CrossMark

Panel bootstrap tests of slope homogeneity

Johan Blomquist[1] · Joakim Westerlund[2,3]

Abstract This paper proposes two bootstrap-based tests that can be used to infer whether the individual slopes in a panel regression model are homogenous. The first test is suitable when wanting to infer the null of homogeneity versus the general alternative, while the second is suitable when wanting to infer the units of the panel that can be pooled. Both approaches are shown to be asymptotically valid, a property that is verified in small samples using Monte Carlo simulation.

1 Introduction

Most common panel regression models, such as the random and fixed-effects models, assume that the regression slopes are equal across the cross section. In fact, in some

✉ Joakim Westerlund
joakim.westerlund@nek.lu.se

[1] AgriFood Economics Center, Department of Economics, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2] Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden

[3] Financial Econometrics Group, Centre for Research in Economics and Financial Econometrics, Deakin University, Geelong, Australia

🖄 Springer

parts of the literature the equal slope assumption is so common that it is hardly ever questioned. The main reason for this is twofold (see, e.g., Baltagi et al. 2008). First, it greatly simplifies the process of estimation and inference. In fact, as a referee to this journal points out, the main feature of panel data is the ability to selectively pool the information regarding the assumed common slope, while allowing great heterogeneity in other parts of the model. Second, if true exploiting the equality of the slopes leads to large gains in efficiency. The problem is that if the assumption is false, panel techniques based on models with equal slopes become inconsistent, causing misleading inference (see, e.g., Hsiao 2003, Chapter 6). It is therefore crucial to test for slope homogeneity before proceeding with an analysis based on this assumption.

The classical approach to test the equal slope assumption is to employ a simple F-test (see, e.g., Hsiao and Pesaran 2008; Baltagi et al. 2008; Pesaran and Yamagata 2008). Alternatively, one may use the Hausman test of Pesaran et al. (1996), in which two estimators are compared, one is constructed under the assumption of homogenous slopes, while the other is not. A third possibility is to use the Swamy-type test of Pesaran and Yamagata (2008), which is based on the dispersion of individual slope estimates from a suitable pooled estimator. Although very popular, these approaches suffer from at least two important shortcomings.

First, all three tests assume that the regression errors are cross-sectional independent, a restriction that is likely to be violated in practice, especially in macroeconomic and financial applications involving country-level data where strong intraeconomy linkages can be expected. Recognizing this shortcoming, Phillips and Sul (2003) propose a Hausman-type test that is appropriate in the special case when the dependence can be represented by means of a single common factor. The problem is, of course, that the common factor structure need not be correct, which would then invalidate the test. Another possibility, which allows for more general forms of cross-sectional dependence, is to use the seemingly unrelated regression (SUR) approach. The advantage of this approach is that, as long as the cross-sectional dimension, N, is "small" and the time series dimension, T, is "large," the cross-sectional dependence can be allowed to be quite general. The drawback is that unless $T >> N$, the small-sample performance of SUR-based tests is often very poor (see Bun 2004).

The second shortcoming relates to the formulation of the hypotheses tested, which, depending on the problem at hand, may not be very informative. In particular, while the null hypothesis can certainly be formulated as that all the slopes are equal, the alternative hypothesis that there are at least some units for which homogeneity fails is typically too broad for any interesting economic conclusions; it could be that the slopes of all units are different, but it could also be that there is only a small fraction of units for which homogeneity fails. Some studies have considered clustering methods for grouping units such that the slopes are homogeneous within each group, but heterogenous across groups (see, e.g., Kapetanios 2006; Lin and Ng 2012). Unfortunately, these methods are quite complicated to implement, which is probably also the reason for why they are almost never used in practice.

In this paper, we develop several procedures to ascertain the homogeneity of a panel. The point of departure is a quite general data-generating process that allows, for example, serial correlation of unknown form and complex cross-sectional dependencies such as dynamic common factor models. In fact, except for some mild regulatory

conditions, there are virtually no restrictions on the forms of serial and cross-sectional dependence that can be permitted. We do require that the regressors are strictly exogenous, but the approach can be easily extended to allow more general types of regressors. Given this generality, corrections aimed at achieving asymptotically pivotal statistics are not really an option. In this paper, we therefore consider the block bootstrap as a means to obtain tests that are asymptotically valid. Two block bootstrap-based test procedures are considered: one is appropriate when testing the above mentioned hypothesis of slope homogeneity versus at least some heterogeneity, while the other can be used to sequentially determine the units for which the slopes are equal. Both procedures are easy to implement and work well even when N and T are similar in magnitude.

The plan of the rest of the paper is as follows. Section 2 describes the model and assumptions, which are used in Sect. 3 to study the properties of the two test procedures considered. Section 4 presents the results of a small Monte Carlo study. Section 5 concludes.

2 Model and assumptions

Let $y_{i,t}$ be generated as

$$y_{i,t} = \theta_i + \beta_i' x_{i,t} + \varepsilon_{i,t}, \tag{1}$$

where θ_i is regarded as a unit-specific intercept or fixed effect, $x_{i,t}$ is a $m \times 1$ vector of regressors, and β_i is a conformable vector of unknown slope coefficients. The purpose of this paper is to infer to which extent β_i can be regarded as equal across i.

The assumptions that we will be working under are stated below, where $\to_p$ and $\to_d$ signifies convergence in probability and distribution, respectively, and $C \in (0, \infty)$ is a generic constant not depending on N or T. Let $Q_{i,T} = T^{-1}(x_i' M_\tau x_i)$, where $x_i = (x_{i,1}, \ldots, x_{i,T})'$ and $M_\tau = I_T - \tau_T(\tau_T' \tau_T)^{-1} \tau_T'$, I_T is the $T \times T$ identity matrix, and $\tau_T = (1, \ldots, 1)'$ is a $T \times 1$ vector of ones. It is also convenient to let $\tilde{z}_t = (\tilde{z}_{1,t}', \ldots, \tilde{z}_{N,t}')'$, where $\tilde{z}_{i,t} = (x_{i,t} - \mu_i)\varepsilon_{i,t}$ and $\mu_i = E(x_{i,t})$. The sigma-field generated by $\{\varepsilon_{i,n}\}_{n=1}^t$ ($\{x_{i,n}\}_{n=1}^t$) is henceforth going to be denoted $\mathcal{F}_{\varepsilon i,t}$ ($\mathcal{F}_{xi,t}$),

Assumption REGR. $Q_{i,T} \to_p Q_i$ as $T \to \infty$, where Q_i is a positive definite matrix.

Assumption ERR.

(i) $E(\varepsilon_{i,t} x_{j,n}) = 0$ for all i, t, j, and n;
(ii) $E(\varepsilon_{i,t}^2 | \mathcal{F}_{\varepsilon i,t-1} \cup \mathcal{F}_{xi,t}) = E(\varepsilon_{i,t}^2) = \sigma_i^2 \leq C$ for all t;
(iii) $\Sigma = \text{var}(T^{-1/2} \sum_{t=1}^T \tilde{z}_t)$ is positive definite;
(iv) $\{(x_{i,t}, \varepsilon_{i,t})\}_{t=1}^T$ is an α-mixing sequence with mixing coefficient $\alpha_i(n)$, which is such that $\sup_{i=1,\ldots,N} \alpha_i(n) \leq \alpha(n)$, where $\alpha(n) = Cn^{-a}$ for some $a > r/(r-2)$ and $r > 2$;
(v) $E(||x_{i,t}||^{2r+p}) \leq C$ and $E(|\varepsilon_{i,t}|^{2r+p}) \leq C$ for some $p > 0$, and all i and t.

Assumptions ERR (i) and (ii) rule out regressors that are non-strictly exogenous (including lagged dependent variables) and heteroskedasticity across time. As we dis-

cuss in Remark 1 of Sect. 3, these assumptions are stronger than needed. However, since the main motivation of the paper is to account for cross-sectional (and serial) correlation, we have preferred to keep Assumptions ERR (i) and (ii) as they stand (see, e.g., Kapetanios 2008; Hidalgo 2003, for similar assumptions). These assumptions are not needed for the specific tests that we propose, but for the validity of our residual-based bootstrap, which can be easily modified to accommodate both endogenous regressors and errors that are heteroskedastic across time. The reason for sticking with the residual-based bootstrap is that it makes for relatively straightforward proofs. Assumption ERR (iv) allows for heterogeneous forms of serial correlation across the cross section, but imposes a uniform bound on the mixing coefficients. Many commonly encountered stochastic processes can be accommodated under this assumption (see Davidson 1994, Section 14.4). Note also how Assumption ERR does not impose any particular cross-correlation structure. In particular, it is not necessary to know whether the dependence is weak or strong (see Chudik et al. 2011; Pesaran and Tosetti 2010). The dependence also does not have to be static but may be dynamic in nature, when $\varepsilon_{i,t}$ is generated by a dynamic common factor model. The moment requirements in Assumption ERR (v) rule out time trends in the regressors.

3 The bootstrap test procedures

Let us denote by q the number of units that cannot be pooled; that is, q is the number of units for which $\beta_i \neq \beta$, where β is the common value of β_i. The purpose of this paper is to make inference regarding q. Let us therefore denote by $0 = q_1 < \cdots < q_K < N$ a set of K user-defined numbers, representing the number non-poolable units to be considered in the testing. Let $H_0(q_k)$ denote the null hypothesis that $q = q_k$, where $k = 1, \ldots, K$, and let $H_1(q_{k+1})$ denote the alternative hypothesis that $q \geq q_{k+1}$.

3.1 A pooled test for testing $q = 0$ versus $q \geq 1$

In this subsection, we consider the relatively simple problem of testing $H_0(0)$ versus $H_1(1)$; that is, the null hypothesis of homogeneity ($q = 0$) is tested versus the alternative that there is at least one unit that cannot be pooled ($q \geq 1$). The reason for considering this testing problem separately is that under the null hypothesis the slopes of all the units are equal, which makes it possible to consider pooled test statistics in the spirit of much of the previous literature (see, e.g., Hsiao and Pesaran 2008; Baltagi et al. 2008; Pesaran and Yamagata 2008). The test statistic that we consider is a bootstrap Swamy-type homogeneity test and is given by

$$S = \sum_{i=1}^{N} T (\hat{\beta}_i - \hat{\beta}_{\text{WFE}})' \frac{Q_{i,T}}{\hat{\sigma}_i^2} (\hat{\beta}_i - \hat{\beta}_{\text{WFE}}),$$

where $\hat{\beta}_i$ is the least-squares (LS) estimator of β_i when applied to cross-sectional unit i, $\hat{\sigma}_i^2 = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2$, $\hat{\varepsilon}_{i,t} = y_{i,t} - \hat{\theta}_i - \hat{\beta}_i' x_{i,t}$, and $\hat{\beta}_{\text{WFE}}$ is the weighted fixed-effects (WFE) estimator, as given by

$$\hat{\beta}_{\text{WFE}} = \left(\sum_{i=1}^{N} \frac{x_i' M_\tau x_i}{\hat{\sigma}_i^2} \right)^{-1} \sum_{i=1}^{N} \frac{x_i' M_\tau y_i}{\hat{\sigma}_i^2}.$$

As Hsiao (2003) and Pesaran and Yamagata (2008) show, under H_0 and the assumption of independently identically distributed (iid) errors, $S \to_d \chi^2[m(N-1)]$ as $T \to \infty$ with N held fixed. However, under the more general assumptions laid out above, this is no longer the case. The approach opted for here is therefore based on the bootstrap.

Algorithm BOOT.

1. Estimate (1) using LS for each cross-sectional unit and organize the residuals, $\hat{\varepsilon}_{i,t}$, in a $T \times N$ matrix $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_T)'$, where $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1,t}, \ldots, \hat{\varepsilon}_{N,t})'$.
2. Obtain $\hat{\beta}_{\text{WFE}}$ and $\hat{\theta}_i = \bar{y}_i - \hat{\beta}_{\text{WFE}} \bar{x}_i$, where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{i,t}$ with a similar definition of $\bar{x}_i$.
3. Choose a block length, l. Let $J_t = (\hat{\varepsilon}_t, \hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+l-1})'$ be the block of l consecutive estimated errors starting at date t, and let $I_1, I_2, \ldots$ be a sequence of iid random variables with a discrete uniform distribution on $\{1, \ldots, (T-l+1)\}$. The $T \times N$ matrix of pseudo errors, ε^*, is such that the first l rows are determined by J_{I_1}, the next l rows are given by J_{I_2}, and so on. The procedure is stopped when T rows have been generated.
4. Simulate pseudo-data under H_0 as $y_{i,t}^* = \hat{\theta}_i + \hat{\beta}_{\text{WFE}}' x_{i,t} + \varepsilon_{i,t}^*$.
5. Compute the bootstrap test statistic, S^*, where S^* is calculated exactly as S, but with (y_i, x_i) replaced by (y_i^*, x_i).
6. Repeat steps 3–5 B times.
7. Select the bootstrap critical value as the $(1-\alpha)$-quantile of the ordered S^* statistics.

Remark 1 The above algorithm is an example of a residual-based bootstrap and is similar to the algorithms used in Bun (2004), Kapetanios (2008), Hidalgo (2003), and Zhou and Shao (2013), to mention a few. Alternatively, we may follow, for example, Freedman (1981) and Fitzenberger (1997) and block bootstrap $(y_{i,t}, x_{i,t}')$. Hence, instead of resampling $\hat{\varepsilon}_t$ in step 3 in Algorithm BOOT, we resample $(y_t', x_t')'$, a $N \times (1+m)$ matrix. The bootstrap test statistic is S based on (y_i^*, x_i^*), the bootstrapped version of (y_i, x_i). The main advantage of this resampling scheme is that it does not require homoskedasticity and strictly exogenous regressors.[1]

Remark 2 Algorithm BOOT requires a choice of block length, l. A common approach is to set l as a deterministic function of T. Alternatively, one may follow, for example, Gonçalves and White (2005) and Gonçalves (2011) and set l according to the data-dependent rule of Andrews (1991) or Newey and West (1994), originally proposed for the purpose of bandwidth selection in long-run variance estimation (see Fitzenberger 1997, Section 3.4, for a discussion). In Sect. 4, we use Monte Carlo simulation to evaluate the effect of various rules for selecting l.

---

[1] As pointed out by MacKinnon (2007), the advantage of bootstrapping $(y_{i,t}, x_{i,t}')$ rather than residuals often comes at the cost of poor small-sample performance, a finding that is supported by our preliminary Monte Carlo results.

The asymptotic validity of the bootstrap procedure requires the following assumption on the block length.

Assumption BL. $l \to \infty$ and $l = o(\sqrt{T})$ as $T \to \infty$.

Theorem 1 *Suppose that Assumptions REGR and ERR hold and that $T \to \infty$ with N held fixed. Under $H_0(0)$,*

$$S \to_d \sum_{j=1}^{kN} \lambda_j U_j^2,$$

where $\lambda_j \in [0, \infty)$ and $U_j \sim N(0, 1)$ independently across j. Under $H_1(1)$,

$$S \to \infty.$$

The corresponding result for S^* is given in Theorem 2.

Theorem 2 *Suppose that Assumptions REGR, ERR, and BL hold and that $T \to \infty$ with N held fixed. Under $H_0(0)$ and $H_1(1)$,*

$$S^* \to_{d*} \sum_{j=1}^{kN} \lambda_j U_j^2 \quad \text{in probability,}$$

where $\to_{d*}$ convergence in distribution conditional on the realization of the sample.

Remark 3 Together Theorems 1 and 2 establish the asymptotic validity of the proposed bootstrap test procedure. There are two requirements. First, S and S^* must converge to the same asymptotic null distribution. Second, under $H_1(1)$, while S^* should converge to the same asymptotic distribution as under $H_0(0)$, S should diverge.

Remark 4 The asymptotic results in Theorems 1 and 2 are based on letting $T \to \infty$, while keeping N fixed. There are several reasons for considering such a large-T and fixed-N asymptotic framework. First, if T is fixed, then the above resampling scheme will not be asymptotically valid. Also, N plays no role in the proofs of Theorems 1 and 2. As there is no resampling in the cross-sectional dimension, there is no reason to expect the performance of the bootstrap test to become more (or less) accurate when N increases.[2] Second, it is difficult to obtain distributional results for $N \to \infty$ without imposing additional conditions on the cross-sectional dependence structure.[3] Third, from a practical viewpoint, a test for slope equality may be of greatest interest in applications in which N is relatively small. In such cases, the condition that $N \to \infty$ may be difficult to justify. There are, of course, applications in which it is more appropriate to rely on large-N asymptotics, but then other test statistics are likely to

---

[2] An alternative bootstrap approach based on cross-sectional resampling has been suggested by Kapetanios (2008). Unlike the bootstrap approach used here, the cross-sectional resampling scheme provides asymptotically valid bootstrap procedures when $N \to \infty$ but T remains fixed, but this is based on cross-sectional independence.

[3] See Gonçalves (2011) for a set of assumptions that can be used in the case of arbitrary cross-sectional dependence and N asymptotics.

be more effective (see Pesaran and Yamagata 2008, for a discussion). However, in such large-N panels T is typically too small for estimation of individual slopes, and therefore, the analysis must be made conditional on the slopes being equal.

Remark 5 In contrast to, for example, Bun (2004), we make no attempt to obtain an asymptotically pivotal test statistics. Nevertheless, it should be noted that there is evidence that bootstrapping of (asymptotically) pivotal test statistics leads to asymptotic refinements (see, e.g., Davidson and MacKinnon 1999). Therefore, if such a test statistic could be obtained, it may be preferable to the approach taken here.

3.2 A sequential test procedure for determining q

The test considered in the previous section is appropriate if one wishes to infer whether there is any evidence against poolability at all. The problem is that in many cases one would like to go further than just concluding that $q > 0$ in case of a rejection, and in this section, we therefore consider a sequential test that can be used to pinpoint q. In so doing, we will assume that $q_k = k - 1$, where $k = 1, \ldots, K$ and $K = N - 1$, such that the number of units to be tested decreases by one at each iteration; later on we discuss how to proceed when $q_1, \ldots, q_K$ are set differently.

To test whether a particular unit i has the same slope coefficient vector as a certain benchmark unit b, we may use the following Wald test statistic:

$$W_i = T(\hat{\beta}_i - \hat{\beta}_b)'(\hat{\sigma}_i^2 Q_{i,T}^{-1} + \hat{\sigma}_b^2 Q_{b,T}^{-1})^{-1}(\hat{\beta}_i - \hat{\beta}_b).$$

The idea is to apply this test statistic in a sequential fashion to determine the set of units with coefficient vector $\beta_b = \beta$. The problem is that in doing so we are likely to end up with spurious rejections due to the multitude of tests; that is, we face the problem of controlling the overall significance level of the approach. To this end, we follow Smeekes (2015), who considers a bootstrap sequential unit root test to determine the stationary units in a panel.

Let $W_{(1)}, W_{(2)}, \ldots, W_{(N-1)}$ denote the $N - 1$ order statistics of $W_1, \ldots, W_{b-1}$, $W_{b+1}, \ldots, W_N$ defined as $W_{(1)} \geq \cdots \geq W_{(N-1)}$. Denote by $\hat{q}$ the estimated number of units that cannot be pooled with the benchmark. The sequential procedure is carried out as follows.

Algorithm SEQ.
1. Set $k = 0$.
2. Test $H_0(k)$ against $H_1(k + 1)$ using $W_{(k+1)}$ as a test statistic. Reject $H_0(k)$ if $W_{(k+1)} > c_\alpha(W_{(k+1)})$, where $c_\alpha(W_{(k+1)})$ is the appropriate critical value at significance level α.
3. If $H_0(k)$ is not rejected, set $\hat{q} = k$, whereas if $H_0(k)$ is rejected, set $k = k + 1$ and go back to step 2.
4. Preform steps 2 and 3 until $H_0(k)$ cannot be rejected anymore, and set $\hat{q} = k$. If all null hypotheses up to $H_0(N - 1)$ are rejected, set $\hat{q} = N$.

We now focus on how to obtain appropriate critical values, $c_\alpha(W_{(k+1)})$. Let $\mathcal{D}_k = \{i : W_i \geq W_{(k)}\}$ denote the set of units for which W_i is larger than the kth-order

statistic. The complement of $\mathcal{D}_k$ is henceforth denoted $\mathcal{D}_k^c$. Let us also denote by $a_{(1:\mathcal{F})}$ the largest element of the set $\{a_i : i \in \mathcal{F}\}$. The following bootstrap algorithm will be used to obtain $c_\alpha(W_{(k+1)})$.

Algorithm SEQBOOT.

1. For each unit, estimate (1) by LS and organize the residuals in a $T \times N$ matrix $\hat{\varepsilon}$.
2. Choose a block length, l, and obtain the matrix of bootstrap errors, ε^*, as described in step 3 of Algorithm BOOT.
3. Simulate pseudo-data under $H_0(k)$ as $y_{i,t}^* = \hat{\theta}_i + \hat{\beta}_b' x_{i,t} + \varepsilon_{i,t}^*$.
4. Obtain W_i^* by applying W_i to $(y_{i,t}^*, x_{i,t})$ for all $i \in \mathcal{D}_k^c$ and obtain the bootstrap test statistic as $W_{(k+1)}^* = W_{(1:\mathcal{D}_k^c)}^*$.
5. Repeat steps 2–4 B times.
6. Select the bootstrap critical value $c_\alpha^*(k+1)$ as the $(1-\alpha)$-quantile of the ordered $W_{(k+1)}^*$ statistics.

Remark 6 Note that the sequential procedure will not only estimate the number of poolable units, $\hat{q}$, but will also identify those units. The set of units that are poolable (non-poolable) equals $\mathcal{D}_{\hat{q}}^c$ ($\mathcal{D}_{\hat{q}}$).

In this study, the benchmark is a single unit. This is not necessary. In fact, as pointed out by Kapetanios (2003), $\hat{\beta}_b$ may be based on a pooled estimator. The problem with such an approach is that the pooled estimator may not make sense under the alternative of different slopes. Suppose, for example, that there are two groups of units, whose slopes are equal within each group but heterogeneous across groups. In this case, the pooled estimate will likely lie between the true coefficients, and therefore, the sequential procedure is likely to find evidence against poolability for all units.[4] The use of a benchmark unit overcomes this problem.

The question is how to choose the benchmark unit. In some applications, there is a natural candidate, as when evaluating policy for a particular unit or when there is a "dominant" unit (see Pesaran and Chudik 2013). In other applications, the choice of benchmark may be less obvious. However, in many cases the researcher will have some a priori information as to the units that are most likely to be poolable, and in such circumstances, the benchmark unit may be picked at random from that set.

Denote by $\mathcal{Z} = \{i : \beta_i = \beta_b, i \neq b\}$ the set of units for which the null hypothesis $\beta_i = \beta_b$ is true. In addition, let G_h denote the asymptotic distribution (as $T \to \infty$) of W_h for $h \in \mathcal{Z}$. Lemma 1 establishes the asymptotic distributions of the relevant test statistics when testing $H_0(k)$ against $H_1(k+1)$.

Lemma 1 *Suppose $H_0(k)$ is tested against $H_1(k+1)$ using Algorithms SEQ and SEQBOOT. Under Assumptions REGR and ERR, as $T \to \infty$ with N held fixed*

$$W_{(k+1)} \to \infty \quad if \, k+1 \leq q,$$
$$W_{(k+1)} \to_d G_{(1:\mathcal{Z})} \quad if \, k+1 > q,$$

---

[4] Kapetanios (2003) assumes the existence of a consistent pooled estimator of some common parameter β, regardless of whether or not the units are poolable. Typically, this requires that a random coefficient assumption is satisfied, or alternatively, that the fraction of non-poolable units tends to zero as $N \to \infty$.

$$W_i^* \to_{d^*} G_i \quad \text{for any } q \text{ and all } i \in \mathcal{D}_k^{\mathsf{c}},$$

$$W_{(k+1)}^* \to_{d^*} G_{\left(1:\mathcal{D}_k^c\right)} \quad \text{for any } k \text{ and } q.$$

Lemma 1 states the asymptotic validity of the bootstrap approach. Here, $G_{(1:\mathcal{Z})}$ represents the asymptotic distribution of $W_{(1:\mathcal{Z})}$. Under Assumptions REGR and ERR, $G_{(1:\mathcal{Z})}$ is unknown and an analytical approach to the sequential test is not feasible. Fortunately, the desired distribution can be obtained via bootstrapping.

Theorem 3 *Under the conditions of Lemma 1,*

$$\lim_{T \to \infty} P(\hat{q} = k) = 0 \quad \text{if } k + 1 \leq q,$$

$$\lim_{T \to \infty} P(\hat{q} = k) = 1 - \alpha \quad \text{if } k = q,$$

$$\lim_{T \to \infty} P(\hat{q} = k) \leq \alpha \quad \text{if } k - 1 \geq q.$$

Remark 7 The first part of Theorem 3 shows that asymptotically the probability of underrejection is zero; that is, the probability of finding the set of poolable units tends to one as $T \to \infty$. Also, as is evident from the third part, since the family-wise error rate (FWE) is at most α, the procedure is able to control the overall significance level.

While asymptotically irrelevant, the finite sample performance of the sequential test depends on the number of hypotheses tested and the true number of poolable units. In particular, for the overall procedure to have satisfactory power when the number of non-poolable units is large, it is necessary that the power at each step of the procedure be close to one. If this is not the case, and if the number of tests is large, the probability of correctly labeling all non-poolable units is likely to be quite low. Consequently, the sequential procedure is mainly suited for panels where N is relatively small. In Sect. 4, we elaborate on this issue.

As a partial solution to the problem of low power, we may follow the suggestion of Smeekes (2015) and apply the sequential procedure to another set of pre-specified null hypotheses. More specifically, instead of testing each unit against the benchmark, we may skip some units. That is, instead of considering $q_k = k - 1$ for $k = 1, \ldots, N - 1$, we may consider another set of numbers $q_1, \ldots, q_K$, where q_k is not necessarily equal to $k - 1$. In this case, in the first iteration of Algorithm SEQ, $H_0(q_1)$ is tested against $H_1(q_2)$. If $H_0(q_1)$ is rejected, in the second iteration $H_0(q_2)$ is tested against $H_1(q_3)$, and so on. By taking $q_{k+1} - q_k > 1$, we reduce the number of tests conducted, which is likely to lead to an increase in the overall power of the testing procedure. There is, however, one major drawback of this approach. Since not every unit is tested against the benchmark, unless $q \in \{q_k : k = 1, \ldots, K\}$, $\hat{q} = q_k$ cannot be interpreted as the number of non-poolable units, nor can $\mathcal{D}_{\hat{q}}^c$ be interpreted as the set of poolable units. Instead, the finding that $\hat{q} = q_k$ should be interpreted as that $q \in [q_{k-1}, q_{k+1}]$. The effect of skipping units is illustrated in the next section.

4 Monte Carlo simulations

In this section, we investigate the small-sample properties of the above tests using Monte Carlo simulations. We begin by considering S^*. In this case, we focus on the performance of the test across different specifications of the dependence structure of the errors and across different block-length selection rules. We then proceed to evaluate the sequential bootstrap approach. Here, we will focus on the ability of the test to correctly identify the set of poolable units and to control the FWE.

4.1 Simulation design

The following data-generating process will be used to analyze the performance of S^*:

$$y_{i,t} = \theta_i + \beta_i x_{i,t} + \varepsilon_{i,t},$$
$$x_{i,t} = \lambda_x f_{x,t} + v_{i,t},$$
$$f_{x,t} = \rho_x f_{x,t-1} + e_{x,t},$$
$$v_{i,t} = \rho_x v_{i,t-1} + u_{i,t},$$

where $\theta_i \sim N(1, 1)$, $\lambda_x = \sqrt{0.5}$, $\rho_x = 0.5$, $e_{x,t} \sim N(0, 1 - \rho_x^2)$, and $u_{i,t} \sim N(0, (1 - \rho_x^2)(1 - \lambda_x^2))$. Three specifications for $\varepsilon_{i,t}$ are considered.

E1. $\varepsilon_{i,t} \sim N(0, 1)$.
E2. $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})'$ is generated according to the following spatial first-order autoregressive model:

$$\varepsilon_t = 0.5 W_N \varepsilon_t + e_t,$$

where $e_t \sim N(0, I_N)$. The spatial weight matrix W_N is constructed as a first-order continuity matrix in which each unit, except for the first and last, has one left and one right neighbor (see Anselin et al. 2008). To make the simulation results invariant to the variance of the error terms, $\varepsilon_{i,t}$ is scaled by the square root of the ith diagonal element of the covariance matrix of ε_t.
E3. $\varepsilon_{i,t}$ is generated according to the following dynamic common factor model:

$$\varepsilon_{i,t} = \lambda_\varepsilon f_{\varepsilon,t} + \eta_{i,t},$$
$$f_{\varepsilon,t} = \rho_\varepsilon f_{\varepsilon,t-1} + e_{\varepsilon,t},$$

where $\lambda_\varepsilon = \sqrt{0.5}$, $\eta_{i,t} \sim N(0, 1 - \lambda_\varepsilon^2)$, and $e_{\varepsilon,t} \sim N(0, 1 - \rho_\varepsilon^2)$. When $\rho_\varepsilon = 0$, $\varepsilon_{i,t}$ is serially independent, whereas when $\rho_\varepsilon \neq 0$, then this is no longer the case.[5]

Under the null, $\beta_1 = \cdots = \beta_N = 1$, while under the alternative, $\beta_i = 1$ for $i = 1, \ldots, \lfloor N/2 \rfloor$ and $\beta_i \sim N(1, 0.04)$ for $i = \lfloor N/2 \rfloor, \ldots, N$, with $\lfloor x \rfloor$ denoting

---

[5] We also considered a moving average model for $f_{\varepsilon,t}$. The results were, however, very similar to the ones based the autoregressive model considered here and are therefore omitted.

the integer part of x. All tests are conducted at a 5 % nominal level, and we take $N \in \{5, 25, 50\}$ and $T \in \{25, 50, 100\}$.

In case of the sequential testing procedure, we consider $N \in \{10, 20\}$. If $N = 10$, then $q \in \{0, 1, 3, 6\}$, whereas if $N = 20$, then $q \in \{0, 1, 3, 6, 12\}$. In both cases, the benchmark unit is taken to be $b = N$. The errors are generated as in E3 with $\rho_\varepsilon = 0.3$. For all $i \in \mathcal{Z}$, $\beta_i = \beta_b = 1$, and for $i \notin \mathcal{Z}$, $\beta_i \sim U(0, 0.5)$. The significance level is taken to be 5%.[6] To evaluate the ability of the sequential procedure to control the overall significance level, we calculate the empirical FWE as the proportion of tests with at least one false rejection. As a measure of power, we report the proportion of tests in which the estimated set of poolable units, $\mathcal{D}_{\hat{j}}^c$, equals the true set, $\mathcal{Z}$, henceforth referred to as "CP." We also consider skipping units. In this case, we have to decide on a sequence of numbers $q_1, \ldots, q_K$ to be tested. In this section, $q_{k+1} = q_k + \delta$, where $\delta \in \{3, 6\}$.

In a final set of experiments, we compare our sequential bootstrap test with two alternative methods frequently encountered in the literature on multiple testing, the Bonferroni and Holm procedures.[7] Denote by $\hat{p}_i$ the p value of W_i. In the Bonferroni procedure, $H_i : \beta_i = \beta_b$ is rejected if $\hat{p}_i \leq \alpha/(N-1)$. The Holm procedure consists of the following steps.

Holm algorithm.

1. Set $k = 1$.
2. Let $\hat{p}_{(1)} \leq \cdots \leq \hat{p}_{(N-1)}$ denote the ordered p values and $H_{(1)}, \ldots, H_{(N-1)}$ the associated null hypotheses. If $\hat{p}_{(k)} \geq \alpha/(N-k)$, accept $H_{(k)}, H_{(k+1)}, \ldots, H_{(N-1)}$ and stop. If $\hat{p}_{(k)} < \alpha/(N-k)$, proceed to step 3.
3. Reject $H_{(k)}$, set $k = k + 1$, and go to step 2.

Compared with the Bonferroni procedure, the Holm criterion for rejecting $H_{(k)}$ becomes increasingly less strict at larger p values. The Holm procedure is therefore expected to be more powerful.

All results are based on 1,000 replications and 499 bootstrap draws. The first 50 time series observations are discarded to reduce the initial value effect.

4.2 Block-length selection rules

An important consideration in practice is the block length, l. In this section, we consider three rules for selecting l.

In the first rule, $l = \lfloor 2T^{1/3} \rfloor$, which amounts to block lengths of 6, 7, and 9 for sample sizes of 25, 50, and 100, respectively. These block lengths are within the range usually encountered in the literature. Although simple, setting l as a function of T means ignoring the covariance structure of the data (see Hall et al. 1995). A data-driven rule is therefore often preferable.

---

[6] The power results are not size corrected because such a correction is generally not available in practice. Hence, a test is useful for applied work only if it respects roughly the nominal significance level.

[7] For a detailed discussion of these procedures, we refer to Lehmann and Romano (2005, Chapter 9).

The second rule amounts to setting l according to the Newey and West (1994) automatic procedure for bandwidth selection. This approach not only accounts for the serial correlation of the data, but is also relatively easy to implement and has good small-sample properties (see Gonçalves and White 2005; Gonçalves 2011).

The Newey and West (1994) approach is designed for variance estimation; however, it is more appealing to consider a block-length selection rule designed for distribution estimation. The proposals of Hall et al. (1995) and Lahiri et al. (2007) are examples of nonparametric resampling methods that can be used for this purpose. In particular, while Hall et al. (1995) propose a rule based on subsampling, the rule of Lahiri et al. (2007) is based on the jackknife-after-bootstrap (JAB) method of Lahiri (2002). In this paper, we will employ the latter method because it is computationally more efficient and has been found to perform relatively well in simulations (see Lahiri et al. 2007).

4.3 Monte Carlo results

The results of the Monte Carlo simulations are reported in Tables 1, 2, 3, and 4. Table 1 reports the results for S and S^*, whereas Tables 2, 3, and 4 report the results for the sequential procedure.

Table 1 provides the empirical size and power of S (the asymptotic test based on iid innovations) in the absence of serial correlation. Looking first at results for E1, we see that S tends to over-reject when $N \geq T$, which is in line with earlier findings in the literature (see, e.g., Pesaran and Yamagata 2008). By contrast, S^* tends to perform quite well for all combinations of N and T. Regarding the block-length selection rules, we see that the three rules considered lead to very similar performance. The deterministic rule is slightly oversized; however, the distortions go away as T increases. When it comes to the behavior under the alternative hypothesis, all tests seem to have satisfactory power properties, with power rising in both N and T.

While oversized in the presence of serial correlation, cross-sectional dependence alone ($\rho_\varepsilon = 0$) makes S undersized. The bootstrap tests, on the other hand, tend to perform well in all cases considered with good size accuracy and power. As expected, the tests based on the data-driven block-length selection rules generally perform best, especially when T is small. Looking next at the results for the case when $\rho_\varepsilon \neq 0$, we see that the size of S is increasing in ρ_ε. The bootstrap tests, on the other hand, continue to perform quite well when $\rho_\varepsilon = 0.3$, although there is a slight tendency to reject too often when T is relatively small. The distortions are made worse by increasing ρ_ε to 0.6, in which case the largest distortions are obtained by using the JAB rule. In other words, the use of a data-driven block length does not automatically lead to better size accuracy. Specifically, although the block lengths selected by the data-driven rules are increasing in ρ_ε, they do not increase sufficiently, which in turn leads to size distortions. However, while this means that the best performance is sometimes obtained by using the deterministic rule, the difference is not very large with the data-driven rules leading to acceptable performance in most cases considered.

To summarize the results so far, we find that S displays substantial size distortions in the presence of cross-sectional and/or serial dependence. The bootstrap tests, on the other hand, generally show small size distortions and maintain satisfactory power

Table 1 Size and power of S and S^*

Error case	T/N	S			S^*(Det)			S^*(NW)			S^*(JAB)		
		5	25	50	5	25	50	5	25	50	5	25	50
Size													
E1	25	7.3	11.9	11.8	7.7	7.3	5.2	6.3	7.0	4.3	5.9	6.5	4.9
	50	6.0	9.6	7.8	6.1	8.1	4.5	6.0	7.6	3.8	5.3	7.3	4.3
	100	6.3	5.9	5.3	6.4	5.0	4.7	6.0	5.0	4.1	6.3	5.1	4.1
E2	25	1.4	3.0	6.6	6.3	6.2	6.9	5.9	5.7	6.2	5.5	5.6	6.2
	50	0.7	1.6	2.4	4.5	5.8	4.4	4.1	5.2	3.9	4.5	5.6	3.9
	100	0.8	1.5	2.4	4.9	7.1	4.1	5.4	6.5	4.3	5.1	7.2	4.2
E3, $\rho_\varepsilon = 0$	25	1.9	1.4	1.4	4.9	5.5	6.7	4.5	5.2	5.2	4.9	5.0	5.2
	50	1.4	1.4	0.3	5.9	6.3	4.6	5.3	6.1	4.2	5.6	5.9	4.2
	100	1.5	0.2	0.1	4.7	5.5	5.0	4.5	5.0	4.9	5.0	5.0	4.7
E3, $\rho_\varepsilon = 0.3$	25	4.4	3.4	3.4	7.8	7.7	5.5	6.5	6.7	5.4	7.6	7.4	5.5
	50	2.5	3.0	1.6	5.6	6.9	5.9	5.4	7.0	6.1	5.6	7.1	6.9
	100	2.7	0.9	0.9	5.5	5.4	6.0	5.3	5.3	6.5	5.4	6.2	7.3
E3, $\rho_\varepsilon = 0.6$	25	6.8	9.5	8.5	8.4	9.4	7.6	8.7	8.6	7.4	8.6	10.1	8.3
	50	4.0	5.8	7.1	6.2	8.1	8.0	6.0	9.0	8.0	6.4	10.5	10.5
	100	5.2	4.0	4.1	6.2	7.1	7.5	6.7	7.5	7.4	6.9	10.0	10.8
Power													
E1	25	20.6	53.3	70.8	20.9	42.3	54.5	19.6	40.4	53.4	19.0	40.5	53.0
	50	34.7	75.6	93.6	33.5	72.2	90.6	33.7	71.5	90.4	32.3	70.6	89.8
	100	55.8	96.4	100	55.4	95.6	100	55.3	96.0	100	54.9	95.9	100
E2	25	10.2	40.2	59.9	26.6	55.3	64.3	25.1	53.3	62.9	25.3	53.8	62.4
	50	27.4	70.7	91.4	47.0	84.4	94.5	46.6	84.6	94.2	46.1	84.9	94.1
	100	49.5	96.2	99.9	70.6	98.3	100	70.5	98.3	99.9	70.7	98.5	100
E3, $\rho_\varepsilon = 0$	25	14.8	28.8	38.5	24.9	50.7	63.6	23.2	48.2	61.6	22.8	48.9	62.2
	50	27.9	62.9	80.9	38.6	82.3	95.4	38.5	83.0	95.2	38.4	83.0	94.8
	100	52.7	92.9	99.7	64.5	98.8	100	64.5	98.6	100	65.0	98.7	100
E3, $\rho_\varepsilon = 0.3$	25	16.9	35.5	50.1	24.9	48.1	65.0	24.8	48.2	63.1	23.9	48.5	65.0
	50	30.1	70.0	87.7	38.2	82.0	94.4	37.9	81.7	94.5	38.4	82.3	94.6
	100	54.4	94.2	99.8	61.9	98.1	100	61.6	98.1	100	63.0	98.3	100
E3, $\rho_\varepsilon = 0.6$	25	20.9	46.3	66.5	24.8	50.2	67.3	24.4	49.4	67.7	24.9	50.9	68.8
	50	34.2	77.0	92.6	36.9	80.1	93.7	36.9	80.5	93.7	37.7	82.1	94.3
	100	56.6	95.2	99.9	59.3	96.1	100	59.0	96.8	100	60.0	97.2	100

The critical values for S are based on the $\chi^2[m(N-1)]$ distribution. "S^*(Det)," "S^*(NW)," and "S^*(JAB)" refer to S^* when combined with the block-length selection rules of $\lfloor 2T^{1/3} \rfloor$, Newey and West (1994) and Lahiri et al. (2007), respectively

in small samples. In particular, the bootstrap tests are robustness to the specification of the cross-sectional dependence and also perform reasonably well in the presence of error serial correlation. Since the bootstrap tests are also very simple to implement, they should be well suited for applied work.

Table 2 FWE and CP of the sequential test

T/q	N = 10				N = 20				
	0	1	3	6	0	1	3	6	12
FWE									
25	6.6	6.5	5.2	5.2	4.3	6.9	6.5	6.8	4.6
50	6.7	5.0	5.0	4.2	5.9	6.0	5.3	4.5	3.9
100	4.4	5.8	6.2	5.8	6.7	6.1	5.1	6.7	5.5
CP									
25	93.4	44.5	20.9	14.7	95.7	37.0	15.1	7.9	3.8
50	93.3	77.1	60.2	51.0	94.1	72.6	56.7	39.1	32.7
100	95.6	92.7	92.7	89.4	93.3	91.3	87.8	80.3	80.1

q denotes the true number of non-poolable units. FWE is the empirical family-wise error rate, and CP measures the ability to correctly identify the set of poolable units

Table 3 FWE and CP of the sequential test when skipping units

δ	T/q	N = 10				N = 20				
		0	1	3	6	0	1	3	6	12
FWE										
3	25	6.3	–	12.9	12.1	6.4	–	15.6	16.6	15.1
	50	6.4	–	10.6	10.4	5.8	–	10.5	11.0	12.4
	100	4.5	–	7.9	6.6	5.7	–	7.3	9.3	7.7
6	25	5.7	–	–	17.5	6.0	–	–	22.9	26.4
	50	6.2	–	–	9.6	4.9	–	–	17.4	16.7
	100	4.9	–	–	2.0	5.7	–	–	9.9	8.5
CP										
3	25	93.7	–	44.5	39.5	93.6	–	32.9	21.4	16.2
	50	93.6	–	76.9	75.0	94.2	–	71.3	60.9	59.3
	100	95.5	–	91.4	93.0	94.3	–	91.5	88.3	89.5
6	25	94.3	–	–	53.1	94.0	–	–	29.9	24.1
	50	93.8	–	–	83.8	95.1	–	–	69.2	68.3
	100	95.1	–	–	98.0	94.3	–	–	89.8	90.8

δ refers to the number of units being skipped. See Table 3 for an explanation of the rest

We now turn our attention to the results for the sequential testing procedure. Here we focus on the results based on the Newey and West (1994) rule, which are shown in Tables 2, 3, and 4. Looking first at FWE in Table 2, it is evident that the sequential test is able to control the overall significance level very well in small samples. This is true for all values of q. We also see that the ability to identify the true set of poolable units (as measured by CP) depends greatly on q when T is small. However, when T increases, CP approaches 95%, as predicted by our theoretical results.

Table 3 provides the results of the sequential test when skipping units. The first thing to note is that, when T is small, CP is much higher in Table 3 than in Table 2. This result suggests that, by skipping units, we may increase the ability of the sequential

Table 4 FWE and CP of the bootstrap, Bonferroni and Holm procedures

Procedure	T/q	$N = 10$				$N = 20$				
		0	1	3	6	0	1	3	6	12
FWE										
Bootstrap	25	5.7	7.3	6.2	4.4	5.3	5.9	5.1	4.7	4.5
	50	6.2	6.0	6.0	5.9	6.1	5.2	6.2	4.7	4.7
	100	4.4	5.4	6.2	6.0	6.5	6.3	5.3	6.2	5.4
Bonferroni	25	5.1	6.2	5.2	2.0	4.5	5.7	4.3	3.8	3.0
	50	5.0	5.0	3.1	1.7	4.4	4.6	4.8	2.8	3.1
	100	3.0	4.2	3.3	1.8	5.3	4.9	3.6	4.3	2.3
Holm	25	5.1	6.4	6.0	3.7	4.5	5.7	4.6	4.4	3.8
	50	5.0	5.4	4.0	4.6	4.4	4.8	5.0	3.5	4.2
	100	3.0	4.4	4.6	5.0	5.3	5.3	4.2	5.5	4.0
CP										
Bootstrap	25	94.3	35.9	14.4	8.4	94.7	27.8	8.9	3.1	1.7
	50	93.8	70.5	50.9	40.4	93.9	65.1	40.8	25.3	17.2
	100	95.6	90.4	84.0	84.4	93.5	88.3	82.1	71.6	67.7
Bonferroni	25	94.9	34.4	12.0	5.9	95.5	26.4	8.7	2.3	0.6
	50	95.0	70.0	47.5	31.3	95.6	63.8	37.5	20.7	10.1
	100	97.0	91.1	83.2	80.9	94.7	89.0	80.5	67.7	57.5
Holm	25	94.9	34.2	12.4	8.3	95.5	26.4	9.2	2.5	1.4
	50	95.0	69.6	49.7	39.0	95.6	63.6	38.8	23.0	15.1
	100	97.0	90.9	83.6	83.4	94.7	88.6	80.4	69.6	66.3

See Table 3 for an explanation

test to correctly identify the true set of poolable units. On the other hand, Table 3 reveals that the increase in power comes at the expense of higher FWE, at least among the smaller values of T. This trade-off between size and power also becomes evident when comparing the results for $\delta = 3, 6$.

Table 4 compares the bootstrap sequential test with the Bonferroni and Holm procedures. In this case, the errors are generated as in E1. Looking first at FWE, as expected, we see that all procedures are quite successful in controlling the overall significance level. We also see that the ability to identify the true set of poolable units, as measured by CP, differs, sometimes considerably, between the procedures. For example, looking at the results for the case when $q = 12$ and $T = 50$, we see that the bootstrap CP is almost twice that of the Bonferroni approach. In general, the bootstrap and Holm procedures are able to identify more non-poolable units than can the Bonferroni approach, and this advantage becomes more important as q increases, which is in accordance with our expectations. We also see that the bootstrap approach outperforms the Holm procedure in terms of CP (when $q > 0$) in all but three cases.

In conclusion, the bootstrap sequential test can control FWE in small samples. On the other hand, the ability to find the true set of poolable units depends largely on

the sample size and the true number of poolable units. When the number of poolable units is small, and T is small, the sequential procedure will inevitably end up rejecting too few non-poolable units. We show, however, that this problem can be somewhat alleviated by skipping units. Finally, the simulation results indicate that even in the iid case, in which standard multiple testing procedures apply, the bootstrap approach offers improvements in the ability to identify non-poolable units.

5 Conclusions

In this paper, we address the issue of testing for slope homogeneity in panel data models with non-spherical errors. Previous slope homogeneity tests require modeling and estimation of the error dependence structure in order to be practical. This may be problematic in the presence of complicated dependencies of possible unknown form. Therefore, in this paper we propose a bootstrap test that can be implemented with no specification or estimation of the dependence structure. Yet, the test can accommodate both cross-sectional and serial dependence in the error process. The results of a Monte Carlo study indicate that the bootstrap test has small size distortions and satisfactory power in small samples.

Appendix: Proofs

This appendix is concerned with the proofs of Lemma 1 and Theorems 1–3. The convergence is in probability, but we generally do not add this explicitly in order to simplify the notation. The sequence $\{a_T\}$ is at most of order T^κ in probability, denoted $a_T = O_p(T^\kappa)$, if $T^{-\kappa} a_T$ converges in distribution. The sequence is of order smaller than T^κ in probability, denoted by $a_T = o_p(T^\lambda)$, if $T^{-\kappa} a_T \to_p 0$. The bootstrap stochastic order symbols, denoted $O_{p*}(\cdot)$ and $o_{p*}(\cdot)$, are defined in an analogous manner.

We begin by defining the following quantities which will be used throughout this Appendix:

$$\xi_{i,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)\varepsilon_{i,t},$$

$$\tilde{\xi}_{i,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{z}_{i,t} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \mu_i)\varepsilon_{i,t},$$

$$\xi_{i,T}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)\varepsilon_{i,t}^*,$$

$$\tilde{\xi}_{i,T}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{z}_{i,t}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \mu_i)\tilde{\varepsilon}_{i,t}^*,$$

where $\tilde{\varepsilon}_{i,t}^* = y_{i,t}^* - \theta_i - x_{i,t}'\beta_i$. Also, let z_t and $\tilde{z}_t$ be the $mN \times 1$ stacked vectors $z_t = (z_{1,t}', \ldots, z_{N,t}')'$ and $\tilde{z}_t = (\tilde{z}_{1,t}', \ldots, \tilde{z}_{N,t}')'$, with similar definitions of z_t^* and $\tilde{z}_t^*$.

Lemma 2 *Under Assumption ERR, as $T \to \infty$,*

$$\Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t \to_d N(0, I_{mN}).$$

Proof of Lemma 2 Clearly

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \overline{x}_i)\varepsilon_{i,t}$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \mu_i)\varepsilon_{i,t} - (\overline{x}_i - \mu_i) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{i,t}$$

$$= \frac{1}{\sqrt{T}} \tilde{z}_{i,t} - (\overline{x}_i - \mu_i) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{i,t}.$$

By Corollary 3.48 in White (2001), $(\overline{x}_i - \mu_i) = T^{-1} \sum_{t=1}^{T} (x_{i,t} - \mu_i) = o_p(1)$, and by further use of his Theorem 5.20, $T^{-1/2} \sum_{t=1}^{T} \varepsilon_{i,t} = O_p(1)$. It follows that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{z}_{i,t} + o_p(1).$$

The required result now follows by applying to $T^{-1/2} \sum_{t=1}^{T} \tilde{z}_t$ a central limit theorem for mixing processes (see, e.g., White 2001, Theorem 5.20). □

Lemma 2* *Under Assumptions ERR and BL, as $T \to \infty$,*

$$\Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t^* \to_{d*} N(0, I_{mN}) \quad \text{in probability.}$$

Proof of Lemma 2* We have $\varepsilon_{i,t}^* = \tilde{\varepsilon}_{i,t}^* - (\hat{\theta}_i - \theta_i) - x_{i,t}'(\hat{\beta}_i - \beta_i)$. Similarly, $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} - (\hat{\theta}_i - \theta_i) - x_{i,t}'(\hat{\beta}_i - \beta_i)$. Using these relationships, we can write

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_{i,t} - \overline{x}_i)\varepsilon_{i,t}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [(x_{i,t} - \overline{x}_i)\tilde{\varepsilon}_{i,t}^* - (x_{i,t} - \overline{x}_i)\varepsilon_{i,t}],$$

where we have used that fact that $\sum_{t=1}^{T} (x_{i,t} - \overline{x}_i)\hat{\varepsilon}_{i,t} = 0$ by the first-order conditions for $\hat{\beta}_i$. By adding and subtracting appropriately, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_{i,t}^* = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [(x_{i,t} - \mu_i)\tilde{\varepsilon}_{i,t}^* - (x_{i,t} - \mu_i)\varepsilon_{i,t}]$$

$$- \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\overline{x}_i - \mu_i)\tilde{\varepsilon}_{i,t}^* - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\overline{x}_i - \mu_i)\varepsilon_{i,t}$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [(x_{i,t} - \mu_i)\tilde{\varepsilon}_{i,t}^* - (x_{i,t} - \mu_i)\varepsilon_{i,t}] + o_p(1),$$

where the last equality follows from using $(\overline{x}_i - \mu_i) = o_p(1)$ (see White 2001, Corollary 3.48), $T^{-1/2} \sum_{t=1}^{T} \tilde{\varepsilon}_{i,t}^* = O_{p^*}(1)$ (see Fitzenberger 1997, Theorem 3.1), and $T^{-1/2} \sum_{t=1}^{T} \varepsilon_{i,t} = O_p(1)$ (see White 2001, Theorem 5.20). The required result now follows from the same argument as in Fitzenberger 1997, giving

$$\Sigma^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\tilde{z}_t^* - \tilde{z}_t) \to_{d^*} N(0, I_{mN}) \quad \text{in probability}$$

as $T \to \infty$. $\qquad\square$

Proof of Theorem 1 We begin by proving the asymptotic distribution under H_0, in which case

$$\hat{\beta}_i - \hat{\beta}_{\text{WFE}} = \frac{1}{\sqrt{T}} Q_{i,T}^{-1} \xi_{i,T} - \frac{1}{\sqrt{T}} \left(\sum_{i=1}^{N} \hat{\sigma}_i^{-2} Q_{i,T} \right)^{-1} \sum_{i=1}^{N} \hat{\sigma}_i^{-2} \xi_{i,T}.$$

This implies

$$S = \sum_{i=1}^{N} \frac{\xi_{i,T}' Q_{i,T}^{-1} \xi_{i,T}}{\hat{\sigma}_i^2} - \left(\sum_{i=1}^{N} \hat{\sigma}_i^{-2} \xi_{i,T} \right)' \left(\sum_{i=1}^{N} \hat{\sigma}_i^{-2} Q_{i,T} \right)^{-1} \left(\sum_{i=1}^{N} \hat{\sigma}_i^{-2} \xi_{i,T} \right), \quad (2)$$

or, in stacked form,

$$S = \xi_T' (\hat{\Sigma}_T Q_T)^{-1} \xi_T - (D\hat{\Sigma}_T^{-1} \xi_T)' (D\hat{\Sigma}_T^{-1} Q_T D')^{-1} (D\hat{\Sigma}_T^{-1} \xi_T), \quad (3)$$

where $\xi_T = (\xi_{1,T}', \ldots, \xi_{N,T}')'$, $D = \tau_N' \otimes I_k$, $Q_T = \text{diag}(Q_{1,T}, \ldots, Q_{N,T})$, and $\hat{\Sigma}_T$ is a diagonal $Nm \times Nm$ matrix whose diagonal is given by $(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2)' \otimes \tau_k$.

By the properties of the LS residuals and Corollary 3.48 in White (2001), we have that $\hat{\sigma}_i^2 - \sigma_i^2 = o_p(1)$, where $\sigma_i^2 = E(\varepsilon_{i,t}^2)$. Note also that, by Lemma 2 and Assumption REGR, $\xi_T \to_d \xi$ as $T \to \infty$, where $\xi \sim N(0, \Sigma)$, and $Q_T - Q = o_p(1)$, where $Q = \text{diag}(Q_1, \ldots, Q_N)$. It follows that

$$S \to_d Z_1' Z_1 - \left(D\Sigma_0^{-1/2}Q^{1/2}Z_1\right)' \left[D\Sigma_0^{-1/2}Q^{1/2}\left(D\Sigma_0^{-1/2}Q^{1/2}\right)'\right]^{-1}\left(D\Sigma_0^{-1/2}Q^{1/2}Z_1\right)$$

$$=_d Z_1'\left(I_{kN} - H'(HH')^{-1}H\right)Z_1 = Z_1'AZ_1, \tag{4}$$

where $=_d$ signifies equality in distribution, $Z_1 = \Sigma_0^{-1/2}Q^{-1/2}\xi$, $Z_1 \sim N(0, B)$ with $B = \Sigma_0^{-1/2}Q^{-1/2}\Sigma\Sigma_0^{-1/2}Q^{-1/2}$, Σ_0 is $\hat\Sigma_T$ with $\hat\sigma_i^2$ replaced by σ_i^2, $H = D\Sigma_0^{-1/2}Q^{1/2}$ and $A = I_{kN} - H'(HH')^{-1}H$. Since A is symmetric,

$$Z_1'AZ_1 =_d \sum_{j=1}^{kN} \lambda_j U_j^2,$$

where $\lambda_1, \ldots, \lambda_{mN}$ are the eigenvalues of $B^{1/2}AB^{1/2}$ (see Mathai and Provost 1992). Noting that A and B are (non-stochastic) positive semidefinite matrices, we have that $\lambda_j \geq 0$ for all j. This establish the required result under H_0.

To show consistency, we consider an alternative hypothesis of the form $H_1 : \beta_i = \beta + \delta_i$, where δ_i are $m \times 1$ vectors of fixed constants such that $\|\delta_i\| \leq C < \infty$ for all i and $\delta_i \neq \delta_h$ for some pair $i \neq h$. Under this alternative,

$$\sqrt{T}(\hat\beta_i - \hat\beta_{\text{WFE}}) = \sqrt{T}\delta_i + Q_{i,T}^{-1}\xi_{i,T} - \sqrt{T}\left(\sum_{i=1}^{N} \frac{Q_{i,T}}{\hat\sigma_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{Q_{i,T}\delta_i}{\hat\sigma_i^2}$$

$$- \sqrt{T}\left(\sum_{i=1}^{N} \frac{Q_{i,T}}{\hat\sigma_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{\xi_{i,T}}{\hat\sigma_i^2}$$

$$= \sqrt{T}\delta_i + Q_{i,T}^{-1}\xi_{i,T} - \sqrt{T}c_T - \sqrt{T}\left(\sum_{i=1}^{N} \frac{Q_{i,T}}{\hat\sigma_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{Q_{i,T}\delta_i}{\hat\sigma_i^2},$$

with an obvious definition of c_T. The last term on the right is $O_p(1)$, suggesting that

$$\sqrt{T}(\hat\beta_i - \hat\beta_{\text{WFE}}) = \sqrt{T}(\delta_i - c_T) + O_p(1).$$

Since $\delta_i \neq \delta_h$ for some $i \neq h$, it must hold that $\delta_i - c_T \neq 0$ for at least one i. Since Q_i/σ_i^2 is positive definite, this means that $T(\hat\beta_i - \hat\beta_{\text{WFE}})'Q_{i,T}(\hat\beta_i - \hat\beta_{\text{WFE}})/\hat\sigma_i^2 \to \infty$ as $T \to \infty$, and so the proof is complete. □

Proof of Theorem 2 Define $\hat\Sigma_T^*$ as the diagonal matrix whose diagonal is given by $(\hat\sigma_1^{*2}, \ldots, \hat\sigma_N^{*2})' \otimes \tau_k$, and assume for the moment that there exists a diagonal matrix Σ_0^* satisfying $\hat\Sigma_T^* - \Sigma_0^* = o_{p^*}(1)$. Under H_0,

$$\hat\beta_i^* - \hat\beta_{\text{WFE}}^* = \frac{1}{\sqrt{T}}Q_{i,T}^{-1}\xi_{i,T}^* - \frac{1}{\sqrt{T}}\left(\sum_{i=1}^{N}(\hat\sigma_i^{*2})^{-1}Q_{i,T}\right)^{-1} \sum_{i=1}^{N}(\hat\sigma_i^{*2})^{-1}\xi_{i,T}^*,$$

where $\hat{\sigma}_i^{*2}$ is the usual LS estimate of the variance from the bootstrap procedure. This result, together with the same arguments used in the proof of Theorem 1, implies that

$$S^* \rightarrow_{d*} Z_1^{*\prime}(I_{kN} - H^{*\prime}(H^*H^{*\prime})^{-1}H^*)Z_1^* =_{d*} Z_1^{*\prime}A^*Z_1^* \quad \text{in probability,} \quad (5)$$

where $H^* = D(\hat{\Sigma}_T^*)^{-1/2}Q^{1/2}$ and $Z_1^* \sim N(0, B^*)$ with $B^* = (\hat{\Sigma}_T^*)^{-1/2}Q^{-1/2}$ $\Sigma(\hat{\Sigma}_T^*)^{-1/2}Q^{-1/2}$. Hence, as in the proof of Theorem 1, if we can show that $\hat{\Sigma}_T^* - \Sigma_0 = o_{p*}(1)$, then

$$Z_1^{*\prime}A^*Z_1^* =_{d*} \sum_{j=1}^{kN} \lambda_j U_j^2 \quad \text{in probability,} \quad (6)$$

and the second result of the theorem follows. We now verify that $\hat{\Sigma}_T^* - \Sigma_0 = o_{p*}(1)$. Under H_0,

$$\frac{T-k-1}{T}(\hat{\sigma}_i^*)^2 = \frac{1}{T}(y_i^* - x_i\hat{\beta}_i^*)'M_\tau(y_i^* - x_i\hat{\beta}_i^*) = \frac{1}{T}(M_\tau\varepsilon_i^*)'M_{x_i}(M_\tau\varepsilon_i^*)$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t}^* - \bar{\varepsilon}_i^*)^2 + o_{p*}(1) = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t}^*)^2 + o_{p*}(1),$$

where $M_{x_i} = I_T - x_i(x_i'M_\tau x_i)^{-1}x_i'$. The second equality follows from straightforward algebra under the null hypothesis. We can use an argument similar to Lemma 2 to show that $T^{-1/2}x_i'M_\tau\varepsilon_i^* = O_{p*}(1)$, and the third equality follows. Finally, by Theorem 3.1 in Fitzenberger (1997), since $\varepsilon_{i,t}^* = \tilde{\varepsilon}_{i,t}^* + \hat{\varepsilon}_{i,t} - \varepsilon_{i,t}$ and $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t} = 0$, we have $\bar{\varepsilon}_i^* = T^{-1}\sum_{t=1}^{T}(\tilde{\varepsilon}_{i,t}^* - \varepsilon_{i,t}) = O_{p*}(T^{-1/2})$, and so the fourth equality follows. It remains to consider $T^{-1}\sum_{t=1}^{T}(\varepsilon_{i,t}^*)^2$. By MBB–Lemma A.3 in Fitzenberger (1997), $T^{-1}\sum_{t=1}^{T}[(\tilde{\varepsilon}_{i,t}^*)^2 - \varepsilon_{i,t}^2] = o_{p*}(1)$. We now show that $T^{-1}\sum_{t=1}^{T}[(\varepsilon_{i,t}^*)^2 - (\tilde{\varepsilon}_{i,t}^*)^2] = o_{p*}(1)$. For this purpose, it is convenient to define $v_{i,t} = w_{i,t}(w_i'w_i)^{-1}\sum_{s=1}^{T}w_{i,s}'\varepsilon_{i,s}$, where $w_i = (\tau_T, x_i)$. Also, by the properties of LS residuals, $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} - v_{i,t}$. Hence,

$$\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{i,t}^*)^2 = \frac{1}{T}\sum_{t=1}^{T}(\tilde{\varepsilon}_{i,t}^*)^2 + \frac{2}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_{i,t}^*(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t}) + \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_{i,t} - \varepsilon_{i,t})^2$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\tilde{\varepsilon}_{i,t}^*)^2 - \frac{2}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_{i,t}^*v_{i,t} + o_p(1)$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\tilde{\varepsilon}_{i,t}^*)^2 + o_{p*}(1) + o_p(1).$$

The second equality follows from the fact that $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{i,t}^2 \rightarrow_p \sigma_i^2$. As for the third equality, by Corollary 3.48 in White (2001), $T^{-1}\sum_{s=1}^{T}w_{i,s}'\varepsilon_{i,s} = o_p(1)$, by Assumption REGR, $(T^{-1}w_i'w_i)^{-1}$ converges to a nonzero deterministic matrix, and by further use of MBB–Lemma A.3 in Fitzenberger (1997), $T^{-1}\sum_{t=1}^{T}(\tilde{\varepsilon}_{i,t}^*w_{i,t} -$

$\varepsilon_{i,t} w_{i,t}) = o_{p*}(1)$. Finally, by Corollary 3.48 in White (2001), $T^{-1} \sum_{t=1}^{T} \varepsilon_{i,t} w_{i,t} = o_p(1)$. This establishes that $\hat{\Sigma}_T^* - \Sigma_0 = o_{p*}(1)$, and so the proof of the first statement of the theorem is complete.

The second part of the theorem follows from the fact that in the bootstrap the data are generated under the null of equal slope coefficients. Consequently, the result in (6) also holds under the alternative hypothesis. □

Proof of Lemma 1 Consider the following alternative hypothesis $H_1 : \beta_i = \beta_b + \delta_i$, where δ_i is a $m \times 1$ vector of fixed constants such that $\|\delta_i\| \leq C < \infty$. By the properties of the LS estimator, $\hat{\beta}_i - \beta_i = O_p(T^{-1/2})$. It follows that $\hat{\beta}_i - \hat{\beta}_b - \delta_i = o_p(1)$. Also, $(\hat{\sigma}_i^2 Q_{i,T}^{-1} + \hat{\sigma}_b^2 Q_{b,T}^{-1})^{-1}$ converges to a positive definite matrix by Assumption REGR. Consequently, provided that $i \notin \mathcal{Z}$, $W_i \to \infty$ as $T \to \infty$, which in turn implies

$$W_{(1)}, \ldots, W_{(q)} \to \infty.$$

Consider the second result of the lemma. We only need to consider the order statistics for which $i \in \mathcal{Z}$, as $W_{(k+1)} \to \infty$ if $k + 1 \leq q$ by the first part. Hence,

$$W_{(k+1)} \to_p W_{1:\mathcal{Z}}.$$

The second result then follows from application of the continuous mapping theorem (see, e.g., White 2000, Proposition 2.2).

In order verify the third result, we proceed in two steps. First, we establish the exact expression of G_i, the limiting distribution of W_i for $i \in \mathcal{Z}$. Second, we show that the bootstrap analog, W_i^*, converges to the same distribution. Define the $m \times Nm$ matrix $P_i = (0, \ldots, 0, I_m, 0, \ldots, 0, -I_m, 0, \ldots, 0)$, where I_m and $-I_m$ are in positions i and b, respectively. In addition, define $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_N)'$. In this notation, since $\beta_i = \beta_b$ for all $i \in \mathcal{Z}$, W_i can be written as

$$W_i = T(P_i \hat{\beta})' \left(P_i \hat{\Sigma}_T Q_T^{-1} P_i' \right)^{-1} (P_i \hat{\beta}) = \tilde{Z}_2' \left(P_i \hat{\Sigma}_T Q_T^{-1} P_i' \right)^{-1} \tilde{Z}_2,$$

where $\tilde{Z}_2 = \sqrt{T} P_i \hat{\beta} = P_i Q_T^{-1} \xi_T$. By Lemma 2 and Assumption REGR,

$$\tilde{Z}_2 \to_d N \left(0, P_i Q^{-1} \Sigma Q^{-1} P_i' \right)$$

as $T \to \infty$, and $\hat{\Sigma}_T Q_T^{-1} - \Sigma_0^* Q^{-1} = o_p(1)$.

To analyze the bootstrap test, W_i^*, define $\hat{\beta}^* = (\hat{\beta}_1^*, \ldots, \hat{\beta}_N^*)'$. It follows that we can write W_i^* as

$$W_i^* = T(P_i \hat{\beta}^*)' \left(P_i \hat{\Sigma}_T^* Q_T^{-1} P_i' \right)^{-1} (P_i \hat{\beta}^*).$$

Note that $\hat{\beta}^*$ is estimated using $y_{i,t}^*$, which is generated under the hypothesis that $\beta_i = \beta_b$ for all i. This implies that

$$W_i^* = \tilde{Z}_2^{*'} \left(P_i \hat{\Sigma}_T^* Q_T^{-1} P_i' \right)^{-1} \tilde{Z}_2^*,$$

where $\tilde{Z}_2^* = \sqrt{T} P_i \hat{\beta}^* = P_i Q_T^{-1} \xi_T^*$. By Lemma 2 and Assumption REGR, we have that

$$\tilde{Z}_2 \to_d N\left(0, P_i Q^{-1} \Sigma Q^{-1} P_i'\right) \quad \text{in probability.}$$

In addition, from the proof of Theorem 1, $\hat{\Sigma}_T^* - \Sigma_0 = o_{p^*}(1)$, which implies that $\hat{\Sigma}_T^* Q_T^{-1} - \Sigma_0 Q^{-1} = o_{p^*}(1)$. This completes the proof of the third result.

The fourth result follows from the third and the continuous mapping theorem (White 2000, Proposition 2.2). □

Proof of Theorem 3 The first result follows from the first and last results of Lemma 1. To show the second result, note that, by the first part of Lemma 1, $W_i \to_p \infty$ if $i \notin \mathcal{Z}$. Therefore, if $k = q$, $\mathcal{D}_k^c$ equals $\mathcal{Z}$ in the limit with probability one. It follows from second and fourth parts of Lemma 1 that the probability of not rejecting the null, $H_0(k)$, is equal to $1 - \alpha$ (the confidence level in Algorithm SEQBOOT). Finally, as for the third result of Theorem 3, note that

$$\sum_{k=0}^{N-1} \lim_{T \to \infty} P(\hat{q} = g) = \sum_{k=0}^{q} \lim_{T \to \infty} P(\hat{q} = g) + \sum_{k=q+1}^{N-1} \lim_{T \to \infty} P(\hat{q} = g) = 1.$$

From the first two parts of the theorem, $\sum_{k=0}^{q} \lim_{T \to \infty} P(\hat{q} = g) = 1 - \alpha$. □

References

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59:817–858

Anselin L, Le Gallo J, Jayet H (2008) Spatial panel econometrics. In: Mátyás L, Sevestre P (eds) The econometrics of panel data. Springer, Berlin

Baltagi BH (2008) Econometric analysis of panel data. Wiley, Chichester

Baltagi BH, Bresson G, Pirotte A (2008) To pool or not to pool? In: Mátyás L, Sevestre P (eds) The econometrics of panel data. Springer, Berlin

Bun MJG (2004) Testing poolability in a system of dynamic regressions with nonspherical disturbances. Empir Econ 29:89–106

Chudik A, Pesaran MH, Tosetti E (2011) Weak and strong cross-section dependence and estimation of large panels. Econom J 14:C45–C90

Davidson J (1994) Stochastic limit theory. Oxford University Press, Oxford

Davidson R, MacKinnon JG (1999) The size distortion of bootstrap tests. Econom Theory 15:361–376

Fitzenberger B (1997) The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. J Econom 82:235–287

Freedman DA (1981) Bootstrapping regression models. Ann Stat 9:1218–1228

Gonçalves S (2011) The moving blocks bootstrap for panel linear regression models with individual fixed effects. Econom Theory 27:1048–1082

Gonçalves S, White H (2005) Bootstrap standard error estimates for linear regression. J Am Stat Assoc 100:970–979

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

Hidalgo J (2003) An alternative bootstrap to moving blocks for time series regression models. J Econom 117:369–399

Hsiao C (2003) Analysis of panel data. Cambridge University Press, Cambridge

Hsiao C, Pesaran MH (2008) Random coefficient panel data models. In: Mátyás L, Sevestre P (eds) The econometrics of panel data. Springer, Berlin

Kapetanios G (2003) Determining the poolability properties of individual series in panel datasets. Queen Mary, University of London Working Paper No. 499

Kapetanios G (2006) Cluster analysis of panel data sets using non-standard optimisation of information criteria. J Econ Dyn Control 30:1389–1408

Kapetanios G (2008) A bootstrap procedure for panel data sets with many cross-sectional units. Econom J 11:377–395

Lahiri SN (2002) On the jackknife-after-bootstrap method for dependent data and its consistency properties. Econom Theory 18:79–98

Lahiri SN, Furukawa K, Lee Y-D (2007) A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. Stat Methodol 4:292–321

Lehmann EL, Romano JP (2005) Testing statistical hypotheses. Springer, New York

Lin C-C, Ng S (2012) Estimation of panel data models with parameter heterogeneity when group membership is unknown. J Econom Methods 1:42–55

MacKinnon JG (2007) Bootstrap hypothesis testing. Queen's economics department Working Paper No. 1127

Mathai AM, Provost SB (1992) Quadratic forms in random variables. Marcel Dekker, New York

Newey WK, West KD (1994) Automatic lag selection in covariance matrix estimation. Rev of Econ Stud 61:613–653

Pesaran MH, Tosetti E (2010) Large panels with common factors and spatial correlations. J Econom 161:182–202

Pesaran MH, Yamagata T (2008) Testing slope homogeneity in large panels. J Econom 142:50–93

Pesaran MH, Chudik A (2013) Econometric analysis of high dimensional VARs featuring a dominant unit. Econom Rev 32:592–649

Pesaran MH, Smith R, Im KS (1996) Dynamic linear models for heterogeneous panels. In: Mátyás L, Sevestre P (eds) The econometrics of panel data. Springer, Berlin

Phillips PCB, Sul D (2003) Dynamic panel estimation and homogeneity testing under cross section dependence. Econom J 6:217–259

Smeekes S (2015) Bootstrap sequential tests to determine the stationary units in a panel. J Time Ser Anal (forthcoming)

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

White H (2001) Asymptotic theory for econometricians. Academic Press, New York

Zhou Z, Shao X (2013) Inference for linear models with dependent errors. J R Stat Soc Ser B 75:323–343