MP</i>	<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>IBSQT</i>	<i>IBSQT</i> <sub>2</sub>	<i>MP</i>
9	$M(\hat{q})$	0.002	0.002	0.001	0.001	0.001	0.230	0.230	0.118	0.123	0.148
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.829	0.828	0.576	0.597	0.709
	<i>FDR</i>	0.006	0.006	0.036	0.036	0.035	0.246	0.245	0.021	0.025	0.037
	<i>WI</i>	0.994	0.994	1.000	1.000	1.000	0.995	0.995	1.000	1.000	1.000
	<i>HI</i>	0.006	0.006	0.000	0.000	0.000	0.005	0.005	0.000	0.000	0.000
10	$M(\hat{q})$	0.008	0.007	0.001	0.001	0.001	0.231	0.232	0.120	0.125	0.151
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.829	0.833	0.583	0.607	0.718
	<i>FDR</i>	0.031	0.026	0.047	0.049	0.048	0.250	0.251	0.021	0.025	0.042
	<i>WI</i>	0.969	0.974	1.000	1.000	1.000	0.999	0.998	1.000	1.000	1.000
	<i>HI</i>	0.031	0.026	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.000
11	$M(\hat{q})$	0.015	0.018	0.008	0.014	0.002	0.038	0.046	0.120	0.135	0.141
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.097	0.104	0.558	0.600	0.677
	<i>FDR</i>	0.036	0.036	0.038	0.047	0.020	0.044	0.051	0.022	0.032	0.022
	<i>WI</i>	0.964	0.964	0.980	0.970	0.996	0.975	0.959	0.981	0.968	0.994
	<i>HI</i>	0.036	0.036	0.020	0.030	0.004	0.025	0.041	0.019	0.032	0.006
12	$M(\hat{q})$	0.008	0.011	0.002	0.006	0.001	0.073	0.077	0.072	0.091	0.087
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.242	0.241	0.338	0.393	0.417
	<i>FDR</i>	0.026	0.029	0.020	0.031	0.019	0.085	0.086	0.024	0.038	0.020
	<i>WI</i>	0.974	0.971	0.994	0.985	0.999	0.983	0.972	0.996	0.979	0.999
	<i>HI</i>	0.026	0.029	0.006	0.015	0.001	0.017	0.028	0.004	0.021	0.001
13	$M(\hat{q})$	0.001	0.001	0.000	0.000	0.000	0.123	0.124	0.040	0.045	0.015
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.462	0.464	0.184	0.205	0.073
	<i>FDR</i>	0.004	0.005	0.019	0.017	0.019	0.241	0.242	0.059	0.071	0.013
	<i>WI</i>	0.996	0.995	1.000	1.000	1.000	0.976	0.978	0.008	0.009	0.000
	<i>HI</i>	0.004	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	$M(\hat{q})$	0.000	0.000	0.000	0.000	0.000	0.003	0.004	0.001	0.001	0.001
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.009	0.010	0.002	0.003	0.003
	<i>FDR</i>	0.000	0.000	0.010	0.011	0.010	0.011	0.012	0.013	0.012	0.012
	<i>WI</i>	1.000	1.000	1.000	1.000	1.000	0.026	0.028	0.000	0.000	0.000
	<i>HI</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	$M(\hat{q})$	0.010	0.013	0.005	0.008	0.001	0.029	0.031	0.015	0.020	0.008
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.076	0.079	0.044	0.056	0.029
	<i>FDR</i>	0.047	0.047	0.038	0.045	0.014	0.059	0.060	0.034	0.039	0.014
	<i>WI</i>	0.953	0.953	0.973	0.968	0.998	0.148	0.145	0.028	0.023	0.002
	<i>HI</i>	0.047	0.047	0.027	0.032	0.002	0.018	0.022	0.009	0.017	0.004
16	$M(\hat{q})$	0.013	0.016	0.007	0.011	0.002	0.029	0.035	0.021	0.028	0.018
	<i>CP</i>	0.000	0.000	0.000	0.000	0.000	0.074	0.079	0.073	0.087	0.079
	<i>FDR</i>	0.055	0.051	0.072	0.075	0.051	0.065	0.067	0.043	0.053	0.028
	<i>WI</i>	0.945	0.949	0.972	0.965	0.997	0.150	0.133	0.039	0.031	0.004
	<i>HI</i>	0.055	0.051	0.028	0.035	0.003	0.016	0.032	0.005	0.021	0.002

Table 4: Simulation results for large  $N$  panels, part 1

DGP		$q_0 = 0.5$					$q_0 = 0.9$				
		<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>IBSQT</i>	<i>IBSQT</i> <sub>2</sub>	<i>MP</i>	<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>IBSQT</i>	<i>IBSQT</i> <sub>2</sub>	<i>MP</i>
9	$M(\hat{q})$	0.505	0.505	0.389	0.401	0.479	0.820	0.819	0.768	0.782	0.915
	<i>CP</i>	0.944	0.943	0.770	0.792	0.918	0.880	0.880	0.848	0.861	0.977
	<i>FDR</i>	0.062	0.061	0.010	0.011	0.041	0.029	0.028	0.007	0.008	0.037
	<i>WI</i>	0.970	0.969	0.476	0.661	0.996	1.000	1.000	0.772	0.807	0.999
	<i>HI</i>	0.024	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	$M(\hat{q})$	0.508	0.508	0.375	0.392	0.451	0.830	0.830	0.785	0.807	0.933
	<i>CP</i>	0.928	0.929	0.735	0.763	0.863	0.887	0.888	0.867	0.889	0.988
	<i>FDR</i>	0.081	0.081	0.019	0.023	0.040	0.032	0.032	0.006	0.008	0.046
	<i>WI</i>	0.956	0.956	0.312	0.548	0.963	1.000	1.000	0.889	0.921	0.999
	<i>HI</i>	0.037	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	$M(\hat{q})$	0.338	0.344	0.304	0.330	0.409	0.769	0.771	0.648	0.686	0.858
	<i>CP</i>	0.630	0.632	0.590	0.631	0.784	0.835	0.836	0.717	0.756	0.919
	<i>FDR</i>	0.037	0.041	0.016	0.023	0.030	0.019	0.020	0.004	0.007	0.032
	<i>WI</i>	0.229	0.208	0.084	0.167	0.722	0.897	0.897	0.282	0.385	0.842
	<i>HI</i>	0.048	0.066	0.016	0.034	0.030	0.000	0.000	0.000	0.000	0.000
12	$M(\hat{q})$	0.357	0.375	0.201	0.269	0.199	0.801	0.808	0.686	0.784	0.830
	<i>CP</i>	0.648	0.670	0.386	0.497	0.389	0.863	0.868	0.755	0.856	0.891
	<i>FDR</i>	0.071	0.079	0.026	0.047	0.019	0.026	0.028	0.008	0.015	0.029
	<i>WI</i>	0.389	0.415	0.043	0.184	0.046	0.974	0.980	0.396	0.707	0.799
	<i>HI</i>	0.016	0.040	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000
13	$M(\hat{q})$	0.281	0.293	0.178	0.206	0.110	0.575	0.602	0.457	0.500	0.343
	<i>CP</i>	0.506	0.523	0.338	0.387	0.215	0.627	0.655	0.502	0.548	0.379
	<i>FDR</i>	0.094	0.101	0.044	0.055	0.018	0.018	0.021	0.011	0.013	0.005
	<i>WI</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	<i>HI</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	$M(\hat{q})$	0.108	0.109	0.008	0.015	0.003	0.226	0.245	0.112	0.133	0.014
	<i>CP</i>	0.183	0.185	0.016	0.027	0.006	0.244	0.264	0.122	0.145	0.016
	<i>FDR</i>	0.129	0.130	0.017	0.032	0.006	0.027	0.029	0.016	0.018	0.002
	<i>WI</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	<i>HI</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	$M(\hat{q})$	0.102	0.111	0.055	0.068	0.034	0.281	0.312	0.194	0.240	0.153
	<i>CP</i>	0.185	0.196	0.104	0.124	0.066	0.309	0.341	0.214	0.265	0.170
	<i>FDR</i>	0.044	0.047	0.021	0.030	0.007	0.010	0.011	0.005	0.006	0.003
	<i>WI</i>	0.007	0.013	0.002	0.004	0.000	0.016	0.038	0.002	0.012	0.004
	<i>HI</i>	0.002	0.005	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
16	$M(\hat{q})$	0.100	0.106	0.055	0.066	0.047	0.252	0.275	0.172	0.204	0.141
	<i>CP</i>	0.181	0.189	0.104	0.120	0.092	0.277	0.301	0.190	0.225	0.156
	<i>FDR</i>	0.046	0.048	0.027	0.032	0.016	0.010	0.011	0.004	0.006	0.003
	<i>WI</i>	0.006	0.008	0.001	0.003	0.000	0.008	0.017	0.000	0.003	0.000
	<i>HI</i>	0.001	0.004	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000

Table 5: Simulation results for large  $N$  panels, part 2

Country	$\theta_i$	<i>UR</i>	<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>RW</i>	<i>RW</i> <sub>2</sub>	<i>MP</i>
Argentina	-1.597	*	*	*		*	*
Australia	-0.949						
Belgium	-1.255	*	*	*			*
Brazil	-1.389	*	*	*			*
Canada	-0.922						
Denmark	-0.627						
Finland	-1.807	*	*	*	*	*	*
France	-1.336	*	*	*			*
Germany	-1.308	*	*	*			*
Italy	-1.421	*	*	*			*
Japan	-1.006						
Mexico	-1.419	*	*	*			*
Netherlands	-0.790						
Norway	-0.772						
Portugal	-0.807						
Spain	-0.856						
Sweden	-1.127	*					
Switzerland	-0.900						
UK	-1.096						
Total		9	8	8	1	2	8

Table 6: Tests for PPP on real exchange rates; ‘\*’denotes a rejection of the unit root hypothesis at a 5% level.

	<i>BSQT</i>	<i>BSQT</i> <sub>2</sub>	<i>IBSQT</i>	<i>IBSQT</i> <sub>2</sub>	<i>MP</i>
Total rejections ( $N = 181$ )	60	91	30	60	7
Proportion of rejections	0.33	0.50	0.17	0.33	0.04
Rejections in subsamples					
CLG ( $N = 58$ )	16	25	7	16	2
Proportion	0.28	0.43	0.12	0.28	0.03
HSG ( $N = 87$ )	29	45	34	16	5
Proportion	0.33	0.52	0.18	0.33	0.06
HSD ( $N = 36$ )	15	21	7	15	0
Proportion	0.42	0.58	0.19	0.42	0.00

Table 7: Unit root tests on log real earnings of households in PSID data at 5% level