

Bootstrap sequential tests to determine the order of integration of individual units in a time series panel

Citation for published version (APA):

Smeeke, S. (2015). Bootstrap sequential tests to determine the order of integration of individual units in a time series panel. *Journal of Time Series Analysis*, 36(3), 398-415. <https://doi.org/10.1111/jtsa.12110>

Document status and date:

Published: 01/01/2015

DOI:

[10.1111/jtsa.12110](https://doi.org/10.1111/jtsa.12110)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

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

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 24 Apr. 2024

SPECIAL ISSUE

BOOTSTRAP SEQUENTIAL TESTS TO DETERMINE THE ORDER OF INTEGRATION OF INDIVIDUAL UNITS IN A TIME SERIES PANEL

STEPHAN SREEKES*

Department of Quantitative Economics, Maastricht University, Maastricht, The Netherlands

We propose an approach to investigate the unit root properties of individual units in a time series panel or large multivariate time series, based on testing user-defined increasing proportions of hypothesized $I(0)$ units sequentially. Asymptotically valid critical values are obtained using the block bootstrap. This sequential approach has an advantage over multiple testing approaches as it can exploit the (cross-sectional) dimension of the system, which the multiple testing approaches cannot do effectively. A simulation study and an empirical application are conducted to analyse the relative performance of the approach in comparison with multiple testing approaches. These demonstrate the usefulness of our method, in particular in systems with a relatively small time dimension.

Received 9 February 2014; Revised 27 October 2014; Accepted 27 October 2014

Keywords: Sequential testing; unit roots; time series panels; block bootstrap.
JEL. C15; C23.

1. INTRODUCTION

Over the last decade, a large number of unit root tests have been designed that can be applied in large multivariate time series and panel data. While most of these tests have the null hypothesis that all individual series, or units in the panel terminology, have a unit root, the alternative hypothesis typically differs. Some tests have the alternative hypothesis that some series are stationary.¹ A rejection for such a test is hard to interpret; maybe only a small fraction of the units is stationary, or maybe all units are stationary. These two opposites will usually have very different consequences for the interpretation of the tests, yet there is no way to distinguish. Other tests have as alternative hypothesis that all units are stationary. While this alternative hypothesis might seem to help interpreting a rejection, these tests also have power if not all units are stationary (cf. Westerlund and Breitung, 2013, Fact 2), and hence, a rejection is not evidence that all series are indeed stationary. Pesaran (2012) considers the interpretation of a rejection of a panel unit root test and argues that a rejection (for large enough cross-sectional dimension) should be interpreted as the finding that a fraction $\delta > 0$ of all units is stationary; therefore, in the event of a rejection, the rejection should be supplemented with an estimate of δ to measure the economic importance of a rejection.

It is therefore important to consider methods that provide more information than just a rejection for the whole panel. In particular, methods that provide an estimate of the proportion of stationary units or that find which units are stationary are required. An estimate of the proportion may be sufficient, but often, one needs to classify the stationary and nonstationary units as well, for example, if the units in the panel can be split into subsets based on certain characteristics and the order of integration may depend on those. Note that such methods are not intended as a substitute for conventional panel unit root tests; rather, they can be used as a supplement to a conventional panel unit root test as recommended by Pesaran (2012).

* Correspondence to: Stephan Smeekes, Department of Quantitative Economics, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands. E-mail: s.smeekes@maastrichtuniversity.nl

¹ Throughout this article, we use 'stationary' and integrated of order zero— $I(0)$ —interchangeably. In the same way, 'nonstationary' is used for $I(1)$. While true under our specific assumptions, this is not the case in general, but for ease of exposition, we maintain our terminology.

One way to estimate the fraction of stationary units is to perform individual unit root tests on each unit. However, this approach does not control size as multiple tests are performed. Pesaran (2007a) shows that, allowing for cross-sectional dependence, under the null hypothesis of a unit root, the fraction of rejections will on average equal the significance level. This provides some guidance when all units have a unit root, but if a higher fraction of rejections is observed, it is more difficult to determine which rejections are true and which are false.

Recently, more sophisticated methods have been proposed. Ng (2008) proposes to estimate the proportion of stationary units using the trend in cross-sectional averages of variances, but this method cannot test which units are stationary. Hanck (2009) and Moon and Perron (2012) apply methods from the literature on multiple testing. Hanck (2009) uses the bootstrap approach of Romano and Wolf (2005) to control the family-wise error rate (FWE), which is mainly suitable for small cross-sectional dimension N . Moon and Perron (2012) consider various methods to control the false discovery rate (FDR), which is better suited to panels with larger N , and find that the bootstrap method of Romano *et al.* (2008) performs best.

We propose a sequential testing approach to determine the stationary units, thereby avoiding issues of controlling size in multiple testing. Our approach can be used to estimate the proportion of stationary units, and it can also be used to determine which units are stationary. Our approach is similar in spirit to that of Kapetanios (2003) and Chortareas and Kapetanios (2009). In contrast to their approach, which bases sequential tests on averages of individual unit root tests, our method proceeds by sequentially testing user-specified quantiles of the unit root test statistics directly. As shown in the paper, their average-based test only allows for effective use of the cross-sectional dimension to detect the stationary units if the majority of the units is indeed stationary. However, the order statistics used in our method can exploit the cross-sectional dimension effectively even when fewer units are stationary, in particular as the order statistics used do not have to be consecutive ('wider' quantiles can be selected).

Moreover, in contrast to the non-bootstrap methods, our block bootstrap method can handle complex error dynamics that include but extend beyond common factor structures (cf. Palm *et al.*, 2011). Compared to the existing bootstrap approaches of Hanck (2009) and Moon and Perron (2012), we find our method to be more powerful, in the sense that it picks up more of the stationary units, but not at the expense of a higher probability of incorrectly finding units to be stationary. Unlike the multiple testing methods, our approach effectively exploits the cross-sectional dimension to gain power, which mainly manifests if T is relatively small compared to N such that power cannot be derived from the time dimension.

The paper is structured as follows. Section 2 introduces the model. In Section 3, the sequential method and its asymptotic properties are considered. A simulation study comparing the properties of our method to related approaches is performed in Section 4. In Section 5, we present an empirical application. Section 6 concludes. Proofs are given in the Appendix.

A word on notation. 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 as $\xrightarrow{d^*}_p (\xrightarrow{p^*}_p)$, where this convergence is taken to take place in probability. $|x|$ applied to a vector or matrix x denotes its Euclidean or Frobenius norm respectively, while $|\mathcal{G}|$ applied to any set of positive integers \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

$$y_t = \beta z_t + x_t, \quad x_t = P x_{t-1} + u_t, \quad u_t = \Psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad (1)$$

where $\beta = (\beta_1, \dots, \beta_N)'$, z_t are deterministic components with $z_t = 1$ or $z_t = (1, t)'$, and $P = \text{diag}(\rho_1, \dots, \rho_N)$. Furthermore, the following assumptions on $\Psi(L)$ and ε_t are needed.

Assumption 1. (i) Let $\sum_{j=0}^{\infty} j |\Psi_j| < \infty$ and $\Psi_i(1) \neq 0$, where $\Psi_i(L)$ is the i th row of $\Psi(L)$; (ii) ε_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 assumption ensures that if $\rho_i = 1$, unit i has a unit root, while if $|\rho_i| < 1$, unit i is stationary. This data generating process (DGP) allows for a wide range of dependencies, including dynamic interdependencies, common factors and cointegration between units. Common factors can be induced by restrictions on $\Psi(L)$, while cointegration occurs if $\Psi(1) = \alpha\beta'$, where α and β are $N \times r$ matrices with $r < N$. This DGP is a special case of the DGP used by Palm *et al.* (2011) and is considered here rather than the original one as in that model non-stationarity for each unit can come from two different sources ('idiosyncratic' components and common factors), whereas in the current specification only the parameter ρ_i determines whether unit i is $I(0)$ or $I(1)$.²

Let $k_0 = 0, 1, \dots, N$ be the number of stationary units. Furthermore, let $q_0 = k_0/N$ denote the proportion of stationary units. Formally, define $\mathcal{S} = \{i \in \mathcal{N}_N : |\rho_i| < 1\}$ as the set of k_0 stationary units, such that $k_0 = |\mathcal{S}|$, and $\mathcal{U} = \{i \in \mathcal{N}_N : \rho_i = 1\}$ as the set of $N - k_0$ nonstationary units. Occasionally, we will write \mathcal{S}_{k_0} and \mathcal{U}_{k_0} in order to emphasize the number of units in each set. In the next section, we will discuss a method aimed at finding both k_0 (or q_0) and the members of \mathcal{S} .

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 .³ 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. $|\mathcal{S}| \geq k_{j+1}$. The algorithm now tests $H_0(q_j)$ against $H_1(q_{j+1})$ sequentially for $j = 1, \dots, r$, terminating once $H_0(q_j)$ cannot be rejected.

In every stage of the sequential procedure, a test statistic is needed; let $\tau(q_j, q_{j+1})$ denote that test statistic to test $H_0(q_j)$ vs. $H_1(q_{j+1})$. We propose to calculate appropriate individual unit root test statistics on each unit and then use the relevant quantile of these as test statistic. For this purpose, let θ_i be any unit root test statistic applied to unit i that satisfies the following assumption.

Assumption 2. Let θ_i be a statistic applied to unit $i = 1, \dots, N$ that rejects the null hypothesis of a unit root for small values of θ_i . Furthermore, as $T \rightarrow \infty$, $\theta_i \xrightarrow{d} \xi_i$ if $\rho_i = 1$ (where ξ_i is non-degenerate), and $\theta_i \xrightarrow{p} -\infty$ if $\rho_i < 1$.

Many unit root tests statistics satisfy these conditions, including the popular augmented Dickey–Fuller (ADF) tests, in which case ξ_i is the Dickey–Fuller distribution, or one of its detrended variants.⁴ It is asymptotically

² As discussed in Remark 2 of Palm *et al.* (2011), the major difference is the absence of exogenous factors in our DGP. Given that common factors are not the focus of this work and could be added if desired, we do not believe this to be restrictive.

³ For any finite N , multiple values for q_j lead to the same value of k_j . We assume q_1, \dots, q_r are chosen such that $k_1 < k_2 < \dots < k_r$; otherwise, the selection does not make sense.

⁴ The assumption of left-tail tests is made for expositional simplicity only; trivial adaptations make the algorithm suitable for right-tailed or two-tailed tests.

not necessary, though very desirable, that the asymptotic distribution ξ_i is free of nuisance parameters; see the discussion in Remark 2. In general, ξ_i and ξ_j will be correlated if there is cross-sectional dependence in the data.

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,

$$\tau(q_j, q_{j+1}) = \theta_{(k_{j+1})} = \theta_{(\lfloor q_{j+1}N \rfloor)}. \tag{2}$$

This choice of test statistic, though natural, is not used often as for the order statistics in particular, finding asymptotically pivotal quantities is notoriously difficult owing to the dependence of $\theta_1, \dots, \theta_N$ caused by the cross-sectional dependence in the panel. For this reason, we propose to use the bootstrap to obtain critical values, which we label the bootstrap sequential quantile test (BSQT). The block bootstrap method employed in this article is the same as in Palm *et al.* (2011). The major practical advantage of this methodology is that the block bootstrap can be applied without knowing the form of the cross-sectional dependence and can therefore be applied to our general model without the need to further specify the nature of the dependence. We can use our method not only to estimate q_0 but also to determine the members of \mathcal{S} . The algorithm looks as follows.

Algorithm 1.

- (1) Calculate the unit root test statistic θ_i from the series $\{y_{i,t}\}$ for $i = 1, \dots, N$.
- (2) For $i = 1, \dots, N$ obtain the detrended series $\tilde{y}_{i,t} = y_{i,t} - \hat{\beta}'_i z_{i,t}$, where $\hat{\beta}_i$ is an estimator of β_i . Calculate

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

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

- (3) For bootstrap iteration b , draw j_0^b, \dots, j_{k-1}^b i.i.d. from the uniform distribution on $\{1, 2, \dots, T-l\}$, where l is the block length and $k = \lfloor (T-2)/l \rfloor + 1$ is the number of blocks.
- (4) Construct the bootstrap errors $u_2^{*b}, \dots, u_T^{*b}$ as $u_t^{*b} = \hat{u}_{j_m^b + s}$, where $m = \lfloor (t-2)/l \rfloor$ and $s = t - ml - 1$. For $i = 1, \dots, N$, let $y_{i,t}^{*b} = y_{i,t-1}^{*b} + u_{i,t}^{*b}$ where $y_{i,1}^{*b} = \tilde{y}_{i,1}$.
- (5) Obtain θ_i^{*b} from $\{y_{i,t}^{*b}\}$ in the same way as in step 1.
- (6) Repeat Steps 3 to 5 B times, obtaining a collection of bootstrap unit root test statistics $\{\theta_i^{*b}\}$ for $i = 1, \dots, N$ and $b = 1, \dots, B$.
- (7) Start the sequential testing by testing $H_0(q_1)$ against $H_1(q_2)$ with test statistic $\tau(q_1, q_2) = \theta_{(k_2)}$. For $b = 1, \dots, B$ obtain bootstrap test statistics $\tau^{*b}(q_1, q_2) = \theta_{(k_2)}^{*b}$. Obtain the bootstrap critical value $c_\alpha^*(q_1, q_2) = \max \left\{ c : B^{-1} \sum_{b=1}^B I(\tau^{*b}(q_1, q_2) < c) \leq \alpha \right\}$, or equivalently as the α -quantile of the ordered $\tau^{*b}(q_1, q_2)$ statistics. Reject $H_0(q_1)$ if $\tau(q_1, q_2) < c_\alpha^*(q_1, q_2)$.
- (8) If $H_0(q_1)$ is not rejected, stop and set $\hat{q} = q_1 = 0$. Otherwise, store the stationary units $\mathbb{S}_{k_2} = \{i : \theta_i \leq \theta_{(k_2)}\}$, and proceed with the next stage to test $H_0(q_2)$ against $H_1(q_3)$. Continue with the sequential tests until $H_0(q_j)$ cannot be rejected for some $j \in \{1, \dots, r\}$. In that case, let $\hat{q} = q_j$ and let the stationary units be the set $\mathbb{S}_{k_j} = \{i : \theta_i \leq \theta_{(k_j)}\}$, i.e. the set of units that have already been classified as stationary in the previous stages. If $H_0(q_r)$ is rejected, set $\hat{q} = 1$.

In stage j , test $H_0(q_j)$ against $H_1(q_{j+1})$ with statistic $\tau(q_j, q_{j+1}) = \theta_{(k_{j+1})}$, and let

$$\tau^{*b}(q_j, q_{j+1}) = \theta_{\left(k_{j+1}-k_j; \mathbb{S}_{k_j}^{\xi_j}\right)^{*b}}, \quad \text{for } b = 1, \dots, B, \tag{3}$$

where, for any set of positive integers \mathcal{G} , $z_{(k:\mathcal{G})}$ denotes the k th smallest element of the set $\{z_i : i \in \mathcal{G}\}$. Reject $H_0(q_j)$ if $\tau(q_j, q_{j+1}) < c_\alpha^*(q_j, q_{j+1})$, where $c_\alpha^*(q_j, q_{j+1}) = \max \left\{ c : B^{-1} \sum_{b=1}^B I(\tau^{*b}(q_j, q_{j+1}) < c) \leq \alpha \right\}$.

It is crucial in (3) that one obtains the bootstrap test statistic as the $(k_{j+1} - k_j)$ th smallest value of the θ_i^{*b} for only the units in the set $S_{k_j}^c$, representing those units still under consideration for which the unit root has not yet been rejected in previous steps. For the original sample, we have that $\theta_{(k_{j+1})} = \theta_{(k_{j+1}-k_j; S_{k_j}^c)}$ by construction, as the same ranking is used both to find the order statistic and to restrict the set of units under consideration. However, in the bootstrap $\theta_{(k_{j+1})}^{*b} \neq \theta_{(k_{j+1}-k_j; S_{k_j}^c)}^{*b}$, as the set $S_{k_j}^c$ is determined by the ranking of the original statistics rather than the bootstrap statistics; that is, we first exclude the ‘rejected’ units (i.e. the units found to be stationary in previous steps) based on their ranking for the original sample statistics and then rank only the remaining units based on the bootstrap statistics (cf. Romano *et al.*, 2008, Rmk 1). By restricting the set of units in the bootstrap to $S_{k_j}^c$, one removes the rejected units from consideration and thereby imposes $H_0(q_j)$, the null hypothesis that k_j units are stationary. Instead, taking the k_{j+1} th smallest value of $\theta_1^{*b}, \dots, \theta_N^{*b}$ implies a unit root in all units and therefore tests $H_0(0)$ against $H_1(q_{j+1})$ rather than $H_0(q_j)$ against $H_1(q_{j+1})$, resulting in a too liberal critical value that fails to control size asymptotically.⁵

One might also be tempted to take as test statistic $\theta_{([q_j N])}$, but this would not test the correct set of hypotheses. Consider testing $H_0(1/4)$ against $H_1(1/2)$. If one takes the order statistic corresponding to 1/4, one can only assess whether 1/4 of the units is indeed stationary, while assessing nothing about 1/2 of the units. However, we already determined that those 1/4 of the units were stationary in the previous step. As such, it would be incorrect to conclude from a rejection that 1/2 of the units are stationary, and it would be incorrect to proceed to the next step of the sequential procedure.

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 statistical arguments and the specific interpretation needed for each particular application.

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.⁶ In that sense, if the outcome of the method is that $\hat{q} = q_j$, this should not necessarily be interpreted to say that q_j is the true stationary proportion, but it can rather be interpreted that the true stationary proportion should be found in the interval (q_{j-1}, q_{j+1}) . 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 impossible 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 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.

Note also that it is in no way necessary to be able to order or rank the cross-sectional units. Through the use of the order statistics, the method takes care of ranking the units based on their statistics θ_i automatically,

⁵ From this point on, we drop the superscript b to indicate which iteration the bootstrap quantity derives from. This is only relevant from a computational point of view, hence its inclusion in the algorithm. To simplify the notation, we now revert to the traditional way of only using a ‘*’ to indicate bootstrap quantities.

⁶ 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.

leading to a ranking from ‘most significantly’ to ‘least significantly’ stationary units. Hence, no knowledge on a cross-sectional ordering, which may typically be hard to construct, is necessary, and the choice of quantiles should only be based on the question of which proportions of stationary units are relevant to the question at hand, bearing in mind that finding a stationary proportion q_j should be interpreted as the true stationary proportion being in the interval (q_{j-1}, q_{j+1}) . Finally, note that the quantiles do not have to be equally spaced; whether 0% or 10% of the units are stationary may have very different impact on the economic interpretation of the results than whether 10% or 20% of the units are stationary.

A very specific option for setting the quantiles is such that each unit is tested sequentially. This amounts to setting $q_j = (j - 1)/N, j = 1, \dots, N$. This approach has a fundamentally different interpretation than when setting ‘fixed’ quantiles. First, note that all problems mentioned earlier 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 N . An upper limit on N for which the unit-by-unit approach is still applicable cannot be determined in general, as it depends much on the corresponding value of T . One can think of abandoning it in favour of wider quantiles though when either N is that large that single units are not of particular interest anymore or when T is not large compared to N , e.g. when using standard multivariate time series analysis becomes undesirable.

3.2. Asymptotic properties

We now proceed with the asymptotic analysis of the method. As the bootstrap method used in this article is the same as in Palm *et al.* (2011), and our DGP is a special case of theirs, their results on the validity of the bootstrap can be used here as well. 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 mainly needed to make statements about the asymptotic properties of the whole sequential method.

Theorem 1.

- (i) Let $y_{i,t}$ be generated by (1), 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 (2). Let $k_0 = [q_0 T]$ be the true number of I(0) units, and let \mathcal{U}_{k_0} be the set of $N - k_0$ I(1) units. Then, as $T \rightarrow \infty$, we have that

$$\tau(q_j, q_{j+1}) \xrightarrow{p} -\infty \quad \text{if } q_{j+1} \leq q_0, \tag{4a}$$

$$\tau(q_j, q_{j+1}) \xrightarrow{d} \xi_{(k_{j+1}-k_0:\mathcal{U}_{k_0})} \quad \text{if } q_{j+1} > q_0. \tag{4b}$$

- (ii) In addition to the preceding conditions, let θ_i^* and $\tau^*(q_j, q_{j+1})$ be defined as in Algorithm 1, and let $1/l + l/T^{1/2} \rightarrow 0$ as $T \rightarrow \infty$. Then, as $T \rightarrow \infty$, for any q_0 ,

$$\theta_i^* \xrightarrow{d^*} \xi_i \quad \text{for all } i \in \mathbb{S}_{k_j}^c \text{ (jointly)}, \tag{5a}$$

$$\tau^*(q_j, q_{j+1}) \xrightarrow{d^*} \xi_{(k_{j+1}-k_j:\mathbb{S}_{k_j}^c)}, \tag{5b}$$

$$P(\mathbb{S}_{k_0}^c = \mathcal{U}_{k_0}) \rightarrow 1. \tag{5c}$$

This theorem shows the asymptotic validity of the bootstrap order statistics. Part (i) of the theorem states that if $q_{j+1} \leq q_0$, such that the alternative hypothesis in step j is true, and the order statistic used for $\tau(q_j, q_{j+1})$ comes from a stationary unit, that $\tau(q_j, q_{j+1})$ diverges to $-\infty$, while if $q_{j+1} > q_0$ and the order statistic used

for $\tau(q_j, q_{j+1})$ comes from a nonstationary unit, that $\tau(q_j, q_{j+1})$ then converges to the $(k_{j+1} - k_0)$ th smallest statistic of the set of random variables ξ_i for those units that have a unit root. The second part of the theorem shows that the bootstrap correctly reproduces the asymptotic distributions of the original statistics: (5a) shows the distributions of the statistics θ_i are correctly reproduced, from which (5b) follows. The asymptotic validity of the bootstrap then follows from (5c), which shows that if the true proportion of stationary units q_0 is part of the sequence of tested quantiles, the correct units will have been classified as stationary in the previous steps of the algorithm.

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.* (2011). As stated before, this cross-sectional dependence can be of very general form. 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

$$\lim_{T \rightarrow \infty} P(\hat{q} = q_j) = 0 \quad \text{if } q_{j+1} \leq q_0, \quad (6a)$$

$$\lim_{T \rightarrow \infty} P(\hat{q} = q_j) = 1 - a_{j,k_0}^\alpha \quad \text{if } q_j < q_0 < q_{j+1}, \quad (6b)$$

$$\lim_{T \rightarrow \infty} P(\hat{q} = q_j) = 1 - \alpha \quad \text{if } q_j = q_0, \quad (6c)$$

$$\lim_{T \rightarrow \infty} P(\hat{q} = q_j) \leq a_{j-1,k_0}^\alpha \quad \text{if } q_{j-1} < q_0 < q_j, \quad (6d)$$

$$\limsup_{T \rightarrow \infty} P(\hat{q} = q_j) \leq \alpha \quad \text{if } q_{j-1} \geq q_0, \quad (6e)$$

where $a_{j,k_0}^\alpha = P\left\{\xi_{(k_{j+1}-k_0):L_{k_0}} < c_{\alpha,\infty}(q_j, q_{j+1})\right\}$, and $c_{\alpha,\infty}(q_j, q_{j+1})$ is such that

$$P\left\{\xi_{(k_{j+1}-k_j):S_{k_j}^c} < c_{\alpha,\infty}(q_j, q_{j+1})\right\} = \alpha. \quad (7)$$

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

$$\alpha \leq a_{j,m}^\alpha \leq a_{j,m+1}^\alpha \leq 1. \quad (8)$$

Several conclusions follow from the preceding results. 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 α ; compare for instance Swensen (2006, Cor. 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 $q_j = (j - 1)/N$. However, as has been discussed before, this method will not always work well in finite samples and is unsuitable for applications where N is large.

Corollary 1 also gives an insight into what happens asymptotically if the true proportion is in between the selected quantiles. Note that a_{j,k_0}^α is the probability that the null hypothesis that $k_j = [q_j N]$ units are stationary is rejected in favour of the alternative that k_{j+1} units are stationary, if the true number of stationary units k_0 is such that $k_j < k_0 < k_{j+1}$. Furthermore, (8) shows the (intuitively appealing) result that this probability of rejection increases when the true number of stationary units increases (from m to $m + 1$ in the corollary, which by induction can be extended to any increase). Then, combining equations (6b) and (6d) with (8) 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.

The corollary also provides an explanation for the earlier statement that the result that $\hat{q} = q_j$ could be interpreted as saying that q_0 lies in the interval (q_{j-1}, q_{j+1}) . 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 $\hat{q} = q_j$, then $\hat{q}_{-1} = q_{j-1}$ and $\hat{q}_{+1} = q_{j+1}$. Now note that

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

By (6a) and (6e), it follows that

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

from which we can conclude that

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

This confirms that we can interpret finding a certain \hat{q} as evidence that q_0 is between \hat{q}_{-1} and \hat{q}_{+1} . It should be stressed here though that clearly this is not a proper interval estimator; it is after all not based on the distribution of a proper (consistent) point estimator. It simply provides a way to interpret the proportion of stationary units found by the sequential procedure.

Remark 1. Note that results (6b) and (8) are not only useful in our sequential context but also describe the asymptotic power of ‘conventional’ bootstrap panel unit root tests based on order statistics, such as the median, minimum and maximum.⁷ 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 2. Although any unit root test can be used for θ_i , in small samples, one should only use statistics whose marginal distribution is free of nuisance parameters. If the marginal distributions of the individual test statistics depend on nuisance parameters, they may not live on the same scale, and the same values for two different individual tests may not imply the same level of significance. As a consequence, any ranking of them becomes unreliable, and in small samples even meaningless.⁸ Asymptotically, this is not an issue: under Assumption 2, the statistics θ_i diverge under the alternative hypothesis, and as such, there is a clear distinction between the stationary and nonstationary units for which $\theta_i = O_p(1)$. In such a case, the stationary units always get ranked first, and it does not matter how the nonstationary units are ranked among themselves. In small samples, such a sharp distinction cannot be made, which is why it is crucial to be able to rank them as accurately as possible. This is achieved by making all the individual tests live on the same scale under the null hypothesis, which is the case if their marginal distributions are the same; also see the discussion in Romano and Wolf (2005, p. 1255).

⁷ See, for example, Di Iorio and Fachin (2009) for an application of such statistics to cointegration testing.

⁸ This occurs, for example, if the variances differ. Suppose that the variance of θ_j is greater than that of θ_i . If in that case θ_i is smaller than θ_j , θ_i does not necessarily indicate more evidence against the unit root null than θ_j , but it does get ranked before it.

If it is not possible to work with statistics that have limit distributions free of nuisance parameters, one may work with p -values instead. If nuisance parameters are present and p -values may be difficult to obtain, the bootstrap (which is performed anyway) can be used to calculate p -values. In the current setting, this means a second bootstrap iteration within the first bootstrap is required to calculate the p -values of the bootstrap test statistics.

Remark 3. Kapetanios (2003) and Chortareas and Kapetanios (2009) apply the IPS group-mean statistic of Im *et al.* (2003) as test statistic, deleting the ‘most stationary’ unit every round. A potential 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. We will investigate this in the simulation study. Moreover, one should take care that the original IPS test does not allow for cross-sectional dependence, so it has to be augmented to deal with cross-sectional dependence. Chortareas and Kapetanios (2009) propose to do so using the instrumental variables approach of Chang (2002) and Chang and Song (2009) or the cross-sectional averages approach of Pesaran (2007b). Alternatively, one could use the bootstrap approach of Palm *et al.* (2011), which we do in this article.

Remark 4. Several extensions to the BSQT method can be devised, of which two are of particular interest. First, one can choose not to delete the units found stationary from the bootstrap statistics used in the next round but instead insert them as stationary units. That is, in step 4 of Algorithm 1, one could generate the bootstrap series for the stationary units as $y_{i,t}^{*b} = \hat{\rho}_i y_{i,t-1}^{*b} + u_{i,t}^{*b}$, where $\hat{\rho}_i$ is taken from step 2. This could lead to a more powerful method as the finite sample effects of the stationary units on the distribution of the order statistics are taken into account, but it would be computationally more demanding as for each unit a second, stationary, bootstrap sample would need to be generated. Second, one may iterate the sequential approach in order to find a more precise estimate of the number of stationary units. If it is found initially that $\hat{q} = q_j$, the method can be applied again to the interval (q_{j-1}, q_{j+1}) , and this process can be repeated until it converges. Both extensions are discussed in detail in Smeekees (2011), but as simulation studies conducted there show that the extensions do not significantly improve on the basic method, they are omitted here.

Remark 5. Multiple testing methods have been used in a panel context by Hanck (2009), Deckers and Hanck (2014) and Moon and Perron (2012) among others as an alternative to our approach. Hanck (2009) tests for which countries purchasing power parity holds, using the bootstrap method of Romano and Wolf (2005) to control the family-wise error rate (FWE), defined as the probability of making at least one false rejection. Smeekees (2011) shows that this bootstrap method is very similar to the BSQT approach applied unit-by-unit. Deckers and Hanck (2014) and Moon and Perron (2012) control the false discovery rate (FDR), defined as the expectation of the proportion of rejections that are false. This generalized error rate is more liberal than the FWE, and therefore more suitable in panels with larger N where FWE is a too strict criterion. Deckers and Hanck (2014) and Moon and Perron (2012) find favourable results for the bootstrap approach of Romano *et al.* (2008) for controlling FDR. A downside of the FDR controlling bootstrap approach is that the critical value calculation is a complex and computationally demanding procedure if N is large (also see Remark 7).

4. SIMULATIONS

4.1. Simulation DGP

We next perform a simulation study to investigate the small sample performance of our approach. We generate data as

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

Table I. Parameter combinations simulation DGPs

DGP	T	N	λ_i	ϕ_i	ψ_i
1	100	50	0	0	0
2	100	50	0	$U[-0.5, 0.5]$	$U[-0.5, 0.5]$
3	100	50	$U[-1, 3]$	0	0
4	100	50	$U[-1, 3]$	$U[-0.5, 0.5]$	$U[-0.5, 0.5]$
5	25	200	0	0	0
6	25	200	0	$U[-0.5, 0.5]$	$U[-0.5, 0.5]$
7	25	200	$U[-1, 3]$	0	0
8	25	200	$U[-1, 3]$	$U[-0.5, 0.5]$	$U[-0.5, 0.5]$

where the individual effects $\mu_i \sim N(0, 1)$. For the common factor we let $f_t = 0.5f_{t-1} + v_{i,t}$, where $v_{i,t} \sim$ i.i.d. $N(0, 1)$. For the factor loadings λ_i , we take either $\lambda_i = 0$, in which case there is no cross-sectional dependence, or $\lambda_i \sim U[-1, 3]$, in which case the cross-sectional dependence is generated by a factor structure. The idiosyncratic components $w_{i,t}$ are modelled as an autoregressive moving-average (ARMA) process, $w_{i,t} = \phi_i w_{i,t-1} + \varepsilon_{i,t} + \psi_i \varepsilon_{i,t-1}$. Here, we take $\phi_i = 0$ or $\phi_i \sim U[-0.5, 0.5]$ for all i , and $\psi_i = 0$ or $\psi_i \sim U[-0.5, 0.5]$ for all i . Finally, we take $\varepsilon_{i,t} \sim$ i.i.d. $N(0, 1)$.⁹

We consider a sample size of $T = 100$ and $N = 50$, which represents a large macro-panel, and a sample size of $T = 25$ and $N = 200$, which represents a micro-panel.¹⁰ Furthermore, for the stationary units in the panel, $i \leq k_0$, where $k_0 = [q_0 N]$, we take $\rho_{i,T} \sim U[0, 1 - cT^{-1}]$ with $c = 10$, and for $i > k_0$, we take $\rho_{i,T} = 1$.¹¹ As the true proportions of stationary units, we take $q_0 = 0, 0.2, 0.5, 0.9$. Table I gives all the combinations of parameters we use. All computations were performed in GAUSS 10 and are based on 1000 Monte Carlo simulations and 499 bootstrap replications.

We consider the BSQT procedure with four equally spaced quantiles to be tested for the macro-panel where $N = 50$, and eight equally spaced quantiles for the micro-panel where $N = 200$. Next to the BSQT method, we consider the bootstrap approach of Moon and Perron (2012), denoted by MP, the approach proposed by Chortareas and Kapetanios (2009), denoted by SPSM, and the approach proposed by Ng (2008), denoted by Ng. For the MP method, we apply the block bootstrap based on residuals as described earlier. Although Moon and Perron (2012) propose their method with the block bootstrap based on first differences, we modify their method to prevent any differences found being caused by differences in bootstrap method. For the same reason, we correct for cross-sectional dependence using the residual-based block bootstrap in the sequential approach by Chortareas and Kapetanios (2009). For BSQT and MP, we use the ADF t -test with ordinary least squares (OLS) demeaning. As in Moon and Perron (2012), lag lengths were selected by the modified Akaike information criterion (MAIC) of Ng and Perron (2001) with a maximum of four lags. Block lengths are taken as $l = 1.75T^{1/3}$ as in Palm *et al.* (2011). The level of all tests (or FDR when appropriate) is taken to be 5%.

Remark 6. With the ARMA specification with $\phi_i \sim U[-0.5, 0.5]$ and $\psi_i \sim U[-0.5, 0.5]$, there will be several cases where the parameters are close to zero or the roots cancel out. Furthermore, the random choice of these parameters (and the factor loadings as well) will lead to fluctuations across different DGPs, making the direct

⁹ Our DGP could be extended to allow for dynamic cross-dependencies, cointegration and so forth, as Palm *et al.* (2011) 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 analyse the finite sample performance of the methods discussed here given the focus on the (non)stationarity properties of individual units.

¹⁰ We also performed a similar simulation study for small N panels, which showed that the BSQT method applied unit-by-unit performs well in that setting. This simulation study can be found as Supporting Information on the journal website. For more simulation results including variations of the tests and different criteria of evaluation, also see Smeekes (2011).

¹¹ The choice to generate the first k_0 units as stationary is arbitrary as the method does not depend at all on the ordering of the cross-sectional units.

effects of serial correlation more difficult to interpret. The goal here however is not to analyse the direct effects of serial correlation but rather to analyse the performance of the methods if the units all have different forms of serial correlation, in particular in light of Remark 2. The effects of serial correlation in itself have been analysed by Paparoditis and Politis (2003) and Palm *et al.* (2011). Simulations with more conventional ARMA parameters in the current setup yielded qualitatively similar results and are available on request.

4.2. Simulation results

Results are given in Tables II and III. As the actual classification of the units into I(0) and I(1) units is the most important aspect of our method, this is the focus of the simulation study. We report the average proportion of units incorrectly found to be stationary (as ‘size’ measure), denoted by ICP, and the average proportion of correctly found stationary units (as ‘power’ measure), denoted by CP. These are the same measures as reported in Chortareas and Kapetanios (2009). To allow a comparison to the multiple testing literature, we also report the FDR. However, note that FDR is not very suitable for judging the performance of the BSQT method in particular, as that cannot select any number of stationary units but is restricted to the user-defined quantiles. As such, we also report the

Table II. Simulation results, part 1

DGP		$q_0 = 0$				$q_0 = 0.2$			
		BSQT	MP	SPSM	Ng	BSQT	MP	SPSM	Ng
1	ICP	0.005	0.001	0.000	0.054	0.079	0.008	0.001	0.068
	CP	—	—	—	—	0.798	0.752	0.441	0.564
	k -FWE	0.018	0.000	0.000	0.059	0.000	0.000	0.000	0.048
	FDR	0.018	0.039	0.005	0.384	0.242	0.034	0.006	0.155
2	ICP	0.007	0.001	0.001	0.075	0.082	0.010	0.003	0.084
	CP	—	—	—	—	0.834	0.767	0.490	0.609
	k -FWE	0.028	0.000	0.000	0.104	0.001	0.000	0.000	0.051
	FDR	0.028	0.041	0.017	0.430	0.251	0.043	0.014	0.185
3	ICP	0.013	0.001	0.007	0.010	0.023	0.008	0.012	0.008
	CP	—	—	—	—	0.076	0.606	0.096	0.077
	k -FWE	0.028	0.002	0.013	0.009	0.028	0.008	0.016	0.003
	FDR	0.028	0.023	0.028	0.066	0.041	0.021	0.024	0.020
4	ICP	0.013	0.001	0.005	0.005	0.026	0.002	0.007	0.013
	CP	—	—	—	—	0.142	0.315	0.082	0.100
	k -FWE	0.041	0.000	0.009	0.003	0.012	0.000	0.002	0.007
	FDR	0.041	0.028	0.033	0.047	0.069	0.017	0.024	0.034
5	ICP	0.001	0.000	0.000	0.000	0.038	0.001	0.000	0.002
	CP	—	—	—	—	0.466	0.089	0.025	0.007
	k -FWE	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.003
	FDR	0.007	0.009	0.000	0.001	0.240	0.016	0.008	0.004
6	ICP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	CP	—	—	—	—	0.001	0.003	0.000	0.000
	k -FWE	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	FDR	0.000	0.005	0.000	0.000	0.001	0.011	0.000	0.000
7	ICP	0.015	0.001	0.012	0.001	0.017	0.001	0.013	0.005
	CP	—	—	—	—	0.090	0.031	0.031	0.007
	k -FWE	0.052	0.004	0.034	0.002	0.029	0.001	0.030	0.007
	FDR	0.052	0.010	0.056	0.002	0.060	0.009	0.045	0.006
8	ICP	0.009	0.001	0.005	0.001	0.014	0.003	0.010	0.004
	CP	—	—	—	—	0.064	0.034	0.018	0.008
	k -FWE	0.046	0.001	0.015	0.001	0.024	0.004	0.023	0.006
	FDR	0.046	0.114	0.044	0.002	0.057	0.071	0.045	0.006

Note: ICP and CP refer to respectively the average proportion of units incorrectly and correctly classified as stationary. The construction of k -FWE is explained in Section 4.2.

Table III. Simulation results, part 2

DGP		$q_0 = 0.5$				$q_0 = 0.9$			
		BSQT	MP	SPSM	Ng	BSQT	MP	SPSM	Ng
1	ICP	0.078	0.037	0.014	0.107	0.283	0.431	0.070	0.264
	CP	0.928	0.870	0.732	0.819	0.884	0.991	0.904	0.952
	k -FWE	0.021	0.000	0.000	0.012	0.000	0.000	0.000	0.000
	FDR	0.074	0.038	0.017	0.088	0.029	0.044	0.008	0.028
2	ICP	0.079	0.042	0.014	0.128	0.280	0.292	0.081	0.316
	CP	0.938	0.909	0.774	0.832	0.878	0.955	0.857	0.946
	k -FWE	0.041	0.000	0.000	0.026	0.000	0.000	0.000	0.000
	FDR	0.072	0.042	0.016	0.101	0.029	0.031	0.010	0.034
3	ICP	0.038	0.024	0.011	0.033	0.155	0.267	0.043	0.188
	CP	0.638	0.799	0.375	0.253	0.775	0.848	0.638	0.782
	k -FWE	0.031	0.006	0.004	0.004	0.000	0.000	0.000	0.000
	FDR	0.031	0.022	0.010	0.026	0.017	0.028	0.006	0.021
4	ICP	0.061	0.015	0.016	0.052	0.214	0.244	0.061	0.250
	CP	0.643	0.528	0.354	0.360	0.849	0.867	0.772	0.891
	k -FWE	0.024	0.000	0.003	0.007	0.000	0.000	0.000	0.000
	FDR	0.060	0.021	0.021	0.047	0.023	0.026	0.008	0.027
5	ICP	0.057	0.006	0.013	0.029	0.105	0.017	0.063	0.378
	CP	0.533	0.257	0.338	0.210	0.631	0.398	0.561	0.769
	k -FWE	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000
	FDR	0.091	0.020	0.035	0.039	0.018	0.005	0.012	0.044
6	ICP	0.033	0.000	0.001	0.003	0.058	0.001	0.057	0.257
	CP	0.160	0.007	0.009	0.004	0.245	0.021	0.244	0.492
	k -FWE	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	FDR	0.131	0.005	0.017	0.004	0.025	0.002	0.024	0.039
7	ICP	0.020	0.002	0.013	0.022	0.030	0.007	0.022	0.294
	CP	0.177	0.055	0.068	0.071	0.280	0.141	0.240	0.513
	k -FWE	0.018	0.001	0.014	0.027	0.000	0.000	0.000	0.000
	FDR	0.048	0.009	0.031	0.028	0.009	0.003	0.007	0.035
8	ICP	0.021	0.003	0.015	0.013	0.039	0.009	0.033	0.289
	CP	0.163	0.063	0.070	0.043	0.281	0.133	0.261	0.515
	k -FWE	0.016	0.001	0.013	0.012	0.000	0.000	0.000	0.000
	FDR	0.052	0.031	0.042	0.016	0.013	0.009	0.011	0.038

Note: See Table II.

k -FWE, which is defined as the probability of making at least k false rejections. The asymptotic results and subsequent discussion show that we may reasonably expect the BSQT method to select q_{j+1} if $q_j < q_0 < q_{j+1}$, but that it should only select q_{j+2} with at most probability α . Therefore, we take $k = \max_j (k_{j+1} - k_0) I(k_j < k_0 < k_{j+1})$. Hence, k -FWE chosen in this way is a measure for BSQT of how often a proportion larger than q_{j+1} is selected.

The BSQT method performs well for most DGPs; in general, it has a fairly low probability of classifying units as $I(0)$ incorrectly, while it has reasonable power. Only for the macro-panel with $q_0 = 0.9$ ICP takes high values, but this is due to the design of the quantiles. Here, BSQT can select either 0.75 or 1; hence, it is expected that the method will regularly select too many stationary units. In general, the results for k -FWE show that BSQT does not select more stationary units that we may expect (i.e. it does not often take $\hat{q} = q_{j+2}$ or higher if $q_j < q_0 \leq q_{j+1}$). For the micro-panel, the power is significantly lower, which is to be expected given the rather small number of time series observations, although the BSQT method still manages to pick up a fair proportion of the stationary units.

The MP method performs very well for the macro-panel; it is able to identify most of the stationary units, has a small probability of an incorrect $I(0)$ classification and controls FDR as desired. However, for the micro-panel, while FDR is still controlled and ICP is still low, the proportion of stationary units picked up decreases quite

strongly. This is as expected, 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. Note that in the micro-panel, BSQT picks up a considerably higher proportion of the stationary units than MP owing to its ability to exploit the cross-sectional dimension.

The SPSM method performs well in terms of size, while providing mixed results on power. It does show the ability to exploit the cross-sectional dimension (the test statistic being the average over the cross-section), as it retains a reasonable power in the micro-panel case, in particular if q_0 is high, although its power is still lower than that of BSQT. However, for small to medium q_0 , SPSM has a fairly low probability of identifying the stationary units and is clearly less powerful than the BSQT method. This is due to the pooled form of the estimator as discussed in Remark 3. For large q_0 , the form of the test statistic is beneficial to SPSM: if most units are $I(0)$, the influence of the $I(1)$ units on the average is very small, and the statistic will clearly indicate a rejection. Hence, for $q_0 = 0.9$, in the early stages of the sequential algorithm, SPSM will have strong evidence against the unit root, and therefore, the algorithm is unlikely to get 'stuck' in the early stages. In general, we can therefore say that SPSM is outperformed by BSQT with the exception of the macro-panel with $q_0 = 0.9$, where it has a much better probability of correct $I(0)$ classification while having similar probability of incorrect $I(0)$ classification.

The Ng method does not seem to perform as well as the other methods. In general, it has a higher probability of classifying an $I(1)$ unit as $I(0)$, while it does not pair this necessarily with a higher probability of a correct $I(0)$ classification. While the estimator of the proportion of stationary units itself performs reasonably, classifying the units is less successful.¹² The exceptions are the panels with $q_0 = 0.9$; however, for the macro-panel, it is still inferior to SPSM, while for the micro-panel, the high CP rate is accompanied by a high ICP rate.

Both cross-sectional dependence and serial correlation appear to have a negative effect on the methods' ability to pick up the stationary units, although the extent of the effect depends on the specific DGP and the specific method. The randomly chosen ARMA parameters and factor loadings add a further source of variability that may explain the fluctuations seen across the different DGPs (though not between methods, which are all applied to the same data). Several patterns are clearly visible though. All methods pick up fewer stationary units if there is a common factor; the effect of common factors is a familiar feature in the analysis of power of panel unit root tests and is caused by the reduced effect of pooling the data in the presence of common factors. This effect is magnified in the micro-panel as pooling is even more important there for power given the small number of time series observations, and the pooling is less effective with the strong dependence across the units. Also, serial correlation has a negative effect on power; adding more lags to deal with the serial correlation has a negative effect on power. In the micro-panel, this is most pronounced as the effect of a larger lag length is amplified through the small T dimension. One might therefore in such panels choose to restrict the allowed lag lengths and find other ways to make the marginal distributions equal, such as by scaling the bootstrap critical values (see Section 5) or by using p -values.

Concluding, BSQT performs well in panels with large N , even if T is small. In panels where T is large and N relatively large, the BSQT method performs nearly as well as the MP method. In panels with small T but large N , BSQT keeps performing reasonably well, whereas the performance of the MP method deteriorates significantly. Overall, the BSQT method also performs better than SPSM. The Ng method is dominated by all the other methods across all DGPs considered. In practice, the BSQT method therefore seems the most useful if T is relatively small compared to N . If T is large compared to N , the MP method is the best method. For such panels, the BSQT method remains to do reasonably well however, whereas if T is small, the MP method seriously lacks power and cannot be used reliably.

5. EMPIRICAL APPLICATION

We consider an application based on income data from the Panel Study of Income Dynamics (PSID). Pesaran (2007b) assesses the validity of the claim of Meghir and Pistaferri (2004) that the log of real earnings of households

¹² Unreported simulations show that on average the Ng estimator is close to the true proportion, although it does have a fairly large variance.

in the 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 dropouts (HSD, $N = 36$). As $T = 22$, this is a typical example of large N , small T ‘micro’ panel.

Pesaran (2007b) finds mixed evidence regarding the unit root; the CIPS rejects for the full sample, but not for all subsamples.¹³ This might lead to the conclusion that there is a relation between education level and stationarity properties, and that the rejection for the sub-group 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 depends on a variety of ‘coincidental’ factors.

There is however no way to find out which of the assertions is true with standard panel unit root tests only. Therefore, we apply several methods designed to determine the proportion of stationary units to the data used in Pesaran (2007b).¹⁴ Next to the BSQT approach, we consider the bootstrap approach of Romano *et al.* (2008) to control for FDR employed in Moon and Perron (2012) (MP) and the sequential approach of Chortareas and Kapetanios (2009) with the block bootstrap employed to deal with cross-sectional dependence (SPSM). 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 size (or FDR) is set to 5%. We do not consider the method by Ng (2008), as the fundamental differences with the other methods considered makes it difficult to compare results directly and to determine which aspect of the methods causes the differences. Moreover, the method by Ng (2008) turned out to be extremely sensitive to the exact specification in our application, which requires a deeper investigation than what is desired here given it is not the central method of interest.

For the individual unit root test statistics θ_i , we consider the union of rejections statistic proposed by Smeekes and Taylor (2012), which has the form of a scaled minimum of four statistics¹⁵:

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

DF-OLS_{*i*} and DF-QD_{*i*} are the ADF test performed on unit *i* with OLS and quasi-differenced (QD) detrending respectively, while superscript μ and τ indicate whether demeaning or detrending is used respectively. The bootstrap critical values such as $c_{i,QD}^{\mu*}(\alpha)$ used in the scaling factors are determined in a preliminary bootstrap step as the individual level α critical values of the four tests. This test statistic is the bootstrap adaptation of the method proposed by Harvey *et al.* (2012) of rejecting the null if either one of the four tests rejects (corrected to have the correct size). This test can therefore deal with uncertainty about the initial condition and the presence of deterministic trends, thus obviating the need to report multiple tests, which is subject to problems with size control and interpretation.

In a panel context, this becomes even more important, as some units may exhibit linear trends while others do not. In such a case, it would be most beneficial to the power of the methods to only include the linear trend for those units that actually contain a 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 testing framework is the use of the scaling factor x_i . Smeekes and Taylor (2012) take $x_i = c_{i,QD}^{\mu*}$, which scales the test statistics of unit *i* towards the critical values of the DF-QD_{*i*} ^{μ} statistic. However, as they argue in Remark 6, any $x_i < 0$ suffices. We therefore take $x_i = -1$ for all $i = 1, \dots, N$, which

¹³ Rejections for sub-groups depend on the specification of the CIPS test; see Pesaran (2007b) for details.

¹⁴ The dataset used in Pesaran (2007b) is available from the Journal of Applied Econometrics Data Archive (www.econ.queensu.ca/jae).

¹⁵ We use the ‘A’ version of their statistic, which is asymptotically valid.

Table IV. Rejections of a unit root in log real household earnings

	BSQT	MP	SPSM
Total rejections ($N = 181$)	60	2	41
Proportion of rejections	0.33	0.01	0.23
Rejections in subsamples			
CLG ($N = 58$)	16	2	12
Proportion	0.28	0.03	0.21
HSG ($N = 87$)	29	0	21
Proportion	0.33	0.00	0.24
HSD ($N = 36$)	15	0	8
Proportion	0.42	0.00	0.22

Note: Rejections are obtained using a 5% level/FDR.

means that all statistics are scaled towards -1 . This step further ensures that the marginal distributions of the θ_i are ‘as free of nuisance parameters as possible’ as discussed in Remark 2, as the statistics are not scaled towards individual critical values but to a fixed constant.¹⁶

As quantiles to be tested for the BSQT method, we take $q_j = \frac{j-1}{6}$, $j = 1, \dots, 6$. Our motivation for these quantiles is that the sub-groups take up approximately 1/2, 1/6 and 1/3 of the sample. As such, we can cover whole sub-groups with this choice of quantiles, which makes it possible to investigate if the stationarity properties of the units are indeed dependent on the properties of the sub-groups they are in. We take a block size of 5 (corresponding to $1.75T^{1/3}$),¹⁷ while we select lags using MAIC with the Perron and Qu (2007) modification with a maximum lag length of three to avoid losing too many observations. Results are based on 4999 bootstrap replications, and all calculations were performed in GAUSS 10. The code to conduct the BSQT method in combination with unions of ADF statistics is available from <http://researchers-sbe.unimaas.nl/stephansmEEKES/code/>.

Table IV presents the results. The BSQT method indicates that about one third of the units is stationary. As expected from the power results in the simulation study, the SPSM method classifies fewer units as stationary 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 (2007b), nor with the methods mentioned earlier.¹⁸ 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 and SPSM methods can overcome the small T by exploiting the cross-sectional dimension, which the MP method is unable to do.

The results for the subsamples are obtained by ordering the units for which rejections were found in the complete sample into the three sub-groups.¹⁹ We find 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.

Remark 7. The computation time required for this applications gives a good overview of the computational intensity of the methods. Barring the initial bootstrap step needed to construct the union test statistics (which is common to all methods), BSQT and SPSM took about 4 min to run, while the MP method took well over an hour

¹⁶ The tests were also performed after transforming to p -values, which does not change the results.

¹⁷ The results for BSQT are fairly robust to the block length. Data-dependent block length selection remains an open issue for nonstationary time series. Palm *et al.* (2011, Section 5.3) find that the method of Giacomini *et al.* (2013) performs reasonably well, whereas Moon and Perron (2012) use a plug-in rule for bandwidth selection in variance estimation. However, the optimality properties of these methods with I(1) data are unknown. For applications, it is therefore advisable to investigate the sensitivity of the methods to different fixed block lengths.

¹⁸ The exact number of rejections for the MP method is sensitive to block lengths, number of lags included and even the seed of the random number generator. However, regardless of these parameters, the number of rejections remains rather small. This sensitivity is in itself another undesirable feature in such small T , large N panels.

¹⁹ It is unnecessary to apply the methods to each subsample separately, as the stationary units found from the whole sample can be split according to the subsample they belong to. Moreover, it would actually harm the methods as the decrease in the cross-sectional dimension would reduce power for BSQT and SPSM.

(on a Windows 7 computer with an Intel Core 2 Duo 3.00 GHz processor). The high computation time for the MP method is not due to the actual bootstrapping but due to the complicated way that the critical values are determined. The BSQT method is much simpler to implement and, in particular if N is large, much less time consuming.

6. CONCLUSION

We have proposed a new method 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 (but finite) 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.

Through a simulation study, we demonstrated the good performance of this method in finite samples, where the sequential method performed well in comparison to alternative methods. It has a significant advantage over multiple testing approaches if N is rather large and T is small. The sequential method based on quantiles proposed here is also shown to be more powerful than sequential methods based on averages. In practice, therefore, the BSQT method is particularly useful if T is small relative to N , a situation in which most other currently available methods fail.

We also illustrated the method by an empirical application of testing for unit roots in log earnings data of households. This application with large N and small T again demonstrated the usefulness of this method.

The method developed in this article is not restricted to unit root testing in panels; this sequential approach 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. Moreover, the method is simple to implement, unlike some of the multiple testing procedures.

There are several limitations and extensions to this article. The most important limitation is that we only consider T asymptotics owing to the non-availability of joint asymptotics for the bootstrap in this setup. In particular, because it is argued that our method does well for large (but in our analysis invariably finite) N , it would be important to develop joint asymptotics for the bootstrap such that the large N case can be analysed properly. Another extension would be to consider different unit root tests or even stationarity tests as mentioned in the paper, as such tests might have other, more favourable properties when applied in a sequential context. A final extension worth mentioning is to develop a complete strategy to analyse the problem of unit roots in panel data as hinted at by Pesaran (2012), consisting of starting with a conventional panel unit root test followed by, in case of a rejection, a method to classify the individual units. As not only several methods for classifying units but even considerably more panel unit root tests are available, many different combinations would be possible, and a study into these combinations could be very interesting.

ACKNOWLEDGEMENTS

This research was supported financially by the Netherlands Organisation for Scientific Research (NWO). I am grateful to two anonymous referees, Eric Beutner, Stefano Fachin, Franz Palm, Robert Taylor and Jean-Pierre Urbain as well as participants at the Panel Data Conference 2010 in Amsterdam for helpful comments and suggestions.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

REFERENCES

- Chang Y. 2002. Nonlinear IV unit root tests in panels with cross-sectional dependency. *Journal of Econometrics* **110**: 261–292.
Chang Y, Song W. 2009. Testing for unit roots in small panels with short-run and long-run cross-sectional dependencies. *Review of Economic Studies* **76**: 903–935.

- Chortareas G, Kapetanios G. 2009. Getting PPP right: identifying mean-reverting real exchange rates in panels. *Journal of Banking and Finance* **33**: 390–404.
- Deckers T, Hanck C. 2014. Multiple testing for output convergence. *Macroeconomic Dynamics* **18**: 199–214.
- Di Iorio F, Fachin S. 2009. A residual-based bootstrap test for panel cointegration. *Economics Bulletin* **29**: 3222–3232.
- Giacomini R, Politis DN, White H. 2013. A Warp-Speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory* **29**: 567–589.
- Hanck C. 2009. For which countries did PPP hold? A multiple testing approach. *Empirical Economics* **37**: 93–103.
- Harvey DI, Leybourne SJ, Taylor AMR. 2012. Testing for unit roots in the presence of uncertainty over both the trend and initial condition. *Journal of Econometrics* **169**: 188–195.
- Im KS, Pesaran MH, Shin Y. 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*.
- Meghir C, Pistaferri L. 2004. Income variance dynamics and heterogeneity. *Econometrica* **72**: 1–32.
- Moon HR, Perron B. 2012. Beyond panel unit root tests: using multiple testing to determine the non stationarity properties of individual series in a panel. *Journal of Econometrics* **169**: 29–33.
- Ng S. 2008. A simple test for nonstationarity in mixed panels. *Journal of Business and Economic Statistics* **26**: 113–127.
- Ng S, Perron P. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* **69**: 1519–1554.
- Palm FC, Smeekes S, Urbain JP. 2011. Cross-sectional dependence robust block bootstrap panel unit root tests. *Journal of Econometrics* **163**: 85–104.
- Paparoditis E, Politis DN. 2003. Residual-based block bootstrap for unit root testing. *Econometrica* **71**: 813–855.
- Perron P, Qu Z. 2007. A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters* **94**: 12–19.
- Pesaran MH. 2007a. A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* **138**: 312–355.
- Pesaran MH. 2007b. A simple panel unit root test in the presence of cross-sectional dependence. *Journal of Applied Econometrics* **22**: 265–312.
- Pesaran MH. 2012. On the interpretation of panel unit root tests. *Economics Letters* **116**: 545–546.
- Romano JP, Shaikh AM, Wolf M. 2008. Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test* **17**: 417–442.
- Romano JP, Wolf M. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* **73**: 1237–1282.
- Shaked M, Shanthikumar JG. 2007. *Stochastic Orders*. New York: Springer.
- Smeekes S. 2011. Bootstrap sequential tests to determine the stationary units in a panel. *METEOR Research Memorandum 11/003, Maastricht University*.
- Smeekes S, Taylor AMR. 2012. Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory* **28**: 422–456.
- Swensen AR. 2006. Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* **74**: 1699–1714.
- Westerlund J, Breitung J. 2013. Lessons from a decade of IPS and LLC. *Econometric Reviews* **32**: 547–591.
- White H. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.

APPENDIX A: PROOFS

Proof of Theorem 1

Note that $\theta_i \xrightarrow{P} -\infty$ for $i \in \mathcal{S}$ as $T \rightarrow \infty$ by Assumption 2. Then

$$\theta_{(1)}, \dots, \theta_{(k_0)} \xrightarrow{P} -\infty. \quad (\text{A1})$$

If $q_{j+1} \leq q_0$ and so $k_{j+1} \leq k_0$, then also $\theta_{(k_{j+1})} \xrightarrow{P} -\infty$, which proves (4a).

If $k_{j+1} > k_0$, then by (A1), all order statistics below k_0 have no effect on the distribution of the remaining order statistics. Therefore, we only have to consider those θ_i for which $i \in \mathcal{U}_{k_0}$, and hence $\theta_{(k_{j+1})} \xrightarrow{P} \theta_{(k_{j+1}-k_0; \mathcal{U}_{k_0})}$. The result in (4b) then follows directly from the continuous mapping theorem (cf. White, 2000, Lem 2).

The result in (5a) follows directly from Paparoditis and Politis (2003) and Palm *et al.* (2011), as their proofs of asymptotic validity for the DF coefficient test can straightforwardly be extended to other unit root statistics. The

crucial result for 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.* (2011). As this result still applies here, it can be used to establish the asymptotic validity of other unit root test statistics.

The result in (5b) follows from the fact that all bootstrap units are generated with a unit root, and the application of the continuous mapping theorem. Finally, (5c) follows as (A1) implies all I(0) units will be ranked ahead of the I(1) units and thus must have led to rejections in the previous steps by (4a) coupled with (5b). \square

Proof of Corollary 1

Equation (6a) directly follows from (4a) coupled with (5b). To prove (6b), note that for $q_j < q_0 < q_{j+1}$,

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

where the third equality follows from (A1). Equation (6c) is the probability of not rejecting the true null hypothesis $H_0(q_j)$, which is equal to $1 - \alpha$ by (5c). To show (6d), note that if $q_{j-1} < q_0 < q_j$,

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

which follows in the same way as (6b). To show (6e), assume without loss of generality that $q_{j-1} = q_0$. The result then follows from the fact that, asymptotically,

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

The middle inequality in (8) follows if $P\{\xi_{(k_{j+1}-m:\mathcal{U}_m)} < x\} \leq P\{\xi_{(k_{j+1}-m-1:\mathcal{U}_{m+1})} < x\}$ for any $x \in \mathbb{R}$, which is equivalent to $\xi_{(k_{j+1}-m-1:\mathcal{U}_{m+1})} \leq_{st} \xi_{(k_{j+1}-m:\mathcal{U}_m)}$, where ‘ $X \leq_{st} Y$ ’ denotes that X is smaller in the usual stochastic order than Y (cf. Shaked and Shanthikumar, 2007, p. 3). This 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 (8) then follows by setting $m = k_j$ along with (7), while the last follows by definition. \square