Bootstrap Sequential Tests to Determine the Stationary Units in a Panel

Stephan Smeekes *

Department of Quantitative Economics
Maastricht University
February 26, 2010

Abstract

We propose sequential tests using the bootstrap to investigate the stationarity properties of individual units in a panel. We first construct asymptotically valid methods based on sequentially testing for unit roots unit-by-unit, suitable for panels with a small cross-sectional dimension N. These methods are closely related to bootstrap methods that control size in multiple testing. We modify these unit by unit methods to be applicable in panels with large N; for this purpose user-defined quantiles instead of individual units are tested sequentially. These quantile-based methods tests have a significant advantage over multiple testing approaches if N is rather large but the time series dimension T is small, as they can fully utilize the cross-sectional dimension, which the multiple testing approaches cannot. A simulation study is conducted to analyze the relative performance of the methods in comparison with multiple testing approaches. The tests are 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 methods, in particular in panels with large N and small T.

Keywords: Sequential testing; unit root; 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

*Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: S.Smeekes@maastrichtuniversity.nl. This research was financially supported by the Netherlands Organisation for Scientific Research (NWO). I would like to thank Franz Palm and Jean-Pierre Urbain for helpful comments and suggestions, and Christoph Hanck for sharing his data. The usual disclaimer applies.

unit root. The alternative hypothesis differs; some tests have the alternative hypothesis that at least some series are stationary. A rejection for such a test is hard to interpret; it could 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 tell them apart. 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; 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 trend coefficient in cross-sectional averages of variances. This method delivers for an estimation of the fraction of (non)stationary units, but cannot test which units are stationary. Hanck (2009) and Moon and Perron (2009) 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 (2009) on the other hand aim to control the false discovery rate (FDR), an approach that is better suited to samples with larger N. Moon and Perron (2009) consider both asymptotic methods and the bootstrap method of Romano, Shaikh, and Wolf (2008a) to control the FDR, and find in general that the bootstrap method works best.

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. We first propose a "unit-by-unit" version that is suitable for a relatively small N. We will show that this method is closely related to the approach of Romano and Wolf (2005) to control FWE; as a side-product of this analysis we propose a modification of their method that can be more powerful. Next we propose an extension of the unit-by-unit test that is suitable for 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 utilize the cross-sectional dimensional to increase power.

The structure of the paper is as follows. In Section 2 the DGP is introduced. The unit-by-unit sequential tests are constructed in Section 3, and compared to the FWE controlling methods in multiple testing. The sequential tests are extended to quantile-based tests in

Section 4. 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 $[x]$. Convergence in distribution (probability) is denoted by $\xrightarrow{d}$ ($\xrightarrow{p}$). Bootstrap quantities (conditional on the original sample) are indicated by appending a superscript $*$ to the standard notation. Convergence in distribution (probability) of bootstrap statistics is denoted $\xrightarrow{d^*}$ ($\xrightarrow{p^*}$), where this convergence is taken to take place in probability. $W(r)$ denotes a univariate standard Brownian motion. $|x|$ applied to a complex number x denotes its absolute value, while $|\mathscr{G}|$ applied to a set $\mathscr{G}$ denotes the cardinality of the set. $\mathscr{G}^c$ denotes the compliment of the set $\mathscr{G}$ taken with respect to the set $\mathscr{N}_N = \{i \in \mathbb{N} : i \leq N\}$, i.e. $\mathscr{G}^c \cup \mathscr{G} = \mathscr{N}_N$ while $\mathscr{G}^c \cap \mathscr{G} = \emptyset$.

2 The Model

Suppose we have a panel of observations $y_{i,t}$ on $i = 1, \ldots, N$ and $t = 1, \ldots, T$, with $y_t = (y_{1,t}, \ldots, y_{N,t})'$ generated as

$$y_t = \Lambda F_t + w_t, \qquad y_0 = 0, \tag{1}$$

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

We let the factors and the idiosyncratic components be generated by

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

where $\Phi = diag(\phi_1, \ldots, \phi_d)$ and $\Gamma = diag(\gamma_1, \ldots, \gamma_N)$.

Furthermore we let f_t and v_t be constructed as

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

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

Assumption 1.

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

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

This is the same DGP as used by Palm, Smeekes, and Urbain (2008). Now define ρ_i as

$$\rho_i = \lim_{t\to\infty} \frac{\mathrm{E}(y_{i,t-1}y_{i,t})}{\mathrm{E}(y_{i,t-1}^2)}. \tag{4}$$

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 the two components. That is, $\rho_i = 1$ if $\gamma_i = 1$ or if $\phi_j = 1$ and $\lambda_{i,j} \neq 0$ for some $j = 1, \ldots, d$, where $\lambda_{i,j}$ is the j-th element of λ_i.

Let $k_0 = 0, 1, \ldots, N$ be the number of stationary units. Formally, define $\mathscr{S} = \{i \in \mathscr{N}_N : |\rho_i| < 1\}$ and $\mathscr{U} = \{i \in \mathscr{N}_N : \rho_i = 1\} = \mathscr{S}^c$. Then $k_0 = |\mathscr{S}|$. Furthermore, let $q_0 = k_0/N$, 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. (2008). 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 unit-by-unit

Initially, we focus on sequential methods to estimate the number of stationary units, that is to estimate $k_0 = |\mathscr{S}|$. This will later be trivially extended to determine which units are stationary. We set up the null and alternative hypotheses to be used in the sequential procedure in terms of the number of stationary units. For that purpose let $H_0(k)$ denote the null hypothesis that k out of N units are stationary, or in other words $H_0(k) : |\mathscr{S}| = k$. Let $H_1(k+1)$ denote the alternative hypothesis that at least $k+1$ units are stationary, i.e. $|\mathscr{S}| \geq k+1$.

Let $\tau_{T,k}$ be a test statistic to test $H_0(k)$ vs. $H_1 : (k+1)$, which rejects $H_0(k)$ if $\tau_{T,k} < c_{\alpha,k}$, and satisfies the conditions

$$\lim_{T\to\infty} \mathrm{P}(\tau_{T,k} < c_{\alpha,k}) = \alpha \qquad \text{for } k = k_0, \tag{5a}$$

$$\lim_{T\to\infty} \mathrm{P}(\tau_{T,k} < c_{\alpha,k}) = 1 \qquad \text{for } k < k_0. \tag{5b}$$

These conditions simply state that if $H_0(k)$ is true, the test should have asymptotic size α; while if $H_0(k)$ is not true, the test should reject with probability 1 (i.e. the test should be consistent). The sequential testing procedure that we consider for determining the members of $|\mathscr{S}|$ and correspondingly k_0, can now be described as below.

Sequential Test 1 (Sequential Unit-by-unit Test)**.**

1. Test $H_0(0)$ versus $H_1(1)$. Calculate $\tau_{T,0}$ and reject $H_0(0)$ if $\tau_{T,0} < c_{\alpha,0}$.
2. If $H_0(0)$ is not rejected, set $\hat{k} = 0$. If $H_0(1)$ is rejected, test $H_0(1)$ versus $H_1(2)$.
3. Keep testing until the null hypothesis $H_0(k^*)$ cannot be rejected. In that case, set $\hat{k} = k^*$. If all null hypotheses up to $H_0(N-1)$ can be rejected, set $\hat{k} = N$.

Remark 2: Kapetanios (2003) considered sequential tests for the number of stationary units in a panel based on the same principle. However, as we will see later, our implementation differs significantly, in particular through the choice of $\tau_{T,k}$.

As mentioned before, condition (5) is a standard condition that any test has to satisfy. In our sequential framework, this condition implies the following result.[1]

Proposition 1. *Let $\tau_{T,k}$ satisfy condition 5. Then*

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{P}(\hat{k} = k) = 0 \qquad \textit{for } k < k_0, \\
&\lim_{T\to\infty} \mathrm{P}(\hat{k} = k) = 1 - \alpha \qquad \textit{for } k = k_0, \\
&\limsup_{T\to\infty} \mathrm{P}(\hat{k} = k) \leq \alpha \qquad \textit{for } k > k_0.
\end{aligned}$$

3.1 Bootstrap $\tau_{T,k}$ tests

We now focus on how to construct the test statistic $\tau_{T,k}$. In order to make $\tau_{T,k}$ satisfy (5), we use the bootstrap. Let $\theta_{T,i}$ be any unit root test statistic applied to unit i that satisfies the following assumption.

Assumption 2. Let $\theta_{T,i}$ be any unit root test applied to unit $i = 1, \ldots, N$ that rejects the null hypothesis of a unit root if $\theta_{T,i} < c$ for some c. Furthermore assume the following.

(i) G_i is the asymptotic null distribution of $\theta_{T,i}$, i.e. if $\rho_i = 1$, then

$$\theta_{T,i} \xrightarrow{d} G_i, \qquad \text{as } T \to \infty. \tag{6}$$

(ii) $\theta_{T,i}$ is consistent, i.e. if $\rho_i < 1$, then

$$\theta_{T,k} \xrightarrow{p} -\infty, \qquad \text{as } T \to \infty. \tag{7}$$

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

$$\theta_{T,(1)} \leq \ldots \leq \theta_{T,(N)}.$$

[1] This result is not specific to sequential testing in panels, but holds for sequential testing in general. See Swensen (2006, Corollary 1) for a similar result in the context of sequential testing for cointegration rank.

Then we take

$$\tau_{T,k} = \theta_{T,(k+1)}.$$

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 $\theta_{T,1}, \dots, \theta_{T,N}$ will not be 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 two variants of our sequential tests, which we call the Bootstrap Sequential Unit-by-unit Test ($BSUT$). The first version, $BSUT_1$, sequentially removes the units found stationary to obtain critical values. We now not only use this method to estimate k_0, but also to determine the members of $\mathscr{S}$.

Bootstrap Algorithm 1 ($BSUT_1$)**.** To test $H_0(k)$ vs. $H_1(k+1)$ in Sequential Test 1, let $\tau_{T,k} = \theta_{T,(k+1)}$ and let

$$\mathbb{S}_k = \{i : \theta_{T,i} \leq \theta_{T,(k)}\}. \tag{8}$$

Then perform the following bootstrap steps.

1. For each unit estimate

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

and calculate

$$\hat{u}_{i,t} = y_{i,t} - \hat{\rho}_i y_{i,t-1} - \frac{1}{T-1} \sum_{t=2}^{T} (y_{i,t} - \hat{\rho}_i y_{i,t-1}). \tag{10}$$

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

2. 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.

3. Construct the bootstrap errors $u_2^*, \dots, u_T^*$ as follows.

$$u_t^* = \hat{u}_{i_m + s}, \tag{11}$$

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

4. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*, \tag{12}$$

where

$$\rho_i^* = 1 \quad \text{for all } i = 1, \ldots, N. \tag{13}$$

5. Obtain θ_i^* for all $i \in \mathbb{S}_k^c$, and let $\tau_{T,k}^* = \min_{i \in \mathbb{S}_k^c} \theta_{T,i}^*$.
6. Repeat Steps 2 to 5 B times, obtaining bootstrap test statistics $\tau_{T,k}^{*b}, b = 1, \ldots, B$, and select the bootstrap critical value $c_{\alpha,k}^*$ as $c_{\alpha,k}^* = \max\{c : B^{-1} \sum_{b=1}^{B} I(\tau_{T,k}^{*b} < c) \leq \alpha\}$, or equivalently as the α-quantile of the ordered $\tau_{T,k}^{*b}$ statistics.

The second method, $BSUT_2$, is very similar to the first, but may be somewhat more powerful. 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 of the stationary units is not discarded unlike in the first method. Asymptotically the stationary units do not influence the distribution of the higher order statistics, but in finite samples there will be an impact. $BSUT_2$ tries to mimic this impact, while $BSUT_1$ ignores it.

Bootstrap Algorithm 2 ($BSUT_2$)**.** To test $H_0(k)$ vs. $H_1(k+1)$ in Sequential Test 1, let $\tau_{T,k} = \theta_{T,(k+1)}$ and let

$$\mathbb{S}_k = \{i : \theta_{T,i} \leq \theta_{T,(k)}\}. \tag{14}$$

Perform bootstrap steps 1 to 3 of Algorithm 1 to obtain bootstrap errors u_t^*.

4. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*, \tag{15}$$

where

$$\rho_i^* = \begin{cases} \hat{\rho}_i & \text{if } i \in \mathbb{S}_k \\ 1 & \text{if } i \in \mathbb{S}_k^c \end{cases}. \tag{16}$$

5. Obtain θ_i^* for all units, and let $\tau_{T,k}^* = \theta_{T,(k+1)}^*$.
6. Repeat Steps 2 to 5 B times, obtaining bootstrap test statistics $\tau_{T,k}^{*b}, b = 1, \ldots, B$, and select the bootstrap critical value $c_{\alpha,k}^*$ as $c_{\alpha,k}^* = \max\{c : B^{-1} \sum_{b=1}^{B} I(\tau_{T,k}^{*b} < c) \leq \alpha\}$, or equivalently as the α-quantile of the ordered $\tau_{T,k}^{*b}$ statistics.

It now follows directly from $BSUT_1$ and $BSUT_2$ and the sequential nature of the procedure, that the set of units deemed stationary is simply equal to $\mathbb{S}_{\hat{k}}$ as defined in (8).

We also need the following assumption on the block length.

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

While the statistics $\theta_{T,i}$ could be any unit root test statistics, we will from now on assume $\theta_{T,i}$ is chosen as one of the Dickey-Fuller (DF) statistics. We allow for a prior detrending procedure that could either be standard OLS or GLS (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 or not (also see Remark 3 below).

Theorem 1.

(i) Let $y_{i,t}$ be generated by (1)-(3) and let Assumption 1 hold. Let $\theta_{T,1}, \ldots, \theta_{T,N}$ satisfy Assumption 2 and let $\tau_{T,k}$ be defined as above. Then, if $|\mathscr{S}| = k_0$, we have that

$$\tau_{T,k} \xrightarrow{p} -\infty \qquad \text{for } k < k_0, \tag{17a}$$

$$\tau_{T,k} \xrightarrow{d} \min_{i \in \mathscr{U}} G_i \qquad \text{for } k = k_0. \tag{17b}$$

(ii) In addition to the conditions used in part (i), let $\theta^*_{T,i}$ and $\tau^*_{T,k}$ be defined as in $BSUT_1$ or $BSUT_2$ and let Assumption 3 hold. Then, for any $k = 0, 1, \ldots, N-1$,

$$\theta^*_{T,i} \xrightarrow{d^*} G_i \text{ in probability} \qquad \text{for } i \in \mathbb{S}^c_k, \tag{18}$$

and

$$\tau^*_{T,k} \xrightarrow{d^*} \min_{i \in \mathbb{S}^c_k} G_i \text{ in probability} \qquad \text{for } k < k_0, \tag{19a}$$

$$\tau^*_{T,k} \xrightarrow{d^*} \min_{i \in \mathscr{U}} G_i \text{ in probability} \qquad \text{for } k = k_0. \tag{19b}$$

It now follows directly from (17a) and (19a) that (5b) holds, while (17b) and (17a) show that (5a) hold.

Remark 3: Although any unit root test can be used for $\theta_{T,i}$, 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. 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).

Remark 4: Kapetanios (2003) used the IPS group-mean test of Im, Pesaran, and Shin (2003) as $\tau_{T,k}$, 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 our test based on order statistics, in particular if only a few units are stationary.

3.2 $BSUT$ tests 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 Romano, Shaikh, and Wolf (2008b). In a panel context these methods have been used by Hanck (2009), Deckers and Hanck (2009) and Moon and Perron (2009) 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 method $BSUT_1$, 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.

1. Test $H_0(0)$ against $H_1(1)$. That is, obtain $\tau_{T,0}$. Reject $H_0(0)$ if $\tau_{T,0} < c_{\alpha,0}$.

2. If $H_0(0)$ is not rejected, set $\hat{k} = 0$. If $H_0(1)$ is rejected, test $H_0(1)$ against $H_1(2)$, still using the critical value $c_{\alpha,0}$. That is, reject $H_0(1)$ if $\tau_{T,1} < c_{\alpha,0}$.

3. Keep testing using critical value $c_{\alpha,0}$ until the null hypothesis $H_0(k_1)$ cannot be rejected. If $0 < k_1 < N$, continue to the second step: Test $H_0(k_1)$ against $H_1(k_1+1)$ by rejecting $H_0(k_1)$ if $\tau_{T,k_1} < c_{\alpha,k_1}$. If $H_0(k_1)$ cannot be rejected, set $\hat{k} = k_1$. If $H_0(k_1)$ is rejected, test $H_0(k_1 + 1)$ against $H_0(k_1 + 2)$ using critical value c_{α,k_1}.

4. Keep testing using critical value c_{α,k_0} until the null hypothesis $H_0(k_2)$ cannot be rejected. If $k_1 < k_2 < N$, continue to the next step, and so on. If all null hypotheses up to $H_0(N - 1)$ can be rejected, set $\hat{k} = N$.

The critical values $c_{\alpha,k}$ are obtained in the same way as in $BSUT_1$. Very informally one could see the RW method as a "shortcut" to our $BSUT_1$ method: instead of calculating critical values at every step, first the critical value from the previous step is used, given that $c_{\alpha,k} \leq c_{\alpha,k+1}$ for all k.

Given the similarity between the two methods, $BSUT_2$ immediately lends itself to construct an extension of RW: in analogy to the difference between $BSUT_1$ and $BSUT_2$, we could change how the critical values $c_{\alpha,k}$ in the RW procedure are found; instead of deleting the units deemed stationary, they can be incorporated them into the bootstrap as stationary units. We will denote this modified RW procedure by RW_2. As for the $BSUT$ methods, we might expect the RW_2 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 (2009) 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 $N \to \infty$. 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 (2009) 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.

3.3 Simulation study in small N panels

We now perform a simulations study in panels with small N. We compare the $BSUT_1$, $BSUT_2$, RW and RW_2 procedures. We also add the bootstrap approach of Moon and Perron (2009) which is denoted by MP. For all methods we apply the block bootstrap based on residuals as described above. Note that Moon and Perron (2009) 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 (2009), lag lengths were selected by MAIC (Ng and Perron, 2001) with a maximum of 4 lags.

We use the following DGP in the simulation study.

$$\begin{aligned} y_{i,t} &= \mu_i + x_{i,t} \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \\ x_{i,t} &= \rho_{i,T} x_{i,t-1} + w_{i,t}, \quad x_0 = 0, \end{aligned} \tag{20}$$

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

$$w_{i,t} = \lambda_i f_t + u_{i,t}, \tag{21}$$

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

For the common factor we let

$$f_t = 0.5 f_{t-1} + \nu_{i,t}, \tag{22}$$

where $\nu_{i,t} \sim i.i.d.N(0, 1)$. For the factor loadings, λ_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 $u_{i,t}$ are modeled as an $ARMA(1,1)$ process,

$$u_{i,t} = \phi_i x_{i,t-1} + \varepsilon_{i,t} + \theta_i \varepsilon_{i,t-1}. \tag{23}$$

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 $N = 10$ as this is a small panel, while we take $T = 50$ and $T = 100$. Table 1 panel A gives all the combinations of parameters we use. For k_0, the number of stationary units in the panel, we take $k_0 = 0, 2, 5, 9$ for all parameter combinations, where for $i \leq k_0$ we take $\rho_{i,T} \sim U[0, 1 - cT^{-1}]$ with $c = 10$, and for $i > k_0$ we take $\rho_{i,T} = 1$. The level of all tests (or FWE/FDR when appropriate) is taken to be 5%.

The simulation results are given in Tables 2 to 5. We report the average $\hat{k}$, denoted by $M(\hat{k})$, the standard deviation of $\hat{k}$, denoted by $S(\hat{k})$, the average fraction 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.

The four tests that asymptotically control FWE can be seen to control FWE in finite samples as well; FWE is always close to α across all models. While the FWE 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. Whether there is cross-sectional dependence does not seem to influence the results much.

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 FDR in such a small panel or wishes to control FWE.

The $BSUT_1$, $BSUT_2$, 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 $BSUT_1$ 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 $BSUT$ 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 lies not in small N panels,

DGP	T	N	λ_i	ϕ_i	θ_i
Panel A: Small N simulations					
1	50	10	0	0	0
2	50	10	0	U[-0.5,0.5]	0
3	50	10	0	0	U[-0.5,0.5]
4	50	10	0	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]	0
7	100	10	0	0	U[-0.5,0.5]
8	100	10	0	U[-0.5,0.5]	U[-0.5,0.5]
9	50	10	U[-1,3]	0	0
10	50	10	U[-1,3]	U[-0.5,0.5]	0
11	50	10	U[-1,3]	0	U[-0.5,0.5]
12	50	10	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
13	100	10	U[-1,3]	0	0
14	100	10	U[-1,3]	U[-0.5,0.5]	0
15	100	10	U[-1,3]	0	U[-0.5,0.5]
16	100	10	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
Panel B: Large N simulations					
1	100	50	0	0	0
2	100	50	0	U[-0.5,0.5]	0
3	100	50	0	0	U[-0.5,0.5]
4	100	50	0	U[-0.5,0.5]	U[-0.5,0.5]
5	25	200	0	0	0
6	25	200	0	U[-0.5,0.5]	0
7	25	200	0	0	U[-0.5,0.5]
8	25	200	0	U[-0.5,0.5]	U[-0.5,0.5]
9	100	10	U[-1,3]	0	0
10	100	10	U[-1,3]	U[-0.5,0.5]	0
11	100	10	U[-1,3]	0	U[-0.5,0.5]
12	100	10	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]
13	25	200	U[-1,3]	0	0
14	25	200	U[-1,3]	U[-0.5,0.5]	0
15	25	200	U[-1,3]	0	U[-0.5,0.5]
16	25	200	U[-1,3]	U[-0.5,0.5]	U[-0.5,0.5]

Table 1: Parameter combinations simulation DGPs

DGP		$k_0 = 0$					$k_0 = 2$				
		$BSUT_1$	$BSUT_2$	RW	RW_2	MP	$BSUT_1$	$BSUT_2$	RW	RW_2	MP
1	$M(\hat{k})$	0.039	0.039	0.036	0.039	0.038	0.292	0.298	0.291	0.294	0.322
	$S(\hat{k})$	0.209	0.209	0.197	0.213	0.211	0.518	0.525	0.502	0.527	0.600
	CP	0.000	0.000	0.000	0.000	0.000	0.122	0.124	0.121	0.122	0.130
	FWE	0.036	0.036	0.034	0.035	0.034	0.048	0.049	0.048	0.049	0.058
	FDR	0.036	0.036	0.034	0.035	0.034	0.041	0.041	0.043	0.040	0.043
2	$M(\hat{k})$	0.044	0.049	0.045	0.044	0.048	0.621	0.639	0.630	0.642	0.697
	$S(\hat{k})$	0.215	0.225	0.217	0.219	0.236	0.603	0.625	0.603	0.631	0.727
	CP	0.000	0.000	0.000	0.000	0.000	0.290	0.297	0.295	0.297	0.311
	FWE	0.042	0.047	0.043	0.041	0.043	0.037	0.038	0.037	0.041	0.067
	FDR	0.042	0.047	0.043	0.041	0.043	0.024	0.023	0.024	0.025	0.036
3	$M(\hat{k})$	0.037	0.037	0.038	0.037	0.037	0.359	0.376	0.367	0.378	0.417
	$S(\hat{k})$	0.194	0.199	0.201	0.199	0.209	0.553	0.597	0.566	0.596	0.682
	CP	0.000	0.000	0.000	0.000	0.000	0.163	0.169	0.167	0.168	0.185
	FWE	0.036	0.035	0.036	0.035	0.033	0.034	0.038	0.034	0.041	0.042
	FDR	0.036	0.035	0.036	0.035	0.033	0.030	0.028	0.028	0.031	0.028
4	$M(\hat{k})$	0.054	0.055	0.057	0.056	0.056	0.341	0.355	0.343	0.344	0.384
	$S(\hat{k})$	0.226	0.228	0.232	0.230	0.234	0.528	0.552	0.527	0.534	0.625
	CP	0.000	0.000	0.000	0.000	0.000	0.145	0.147	0.145	0.146	0.154
	FWE	0.054	0.055	0.057	0.056	0.055	0.049	0.059	0.050	0.051	0.069
	FDR	0.054	0.055	0.057	0.056	0.055	0.039	0.045	0.040	0.039	0.046
5	$M(\hat{k})$	0.038	0.042	0.037	0.044	0.042	1.732	1.740	1.732	1.743	1.931
	$S(\hat{k})$	0.191	0.206	0.189	0.210	0.206	0.568	0.573	0.566	0.567	0.631
	CP	0.000	0.000	0.000	0.000	0.000	0.844	0.846	0.845	0.849	0.900
	FWE	0.038	0.041	0.037	0.043	0.041	0.043	0.047	0.042	0.045	0.113
	FDR	0.038	0.041	0.037	0.043	0.041	0.016	0.017	0.016	0.017	0.042
6	$M(\hat{k})$	0.044	0.048	0.047	0.047	0.050	0.816	0.819	0.821	0.819	0.911
	$S(\hat{k})$	0.205	0.218	0.216	0.225	0.231	0.539	0.559	0.532	0.546	0.667
	CP	0.000	0.000	0.000	0.000	0.000	0.390	0.389	0.391	0.392	0.417
	FWE	0.044	0.047	0.046	0.044	0.047	0.035	0.040	0.040	0.034	0.071
	FDR	0.044	0.047	0.046	0.044	0.047	0.020	0.021	0.024	0.018	0.035
7	$M(\hat{k})$	0.056	0.058	0.054	0.051	0.054	1.727	1.737	1.728	1.736	1.946
	$S(\hat{k})$	0.238	0.246	0.235	0.233	0.239	0.625	0.629	0.631	0.651	0.741
	CP	0.000	0.000	0.000	0.000	0.000	0.823	0.827	0.823	0.823	0.876
	FWE	0.054	0.055	0.052	0.048	0.051	0.076	0.078	0.078	0.085	0.167
	FDR	0.054	0.055	0.052	0.048	0.051	0.030	0.031	0.031	0.033	0.062
8	$M(\hat{k})$	0.058	0.061	0.059	0.060	0.056	1.722	1.734	1.724	1.741	1.936
	$S(\hat{k})$	0.234	0.243	0.236	0.242	0.234	0.599	0.614	0.588	0.615	0.742
	CP	0.000	0.000	0.000	0.000	0.000	0.829	0.829	0.830	0.833	0.875
	FWE	0.058	0.060	0.059	0.059	0.055	0.063	0.073	0.063	0.071	0.149
	FDR	0.058	0.060	0.059	0.059	0.055	0.024	0.028	0.025	0.027	0.058

Table 2: Simulation results small N; cross-sectional independence; $k_0 = 0, 2$

DGP		$k_0 = 5$					$k_0 = 9$				
		$BSUT_1$	$BSUT_2$	RW	RW_2	MP	$BSUT_1$	$BSUT_2$	RW	RW_2	MP
1	$M(\hat{k})$	2.785	2.877	2.796	2.856	3.503	4.109	4.489	4.135	4.499	6.757
	$S(\hat{k})$	1.170	1.197	1.168	1.187	1.405	1.778	1.942	1.748	1.919	2.435
	CP	0.549	0.566	0.551	0.562	0.667	0.456	0.497	0.459	0.498	0.728
	FWE	0.040	0.046	0.038	0.043	0.151	0.006	0.018	0.006	0.018	0.201
	FDR	0.012	0.014	0.012	0.013	0.036	0.001	0.003	0.001	0.002	0.022
2	$M(\hat{k})$	2.019	2.065	2.014	2.049	2.548	5.042	5.288	5.035	5.326	7.604
	$S(\hat{k})$	1.006	1.050	1.006	1.042	1.364	1.731	1.787	1.705	1.773	2.069
	CP	0.400	0.408	0.399	0.405	0.489	0.559	0.586	0.558	0.590	0.816
	FWE	0.020	0.024	0.019	0.025	0.095	0.010	0.014	0.011	0.017	0.258
	FDR	0.007	0.008	0.007	0.008	0.024	0.001	0.002	0.002	0.002	0.027
3	$M(\hat{k})$	1.864	1.966	1.887	1.949	2.531	5.235	5.425	5.249	5.415	7.531
	$S(\hat{k})$	1.117	1.217	1.110	1.221	1.501	1.514	1.577	1.521	1.564	1.907
	CP	0.367	0.385	0.372	0.382	0.488	0.581	0.602	0.582	0.601	0.816
	FWE	0.029	0.037	0.027	0.036	0.079	0.006	0.010	0.007	0.009	0.187
	FDR	0.011	0.011	0.010	0.011	0.020	0.001	0.001	0.001	0.001	0.019
4	$M(\hat{k})$	2.529	2.582	2.510	2.585	3.286	4.590	4.765	4.581	4.779	7.085
	$S(\hat{k})$	1.060	1.110	1.055	1.109	1.437	1.518	1.589	1.500	1.606	2.156
	CP	0.492	0.501	0.491	0.503	0.614	0.509	0.529	0.508	0.530	0.767
	FWE	0.065	0.073	0.056	0.069	0.186	0.006	0.008	0.005	0.006	0.186
	FDR	0.019	0.020	0.016	0.019	0.045	0.001	0.001	0.001	0.001	0.019
5	$M(\hat{k})$	3.806	3.838	3.815	3.845	4.635	7.877	7.984	7.885	7.996	9.388
	$S(\hat{k})$	0.779	0.785	0.754	0.771	0.833	1.123	1.100	1.140	1.087	0.667
	CP	0.754	0.759	0.755	0.761	0.880	0.871	0.882	0.872	0.884	0.990
	FWE	0.037	0.042	0.040	0.041	0.191	0.035	0.042	0.040	0.044	0.482
	FDR	0.008	0.009	0.009	0.009	0.039	0.004	0.004	0.004	0.005	0.048
6	$M(\hat{k})$	4.651	4.674	4.642	4.672	5.263	6.496	6.602	6.485	6.578	8.698
	$S(\hat{k})$	0.723	0.694	0.722	0.721	0.677	1.147	1.150	1.114	1.188	1.142
	CP	0.919	0.924	0.918	0.922	0.992	0.720	0.732	0.719	0.729	0.929
	FWE	0.052	0.052	0.049	0.056	0.235	0.013	0.015	0.012	0.015	0.333
	FDR	0.009	0.009	0.009	0.010	0.046	0.002	0.002	0.002	0.002	0.033
7	$M(\hat{k})$	2.456	2.565	2.453	2.564	3.536	5.844	5.928	5.837	5.923	8.375
	$S(\hat{k})$	0.944	1.015	0.943	1.037	1.356	1.093	1.148	1.086	1.139	1.421
	CP	0.484	0.504	0.484	0.505	0.666	0.647	0.656	0.646	0.655	0.894
	FWE	0.032	0.040	0.032	0.038	0.172	0.020	0.024	0.021	0.026	0.331
	FDR	0.010	0.010	0.010	0.010	0.039	0.003	0.003	0.003	0.003	0.034
8	$M(\hat{k})$	4.363	4.395	4.346	4.401	5.165	8.537	8.599	8.534	8.606	9.504
	$S(\hat{k})$	0.760	0.791	0.755	0.780	0.799	0.847	0.798	0.865	0.802	0.527
	CP	0.861	0.866	0.858	0.867	0.967	0.942	0.949	0.942	0.949	0.998
	FWE	0.056	0.062	0.053	0.065	0.260	0.061	0.059	0.059	0.063	0.518
	FDR	0.011	0.011	0.010	0.012	0.052	0.006	0.006	0.006	0.006	0.052

Table 3: Simulation results small N; cross-sectional independence; $k_0 = 5, 9$

		$k_0=0$					$k_0=2$				
DGP		$BSUT_1$	$BSUT_2$	RW	RW_2	MP	$BSUT_1$	$BSUT_2$	RW	RW_2	MP
9	$M(\hat{k})$	0.028	0.032	0.030	0.032	0.038	0.713	0.716	0.706	0.714	0.768
	$S(\hat{k})$	0.177	0.202	0.176	0.197	0.234	0.594	0.613	0.595	0.593	0.713
	CP	0.000	0.000	0.000	0.000	0.000	0.335	0.334	0.331	0.336	0.346
	FWE	0.026	0.028	0.029	0.028	0.030	0.041	0.043	0.042	0.039	0.063
	FDR	0.026	0.028	0.029	0.028	0.030	0.025	0.026	0.027	0.024	0.034
10	$M(\hat{k})$	0.065	0.063	0.064	0.070	0.071	1.268	1.271	1.265	1.268	1.367
	$S(\hat{k})$	0.308	0.321	0.310	0.348	0.358	0.808	0.826	0.814	0.813	0.926
	CP	0.000	0.000	0.000	0.000	0.000	0.621	0.619	0.618	0.619	0.642
	FWE	0.055	0.050	0.053	0.053	0.054	0.024	0.028	0.027	0.027	0.061
	FDR	0.055	0.050	0.053	0.053	0.054	0.011	0.014	0.012	0.013	0.026
11	$M(\hat{k})$	0.057	0.060	0.058	0.057	0.071	0.715	0.719	0.696	0.709	0.809
	$S(\hat{k})$	0.264	0.297	0.284	0.282	0.358	0.651	0.686	0.627	0.667	0.874
	CP	0.000	0.000	0.000	0.000	0.000	0.334	0.331	0.328	0.332	0.348
	FWE	0.050	0.048	0.048	0.047	0.049	0.040	0.044	0.034	0.036	0.072
	FDR	0.050	0.048	0.048	0.047	0.049	0.026	0.027	0.023	0.023	0.041
12	$M(\hat{k})$	0.041	0.037	0.038	0.039	0.046	1.388	1.389	1.385	1.394	1.538
	$S(\hat{k})$	0.198	0.199	0.196	0.204	0.325	0.745	0.763	0.742	0.758	0.892
	CP	0.000	0.000	0.000	0.000	0.000	0.668	0.665	0.667	0.668	0.698
	FWE	0.041	0.035	0.037	0.037	0.036	0.048	0.052	0.048	0.053	0.112
	FDR	0.041	0.035	0.037	0.037	0.036	0.019	0.021	0.020	0.022	0.047
13	$M(\hat{k})$	0.044	0.046	0.046	0.047	0.046	1.340	1.366	1.358	1.370	1.561
	$S(\hat{k})$	0.215	0.228	0.223	0.238	0.236	0.626	0.643	0.628	0.636	0.767
	CP	0.000	0.000	0.000	0.000	0.000	0.648	0.656	0.656	0.659	0.717
	FWE	0.042	0.042	0.043	0.042	0.041	0.044	0.049	0.046	0.048	0.109
	FDR	0.042	0.042	0.043	0.042	0.041	0.019	0.020	0.019	0.020	0.044
14	$M(\hat{k})$	0.033	0.042	0.034	0.033	0.039	1.764	1.774	1.757	1.771	1.950
	$S(\hat{k})$	0.190	0.220	0.192	0.195	0.227	0.599	0.616	0.598	0.604	0.739
	CP	0.000	0.000	0.000	0.000	0.000	0.857	0.859	0.855	0.859	0.895
	FWE	0.031	0.038	0.032	0.030	0.033	0.048	0.052	0.046	0.051	0.120
	FDR	0.031	0.038	0.032	0.030	0.033	0.017	0.019	0.016	0.018	0.046
15	$M(\hat{k})$	0.067	0.071	0.070	0.073	0.101	1.220	1.222	1.212	1.227	1.444
	$S(\hat{k})$	0.361	0.392	0.368	0.389	0.599	0.793	0.803	0.788	0.803	1.094
	CP	0.000	0.000	0.000	0.000	0.000	0.583	0.583	0.581	0.586	0.639
	FWE	0.046	0.046	0.048	0.047	0.047	0.049	0.050	0.045	0.050	0.107
	FDR	0.046	0.046	0.048	0.047	0.047	0.023	0.024	0.022	0.023	0.047
16	$M(\hat{k})$	0.045	0.047	0.043	0.050	0.064	0.988	0.990	0.990	1.005	1.078
	$S(\hat{k})$	0.239	0.251	0.235	0.267	0.417	0.422	0.477	0.419	0.479	0.669
	CP	0.000	0.000	0.000	0.000	0.000	0.465	0.464	0.466	0.469	0.475
	FWE	0.039	0.039	0.037	0.040	0.039	0.050	0.047	0.051	0.052	0.080
	FDR	0.039	0.039	0.037	0.040	0.039	0.027	0.026	0.027	0.028	0.045

Table 4: Simulation results small N; common factor; $k_0 = 0, 2$

		$k_0 = 5$					$k_0 = 9$				
DGP		$BSUT_1$	$BSUT_2$	RW	RW_2	MP	$BSUT_1$	$BSUT_2$	RW	RW_2	MP
9	$M(\hat{k})$	1.936	1.976	1.915	1.981	2.576	2.840	3.031	2.852	3.032	4.866
	$S(\hat{k})$	1.040	1.078	1.051	1.099	1.510	1.577	1.750	1.591	1.765	2.835
	CP	0.380	0.387	0.376	0.387	0.486	0.314	0.336	0.316	0.336	0.528
	FWE	0.033	0.040	0.036	0.043	0.120	0.010	0.011	0.010	0.012	0.113
	FDR	0.012	0.013	0.013	0.014	0.031	0.003	0.003	0.003	0.003	0.014
10	$M(\hat{k})$	2.112	2.189	2.124	2.182	2.789	4.462	4.778	4.415	4.763	6.787
	$S(\hat{k})$	1.364	1.425	1.374	1.418	1.868	2.306	2.425	2.296	2.440	3.011
	CP	0.415	0.430	0.417	0.428	0.521	0.495	0.529	0.489	0.527	0.728
	FWE	0.035	0.039	0.036	0.040	0.123	0.011	0.016	0.013	0.019	0.239
	FDR	0.011	0.011	0.011	0.011	0.034	0.001	0.002	0.002	0.003	0.025
11	$M(\hat{k})$	2.788	2.818	2.799	2.827	3.465	4.903	5.053	4.917	5.066	6.976
	$S(\hat{k})$	1.625	1.642	1.620	1.647	1.971	2.319	2.426	2.325	2.409	3.038
	CP	0.549	0.555	0.552	0.556	0.651	0.542	0.558	0.544	0.560	0.742
	FWE	0.039	0.040	0.038	0.042	0.162	0.025	0.032	0.024	0.030	0.300
	FDR	0.011	0.011	0.011	0.012	0.038	0.003	0.004	0.003	0.004	0.031
12	$M(\hat{k})$	3.057	3.132	3.090	3.142	3.901	5.177	5.374	5.151	5.375	7.353
	$S(\hat{k})$	1.309	1.333	1.286	1.335	1.563	2.112	2.184	2.105	2.191	2.492
	CP	0.603	0.617	0.610	0.618	0.735	0.574	0.596	0.571	0.596	0.790
	FWE	0.042	0.047	0.041	0.050	0.171	0.010	0.012	0.008	0.013	0.247
	FDR	0.010	0.010	0.010	0.011	0.038	0.001	0.002	0.001	0.002	0.025
13	$M(\hat{k})$	4.158	4.219	4.173	4.231	4.978	4.955	5.245	4.957	5.253	8.185
	$S(\hat{k})$	0.971	0.953	0.964	0.961	0.961	1.626	1.798	1.653	1.811	1.845
	CP	0.821	0.832	0.824	0.834	0.937	0.549	0.580	0.549	0.581	0.875
	FWE	0.047	0.051	0.045	0.053	0.212	0.015	0.023	0.016	0.022	0.307
	FDR	0.010	0.011	0.009	0.011	0.046	0.002	0.003	0.003	0.003	0.031
14	$M(\hat{k})$	3.219	3.266	3.206	3.273	4.099	7.538	7.676	7.565	7.666	9.241
	$S(\hat{k})$	1.042	1.033	1.037	1.040	1.205	1.240	1.166	1.241	1.200	0.806
	CP	0.636	0.645	0.633	0.644	0.775	0.834	0.849	0.837	0.848	0.977
	FWE	0.039	0.040	0.039	0.047	0.156	0.028	0.033	0.028	0.034	0.452
	FDR	0.009	0.009	0.009	0.011	0.038	0.003	0.004	0.003	0.004	0.045
15	$M(\hat{k})$	3.803	3.841	3.807	3.865	4.765	7.087	7.205	7.085	7.208	8.969
	$S(\hat{k})$	0.983	0.986	0.993	0.984	1.363	1.535	1.492	1.525	1.488	1.144
	CP	0.748	0.755	0.748	0.759	0.872	0.784	0.796	0.784	0.796	0.952
	FWE	0.053	0.053	0.056	0.055	0.213	0.032	0.037	0.030	0.042	0.404
	FDR	0.012	0.012	0.012	0.012	0.056	0.004	0.004	0.003	0.005	0.040
16	$M(\hat{k})$	2.573	2.635	2.560	2.621	3.509	7.634	7.811	7.643	7.780	9.389
	$S(\hat{k})$	0.929	0.975	0.922	0.966	1.422	1.526	1.435	1.501	1.452	0.764
	CP	0.508	0.518	0.505	0.517	0.660	0.844	0.863	0.846	0.860	0.988
	FWE	0.026	0.035	0.027	0.029	0.132	0.034	0.047	0.030	0.038	0.499
	FDR	0.008	0.010	0.008	0.008	0.035	0.003	0.005	0.003	0.004	0.050

Table 5: Simulation results small N; common factor; $k_0 = 5, 9$

but in large N panels, as the $BSUT$ tests directly allow for an extension to large N that RW does not. This we will discuss next.

4 Sequential testing for unit roots using quantiles

If N is large, the $BSUT$ tests suffer from the curse of dimensionality and will significantly lose power. Similarly, the RW test will not work properly anymore as controlling FWE does not make much sense for large N. The $BSUT$ tests however can easily be adapted to construct methods that can be used if N is large. Instead of testing per unit, user-defined quantiles are tested sequentially.

4.1 Bootstrap sequential testing using quantiles

Let $H_0(q) : N^{-1}|\mathscr{S}| = q$ denote the null hypothesis that a proportion q of the units is stationary, and $H_1(q') : N^1|\mathscr{S}| \geq q'$ the alternative hypothesis that at least a proportion q' of the units is stationary, where $q' > q$. Let $0 = q_1 < \ldots < q_r < 1$ denote a set of r user-defined numbers. We can then outline the general procedure as follows.

Sequential Test 2 (Sequential Quantile Test)**.**

1. Test $H_0(q_1)$ against $H_1(q_2)$ using test statistic $\tau_T(q_1, q_2)$. Reject $H_0(q_1)$ if $\tau_T(q_1, q_2) < c_\alpha(q_1, q_2)$.
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 $H_0(q_i)$ cannot be rejected. In that case, set $\hat{q} = q_i$. If all null hypotheses up to $H_0(q_r)$ are rejected, set $\hat{q} = 1$.

The test statistic used to test $H_0(q_j)$ versus $H_0(q_{j+1})$, denoted by $\tau_T(q_j, q_{j+1})$, is now obtained as

$$\tau_T(q_j, q_{j+1}) = \theta_{T,([q_{j+1}N])} = \theta_{T,(k_{j+1})}, \tag{24}$$

where $k_{j+1} = [q_{j+1}N]$. This choice reflects the different alternative hypothesis compared to the unit-by-unit method. The bootstrap critical values for the quantile tests, which we denote as $BSQT_1$ and $BSQT_2$ (for Bootstrap Sequential Quantile Test) can now be obtained in the same way as for $BSUT_1$ and $BSUT_2$.

Bootstrap Algorithm 3 ($BSQT_1$)**.** To test $H_0(q_j)$ vs. $H_1(q_{j+1})$ in Sequential Test 2, let $\tau_T(q_j, q_{j+1}) = \theta_{T,([q_{j+1}N])} = \theta_{T,(k_{j+1})}$ and let

$$\mathbb{S}_{q_j} = \{i : \theta_{T,i} \leq \theta_{T,(k_j)}\}. \tag{25}$$

Perform bootstrap steps 1 to 3 of Algorithm 1 to obtain bootstrap errors u_t^*.

4. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*, \tag{26}$$

where

$$\rho_i^* = 1 \quad \text{for all } i = 1, \ldots, N. \tag{27}$$

5. Obtain θ_i^* for all $i \in \mathbb{S}_{q_j}^c$, and order the statistics as

$$\theta_{T,(1)}^* \leq \ldots \leq \theta_{T,(N-k_j)},$$

and let $\tau_T^*(q_j, q_{j+1}) = \theta_{T,(k_{j+1}-k_j)}^*$.

6. Repeat Steps 2 to 5 B times, obtaining bootstrap test statistics $\tau_T^{*b}(q_j, q_{j+1})$, for $b = 1, \ldots, B$, and select the bootstrap critical value $c_\alpha^*(q_j, q_{j+1})$ as $c_\alpha^*(q_j, q_{j+1}) = \max\{c : B^{-1}\sum_{b=1}^{B} I(\tau_T^{*b}(q_j, q_{j+1}) < c) \leq \alpha\}$, or equivalently as the α-quantile of the ordered $\tau_T^{*b}(q_j, q_{j+1})$ statistics.

Bootstrap Algorithm 4 ($BSQT_2$)**.** To test $H_0(q_j)$ vs. $H_1(q_{j+1})$ in Sequential Test 2, let $\tau_T(q_j, q_{j+1}) = \theta_{T,([q_{j+1}N])} = \theta_{T,(k_{j+1})}$ and let

$$\mathbb{S}_{q_j} = \{i : \theta_{T,i} \leq \theta_{T,(k_j)}\}. \tag{28}$$

Perform bootstrap steps 1 to 3 of Algorithm 1 to obtain bootstrap errors u_t^*.

4. Let

$$y_{i,t}^* = \rho_i^* y_{i,t-1}^* + u_{i,t}^*, \tag{29}$$

where

$$\rho_i^* = \begin{cases} \hat{\rho}_i & \text{if } i \in \mathbb{S}_{q_j} \\ 1 & \text{if } i \in \mathbb{S}_{q_j}^c \end{cases}. \tag{30}$$

5. Obtain θ_i^* for all units, and let $\tau_T^*(q_j, q_{j+1}) = \theta_{T,(k_{j+1})}^*$.

6. Repeat Steps 2 to 5 B times, obtaining bootstrap test statistics $\tau_T^{*b}(q_j, q_{j+1})$, for $b = 1, \ldots, B$, and select the bootstrap critical value $c_\alpha^*(q_j, q_{j+1})$ as $c_\alpha^*(q_j, q_{j+1}) = \max\{c : B^{-1}\sum_{b=1}^{B} I(\tau_T^{*b}(q_j, q_{j+1}) < c) \leq \alpha\}$, or equivalently as the α-quantile of the ordered $\tau_T^{*b}(q_j, q_{j+1})$ statistics.

The validity of the $BSQT$ tests follows directly from the validity of the $BSUT$ tests, given the relation between quantiles and order statistics.

Remark 5: Note that in $BSQT_1$ the bootstrap test statistic must be selected as $\tau_T^*(q_j, q_{j+1}) = \theta^*_{T,(k_{j+1}-k_j)}$. This must be done as up to this point, k_j units have been found stationary. The k_{j+1}-th order statistic in the original, complete sample then corresponds to the $(k_{j+1} - k_j)$-th order statistic in the smaller bootstrap sample.

Remark 6: The validity of the $BSQT$ methods (as well as of the $BSUT$ methods) can only be shown for finite N, as Palm et al. (2008) only provide results for finite N. The extension to infinite N is very difficult with a general DGP as ours. However, neither in Palm et al. (2008) nor in our case is there any restriction on N; N may be very large, just not infinity. Therefore these validity results are still useful for our purpose. The same caveat holds for the FDR controlling bootstrap method of Romano et al. (2008a) as well.

The major advantage of these tests compared to the $BSUT$, RW and MP tests is that it can use the panel dimension, as the information in all units between two consecutive quantiles to be tested is utilized. This is different from the other tests, which proceed unit by unit, and can therefore not utilize the cross-sectional dimension. Hence, the $BSQT$ methods reduce the dimension of the testing problem to the number of quantiles, and therefore do not suffer from the curse of dimensionality as the other tests do.

The downside of our tests is that they cannot exactly indicate the number of stationary units if the true proportion of stationary units lies between the tested quantiles. However, one can argue about how often one is interested in the exact number of stationary variables, or just in more global information such as "a quarter of the units is stationary". This seems particularly relevant in panels with large N, as identifying individual stationary members does not seem of particular interest. Besides the user can define how precisely the method should work by selecting the quantiles, although there is a clear tradeoff between the spread of the tested quantiles and the power of the method; if the quantiles are taken too close to each other, much of the cross-sectional dimension will be lost and the test will lose power.

For the situation where a more precise estimate is required, we propose the following iterative application of the $BSQT$ method, denoted by $BSQT^*$, although without theoretical justification.

Sequential Test 3 (Iterative Sequential Quantile Test)**.**

1. Select a set of quantiles to be tested $q_1 < \ldots < q_r$.
2. Apply $BSQT_1$ or $BSQT_2$ to obtain an estimate $\hat{q}$. Suppose that $\hat{q} = q_m$.
3. Define a new set of quantiles restricted to the set $[q_{m-1}, q_{m+1}]$[2] and apply the $BSQT_1$ or $BSQT_2$ method to that set. Keep performing this test until the interval has decreased to

[2] If at any stage of the algorithm the estimated quantile is at the boundary of the range, only consider half of the interval in the next step.

consecutive units. Apply the $BSUT$ test to this interval to determine the final estimate $\hat{q}^*$.

In general the number of quantiles in the $BSQT^*$ procedures should be relatively small to properly utilize the cross-sectional dimension.

4.2 Simulation study in large N panels

We now investigate the performance of the tests in a simulation study in a panel with large N. The DGP and parameter combinations used are the same as in the previous section, except the dimensions of the panel. 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. All parameter combinations are summarized on Table 1 panel B. The tests considered are $BSQT_1$, $BSQT_2$, $BSQT_1^*$, $BSQT_2^*$ 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 $BSQT$ tests we take four equally spaced quantiles to be tested for the model with $N = 50$, and eight for $N = 200$. For $BSQT^*$ the same quantiles are used in the first step, with four quantiles in the following steps.

Results are given in Tables 6 to 9. 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 $BSQT_1$ and $BSQT_2$ tests 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. 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 $BSQT^*$ 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 $BSQT^*$ 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 $q_0 = q_j$.[3] 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.

[3] 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 $N = 100$

		$q_0 = 0$					$q_0 = 0.2$				
DGP		$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP	$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP
1	$M(\hat{q})$	0.004	0.004	0.000	0.001	0.000	0.220	0.219	0.118	0.122	0.147
	CP	0.000	0.000	0.000	0.000	0.000	0.792	0.788	0.579	0.595	0.708
	FDR	0.014	0.015	0.021	0.024	0.022	0.235	0.234	0.016	0.020	0.034
	WI	0.986	0.985	1.000	1.000	1.000	0.999	0.999	1.000	1.000	1.000
	HI	0.014	0.015	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000
2	$M(\hat{q})$	0.005	0.004	0.001	0.001	0.001	0.211	0.211	0.105	0.108	0.126
	CP	0.000	0.000	0.000	0.000	0.000	0.723	0.723	0.510	0.523	0.607
	FDR	0.020	0.016	0.029	0.031	0.029	0.253	0.253	0.021	0.022	0.032
	WI	0.980	0.984	1.000	1.000	1.000	0.997	0.998	1.000	1.000	1.000
	HI	0.020	0.016	0.000	0.000	0.000	0.003	0.002	0.000	0.000	0.000
3	$M(\hat{q})$	0.007	0.007	0.001	0.001	0.001	0.241	0.241	0.129	0.134	0.159
	CP	0.000	0.000	0.000	0.000	0.000	0.885	0.882	0.625	0.646	0.753
	FDR	0.025	0.026	0.038	0.042	0.041	0.246	0.245	0.026	0.029	0.044
	WI	0.975	0.974	1.000	1.000	1.000	0.996	0.995	1.000	1.000	1.000
	HI	0.025	0.026	0.000	0.000	0.000	0.004	0.005	0.000	0.000	0.000
4	$M(\hat{q})$	0.005	0.004	0.001	0.001	0.001	0.210	0.211	0.112	0.118	0.143
	CP	0.000	0.000	0.000	0.000	0.000	0.741	0.745	0.551	0.574	0.685
	FDR	0.018	0.016	0.035	0.034	0.032	0.236	0.237	0.017	0.020	0.033
	WI	0.982	0.984	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
	HI	0.018	0.016	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
5	$M(\hat{q})$	0.001	0.001	0.000	0.000	0.000	0.124	0.125	0.047	0.053	0.028
	CP	0.000	0.000	0.000	0.000	0.000	0.484	0.485	0.218	0.244	0.135
	FDR	0.005	0.006	0.010	0.011	0.013	0.219	0.221	0.049	0.060	0.018
	WI	0.995	0.994	1.000	1.000	1.000	0.990	0.991	0.005	0.008	0.000
	HI	0.005	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	$M(\hat{q})$	0.001	0.001	0.000	0.000	0.000	0.122	0.123	0.036	0.041	0.011
	CP	0.000	0.000	0.000	0.000	0.000	0.458	0.459	0.163	0.187	0.054
	FDR	0.006	0.008	0.013	0.013	0.015	0.244	0.244	0.061	0.071	0.009
	WI	0.994	0.992	1.000	1.000	1.000	0.976	0.976	0.002	0.006	0.000
	HI	0.006	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	$M(\hat{q})$	0.005	0.004	0.001	0.001	0.001	0.127	0.129	0.060	0.065	0.050
	CP	0.000	0.000	0.000	0.000	0.000	0.471	0.474	0.262	0.283	0.228
	FDR	0.041	0.031	0.103	0.108	0.104	0.256	0.258	0.109	0.119	0.077
	WI	0.959	0.969	1.000	1.000	1.000	0.997	0.997	0.022	0.028	0.000
	HI	0.041	0.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	$M(\hat{q})$	0.003	0.003	0.001	0.001	0.001	0.128	0.128	0.057	0.063	0.056
	CP	0.000	0.000	0.000	0.000	0.000	0.475	0.476	0.260	0.280	0.259
	FDR	0.027	0.027	0.114	0.118	0.126	0.251	0.252	0.080	0.090	0.065
	WI	0.973	0.973	1.000	1.000	1.000	0.998	0.996	0.019	0.030	0.000
	HI	0.027	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6: Simulation results for large N panels; cross-sectional independence; $q_0 = 0, 0.2$

DGP		$q_0 = 0.5$					$q_0 = 0.9$				
		$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP	$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP
1	$M(\hat{q})$	0.504	0.505	0.381	0.394	0.471	0.823	0.823	0.774	0.786	0.923
	CP	0.938	0.939	0.752	0.776	0.901	0.882	0.882	0.854	0.866	0.982
	FDR	0.067	0.067	0.012	0.015	0.040	0.029	0.030	0.007	0.008	0.041
	WI	0.974	0.975	0.373	0.567	0.992	1.000	1.000	0.815	0.854	1.000
	HI	0.021	0.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	$M(\hat{q})$	0.498	0.499	0.358	0.368	0.426	0.824	0.823	0.784	0.799	0.929
	CP	0.907	0.908	0.702	0.719	0.818	0.884	0.883	0.865	0.881	0.988
	FDR	0.086	0.087	0.018	0.020	0.037	0.029	0.028	0.006	0.007	0.041
	WI	0.958	0.956	0.210	0.329	0.886	1.000	1.000	0.883	0.913	1.000
	HI	0.017	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	$M(\hat{q})$	0.507	0.508	0.386	0.401	0.473	0.819	0.820	0.782	0.794	0.927
	CP	0.933	0.933	0.759	0.786	0.903	0.881	0.882	0.864	0.876	0.986
	FDR	0.075	0.075	0.014	0.018	0.044	0.026	0.027	0.005	0.006	0.041
	WI	0.955	0.953	0.424	0.623	0.997	1.000	1.000	0.863	0.887	1.000
	HI	0.036	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	$M(\hat{q})$	0.502	0.503	0.381	0.395	0.467	0.820	0.820	0.773	0.795	0.919
	CP	0.935	0.935	0.756	0.781	0.900	0.881	0.881	0.854	0.877	0.981
	FDR	0.066	0.068	0.008	0.011	0.034	0.028	0.028	0.006	0.007	0.038
	WI	0.979	0.974	0.404	0.622	0.994	1.000	1.000	0.800	0.864	1.000
	HI	0.014	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	$M(\hat{q})$	0.300	0.312	0.205	0.225	0.145	0.584	0.607	0.463	0.504	0.368
	CP	0.542	0.558	0.391	0.427	0.284	0.637	0.661	0.509	0.553	0.407
	FDR	0.090	0.097	0.042	0.049	0.021	0.018	0.020	0.009	0.012	0.005
	WI	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	HI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	$M(\hat{q})$	0.299	0.308	0.204	0.224	0.147	0.583	0.606	0.465	0.506	0.378
	CP	0.542	0.554	0.391	0.425	0.287	0.636	0.660	0.512	0.556	0.417
	FDR	0.089	0.095	0.039	0.048	0.020	0.017	0.020	0.009	0.011	0.005
	WI	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	$M(\hat{q})$	0.307	0.317	0.213	0.227	0.180	0.585	0.613	0.462	0.506	0.363
	CP	0.542	0.556	0.399	0.423	0.345	0.635	0.664	0.506	0.552	0.399
	FDR	0.110	0.116	0.059	0.066	0.043	0.022	0.025	0.015	0.017	0.010
	WI	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
	HI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	$M(\hat{q})$	0.318	0.328	0.222	0.236	0.184	0.578	0.606	0.457	0.501	0.364
	CP	0.563	0.576	0.417	0.439	0.352	0.629	0.657	0.501	0.548	0.401
	FDR	0.109	0.115	0.060	0.065	0.041	0.021	0.023	0.013	0.015	0.008
	WI	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	HI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 7: Simulation results for large N panels; cross-sectional independence; $q_0 = 0.5, 0.9$

DGP		$q_0 = 0$					$q_0 = 0.2$				
		$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP	$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP
9	$M(\hat{q})$	0.013	0.013	0.003	0.004	0.002	0.097	0.098	0.121	0.125	0.149
	CP	0.000	0.000	0.000	0.000	0.000	0.317	0.314	0.576	0.591	0.703
	FDR	0.045	0.043	0.043	0.046	0.043	0.110	0.110	0.024	0.028	0.040
	WI	0.955	0.957	0.991	0.989	0.996	0.966	0.962	0.994	0.993	0.999
	HI	0.045	0.043	0.009	0.011	0.004	0.034	0.038	0.006	0.007	0.001
10	$M(\hat{q})$	0.017	0.017	0.004	0.004	0.001	0.114	0.116	0.113	0.117	0.136
	CP	0.000	0.000	0.000	0.000	0.000	0.382	0.385	0.537	0.552	0.645
	FDR	0.057	0.057	0.039	0.040	0.034	0.127	0.129	0.024	0.027	0.035
	WI	0.943	0.943	0.988	0.986	0.999	0.975	0.970	0.994	0.991	0.999
	HI	0.057	0.057	0.012	0.014	0.001	0.025	0.030	0.006	0.009	0.001
11	$M(\hat{q})$	0.023	0.026	0.005	0.007	0.001	0.122	0.123	0.118	0.123	0.147
	CP	0.000	0.000	0.000	0.000	0.000	0.408	0.409	0.557	0.579	0.691
	FDR	0.081	0.084	0.053	0.057	0.041	0.133	0.135	0.027	0.030	0.040
	WI	0.919	0.916	0.982	0.978	0.998	0.967	0.963	0.994	0.992	0.998
	HI	0.081	0.084	0.018	0.022	0.002	0.033	0.037	0.006	0.008	0.002
12	$M(\hat{q})$	0.015	0.013	0.002	0.003	0.001	0.108	0.109	0.114	0.119	0.133
	CP	0.000	0.000	0.000	0.000	0.000	0.337	0.335	0.536	0.550	0.630
	FDR	0.052	0.047	0.033	0.038	0.025	0.133	0.134	0.029	0.033	0.037
	WI	0.948	0.953	0.995	0.992	1.000	0.964	0.961	0.991	0.985	0.996
	HI	0.052	0.047	0.005	0.008	0.000	0.036	0.039	0.009	0.015	0.004
13	$M(\hat{q})$	0.011	0.013	0.004	0.005	0.000	0.092	0.096	0.048	0.058	0.040
	CP	0.000	0.000	0.000	0.000	0.000	0.359	0.366	0.217	0.248	0.188
	FDR	0.067	0.067	0.041	0.046	0.015	0.106	0.110	0.039	0.049	0.021
	WI	0.933	0.933	0.978	0.973	0.999	0.607	0.603	0.107	0.117	0.067
	HI	0.067	0.067	0.022	0.027	0.001	0.014	0.022	0.003	0.011	0.001
14	$M(\hat{q})$	0.011	0.012	0.003	0.005	0.000	0.078	0.081	0.036	0.043	0.021
	CP	0.000	0.000	0.000	0.000	0.000	0.279	0.283	0.150	0.176	0.100
	FDR	0.065	0.068	0.030	0.042	0.021	0.122	0.124	0.047	0.060	0.021
	WI	0.935	0.932	0.985	0.979	0.999	0.508	0.502	0.083	0.093	0.019
	HI	0.065	0.068	0.015	0.021	0.001	0.014	0.021	0.001	0.006	0.000
15	$M(\hat{q})$	0.015	0.016	0.004	0.006	0.000	0.083	0.086	0.043	0.051	0.034
	CP	0.000	0.000	0.000	0.000	0.000	0.294	0.298	0.176	0.202	0.152
	FDR	0.092	0.088	0.067	0.070	0.038	0.128	0.132	0.066	0.077	0.041
	WI	0.908	0.912	0.976	0.967	1.000	0.523	0.514	0.097	0.095	0.035
	HI	0.092	0.088	0.024	0.033	0.000	0.018	0.028	0.005	0.016	0.001
16	$M(\hat{q})$	0.014	0.016	0.004	0.006	0.000	0.083	0.085	0.046	0.055	0.047
	CP	0.000	0.000	0.000	0.000	0.000	0.305	0.307	0.203	0.232	0.216
	FDR	0.092	0.091	0.076	0.097	0.061	0.121	0.122	0.055	0.065	0.043
	WI	0.908	0.909	0.975	0.965	1.000	0.546	0.538	0.084	0.093	0.043
	HI	0.092	0.091	0.025	0.035	0.000	0.014	0.022	0.001	0.009	0.000

Table 8: Simulation results for large N panels; common factor; $q_0 = 0, 0.2$

DGP		$q_0 = 0.5$					$q_0 = 0.9$				
		$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP	$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP
9	$M(\hat{q})$	0.424	0.426	0.341	0.356	0.434	0.817	0.816	0.763	0.781	0.926
	CP	0.768	0.772	0.664	0.689	0.827	0.879	0.878	0.843	0.863	0.980
	FDR	0.066	0.066	0.019	0.021	0.038	0.027	0.026	0.005	0.006	0.045
	WI	0.531	0.542	0.207	0.305	0.845	0.990	0.990	0.714	0.760	0.962
	HI	0.072	0.071	0.011	0.017	0.021	0.000	0.000	0.000	0.000	0.000
10	$M(\hat{q})$	0.480	0.479	0.371	0.384	0.467	0.821	0.819	0.754	0.768	0.914
	CP	0.895	0.893	0.731	0.755	0.894	0.881	0.880	0.832	0.848	0.972
	FDR	0.057	0.057	0.013	0.015	0.039	0.028	0.028	0.006	0.007	0.041
	WI	0.829	0.823	0.413	0.536	0.914	0.997	0.996	0.664	0.717	0.973
	HI	0.043	0.043	0.003	0.004	0.009	0.000	0.000	0.000	0.000	0.000
11	$M(\hat{q})$	0.453	0.452	0.354	0.370	0.447	0.801	0.802	0.731	0.743	0.899
	CP	0.817	0.816	0.689	0.716	0.848	0.866	0.867	0.808	0.821	0.957
	FDR	0.073	0.072	0.019	0.022	0.043	0.023	0.023	0.005	0.006	0.039
	WI	0.636	0.629	0.281	0.415	0.862	0.987	0.987	0.552	0.586	0.931
	HI	0.079	0.081	0.006	0.013	0.020	0.000	0.000	0.000	0.000	0.000
12	$M(\hat{q})$	0.453	0.454	0.346	0.358	0.434	0.821	0.823	0.750	0.766	0.909
	CP	0.827	0.829	0.680	0.702	0.831	0.881	0.883	0.827	0.843	0.968
	FDR	0.070	0.071	0.014	0.016	0.037	0.029	0.029	0.007	0.008	0.040
	WI	0.707	0.709	0.252	0.360	0.851	0.997	0.999	0.624	0.673	0.979
	HI	0.046	0.048	0.003	0.003	0.004	0.000	0.000	0.000	0.000	0.000
13	$M(\hat{q})$	0.235	0.246	0.153	0.175	0.134	0.498	0.521	0.399	0.438	0.383
	CP	0.433	0.446	0.292	0.329	0.258	0.544	0.567	0.440	0.481	0.421
	FDR	0.058	0.064	0.028	0.034	0.021	0.014	0.016	0.007	0.008	0.007
	WI	0.048	0.065	0.012	0.019	0.009	0.083	0.107	0.009	0.029	0.018
	HI	0.009	0.018	0.001	0.004	0.000	0.000	0.000	0.000	0.000	0.000
14	$M(\hat{q})$	0.242	0.250	0.156	0.176	0.138	0.545	0.571	0.436	0.482	0.412
	CP	0.448	0.458	0.301	0.335	0.267	0.594	0.621	0.480	0.529	0.454
	FDR	0.061	0.066	0.026	0.033	0.018	0.015	0.017	0.008	0.010	0.006
	WI	0.037	0.048	0.003	0.013	0.001	0.068	0.089	0.005	0.021	0.010
	HI	0.003	0.009	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
15	$M(\hat{q})$	0.242	0.251	0.157	0.176	0.137	0.522	0.547	0.416	0.457	0.401
	CP	0.439	0.450	0.296	0.328	0.263	0.569	0.595	0.458	0.501	0.441
	FDR	0.072	0.076	0.037	0.042	0.027	0.015	0.017	0.008	0.010	0.007
	WI	0.057	0.065	0.011	0.025	0.004	0.075	0.106	0.015	0.025	0.011
	HI	0.008	0.017	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
16	$M(\hat{q})$	0.259	0.268	0.173	0.195	0.172	0.513	0.538	0.410	0.451	0.398
	CP	0.470	0.481	0.327	0.365	0.328	0.557	0.584	0.449	0.494	0.438
	FDR	0.068	0.072	0.035	0.041	0.030	0.017	0.019	0.009	0.011	0.008
	WI	0.059	0.067	0.012	0.020	0.011	0.095	0.116	0.022	0.039	0.021
	HI	0.014	0.025	0.003	0.008	0.001	0.000	0.000	0.000	0.000	0.000

Table 9: Simulation results for large N panels; common factor; $q_0 = 0.5, 0.9$

We first look at the MP method. It can be seen that for the "macro" panel, the MP works very well; it is able to identify a large part of the stationary units, and controls FDR well in most cases. If we look at the model with $N = 200$ things are very different however; while FDR is still controlled, the proportion of stationary units picked up decreases quite dramatically; with the decrease often as large 0.5. This is not so unexpected, as the MP test cannot utilize the cross-sectional dimension, and there are only 25 time series observations in this model. With such a large N, the MP procedure suffers from the curse of dimensionality, as acknowledged by Moon and Perron (2009).

The $BSQT$ methods behave quite differently from the MP method. In the macro panel they are able to pick up slightly fewer stationary units than MP if q_0 is not equal to a tested quantile, but if they are equal, the $BSQT$ methods perform at least as good. In the "micro" panel the $BSQT$ methods pick up much more stationary units than MP, although still not nearly all.[4] 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 $BSQT^*$ 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. The $BSQT_2$ ($BSQT_2^*$) method is on average more powerful than the $BSQT_1$ ($BSQT_1^*$) method, although the difference is again very small.

All tests are less able to pick up the stationary units if there is a common factor; this is not surprising for the $BSQT$ and $BSQT^*$ 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.

Concluding, the $BSQT$ tests perform well in panels with large N, even if T is small. In panels where T is large and N relatively large, the $BSQT$ tests perform nearly as well as the MP tests. In panels with small T but very large N, the $BSQT$ keep performing well whereas the performance of the MP tests deteriorates significantly. Hence, in particular in panels with a large cross-sectional dimension, but a small time dimension, the $BSQT$ tests are more reliable than existing methods.

5 Applications

In this section we consider two applications of the tests proposed in this paper. We first test if 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).

[4]It is to be remembered however that this model, with T only equal to 25, is very unfavorable, and picking up about half of the stationary units is not bad at all.

Country	ADF-GLS	$BSUT_1$	$BSUT_2$	RW	RW_2	MP
Argentina	-4.610		*		*	*
Australia	-2.583					
Belgium	-5.194	*	*	*	*	*
Brazil	-2.593					
Canada	-4.228					*
Denmark	-2.631					
Finland	-6.043	*	*	*	*	*
France	-2.955					
Germany	-3.011					
Italy	-3.724					
Japan	-2.843					
Mexico	-7.300	*	*	*	*	*
Netherlands	-3.037					
Norway	-3.513					
Portugal	-3.064					
Spain	-3.113					
Sweden	-4.310					*
Switzerland	-3.490					
Uk	-3.128					
Total		3	4	3	4	6

Table 10: Tests for PPP on real exchange rates; '*'denotes a rejection of the unit root hypothesis at a 5% level.

5.1 Tests for PPP

We test for PPP in a panel of real exchange rates, using the data of Hanck (2009). His data is based on the long annual exchange rate data of Taylor (2002). We have data for 19 countries for 105 years.

For $\theta_{T,i}$ we use the ADF test with GLS detrending, with lag lengths selected by BIC. We allow for an intercept and trend.[5] The results are presented in Table 10. The $BSUT_1$ and RW methods reject for three countries, the $BSUT_2$ and RW_2 methods for four countries.[6] The MP method rejects somewhat more often, in this case for six countries. As discussed before, the additional rejections from the MP method might simply be caused by the liberal nature of the method. In a panel with a relatively small cross-section like this, controlling FDR might not be the most appropriate way to control size.

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

[5] Different specification give similar results.

[6] Our results differ from Hanck (2009) because of different test statistics, bootstrap method and lag selection, but are qualitatively similar.

	$BSQT_1$	$BSQT_2$	$BSQT_1^*$	$BSQT_2^*$	MP
Total rejections ($N = 181$)	91	91	60	80	1
Rejections in subsamples					
CLG ($N = 58$)	25	24	18	25	0
HSG ($N = 87$)	46	40	34	46	1
HSD ($N = 36$)	20	16	8	20	0

Table 11: Unit root tests on log real earnings of households in PSID data at 5% level

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.[7] This might lead one to draw 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. It could however also be 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 all sorts 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 $BSQT$ tests (as well as the MP method) to the data used in Pesaran (2007).[8]. As quantiles to be tested we take $q_j = \frac{j-1}{6}$, $j = 1, \ldots, 6$, while for θ_i we again take the ADF test with GLS detrending allowing for intercept and trend with lag length selection by BIC.[9]

Results are listed in Table 11. 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. Our results indicate that about one third to one half of the units are stationary. Moreover this holds for all the subsamples as well, hence there appears to be no relation between education level and the unit root. Note also how the MP test fails 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 $BSQT$ methods.

[7]Which subgroups are rejected depends on the speciffcation of the $CIPS$ test, see Pesaran (2007) for details.

[8]The dataset used in Pesaran (2007) is available from the Journal of Applied Econometrics Data Archive (www.econ.queensu.ca/jae)

[9]Again, different specifications lead to similar results.

6 Conclusion

We have proposed new methods based on sequential tests to investigate the stationarity properties of individual units in a panel. We constructed tests based on sequential testing unit-by-unit, the $BSUT$ tests, which are mainly suited for use in panels with a small N. We showed that these tests are very similar to the multiple testing procedure developed by Romano and Wolf (2005) and, based on this similarity, we propose a modification of that procedure that can be more powerful.

We can modify the $BSUT$ tests to be applicable in large panels; for this purpose we sequentially test user-defined quantiles instead of all units. These $BSQT$ tests can also be applied iteratively, yielding the $BSQT^*$ tests. We demonstrated the good performance of these tests in finite samples, where the $BSQT$ tests have a significant advantage over the multiple testing approach of Moon and Perron (2009) 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 demonstrate the usefulness of these methods.

The methods developed in this paper are not restricted to unit root testing in panels. In particular the quantile-based tests 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

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.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*, 813–836.

Hanck, C. (2009). For which countries did PPP hold? A multiple testing approach. *Empirical Economics 37*, 93–103.

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.

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 (2009). Beyond panel unit root tests: Using multiple testing to determine the non stationarity properties of individual series in a panel. Working paper.

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 (2008). Cross-sectional dependence robust block bootstrap panel unit root tests. METEOR Research Memorandum 08/048, Maastricht University.

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

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.

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.

White, H. (2000). A reality check for data snooping. *Econometrica 68*, 1097–1126.

A Appendix: Proofs

Proof of Proposition 1. The first statement directly follows from (5b). The second statement is the probability of not rejecting $H_0(k_0)$, which is equal to $\mathrm{P}(\tau_{T,k_0} \geq c_{\alpha,k_0})$. This therefore follows from (5a). Finally, the third property follows from the fact that

$$\sum_{k=0}^{N} \mathrm{P}(\hat{k} = k) = \sum_{k=0}^{k_0-1} \mathrm{P}(\hat{k} = k) + \mathrm{P}(\hat{k} = k_0) + \sum_{k=k_0+1}^{N} \mathrm{P}(\hat{k} = k) = 1 - \alpha + \sum_{k=k_0+1}^{N} \mathrm{P}(\hat{k} = k) = 1,$$

where we apply the first and second statement. □

Proof of Theorem 1. We first show (17). The first part (17a) follows directly as $\theta_{T,i} \xrightarrow{p} -\infty$ for $i \in \mathscr{S}$ as $T \to \infty$ by Assumption 2. By the same result, we have that

$$\theta_{T,(k_0+1)} = \min_{i \in \mathscr{U}} \theta_{T,i},$$

with probability 1 as $T \to \infty$. The result in (17b) then follows directly from the continuous mapping theorem (cf. White, 2000, Lemma 2).

The result in (18) follows directly from Paparoditis and Politis (2003) and Palm et al. (2008), as their proofs of asymptotic validity can straightforwardly be extended to general DF statistics. The DF t-test and detrending are not considered in Paparoditis and Politis (2003) and Palm et al. (2008), but can easily be dealt with.[10]

Finally, the result in (19a) follows from the continuous mapping theorem. For $BSUT_1$ (19b) follows directly, as by the consistency of $\theta_{T,k}$, $\mathbb{S}_k = \mathscr{S}$ with probability 1 as $T \to \infty$. For $BSUT_2$, note that as $T \to \infty$, $\hat{\rho}_i - \rho_i = o_p(1)$ for $i \in \mathbb{S}_k$. 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), we have that $\hat{\rho}_i^* - \rho_i = o_p^*(1)$, from which it then easily follows that, for $i \in \mathbb{S}_k$, $\theta_{T,i}^* \xrightarrow{p^*} -\infty$ as $T \to \infty$. (19b) then follows. □

[10]The crucial result is the invariance principle derived in Lemmas 2 and 4 of Palm et al. (2008). As this result still applies here, it can be used to establish the asymptotic validity of the more general DF statistics.