The moving blocks bootstrap for panel linear regression models with individual fixed effects *

Sílvia Gonçalves[†]
Département de Sciences Économiques, CIREQ and CIRANO, Université de Montréal

March 17, 2010

Abstract

In this paper we propose a bootstrap method for panel data linear regression models with individual fixed effects. The method consists of applying the standard moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) to the vector containing all the individual observations at each point in time. We show that this bootstrap is robust to serial and cross sectional dependence of unknown form under the assumption that n (the cross sectional dimension) is an arbitrary nondecreasing function of T (the time series dimension), where $T \to \infty$, thus allowing for the possibility that both n and T diverge to infinity. The time series dependence is assumed to be weak (of the mixing type) but we allow the cross sectional dependence to be either strong or weak (including the case where it is absent). Under appropriate conditions, we show that the fixed effects estimator (as well as its bootstrap analogue) have convergence rates that depend on the degree of cross section dependence in the panel. Despite this, the same studentized test statistics can be computed without reference to the degree of cross section dependence. Our simulation results show that the moving blocks bootstrap percentile-t intervals have very good coverage properties even when the degree of serial and cross sectional correlation is large, provided the block size is appropriately chosen.

1 Introduction

This paper considers the bootstrap for panel data linear regression models with individual fixed effects. The parameters of interest are the slope coefficients β and the estimation method is the fixed effects ordinary least squares (OLS) estimator $\hat{\beta}$. The main goal of the paper is to develop a bootstrap method that allows for inference on β based on $\hat{\beta}$ in a way that is robust to the potential presence of

*I would like to thank Barbara Rossi for very stimulating discussions during my visit to Duke University in April 2008, which led to the idea of this paper. I am also grateful to Rob Taylor (the co-editor) and two anonymous referees as well as to Tim Conley, Lutz Kilian, James MacKinnon, Roger Moon, Pierre Perron, and Tim Vogelsang, who provided very useful comments on previous versions of this paper. Finally, I would like to thank all the seminar participants at Queen's University, Boston University, Boston College, Vanderbilt University, Toulouse School of Economics and USC, and conference participants at the "Bootstrap and Numerical Methods in Time Series", Granger Center, University of Nottingham, September 2008; the Third Cireq Time Series conference, Montreal, May 2009; the North American Summer Meetings, Boston, June 2009; the Yale 2009 summer conference: "Handling Dependence: Temporal, Cross-sectional, and Spatial", June 2009; the Third Meeting of the Portuguese Economic Journal, Madeira, June 2009; the JSM 2009, Washington, August 2009; the CESG 2009, Ottawa, September 2009; and the Texas A&M conference on "Resampling Methods and High Dimensional Data", March 2010. Financial support from the grant NSERC is gratefully acknowledged.

[†]Département de sciences économiques, CIREQ and CIRANO, Université de Montréal. Address: C.P.6128, succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada. Tel: (514) 343 6556. Email: silvia.goncalves@umontreal.ca

heteroskedasticity as well as serial and cross sectional dependence of unknown form in the regressors and error term of the model. Handling both forms of dependence in panel data models is important because in addition to the usual time series dependence, many panel data sets are characterized by dependencies among individuals. A source of cross section dependence can be the presence of common shocks such as macroeconomic shocks or political shocks. See Andrews (2005) for more discussion of common shocks and their effects on the properties of OLS estimators in the context of cross section regression models.

We propose the panel moving blocks bootstrap (MBB). The panel MBB consists of applying the standard MBB of Künsch (1989) and Liu and Singh (1992) to the vector containing all the individual observations at each point in time. Because it does not resample the individual observations directly, the panel MBB is expected to be robust to arbitrary forms of cross sectional dependence. By relying on the MBB, it is robust to serial dependence of unknown form as long as this dependence satisfies a mixing type condition.

We make two main contributions. First, we study the asymptotic properties of the fixed effects estimator for a panel linear regression model with individual fixed effects where the regressors and errors are subject to heteroskedasticity, and serial and cross sectional dependence of unknown forms. Building on these results, we then prove the first order asymptotic validity of the MBB in this context.

The asymptotic results are derived under the assumption that n (the cross section dimension) is an arbitrary nondecreasing function of T (the time series dimension), where $T \to \infty$, which allows for large n, large T panels. To derive the asymptotic distribution of $\hat{\beta}$ in a context that allows for arbitrary forms of cross sectional dependence, we assume that the cross sectional sums of the individual scores for β (after the fixed effects have been concentrated out) satisfy a central limit theorem when appropriately standardized by n^{ρ}, where $\rho \in [1/2, 1]$. The parameter ρ ensures that the long run variance of the standardized cross sectional sums of the scores is bounded and bounded away from zero. When the scores are subject to strong cross section dependence (due for instance to common shocks that affect all the individuals) the appropriate value of ρ is 1. If instead the scores are cross sectionally independent (or weakly dependent), $\rho = 1/2$. For each individual in the panel, we assume the regressors and error terms to be weakly dependent in the time dimension and impose only a weak exogeneity assumption on the regressors. Under these assumptions, we show that the rate of convergence of the fixed effects estimator is $\sqrt{T} n^{1-\rho}$. For the special case of strong cross sectional dependence, where $\rho = 1$, this result implies that the fixed effects estimator is only $\sqrt{T}$ consistent despite the fact that both n and T are large. Instead, if $\rho = 1/2$ (due for instance to cross sectional independence) we get $\sqrt{nT}$ convergence. Because our assumptions allow for weakly exogeneous regressors, the limiting distribution of $\hat{\beta}$ contains a bias term that is of the order $O_P\left(\frac{n^{1-\rho}}{\sqrt{T}}\right)$. In order to obtain a limiting distribution centered at zero, we require $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$. This imposes a restriction on the growth rate of n with T when $\rho < 1$. In particular, it requires $\frac{n}{T} \to 0$ when $\rho = 1/2$.

Although the rate of convergence of $\hat{\beta}$ (and of its bootstrap analogue) depends on the degree of cross section dependence in the panel (as summarized by ρ), we show that the same t and Wald statistics (as well as their bootstrap analogues) can be computed without reference to ρ. Specifically, we show that Wald statistics studentized with a standard nonparametric heteroskedasticity and autocorrelation consistent (HAC) variance estimator applied to the cross sectional averages of the estimated scores are asymptotically valid, independently of the degree of cross section dependence. This result is entirely analogous to a recent result in Hansen (2007). He shows that under cross sectional independence, the same test statistics studentized with clustered standard errors (as proposed by Arellano (1987)) can be computed without reference to the degree of serial dependence in the panel (which can be mixing or not). This is true despite the fact that in Hansen's (2007) context the rate of convergence of the OLS estimator is either $\sqrt{n}$ (if no mixing in the time series dimension exists) or $\sqrt{nT}$ (under time series mixing).

The idea of applying HAC variance estimators to cross sectional sums is not new. Driskoll and Kraay (1998) proposed this approach for computing standard errors for panel data estimators defined by moment conditions. Although quite general, the Driskoll and Kraay (1998) setup does not cover the fixed effects estimator because it does not allow for individual fixed effects (the moment conditions defining the common parameter of interest are not allowed to depend on individual time series averages). Moreover, their results assume implicitly that the degree of cross section dependence is strong (i.e. their results only cover the case $\rho = 1$).

Recently, Vogelsang (2008) proposed a new asymptotic theory for test statistics studentized with HAC variance estimators of cross sectional sums in the context of panel linear regression models with individual and time effects. Specifically, Vogelsang (2008) derives the limiting distribution of the test statistic assuming that the bandwidth is a fixed proportion of the sample size, following the approach of Kiefer and Vogelsang (2005). His simulation results show that the fixed-b asymptotic distribution is more accurate than the standard normal approximation in finite samples.

We study the finite sample performance of the MBB in the context of a panel linear regression model estimated with the fixed effects estimator, where the errors and the regressors follow a factor structure, thus displaying cross sectional and serial dependence. Our results show that the MBB performs very well, even when there is strong serial correlation in the cross sectional averages of the scores. The performance of the MBB method is robust to arbitrary forms of cross sectional correlation, including the case of cross sectional independence. It outperforms the standard normal approximation based on robust HAC standard errors. It also outperforms the fixed-b asymptotic approximation of Vogelsang (2008) when the serial correlation is strong and the block size is appropriately chosen.

The rest of this paper is organized as follows. In Section 2, we derive the asymptotic properties of the fixed effects estimator. Section 3 contains the bootstrap results. Section 4 reports the Monte Carlo simulation results and Section 5 concludes. Two mathematical appendices are included. Appendix A contains the proofs of the results in Section 2 whereas Appendix B contains the proofs of the results in Section 3.

2 Asymptotic properties of the fixed effects estimator

2.1 The model and the fixed effects estimator

We consider the following panel regression model

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it}, \quad i = 1, \ldots, n; t = 1, \ldots, T, \tag{1}$$

where α_i are individual fixed effects, y_{it} and ε_{it} are scalars, and x_{it} and β are $p \times 1$ vectors.

The parameter of interest is β and its estimator is the fixed effects OLS estimator

$$\hat{\beta}_{nT} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i), \tag{2}$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ and $\bar{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$.

Our goal in this section is to derive the asymptotic properties of $\hat{\beta}$ under general forms of heteroskedasticity and cross sectional/serial dependence in the regressors and in the errors of model (1).

2.2 Asymptotic distribution

Next we provide a set of assumptions that allow us to characterize the asymptotic distribution of $\hat{\beta}$. Throughout this paper, we let n be an arbitrary nondecreasing function of T, allowing for the possibility that n is either fixed as $T \to \infty$ or $n, T \to \infty$ jointly. Henceforth we write $n, T \to \infty$ to denote these two possibilities. In what follows, for any random vector z_{it}, we let $\|z_{it}\|_p \equiv (E|z_{it}|^p)^{1/p}$ denote its L_p norm and $|z_{it}|$ its Euclidean norm.

Assumption 1

(i) $E(\varepsilon_{it}) = 0$ and $E(x_{it}\varepsilon_{it}) = 0$, for all $i = 1, \ldots, n$, $t = 1, \ldots, T$.

(ii) For some $r > 2$, $\|x_{it}\|_{2r} \leq \Delta < \infty$ and $\|\varepsilon_{it}\|_{2r} \leq \Delta < \infty$ for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$.

(iii) For each $i = 1, \ldots, n$, $\{(x_{it}', \varepsilon_{it}) : t = 1, \ldots, T\}$ are the realization of a stationary α-mixing process with mixing coefficients $\alpha_i(k)$ such that $\sup_i \alpha_i(k) \leq \alpha(k)$ where $\alpha(k) = Ck^{-\lambda}$ for some constant C and some $\lambda > \frac{4r}{r-2}$, $r > 2$.

Assumption 1(i) requires that for each unit i in the panel the error term be mean zero and the regressors be contemporaneously uncorrelated with the errors. This weak exogeneity assumption is in principle compatible with dynamic panel models. Assumption 1(ii) imposes uniform bounds on the regressors and error moments of order $2r$ (with $r > 2$). It rules out time trends in the regressors. Assumption 1(iii) restricts the serial dependence in the time series of the regressors and error term for each individual i. See Hahn and Kuersteiner (2004) for a similar set of time series dependence assumptions in the context of bias correction for nonlinear dynamic panel data models. The mixing coefficients $\alpha_i(k)$ are defined in the usual way. Specifically, for each $i = 1, \ldots, n$, let $w_{it} = (x_{it}', \varepsilon_{it})$, and define $\mathcal{G}_{-\infty}^{i,t} = \sigma(\ldots, w_{i,t-1}, w_{it})$ and $\mathcal{G}_{t+k}^{i,+\infty} = \sigma(w_{i,t+k}, w_{i,t+k+1}, \ldots)$ as the σ-fields generated by the corresponding set of random variables. Then, for each individual i, we let

$$\alpha_i(k) \equiv \sup_t \sup_{\left\{A \in \mathcal{G}_{-\infty}^{i,t}, B \in \mathcal{G}_{t+k}^{i,+\infty}\right\}} |P(A \cap B) - P(A) P(B)|.$$

Assumption 1(iii) allows for heterogeneous forms of serial dependence across i, but imposes a uniform bound on the individual mixing coefficients. Time stationarity is imposed for simplicity. Some forms of time heterogeneity could be allowed for but this would require extra conditions controlling the degree of heterogeneity. Assumption 1 does not impose a restriction on the amount of cross sectional dependence in $\{(x'_{it}, \varepsilon_{it})\}$.

Our next assumption requires A_{nT}, the Hessian matrix underlying model (1), to be nonsingular, uniformly in (n, T). As we will see below, the asymptotic covariance matrix of $\hat{\beta}$ depends on the inverse of A_{nT}, thus justifying the need for Assumption 2.

Assumption 2 $A_{nT} \equiv \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} E\left[(x_{it} - \mu_i)(x_{it} - \mu_i)'\right]$ is nonsingular uniformly in n, T, i.e. $|\det(A_{nT})| \geq \epsilon > 0$ for all (n, T) sufficiently large, where $\mu_i \equiv E(x_{it})$.

To describe our next assumption, let

$$s_{nt} \equiv \sum_{i=1}^{n} (x_{it} - \mu_i) \varepsilon_{it}, \text{ for } t = 1, 2, \ldots, T,$$

denote the cross sectional sums of $s_{it} \equiv (x_{it} - \mu_i) \varepsilon_{it}$, the individual scores for β, after concentrating out α_i. We make the following assumption.

Assumption 3 For some parameter $\rho \in [1/2, 1]$, as $n, T \to \infty$,

$$B_{nT,\rho}^{-1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{n^{\rho}} \sum_{i=1}^{n} (x_{it} - \mu_i) \varepsilon_{it} \xrightarrow{d} N(0, I_p),$$

where $B_{nT,\rho} \equiv Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{s_{nt}}{n^{\rho}}\right)$ is $O(1)$ and is uniformly positive definite.

Assumption 3 is a high level assumption that requires the double array formed by the cross sectional sums $\{s_{nt}\}$ to satisfy a central limit theorem when appropriately standardized by n^{ρ}. The parameter ρ reflects the degree of cross sectional dependence in the individual scores, as we now explain.

The presence of (lagged) cross sectional and/or serial dependence in the individual scores $\{s_{it}\}$ will induce serial correlation in $\{s_{nt}\}$. Suppose $\{s_{nt}\}$ is a zero mean weakly dependent stationary array, where n is an arbitrary nondecreasing function of T. Then we can write

$$B_{nT,\rho} \equiv Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{s_{nt}}{n^{\rho}}\right) = \Gamma_{n,\rho}(0) + \sum_{\tau=1}^{T-1} \left(1 - \frac{\tau}{T}\right) \left(\Gamma_{n,\rho}(\tau) + \Gamma'_{n,\rho}(\tau)\right), \tag{3}$$

where for any $\tau \geq 0$,

$$\Gamma_{n,\rho}(\tau) = \frac{1}{n^{2\rho}} E\left(s_{nt} s_{nt+\tau}'\right) = \frac{1}{n^{2\rho}} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left(s_{it} s_{jt+\tau}'\right)$$

is the autocovariance matrix of $\left\{\frac{s_{nt}}{n^\rho}\right\}$ at lag τ. For $B_{nT,\rho}$ to be $O(1)$ and uniformly positive definite, these same restrictions need to be imposed on $\Gamma_{n,\rho}(0)$. This restricts the amount of cross sectional dependence in the panel. If the cross sectional dependence is pervasive and affects all individuals in the panel (such as in the case of common shocks), $\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(s_{it}s_{jt}'\right)$ is of order $O\left(n^2\right)$ and the appropriate value of ρ is 1. If instead the cross sectional dependence is sufficiently weak such that $\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(s_{it}s_{jt}'\right)$ is of order $O(n)$, we need $\rho = 1/2$. This includes the case of cross sectional independence as a special case. More generally, $\rho = 1/2$ corresponds to the case of weak cross sectional dependence, where some mixing type condition holds for $\{s_{it}\}$ in the cross sectional dimension. We allow for intermediate cases where $\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(s_{it}s_{jt}'\right)$ is of order $O\left(n^{2\rho}\right)$, with $1/2 < \rho < 1$.

Under Assumptions 1, 2 and 3, we show in the appendix (cf. Appendix A) that the fixed effects OLS estimator has the representation,

$$\sqrt{T} n^{1-\rho}\left(\hat{\beta}_{nT} - \beta\right) = A_{nT}^{-1} \underbrace{\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{n^\rho} \sum_{i=1}^{n} (x_{it} - \mu_i)\, \varepsilon_{it}}_{\xrightarrow{d} N\left(0, B_{nT,\rho}\right)} + \underbrace{A_{nT}^{-1} \cdot R_{nT,\rho}}_{Bias} + o_P(1),$$

where $A_{nT}^{-1} \cdot R_{nT,\rho}$ is a bias term of order $O_P\left(\frac{n^{1-\rho}}{\sqrt{T}}\right)$ due to the estimation of the fixed effects. In particular,

$$R_{nT,\rho} = -\frac{n^{1-\rho}}{\sqrt{T}} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it}\right)\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} (x_{is} - \mu_i)\right).$$

Under Assumptions 1, 2 and 3, the following assumption suffices for $R_{n,T} \xrightarrow{P} 0$ as $n, T \to \infty$, ensuring that the limiting distribution of $\hat{\beta}_{nT}$ is centered at zero.

Assumption 4 $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$ as $n, T \to \infty$.

When $\rho = 1$ (the strong cross sectional dependent case), $\frac{n^{1-\rho}}{\sqrt{T}} = \frac{1}{\sqrt{T}} \to 0$ as $T \to \infty$, independently of the behavior of n (which can either be fixed or diverge to infinity at any rate relatively to $T \to \infty$). Assumption 4 is then automatically satisfied, and Assumptions 1 through 3 suffice for the limiting distribution of $\hat{\beta}$ to be centered at zero. When $\frac{1}{2} \leq \rho < 1$, Assumption 4 imposes a requirement on the rate at which $n \to \infty$ with $T \to \infty$. For the leading case in which there is weak cross sectional dependence and $\rho = 1/2$, the requirement is that $\frac{n}{T} \to 0$.

Under Assumptions 1 through 4 we can state the following result.

Theorem 2.1 *Under Assumptions 1 and 2, and for any $\rho \in [1/2, 1]$ such that Assumptions 3 and 4* *hold, we have that as* $n, T \to \infty$,

$$B_{nT,\rho}^{-1/2} A_{nT} \sqrt{T} n^{1-\rho} \left(\hat{\beta}_{nT} - \beta \right) \xrightarrow{d} N\left(0, I_p\right).$$

The proof of Theorem 2.1 and of all the results in this Section are in Appendix A. According to Theorem 2.1, $\hat{\beta}$ is $\sqrt{T}n^{1-\rho}$ consistent. Thus, the rate of convergence of $\hat{\beta}$ is inversely related to the amount of cross sectional dependence that exists in $\{s_{it}\}$. This trade off is explained by the fact that the stronger the cross sectional dependence is, the less variation exists in the cross sectional dimension and therefore the slower the rate of convergence of $\hat{\beta}$ is as a function of n. In the limiting case in which $\rho = 1$ (such as when a factor model is driving the cross section dependence), $\hat{\beta}$ is only $\sqrt{T}$ consistent despite the fact that both n and T are large. Instead, when $\rho = 1/2$ (such as when there is cross sectional independence or weak cross sectional dependence), $\rho = 1/2$ and we get $\sqrt{nT}$ convergence.

Under Assumptions 1 through 4, $\hat{\beta}$ is consistent and asymptotically unbiased even for dynamic panel models, where the regressors are only weakly exogenous. If $\rho = 1$, Assumption 4 is redundant and this result is true independently of the rate of growth of n and T, as we argued above. Instead, if $\rho < 1$ Assumption 4 restricts the rate of growth of n with T, requiring that $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$ as $n, T \to \infty$. In particular, this requires $\frac{n}{T} \to 0$ under cross sectional independence. If n and T are of comparable size and the regressors contain lagged dependent variables (as in Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003); see also Bai (2009) and Moon and Weidner (2009) for more recent papers that discuss bias correction in the context of panel models estimated with interactive fixed effects), the term $R_{nT,\rho}$ defined above will not vanish and a bias term will appear in the asymptotic distribution of $\hat{\beta}$. In this case a bias correction procedure is needed. Here we do not require bias correction because we impose Assumption 4.

We can replace Assumption 4 with the following assumption.

Assumption 4′ $\frac{1}{T^2} \frac{1}{n^{2\rho}} \sum_{i,j} \sum_{t,s,u,v} E\left(\left(x_{it,k} - \mu_{i,k} \right) \varepsilon_{is} \left(x_{ju,k} - \mu_{j,k} \right) \varepsilon_{jv} \right) \leq \Delta < \infty$ for $k = 1, \ldots, p$.

We can show that Assumption $4'$ suffices for $R_{nT,\rho} \xrightarrow{P} 0$, thus ensuring that the results in Theorem 2.1 continue to hold under Assumptions 1, 2, 3 and $4'$. Although it does not impose a particular rate of growth of n with T, this condition imposes additional restrictions on the cross sectional dependence and on the exogeneity of the regressors. In particular, we can show that it is satisfied if ε_{it} is independent of x_{js} for all (i, j) and (s, t) (a very strong form of strict exogeneity), and $(x'_{it}, \varepsilon_{it})$ is independent of $\left(x'_{js}, \varepsilon_{js} \right)$ for all $i \neq j$ and all (t, s). The strict exogeneity assumption is overly restrictive when $\rho = 1$.

2.3 Variance estimation

Theorem 2.1 shows that the fixed effects OLS estimator $\hat{\beta}$ is asymptotically distributed as a normal distribution with mean zero and covariance matrix $C_{nT,\rho} \equiv A_{nT}^{-1} B_{nT,\rho} A_{nT}^{-1}$, where A_{nT} is defined in Assumption 2 and $B_{nT,\rho}$ is given in (3). Under Assumption 1, a consistent estimator of A_{nT}

is $\hat{A}_{nT} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'$, as Lemma A.2 shows. Next we provide a consistent estimator of $B_{nT,\rho}$.

We propose the following kernel estimator of $B_{nT,\rho}$,

$$\hat{B}_{nT,\rho} = \hat{\Gamma}_{nT,\rho}(0) + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(\hat{\Gamma}_{nT,\rho}(\tau) + \hat{\Gamma}'_{nT,\rho}(\tau)\right),$$

where $k(\cdot)$ is a kernel function, M is a bandwidth parameter, and for any $\tau \geq 0$,

$$\hat{\Gamma}_{nT,\rho}(\tau) = T^{-1}n^{-2\rho}\sum_{t=1}^{T-\tau}\hat{s}_{nt}\hat{s}'_{nt+\tau},$$

with $\hat{s}_{nt} = \sum_{i=1}^{n}(x_{it}-\bar{x}_i)\hat{\varepsilon}_{it}$, and $\hat{\varepsilon}_{it} = y_{it} - \bar{y}_i - (x_{it}-\bar{x}_i)'\hat{\beta}$ the fixed effects OLS residuals.

$\hat{B}_{nT,\rho}$ is a standard HAC estimator of the long run variance of the standardized cross sectional sums $\left\{\frac{s_{nt}}{n^{\rho}} = n^{-\rho}\sum_{i=1}^{n}(x_{it}-\mu_i)\varepsilon_{it}\right\}$. Because μ_i and ε_{it} are unknown, we replace these with $\bar{x}_i$ and $\hat{\varepsilon}_{it}$. To estimate $B_{nT,\rho}$ we need to take a stand on the degree of cross sectional dependence in the panel since $\hat{B}_{nT,\rho}$ depends on ρ. As we will see in the next section, we can nevertheless do inferences on β without having to commit to a particular value of ρ when constructing a confidence interval or testing hypotheses about β, if we rely on studentized statistics.

In the context of GMM estimators with panel data, Driskoll and Kraay (1998) proposed estimating the long run variance of the cross sectional averages of moment conditions defining a common parameter vector with a standard HAC variance estimator applied to the cross sectional averages of the estimated moment conditions. Nevertheless, their setup does not allow for individual fixed effects. When $\rho = 1$, $\hat{B}_{nT} \equiv \hat{B}_{nT,1}$ is an extension of the Driskoll and Kraay approach to the case of linear panel regression models with individual fixed effects.

To prove the consistency of $\hat{B}_{nT,\rho}$ for $B_{nT,\rho}$ more structure on the array $\left\{\frac{s_{nt}}{n^{\rho}}\right\}$ is required. In particular, we replace Assumption 3 with the following assumption.

Assumption 3′ For some $\rho \in [1/2, 1]$, we have that

(i) $\left\|\frac{s_{nt}}{n^{\rho}}\right\|_r \leq \Delta < \infty$, for some $r > 2$, for all (t, n).

(ii) $\left\{\frac{s_{nt}}{n^{\rho}} : t = 1, \ldots, T\right\}$ is the realization of a zero mean stationary α-mixing double array of size $-\frac{r}{r-2}$, for some $r > 2$.

(iii) $B_{nT,\rho} \equiv Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{s_{nt}}{n^{\rho}}\right)$ is such that $B_{nT,\rho} = O(1)$ and $\det(B_{nT,\rho}) > \epsilon > 0$ for all n, T sufficiently large.

We can show that Assumption 3′ implies Assumption 3. Assumption 3′(i) requires the standardized sums $\left\{\frac{s_{nt}}{n^{\rho}}\right\}$ to be L_r-bounded, uniformly in (t, n). When $\rho = 1$, this assumption is implied by Assumption 1(ii) since in this case moment restrictions on the the individual scores directly translate into moment restrictions on cross sectional averages. When $\rho < 1$, Assumption 3′(i) is satisfied under further restrictions in the cross sectional dependence. In particular, for $\rho = 1/2$, it is implied by a mixing condition on $\{s_{it}\}$ in the cross sectional dimension. Assumption $3'$(ii) restricts the serial dependence on $\left\{\frac{s_{nt}}{n^\rho}\right\}$ by postulating this array to be strong mixing. See Driskoll and Kraay (1998) for a more primitive dependence assumption on $\{s_{it}\}$ that implies Assumption $3'$(ii). In particular, it suffices that for any pair (i,j), s_{it} and $s_{jt+\tau}$ be asymptotically independent as $\tau \to \infty$. Assumption $3'$(iii) is a restatement of the last part of Assumption 3.

Our next assumption describes the class of kernels that will be considered.

Assumption 5 $k(\cdot) \in \mathcal{K}$, where $\mathcal{K} = \left\{ \begin{array}{c} k(\cdot) : \mathbb{R} \to [0,1] \text{ such that } k(x) = k(-x), \forall x \in \mathbb{R},\ k(0) = 1, \\ k(x) \text{ is continuous at 0 and at all but a finite number of points,} \\ \int_{-\infty}^{\infty} |k(x)|\, dx < \infty, \text{ and } \int_{-\infty}^{\infty} |\psi(\xi)|\, d\xi < \infty. \end{array} \right\}$

where $\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{+\infty} k(x)\, e^{i\xi x} dx$.

Assumption 5 corresponds to Assumption 1 of de Jong and Davidson (2000). As they remark, it contains many popular kernels, including the Bartlett, Quadratic Spectral, Parzen, and the Tuckey-Hanning kernels.

Assumption 6 $M \equiv M_{nT} \to \infty$ and $\frac{Mn^{1-\rho}}{\sqrt{T}} \to 0$ as $n, T \to \infty$.

Under Assumption 6, the growth rate of M is a function of ρ. When $\rho = 1$, we require that $M = o\left(\sqrt{T}\right)$ as $T \to \infty$. When $\rho < 1$, M is required to grow at a smaller rate, namely at a rate slower than $\frac{\sqrt{T}}{n^{1-\rho}}$, which diverges to infinity as $n, T \to \infty$ under Assumption 4. When $\rho = 1/2$, a sufficient condition for Assumption 6 is that $n = o\left(\sqrt{T}\right)$ and $M = o\left(T^{1/4}\right)$ as $T \to \infty$. We rely on Assumption 6 to show that estimation of s_{nt} with $\hat{s}_{nt}$ does not introduce a bias term in the estimation of $B_{nT,\rho}$.

Theorem 2.2 *Under Assumptions 1, $3'$, 4, 5 and 6, $\hat{B}_{nT,\rho} - B_{nT,\rho} \xrightarrow{P} 0$ as $n, T \to \infty$.*

If we are willing to strengthen Assumption $4'$ as follows, we can show that Assumptions 1, $3'$, $4''$, 5 and 6 suffice for consistency of $\hat{B}_{nT,\rho}$ for $B_{nT,\rho}$.

Assumption $4''$ $\frac{1}{T^2} \frac{1}{n^{2\rho}} \sum_{i,j} \sum_{t,s,u,v} \left| E\left(\left(x_{it,k} - \mu_{i,k} \right) \varepsilon_{is} \left(x_{ju,k} - \mu_{j,k} \right) \varepsilon_{jv} \right) \right| \leq \Delta < \infty$ for $k = 1, \ldots, p$.

2.4 Hypothesis testing

Consider testing the null hypothesis $H_0 : R\beta = r$ against the alternative $H_1 : R\beta \neq r$, where R is a $q \times p$ matrix of rank q and r is a $p \times 1$ vector.

We propose the following Wald statistic for testing H_0:

$$\mathcal{W}_{nT} = T \left(R\hat{\beta} - r \right)' \left[R\hat{A}_{nT}^{-1} \hat{B}_{nT} \hat{A}_{nT}^{-1} R' \right]^{-1} \left(R\hat{\beta} - r \right),$$

where $\hat{B}_{nT} \equiv \hat{B}_{nT,1}$ is a HAC estimator of the long run variance of the cross sectional *averages* $\left\{\frac{s_{nt}}{n}\right\}$.

Theorem 2.3 *Suppose Assumptions 1, 2 and 5 hold. For any* $\rho \in [1/2, 1]$ *such that Assumptions 3′, 4 and 6 hold, we have that under* $H_0 : R\beta = r$,

$$\mathcal{W}_{nT} = T\left(R\hat{\beta} - r\right)'\left[R\hat{A}_{nT}^{-1}\hat{B}_{nT}\hat{A}_{nT}^{-1}R'\right]^{-1}\left(R\hat{\beta} - r\right) \xrightarrow{d} \chi_q^2,$$

where $\hat{B}_{nT} \equiv \hat{B}_{nT,1}$.

Theorem 2.3 shows that the *same* Wald statistic $\mathcal{W}_{nT}$ can be computed and is asymptotically χ_q^2 independently of the value of ρ underlying the data generating process. In particular, this is true even though we compute $\mathcal{W}_{nT}$ as if the value of ρ was equal to 1. Suppose $\rho < 1$. Then the appropriate Wald statistic is

$$\mathcal{W}_{nT,\rho} = Tn^{2(1-\rho)}\left(R\hat{\beta} - r\right)'\left[R\hat{A}_{nT}^{-1}\hat{B}_{nT,\rho}\hat{A}_{nT}^{-1}R'\right]^{-1}\left(R\hat{\beta} - r\right),$$

where $\hat{B}_{nT,\rho}$ is a consistent estimator of $B_{nT,\rho}$. Because we can write $\hat{B}_{nT,\rho} = n^{2(1-\rho)}\hat{B}_{nT}$, the factor $n^{2(1-\rho)}$ cancels out in $\mathcal{W}_{nT,\rho}$, implying that $\mathcal{W}_{nT,\rho} = \mathcal{W}_{nT}$ for any value of ρ. This explains the invariance of the Wald statistic $\mathcal{W}_{nT}$ to the degree of cross sectional dependence that there exists in the panel.

Recently, Hansen (2007) studies the asymptotic properties of test statistics studentized with the Arellano (1987) clustered standard errors when both n and T are large. Assuming cross sectional independence, Hansen (2007) shows that the OLS estimator is $\sqrt{n}$-convergent when the time series dependence is left unrestricted whereas it is $\sqrt{nT}$ when a mixing type condition is imposed in the time series dimension. Despite this discontinuity in the convergence rates of the OLS estimator, Hansen (2007) shows that the same test statistics can be used and are asymptotically valid in the two cases (no-mixing and mixing in the time series dimension). Theorem 2.3 is the analogue of Hansen's (2007) result when we assume that the serial dependence is mixing and the cross sectional dependence can either be strong or weak.

3 Bootstrap results

The bootstrap fixed effects OLS estimator is defined as

$$\hat{\beta}_{nT}^* = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}^* - \bar{x}_i^*\right)\left(x_{it}^* - \bar{x}_i^*\right)'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}^* - \bar{x}_i^*\right)\left(y_{it}^* - \bar{y}_i^*\right),$$

where $\bar{y}_i^* = T^{-1}\sum_{t=1}^{T} y_{it}^*$ and $\bar{x}_i^* = T^{-1}\sum_{t=1}^{T} x_{it}^*$. It is the fixed effects OLS estimator of β based on the bootstrap data $\left\{z_{it}^* = (y_{it}^*, x_{it}^{*\prime})' : i = 1, \ldots, n, t = 1, \ldots, T\right\}$ obtained with the MBB as follows. Let $Z_{t,n} \equiv (z_{1t}', z_{2t}', \ldots, z_{nt}')'$ denote the $n(p+1) \times 1$ vector containing the n cross sectional observations on z_{it}. Let $\ell = \ell_T \in \mathbb{N}$ $(1 \leq \ell < T)$ denote the length of the blocks and let $B_{t,\ell} = \{Z_{t,n}, Z_{t+1,n}, \ldots, Z_{t+\ell-1,n}\}$ be the block of ℓ consecutive observations starting at observation t; $\ell = 1$ corresponds to the standard i.i.d. bootstrap on the vector $Z_{t,n}$. Assume for simplicity that $T = k\ell$.

The MBB resamples $k = T/\ell$ blocks randomly with replacement from the set of $T - \ell + 1$ overlapping blocks $\{B_{1,\ell}, \ldots, B_{T-\ell+1,\ell}\}$. Thus, if we let $I_1, \ldots, I_k$ be i.i.d. random variables uniformly distributed on $\{0, \ldots, T - \ell\}$, the MBB pseudo-data $\left\{Z^*_{t,n}, t = 1, \ldots, T\right\}$ is the result of arranging the elements of the k resampled blocks $B_{I_1+1,\ell}, \ldots, B_{I_k+1,\ell}$ in a sequence: $Z^*_{1,n} = Z_{I_1+1,n}, Z^*_{2,n} = Z_{I_1+2,n}, \ldots, Z^*_{\ell,n} = Z_{I_1+\ell,n}, Z^*_{\ell+1,n} = Z_{I_2+1,n}, \ldots, Z^*_{kl,n} = Z_{I_k+\ell,n}$. The panel MBB corresponds to the standard MBB applied to the vector that contains the n cross section observations for time t. As we will prove here, this method is robust to both serial and cross sectional dependence of unknown form when applied to the fixed effects estimator.

A word on notation. In this paper, and as usual in the bootstrap literature, P^* (E^* and Var^*) denotes the probability measure (expected value and variance) induced by the bootstrap resampling, conditional on a realization of the original data. In addition, for a sequence of bootstrap statistics Z^*_{nT}, we write $Z^*_{nT} = o_{P^*}(1)$ in probability, or $Z^*_{nT} \overset{P^*}{\to} 0$, as $n, T \to \infty$, in probability, if for any $\varepsilon > 0$, $\delta > 0$, $\lim_{n,T\to\infty} P\left[P^*\left(|Z^*_{nT}| > \delta\right) > \varepsilon\right] = 0$. Similarly, we write $Z^*_{nT} = O_{P^*}(1)$ as $n, T \to \infty$, in probability if for all $\varepsilon > 0$ there exists a $M_\varepsilon < \infty$ such that $\lim_{n,T\to\infty} P\left[P^*\left(|Z^*_{nT}| > M_\varepsilon\right) > \varepsilon\right] = 0$. Finally, we write $Z^*_{nT} \overset{d^*}{\to} Z$ as $n, T \to \infty$, in probability, if conditional on the sample, Z^*_{nT} weakly converges to Z under P^*, for all samples contained in a set with probability converging to one.

For the bootstrap results we strengthen Assumption $3'$ as follows.

Assumption 3$''$ For some $\rho \in [1/2, 1]$, we have that

(i) $\left\|\frac{s_{nt}}{n^\rho}\right\|_{3r} \leq \Delta < \infty$, for some $r > 2$, for all (t, n).

(ii) $\left\{\frac{s_{nt}}{n^\rho} : t = 1, \ldots, T\right\}$ is the realization of a zero mean stationary α-mixing double array of size $-\frac{(2+\delta)r}{r-2}$, for some $r > 2$, and some small $\delta > 0$.

(iii) $B_{nT,\rho} \equiv Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{s_{nt}}{n^\rho}\right)$ is such that $B_{nT,\rho} = O(1)$ and $\det\left(B_{nT,\rho}\right) > \epsilon > 0$ for all n, T sufficiently large.

Theorem 3.1 *Suppose Assumptions 1 and 2 hold. For any $\rho \in [1/2, 1]$ such that Assumptions $3''$ and 4 are verified, if $\ell_T \to \infty$ and $\ell_T = o\left(\sqrt{T}\right)$ as $T \to \infty$,*

$$\sup_{x\in\mathbb{R}^p}\left|P^*\left(\sqrt{T}n^{1-\rho}\left(\hat{\beta}^* - \hat{\beta}\right) \leq x\right) - P\left(\sqrt{T}n^{1-\rho}\left(\hat{\beta} - \beta\right) \leq x\right)\right| \overset{P}{\to} 0,$$

as $n, T \to \infty$.

The proof of Theorem 3.1 and of all the results in this section are in Appendix B. Theorem 3.1 justifies using the order statistics of the bootstrap distribution of $\hat{\beta}^* - \hat{\beta}$ to approximate the quantiles of the distribution of $\hat{\beta} - \beta$. This result is useful for constructing bootstrap percentile confidence intervals for β with asymptotically correct coverage probabilities.

Next we discuss bootstrapping studentized statistics. Specifically, we consider testing $H_0 : R\beta = r$ against $H_1 : R\beta \neq r$ where R and r are as defined in the previous section. The bootstrap Wald statistic we propose is

$$\mathcal{W}_{nT}^{*}=T\left(R\hat{\beta}^{*}-R\hat{\beta}\right)'\left[R\hat{A}_{nT}^{*-1}\hat{B}_{nT}^{*}\hat{A}_{nT}^{*-1}R'\right]^{-1}\left(R\hat{\beta}^{*}-R\hat{\beta}\right),$$

where $\hat{A}_{nT}^{*}$ is the bootstrap analogue of $\hat{A}_{nT}$ and is given by

$$\hat{A}_{nT}^{*}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}^{*}-\bar{x}_{i}^{*}\right)\left(x_{it}^{*}-\bar{x}_{i}^{*}\right)'.$$

To define $\hat{B}_{nT}^{*}$, let $\hat{s}_{nt}^{*}=\sum_{i=1}^{n}\left(x_{it}^{*}-\bar{x}_{i}^{*}\right)\hat{\varepsilon}_{it}^{*}$, where $\hat{\varepsilon}_{it}^{*}=y_{it}^{*}-\bar{y}_{i}^{*}-\left(x_{it}^{*}-\bar{x}_{i}^{*}\right)'\hat{\beta}^{*}$ are the bootstrap fixed effects residuals. Note that for any $j=1,\ldots,k$ and $t=1,\ldots,\ell$, $\hat{s}_{n,(j-1)\ell+t}^{*}=\sum_{i=1}^{n}\left(x_{i,I_j+t}-\bar{x}_{i}^{*}\right)\tilde{\varepsilon}_{i,I_j+t}$, where $\tilde{\varepsilon}_{i,t}=y_{it}-\bar{y}_{i}^{*}-\left(x_{it}-\bar{x}_{i}^{*}\right)'\hat{\beta}^{*}$, where I_j are i.i.d Uniform on $\{0,\ldots,T-\ell\}$. Then,

$$\hat{B}_{nT}^{*}=\frac{1}{k}\sum_{j=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}n^{-1}\hat{s}_{n,(j-1)\ell+t}^{*}\right)\left(\ell^{-1/2}\sum_{t=1}^{\ell}n^{-1}\hat{s}_{n,(j-1)\ell+t}^{*}\right)'. \tag{4}$$

$\hat{B}_{nT}^{*}$ is a consistent estimator of the bootstrap long run variance of the bootstrap cross sectional average of the scores when $\rho=1$. It is the multivariate analogue of the estimator of the MBB variance proposed by Götze and Künsch (1996) for studentizing the sample mean, adapted to the fixed effects context.

Theorem 3.2 *Suppose Assumptions 1 and 2 and 5 hold and there exists $\rho\in[1/2,1]$ such that Assumptions 3″, 4 and 6 are verified. If $\ell_T\to\infty$ such that $\ell_T=o\left(\sqrt{T}\right)$, we have that*

$$\sup_{x\in\mathbb{R}}\left|P^{*}\left(\mathcal{W}_{nT}^{*}\leq x\right)-P\left(\mathcal{W}_{nT}\leq x\right)\right|\xrightarrow{P}0,$$

as $n,T\to\infty$.

Theorem 3.2 justifies using the MBB distribution of $\mathcal{W}_{nT}^{*}$ to compute critical values for $\mathcal{W}_{nT}$ when testing H_0 against H_1. The same bootstrap Wald statistic is first order asymptotically valid under strong and weak cross sectional dependence even though $\mathcal{W}_{nT}^{*}$ is computed as if $\rho=1$. As for $\mathcal{W}_{nT}$, this is true because when $\rho<1$ the convergence rate of $\hat{\beta}^{*}$ for $\hat{\beta}$ is $\sqrt{T}n^{1-\rho}$ and the appropriate bootstrap variance estimator $\hat{B}_{nT,\rho}^{*}$ can be written as $n^{2(1-\rho)}\hat{B}_{nT}^{*}$, resulting in a bootstrap Wald statistic $\mathcal{W}_{nT,\rho}^{*}$ that is equal to $\mathcal{W}_{nT}^{*}$.

4 Monte Carlo results

This section provides simulation evidence of the finite sample performance of the MBB in the context of the following model

$$y_{it}=\alpha_i+x_{it}'\beta+\varepsilon_{it},$$

where ε_{it} and $x_{it}=\left(x_{1,it},x_{2,it},x_{3,it}\right)'$ are serially and cross sectionally correlated, and ε_{it} and x_{it} are mutually independent. Since the distribution of the test statistics based on the fixed effects OLS estimator considered in this paper is exactly invariant to the value of α_i and β, we can set $\alpha_i = \beta = 0$ without loss of generality.

To introduce cross sectional dependence we assume a factor structure for the errors and the regressors. In particular, we let

$$\varepsilon_{it} = \lambda f_{\varepsilon,t} + e_{\varepsilon,it}, \tag{5}$$

where $f_{\varepsilon,t}$ denotes a common time varying factor with factor loading λ and $e_{\varepsilon,it}$ is an idiosyncratic term independent of $f_{\varepsilon,t}$. The same structure is assumed for each of the regressors, i.e. for $l = 1, 2, 3$, we let

$$x_{l,it} = \lambda f_{l,t} + e_{l,it}, \tag{6}$$

where $(f_{\varepsilon,t}, f_{1,t}, f_{2,t}, f_{3,t})$ and $(e_{\varepsilon,it}, e_{1,it}, e_{2,it}, e_{3,it})$ are mutually independent (thus there is strict exogeneity in this model). We let

$$f_{\varepsilon,t} = af_{\varepsilon,t-1} + u_{\varepsilon,t}, \quad u_{\varepsilon,t} \sim N\left(0, 1 - a^2\right) \tag{7}$$

$$e_{\varepsilon,it} = ae_{\varepsilon,it-1} + v_{\varepsilon,it},\ v_{\varepsilon,it} \sim N\left(0, \left(1 - a^2\right)\left(1 - \lambda^2\right)\right), \tag{8}$$

and $u_{\varepsilon,t}$ and $v_{\varepsilon,it}$ are mutually independent. These variables are uncorrelated over time and across units with $f_{\varepsilon,0} \sim N(0,1)$ and $e_{\varepsilon,i0} \sim N\left(0, 1 - \lambda^2\right)$. Thus, the error term for each individual is correlated over time with an autocorrelation coefficient equal to a^τ at lag τ, whereas the error terms of any two individuals (i, j) are equicorrelated according to $\lambda^2 a^\tau$. A similar $AR(1)$ structure is assumed for each regressor.[1]

We let $\lambda \in \left\{0, \sqrt{0.5}\right\}$, where $\lambda = 0$ implies cross sectional independence whereas $\lambda = \sqrt{0.5}$ implies a cross sectional correlation of 0.5 for each regressor and error term (note that this implies that $s_{it} \equiv x_{it}\varepsilon_{it}$ is equicorrelated with correlation equal to $\lambda^4 = 0.25$). Thus, $\lambda = 0$ corresponds to a value of $\rho = 1/2$ whereas $\lambda = \sqrt{0.5}$ corresponds to $\rho = 1$.

We examine the finite sample performance of two-sided symmetric 95% confidence intervals for β_1 based on the studentized statistic

$$t_{\hat{\beta}_1} \equiv \frac{\sqrt{T}\left(\hat{\beta}_{1,nT} - \beta_1\right)}{\sqrt{\hat{C}_{nT}^{(1,1)}}},$$

where $\hat{\beta}_{1,nT}$ is the first element of $\hat{\beta}_{nT}$ and $\hat{C}_{nT}^{(1,1)}$ denotes the element $(1,1)$ of $\hat{C}_{nT} = \hat{A}_{nT}^{-1}\hat{B}_{nT}\hat{A}_{nT}^{-1}$, with $\hat{A}_{nT}$ and $\hat{B}_{nT}$ as given in Section 2. In particular, $\hat{B}_{nT}$ is a HAC estimator of the variance of the cross sectional *averages* of the individual scores based on the Bartlett kernel where the bandwidth is chosen by Andrews' (1991) automatic procedure based on approximating AR(1) models for the elements of $\frac{\hat{s}_{nt}}{n} \equiv n^{-1}\sum_{i=1}^{n} \hat{s}_{it}$.

[1] We also ran simulations for an AR(1)-t_6 model where the normal distributions in (7) and (8) were replaced with Student-t_6 distributions, suitably scaled so as to guarantee that $Var\left(f_{\varepsilon,t}\right) = 1$ and $Var\left(e_{\varepsilon,it}\right) = 1 - \lambda^2$; and for an MA(1)-Gaussian model, where (7) and (8) were generated by MA(1) models. The (unreported) results followed the same patterns as for the AR(1)-Gaussian model.

We consider confidence intervals based on the normal approximation ($N(0,1)$ intervals), on the new fixed-b asymptotic theory of Vogelsang (2008) (Fixed-b), and on the bootstrap (MBB). The $N(0,1)$ intervals rely on the standard normal distribution for computing critical values for $t_{\hat{\beta}_1}$. The Fixed-b intervals rely on the fixed-b asymptotic distribution of Vogelsang (2008) (see also Kiefer and Vogelsang (2005)), where we set $b = \frac{\hat{M}}{T}$ with $\hat{M}$ equal to the chosen data driven bandwidth.

The MBB intervals rely on the bootstrap distribution of

$$t_{\hat{\beta}_1^*} = \frac{\sqrt{T}\left(\hat{\beta}_{1,nT}^* - \hat{\beta}_{1,nT}\right)}{\sqrt{\hat{C}_{nT}^{*(1,1)}}}$$

for computing the critical values of the distribution of $t_{\hat{\beta}_1}$. Here $\hat{C}_{nT}^{*(1,1)}$ is the $(1,1)$-element of $\hat{C}_{nT}^* = \hat{A}_{nT}^{*-1}\hat{B}_{nT}^*\hat{A}_{nT}^{*-1}$, with $\hat{A}_{nT}^*$ and $\hat{B}_{nT}^*$ as given in Section 3. In particular, $\hat{B}_{nT}^*$ is the analogue of the Götze and Künsch (1996) bootstrap variance estimator for the panel context.[2]

To choose the block size, we exploit the asymptotic equivalence between the MBB and the Bartlett kernel variance estimators and use the integer part of the automatic bandwidth chosen by Andrews' automatic procedure. For comparison purposes, we also include the i.i.d. bootstrap where $\ell = 1$.

Figures 1 and 2 contain the results for $\lambda = 0.5$ and for $\lambda = 0$, respectively. We find six plots in each figure, corresponding to two different values of $T \in \{25, 100\}$ and three different values of $a \in \{0, 0.5, 0.9\}$. Each plot depicts the actual coverage rates of each interval as a function of $n \in \{10, 20, \ldots, 100\}$. The results are based on 2,000 random samples for each (n, T) combination and the bootstrap intervals are based on 999 bootstrap replications for each sample. We show results for four types of intervals: the confidence intervals based on the normal approximation ($N(0,1)$), the fixed-b intervals (Fixed-b) based on the Vogelsang (2008) approach, and the MBB intervals implemented with a data-driven block size (MBB) and a block size equal to 1 (MBB1).

Figure 1 shows that when there is no serial correlation ($a = 0$) but individuals are cross sectionally correlated, some finite sample distortions arise for $T = 25$, especially for the $N(0,1)$ intervals (whose rates are in the range 89.5%-91.5%). The Fixed-b intervals outperform the $N(0,1)$ intervals by a small margin, with rates between 91.5%-94% for $T = 25$. The MBB with a data-dependent block size performs the best (the selected ℓ was on average 1.60 across all values of n and T). For $T = 100$, the differences between all methods decrease and they all perform well. When we increase a to 0.5, the performance of all methods deteriorates, but this is more pronounced for the $N(0,1)$ intervals (with rates around 85% when $T = 25$). The Fixed-b intervals outperform the $N(0,1)$ intervals, displaying

[2] An alternative approach in computing the bootstrap statistic $t_{\hat{\beta}_1^*}$ is to replace $\hat{B}_{nT}^*$ with an estimator of the same form as $\hat{B}_{nT}$, where the bootstrap data replaces the original data. This naive approach was recently considered by Gonçalves and Vogelsang (2009) in the pure time series context. Their results show that there is a close link between the naive bootstrap and the fixed-b asymptotic theory, with the naive i.i.d. bootstrap (where the block size equals 1) following almost exactly the fixed-b asymptotic theory. In unreported results, we found that this same patterns hold for the panel context. We also found that the naive block bootstrap approach implemented with a data dependent bandwidth and a data dependent block size was dominated by the MBB based on the Götze and Künsch (1996) variance estimator as implemented in Figures 1 and 2.

rates between 88% and 90% when $T = 25$, followed by the MBB1. Choosing a block size larger than one implies a further coverage error reduction (the average value of the block size was 2.00 across all values of n when $T = 25$ and 3.5 when $T = 100$). When $a = 0.9$ and $\lambda = \sqrt{0.5}$, the degree of undercoverage increases significantly for all methods (except for MBB). Of all methods, the $N(0,1)$ intervals are the most distorted, with coverage rates between 62% and 65% for $T = 25$ (these rates increase to about 75% for $T = 100$, across all values of n). The Fixed-b intervals outperform the i.i.d. bootstrap method (MBB1) and the $N(0,1)$ intervals, for all values of T and n. Overall, the best method is MBB (with a data-driven block size equal on average to 4.4 when $T = 25$ and 12.2 when $T = 100$). The performance of the MBB intervals is very good, even for the smallest sample size, where the actual rates are between 87.6% and 91.7%.

When $\lambda = 0$, a comparison between Figures 1 and 2 shows that the degree of coverage distortions for all methods decreases but the results follow the same patterns as when $\lambda = \sqrt{0.5}$. In particular, the MBB is the best performing method among the ones we consider and its performance is very good across different values of a, T, and n.

5 Conclusion

In this paper we introduce and show the first order asymptotic validity of the moving blocks bootstrap for fixed effects estimators of panel linear regression models with individual fixed effects. We show that this method is robust to heteroskedasticity and cross sectional and serial dependence of unknown forms under the assumption that n is an arbitrary nondecreasing function of T and $T \to \infty$ (thus allowing for the possibility that both n and T diverge to infinity). We derive our results under weak time series dependence, but allow for the possibility that the cross section dependence is either weak (with cross sectional independence as a special case) or strong. We show that although the fixed effects estimator and its bootstrap analogue have convergence rates that are a function of the degree of cross section dependence, the same t and Wald statistics can be computed independently of how much cross section dependence there is in the data. Our simulation results show that the block bootstrap has better finite sample properties than competitors based on the normal approximation or on the fixed-b asymptotic theory, as derived by Vogelsang (2008), provided the block size is appropriately chosen (and given that the bandwidth is chosen in a data dependent fashion).

The crucial condition under which the MBB works is that a mixing condition holds in the time series dimension. If such a condition does not hold, the MBB is not valid. This occurs for instance if the error term includes an individual specific random effect that is uncorrelated with the regressors and the estimated model does not include an individual fixed effect, as in the simulations of Hounkannounon (2008). In this case, all observations for a given individual are equicorrelated over time and this will not satisfy our mixing conditions in the time series dimension.

The MBB as well as the HAC standard errors do not exploit any mixing in the cross sectional dimension. This is an attractive feature because no natural ordering in the cross sectional dimension need exist (other approaches that rely on the availability of a cross sectional ordering have been proposed in the literature on cross sectional dependence, see e.g. Conley (1999), and more recently, Ibragimov and Mueller (2009), Bester, Conley and Hansen (2008) and Bester, Conley, Hansen and Vogelsang (2008)). Nevertheless, if an ordering in the cross sectional dimension exists, the MBB as proposed here may not be the most efficient method. Proposing a bootstrap method that exploits the mixing conditions in both dimensions (cross sectional and time series) is an important area of research.

A Appendix A: proofs of the results in Section 2.

This Appendix is organized as follows. First, we state some auxiliary lemmas and their proofs. Then, we prove the results in Section 2. Throughout we will let $\mu_i \equiv E\left(x_{it}\right)$ for all (i,t). We first state a well known maximal inequality for strong mixing double arrays.

Lemma A.1 *Let* $\{\mathcal{X}_{Nt} : t = 1, 2, \ldots, N = 1, 2, \ldots\}$ *be a zero mean* $\alpha-$*mixing array with mixing coefficients*
$\alpha\left(k\right) \equiv \sup_t \sup_{\left\{A \in \mathcal{G}_{-\infty}^{Nt}, B \in \mathcal{G}_{t+k}^{N,+\infty}\right\}} \left|P\left(A \cap B\right) - P\left(A\right) P\left(B\right)\right|$, *where* $\mathcal{G}_{-\infty}^{Nt} = \sigma\left(\ldots, \mathcal{X}_{Nt}\right)$ *and* $\mathcal{G}_{t+k}^{N,+\infty} = \sigma\left(\mathcal{X}_{N,t+k}, \ldots\right)$. *Then for some constant* K *and for any* $1 < p < r$,

(i) *If* $1 < p < 2$, $\left\|\max_{j \leq N}\left|\sum_{t=1}^{j} \mathcal{X}_{Nt}\right|\right\|_p \leq K\left(\sum_{k=1}^{\infty} \alpha\left(k\right)^{\frac{1}{p}-\frac{1}{r}}\right)\left(\sum_{t=1}^{N}\left\|\mathcal{X}_{Nt}\right\|_r^p\right)^{1/p}$.

(ii) *If* $p \geq 2$, $\left\|\max_{j \leq N}\left|\sum_{t=1}^{j} \mathcal{X}_{Nt}\right|\right\|_p \leq K\left(\sum_{k=1}^{\infty} \alpha\left(k\right)^{\frac{1}{p}-\frac{1}{r}}\right)\left(\sum_{t=1}^{N}\left\|\mathcal{X}_{Nt}\right\|_r^2\right)^{1/2}$.

The next set of results are auxiliary in deriving the asymptotic distribution of $\hat{\beta}_{nT}$.

Lemma A.2 *Under Assumption 1, as* $n, T \to \infty$,

a) $\tilde{A}_{nT} - A_{nT} \to^P 0$, *where* $\tilde{A}_{nT} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \mu_i\right)\left(x_{it} - \mu_i\right)'$.

b) $\frac{1}{n}\sum_{i=1}^{n}\left(\mu_i - \bar{x}_i\right)\left(\mu_i - \bar{x}_i\right)' \to^P 0$.

c) $\hat{A}_{nT} - A_{nT} \to^P 0$, *where* $\hat{A}_{nT} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)'$.

Lemma A.3 *Under Assumptions 1, 3 (or 3′) and 4, as* $n, T \to \infty$,

a) $B_{nT,\rho}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(x_{it} - \mu_i\right)\varepsilon_{it} \to^d N\left(0, I_p\right)$.

b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(\mu_i - \bar{x}_i\right)\varepsilon_{it} \to^P 0$.

c) $B_{nT,\rho}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(x_{it} - \bar{x}_i\right)\varepsilon_{it} \to^d N\left(0, I_p\right)$.

Proof of Lemma A.1. By Corollary 17.6 (Davidson, 1994, p. 265), we can show that $\left\{X_{Nt}, \mathcal{G}_{-\infty}^{Nt}\right\}$ is an L_p-mixingale with mixingale coefficients $\psi(k) = \alpha(k)^{1/p-1/r}$ and mixingale constants $c_{Nt} = O(\|X_{Nt}\|_r)$. We can then apply the maximal inequalities for L_p-mixingales given e.g. in Hansen (1991, 1992).

Proof of Lemma A.2. a) We show that $E\left|\tilde{A}_{nT} - A_{nT}\right| \to 0$ as $n, T \to \infty$, from which the result follows given Markov's inequality. For each i, let $w_{it} \equiv \left(x_{it,k} - \mu_{i,k}\right)\left(x_{it,l} - \mu_{i,l}\right)$, a typical (k,l) element of $(x_{it} - \mu_i)(x_{it} - \mu_i)'$. Define $\phi_{iT} \equiv \sum_{t=1}^{T}(w_{it} - E(w_{it}))$. For each i, under Assumption 1(ii), $\{w_{it} - E(w_{it})\}$ is a zero mean process with $\sup_{i,t}\|w_{it}\|_r \leq \Delta < \infty$. By 1(iii), it is α-mixing of size $-\frac{4r}{r-2}$ uniformly in i. Thus, it follows that

$$E\left|\tilde{A}_{nT,kl} - A_{nT,kl}\right| \leq \frac{1}{nT}\sum_{i=1}^{n}\sup_{1\leq i\leq n} E|\phi_{iT}| \leq \frac{1}{nT}\sum_{i=1}^{n}\sup_{1\leq i\leq n}\|\phi_{iT}\|_2.$$

For each i, Assumption 1 and Lemma A.1 imply that $E|\phi_{iT}| \leq \|\phi_{iT}\|_2 = O\left(\sqrt{T}\right)$ uniformly in i. Thus, it follows that $E\left|\tilde{A}_{nT,kl} - A_{nT,kl}\right| = O\left(T^{-1/2}\right) = o(1)$. b) We can write

$$R_{1,nT} \equiv \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \bar{x}_i)(\mu_i - \bar{x}_i)' = -\frac{1}{n}\sum_{i=1}^{n}T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}(x_{it} - \mu_i)(x_{is} - \mu_i)' = -\frac{1}{n}\sum_{i=1}^{n}T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}z_{it}z_{is}',$$

where we let $z_{it} \equiv x_{it} - \mu_i$. We show that $E|R_{1,nT}| \to 0$ and consequently $R_{1,nT} \to^P 0$ by Markov's inequality. Define $\xi_{iT} \equiv \sum_{t=1}^{T} z_{it}$. It follows that $R_{1,nT} = -\frac{1}{nT^2}\sum_{i=1}^{n}\left(\sum_{t=1}^{T} z_{it}\right)\left(\sum_{s=1}^{T} z_{is}'\right) = -\frac{1}{nT^2}\sum_{i=1}^{n}\xi_{iT}\xi_{iT}'$. The triangle inequality and the Cauchy-Schwartz inequality imply that $E|R_{1,nT}| \leq \frac{1}{nT^2}\sum_{i=1}^{n}E\left|\xi_{iT}\xi_{iT}'\right| \leq \frac{1}{nT^2}\sum_{i=1}^{n}\|\xi_{iT}\|_2^2$. Next we show that $\|\xi_{iT}\|_2 = O\left(T^{1/2}\right)$ uniformly in i using a maximal inequality for mixing processes. This implies that $E|R_{1,nT}| = O\left(\frac{1}{T}\right) = o(1)$ as $T \to \infty$. Specifically, for each i, Assumption 1 implies that z_{it} is a zero mean α-mixing process with $\alpha_i(k) \leq \alpha(k)$. Thus, by Lemma A.1, we have that $\|\xi_{iT}\|_2 \leq K\sum_{k=1}^{\infty}\alpha(k)^{\frac{1}{2}-\frac{1}{r}}\left(\sum_{t=1}^{T}\|z_{it}\|_r^2\right)^{1/2}$ for some $r > 2$. Assumption 1(ii) implies that $\|z_{it}\|_r \leq \Delta < \infty$ whereas Assumption 1(iii) implies that $\sum_{k=1}^{\infty}\alpha(k)^{\frac{1}{2}-\frac{1}{r}} < \infty$, thus proving that $\|\xi_{iT}\|_2 \leq CT^{1/2}$ for some constant C. c) Adding and subtracting appropriately, we can write $\hat{A}_{nT} - A_{nT} = \tilde{A}_{nT} + a_{2,nT} + a_{3,nT} + a_{4,nT} - A_{nT}$, where $a_{2,nT} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \mu_i)(\mu_i - \bar{x}_i)'$, $a_{3,nT} = a_{2,nT}'$, and $a_{4,nT} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\mu_i - \bar{x}_i)(\mu_i - \bar{x}_i)'$. By part a) of this Lemma, $\tilde{A}_{nT} - A_{nT} = o_P(1)$. We can show that $a_{2,nT} = o_P(1)$ as $n, T \to \infty$ given part b). The same holds for $a_{3,nT}$ and $a_{4,nT}$, thus completing the proof.

Proof of Lemma A.3. Part a) is Assumption 3. Alternatively, if we let $\frac{s_{nt}}{n^\rho} \equiv \frac{1}{n^\rho}\sum_{i=1}^{n}(x_{it} - \mu_i)\varepsilon_{it}$, under Assumption 3′ (i)-(iii), the array $\left\{\frac{s_{nt}}{n^\rho} : t = 1, 2, \ldots, T\right\}$ satisfies the assumptions of Theorem 5.20 of White (2001), implying a). Part b) follows by noting that $\mu_i - \bar{x}_i = -\frac{1}{T}\sum_{t=1}^{T}(x_{it} - \mu_i)$, and using Assumption 4. Part c) follows from a) and b).

Proof of Theorem 2.1. We can write

$$\sqrt{T}n^{1-\rho}\left(\hat{\beta}_{nT} - \beta\right) = \hat{A}_{nT}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^\rho}\sum_{i=1}^{n}(x_{it} - \bar{x}_i)\varepsilon_{it}.$$

Under Assumption 1, by Lemma A.2, $\hat{A}_{nT} - A_{nT} \to^P 0$ as $n, T \to \infty$. Assumption 2 guarantees A_{nT}^{-1} exists. Thus, we have that $\sqrt{T} n^{1-\rho} \left(\hat{\beta}_{nT} - \beta\right) = A_{nT}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{n^\rho} \sum_{i=1}^n (x_{it} - \bar{x}_i) \varepsilon_{it} + o_P(1)$. Adding and subtracting appropriately yields

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{n^\rho} \sum_{i=1}^n (x_{it} - \bar{x}_i) \varepsilon_{it} = \underbrace{\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{n^\rho} \sum_{i=1}^n (x_{it} - \mu_i) \varepsilon_{it}}_{\to^d N\left(0, B_{nT,\rho}\right)} + \underbrace{R_{nT,\rho}}_{=o_P(1)},$$

given Assumptions 3 and 4 (or 4′) and Lemma A.2. This proves the result.

Proof of Theorem 2.2. By definition,

$$\hat{B}_{nT,\rho} - B_{nT,\rho} = T^{-1} \sum_{t=1}^T n^{-2\rho} \hat{s}_{nt} \hat{s}_{nt}' + 2 \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right) \left(T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} \hat{s}_{nt} \hat{s}_{nt+\tau}' + T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} \hat{s}_{nt+\tau} \hat{s}_{nt}'\right) - B_{nT,\rho},$$

where $\hat{s}_{nt} = \sum_{i=1}^n (x_{it} - \bar{x}_i) \hat{\varepsilon}_{it}$, where $\hat{\varepsilon}_{it}$ is the fixed effects OLS residual. In particular, $\hat{\varepsilon}_{it} = (y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)' \hat{\beta} = y_{it} - x_{it}' \hat{\beta} - \hat{\alpha}_i$, where $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i' \hat{\beta}$. Since $\varepsilon_{it} = y_{it} - \alpha_i - x_{it}' \beta$, it follows that $\hat{\varepsilon}_{it} = \varepsilon_{it} - x_{it}' \left(\hat{\beta} - \beta\right) - (\hat{\alpha}_i - \alpha_i)$. We can write $\hat{\alpha}_i - \alpha_i = -\bar{x}_i' \left(\hat{\beta} - \beta\right) + \bar{\varepsilon}_i$, implying that $\hat{\varepsilon}_{it} = \varepsilon_{it} - x_{it}' \left(\hat{\beta} - \beta\right) - (\hat{\alpha}_i - \alpha_i) = \varepsilon_{it} - (x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right) - \bar{\varepsilon}_i$. We can write $\hat{s}_{nt} = s_{nt} + r_{nt}$, where $s_{nt} = \sum_{i=1}^n (x_{it} - \mu_i) \varepsilon_{it}$, and $r_{nt} = (a_{nt} + b_{nt}) + c_{nt} \equiv d_{nt} + c_{nt}$. Let $\xi_{iT} = \sum_{t=1}^T (x_{it} - \mu_i) \equiv \sum_{t=1}^T z_{it}$ and $\eta_{iT} = \sum_{t=1}^T \varepsilon_{it}$. Then,

$$\begin{aligned}
a_{nt} &= \sum_{i=1}^n (\mu_i - \bar{x}_i) \varepsilon_{it} = -T^{-1} \sum_{i=1}^n \xi_{iT} \varepsilon_{it}; \\
b_{nt} &= -\sum_{i=1}^n (x_{it} - \mu_i) \bar{\varepsilon}_i - \sum_{i=1}^n (\mu_i - \bar{x}_i) \bar{\varepsilon}_i = -T^{-1} \sum_{i=1}^n z_{it} \eta_{iT} + T^{-2} \sum_{i=1}^n \xi_{iT} \cdot \eta_{iT} \equiv b_{1nt} + b_{2nT}; \\
c_{nt} &\equiv -\sum_{i=1}^n (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right).
\end{aligned}$$

Substituting $\hat{s}_{nt}$ in $\hat{B}_{nT,\rho} - B_{nT,\rho}$ yields $\hat{B}_{nT,\rho} - B_{nT,\rho} = I_{1,nT} + I_{2,nT} + I_{3,nT} + I_{3,nT}'$, where

$$\begin{aligned}
I_{1nT} &\equiv T^{-1} \sum_{t=1}^T n^{-2\rho} s_{nt} s_{nt}' + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right) \left(T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} s_{nt} s_{nt+\tau}' + T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} s_{nt+\tau} s_{nt}'\right) - B_{nT,\rho}; \\
I_{2nT} &\equiv \underbrace{T^{-1} \sum_{t=1}^T n^{-2\rho} s_{nt} r_{nt}'}_{\equiv J_{sr,nT}^0} + \underbrace{T^{-1} \sum_{t=1}^T n^{-2\rho} r_{nt} s_{nt}'}_{\equiv J_{sr,nT}^{0\prime}} + \underbrace{T^{-1} \sum_{t=1}^T n^{-2\rho} r_{nt} r_{nt}'}_{J_{rr,nT}^0}; \text{ and} \\
I_{3,nT} &\equiv \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right) \left(\underbrace{T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} s_{nt} r_{nt+\tau}'}_{J_{sr,nT}^\tau} + \underbrace{T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} r_{nt} s_{nt+\tau}'}_{J_{rs,nT}^\tau} + \underbrace{T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} r_{nt} r_{nt+\tau}'}_{J_{rr,nT}^\tau}\right).
\end{aligned}$$

By Assumption $3'$(i) and (ii), $I_{1nT,\rho} \xrightarrow{P} 0$ given Theorem 2.1 in de Jong and Davidson (2000) applied to the array $\left\{\frac{s_{nt}}{n^{\rho}}\right\}$ provided Assumption 5 holds and $M \to \infty$ such that $M/T \to 0$. Next we show that each of the remaining terms is $o_P(1)$. Specifically, we bound each of the J terms above. Start with $J^0_{sr,nT}$. We can write

$$J^0_{sr,nT} = T^{-1}\sum_{t=1}^{T} n^{-2\rho} s_{nt} d'_{nt} + T^{-1}\sum_{t=1}^{T} n^{-2\rho} s_{nt} c'_{nt} \equiv J^0_{sr.1,nT} + J^0_{sr.2,nT}.$$

Replacing d_{nt} with $a_{nt} + b_{nt}$ yields

$$J^0_{sr.1,nT} = T^{-1}\sum_{t=1}^{T} n^{-2\rho} s_{nt} a'_{nt} + T^{-1}\sum_{t=1}^{T} n^{-2\rho} s_{nt} b'_{nt} \equiv X_{sa,nT} + X_{sb,nT}.$$

By the Cauchy-Schwartz inequality,

$$|X_{sa,nT}| \leq \left(T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} s_{nt}\right|^2\right)^{1/2} \left(T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} a_{nt}\right|^2\right)^{1/2}.$$

Similarly,

$$|X_{sb,nT}| \leq \left(T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} s_{nt}\right|^2\right)^{1/2} \left(T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} b_{nt}\right|^2\right)^{1/2}.$$

Next we provide a bound for each of the sums in $X_{sa,nT}$ and $X_{sb,nT}$ which holds unifomly in n, T. Consider $T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} s_{nt}\right|^2$. Under Assumption $3'$(i), $\|n^{-\rho} s_{nt}\|_r \leq \Delta < \infty$ for some $r > 2$ and therefore $T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} s_{nt}\right|^2 = O_P(1)$. Next consider $T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} a_{nt}\right|^2$. The Minkowski and the Cauchy-Schwartz inequalities imply that $\|n^{-\rho} a_{nt}\|_2 \leq T^{-1} n^{-\rho} \sum_{i=1}^{n} \|\xi_{iT}\|_4 \|\varepsilon_{it}\|_4 \leq \Delta T^{-1} n^{-\rho} \sum_{i=1}^{n} \|\xi_{iT}\|_4$, where we have used Assumption 1(ii) to bound $\|\varepsilon_{it}\|_4$. By definition of the L_4- and the Euclidean norms, $\|\xi_{iT}\|_4 = \left(E\left|\xi_{iT}\right|^4\right)^{1/4} = \left(E\left|\sum_{k=1}^{p} \xi^2_{iT,k}\right|^2\right)^{1/4} \leq \left(\sum_{k=1}^{p} E\left|\xi_{iT,k}\right|^4\right)^{1/4} \leq \sum_{k=1}^{p} \left\|\xi_{iT,k}\right\|_4$. For each $k = 1, \ldots, p$, we can show that $\left\|\xi_{iT,k}\right\|_4 \leq C\sqrt{T}$ for some constant C independent of i. In particular, Lemma A.1 implies that $\left\|\xi_{iT,k}\right\|_4 \leq K \sum_{j=1}^{\infty} \alpha(j)^{\frac{1}{4}-\frac{1}{r'}} \left(\sum_{t=1}^{T} \|z_{it,k}\|^2_{r'}\right)^{1/2}$ for some $r' > 4$. Setting $r' = 2r$ (where $r > 2$) and using the size condition in Assumption 1(iii$'$) [we need $\lambda > \frac{4r}{r-2}$] and the moment condition in Assumption 1(ii), it follows that $\left\|\xi_{iT,k}\right\|_4 = O\left(\sqrt{T}\right)$ uniformly in i. Thus, $\|n^{-\rho} a_{nt}\|_2 \leq C n^{1-\rho} T^{-1/2}$. This implies that $\left(T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} a_{nt}\right|^2\right)^{1/2} = O_P\left(\frac{n^{1-\rho}}{\sqrt{T}}\right)$, and therefore $X_{sa,nT} = O_P\left(\frac{n^{1-\rho}}{\sqrt{T}}\right) = o_P(1)$, under Assumptions 1, 3 and 4. Alternatively, using the definition of the Euclidean norm, we can write

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} \left|n^{-\rho} a_{nt}\right|^2 &= T^{-1}\sum_{t=1}^{T} \left|-n^{-\rho} T^{-1} \sum_{i=1}^{n} \xi_{iT}\varepsilon_{it}\right|^2 = \sum_{k=1}^{p} T^{-3} n^{-2\rho} \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} \xi_{iT,k}\varepsilon_{it}\xi_{jT,k}\varepsilon_{jt} \\ &= \sum_{k=1}^{p} T^{-3} n^{-2\rho} \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n} z_{is,k}\varepsilon_{it} z_{ju,k}\varepsilon_{jt}. \end{aligned}$$

This term will be $O_P\left(T^{-1}\right)$ under Assumption 4$''$, implying that $X_{sa,nT} = O_P\left(T^{-1/2}\right)$. Next we analyze $T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}b_{nt}\right|^2$. Since $b_{nt} = b_{1nt} + b_{2nT}$, we have that $\left\|n^{-\rho}b_{nt}\right\|_2 \leq \left\|n^{-\rho}b_{1nt}\right\|_2 + \left\|n^{-\rho}b_{2nT}\right\|_2$, where

$$\left\|n^{-\rho}b_{1nt}\right\|_2 \leq n^{-\rho}T^{-1}\sum_{i=1}^{n}\left\|z_{it}\eta_{iT}\right\|_2 \leq n^{-\rho}T^{-1}\sum_{i=1}^{n}\left\|z_{it}\right\|_4\left\|\eta_{iT}\right\|_4 = O\left(\frac{n^{1-\rho}}{T^{1/2}}\right),$$

given the mixing and moment conditions imposed under Assumption 1. Similarly,

$$\left\|n^{-\rho}b_{2nT}\right\|_2 \leq n^{-\rho}T^{-2}\sum_{i=1}^{n}\left\|\xi_{iT}\right\|_4\cdot\left\|\eta_{iT}\right\|_4 = O\left(n^{1-\rho}\right)O\left(T^{-2}\right)O\left(T^{1/2}\right)O\left(T^{1/2}\right) = O\left(\frac{n^{1-\rho}}{T}\right).$$

This implies then that $\left(T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}b_{nt}\right|^2\right)^{1/2} = O_P\left(n^{(1-\rho)}T^{-1/2}\right)$ and thus $X_{sb,nT} = O_P\left(n^{(1-\rho)}T^{-1/2}\right) = o_P(1)$ if we impose Assumption 4. Alternatively, we can write

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}b_{1nt}\right|^2 &= T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}T^{-1}\sum_{i=1}^{n}z_{it}\eta_{iT}\right|^2 = \sum_{k=1}^{p}T^{-3}n^{-2\rho}\sum_{i,j}\sum_{t}z_{it,k}\eta_{iT}z_{jt,k}\eta_{jT} \\
&= T^{-1}\sum_{k=1}^{p}\left\{T^{-2}n^{-2\rho}\sum_{i,j}\sum_{t,s,u}z_{it,k}\varepsilon_{is}z_{jt,k}\varepsilon_{ju}\right\}.
\end{aligned}$$

The term in the curly brackets will be $O_P(1)$ under Assumption 4$''$, implying that $T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}b_{1nt}\right|^2 = O_P\left(T^{-1}\right)$. Similarly,

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}b_{2nT}\right|^2 &= \left|n^{-\rho}T^{-2}\sum_{i=1}^{n}\xi_{iT}\cdot\eta_{iT}\right|^2 = \sum_{k=1}^{p}T^{-4}n^{-2\rho}\sum_{i,j}\xi_{iT,k}\eta_{iT}\xi_{jT,k}\eta_{jT} \\
&= T^{-2}\sum_{k=1}^{p}\left\{T^{-2}n^{-2\rho}\sum_{i,j}\sum_{t,s,u,v}z_{it,k}\varepsilon_{is}z_{ju,k}\varepsilon_{jv}\right\}.
\end{aligned}$$

The term in the curly brackets is $O_P(1)$ under Assumption 4$''$, implying that the whole expression is $O_P\left(T^{-2}\right)$. Thus, under this condition, $X_{sb,nT} = O_P\left(T^{-1}\right) = o_P(1)$. To conclude, we get that $J^0_{sr.1,nT} = O_P\left(\frac{n^{1-\rho}}{T^{1/2}}\right)$ under Assumptions 1,3$'$ and 4, or $O_P\left(\frac{1}{T^{1/2}}\right)$ under Assumption 4$''$. Next we study $J^0_{sr.2,nT} = T^{-1}\sum_{t=1}^{T}n^{-2\rho}s_{nt}c'_{nt}$, where $c_{nt} \equiv -\sum_{i=1}^{n}\left(x_{it}-\bar{x}_i\right)\left(x_{it}-\bar{x}_i\right)'\left(\hat{\beta}-\beta\right)$. Let $p=1$ for simplicity and write $W_{nt} = n^{-1}\sum_{i=1}^{n}\left(x_{it}-\bar{x}_i\right)^2$. Then we have that

$$J^0_{sr.2,nT} = \underbrace{\left[n^{1-\rho}\left(\hat{\beta}-\beta\right)\right]}_{O_P\left(\frac{1}{\sqrt{T}}\right)}\cdot T^{-1}\sum_{t=1}^{T}\left(n^{-\rho}s_{nt}\right)\cdot W_{nt},$$

where $n^{1-\rho}\left(\hat{\beta}-\beta\right) = O_P\left(\frac{1}{\sqrt{T}}\right)$ under the assumptions of Theorem 2.1. By the Cauchy-Swchartz inequality,

$$\left|T^{-1}\sum_{t=1}^{T}\left(n^{-\rho}s_{nt}\right)\cdot W_{nt}\right| \leq \left(T^{-1}\sum_{t=1}^{T}\left|n^{-\rho}s_{nt}\right|^2\right)^{1/2}\cdot\left(T^{-1}\sum_{t=1}^{T}\left|W_{nt}\right|^2\right)^{1/2} = O_P(1)\cdot O_P(1),$$

since $\|n^{-\rho}s_{nt}\|_r \leq \Delta < \infty$ under Assumption $3'$ and $\|W_{nt}\| \leq \Delta < \infty$ under Assumption 1(ii) that $\|x_{it}\|_{2r} \leq \Delta < \infty$. It follows that $J^0_{sr.2,nT} = O_P\left(\frac{1}{\sqrt{T}}\right) = o_P(1)$ under our assumptions. This concludes the proof that $J^0_{sr,nT} = O_P\left(\frac{n^{1-\rho}}{T^{1/2}}\right)$ under Assumptions 1,$3'$ and 4, or $O_P\left(\frac{1}{T^{1/2}}\right)$ under Assumptions 1, $3'$ and $4''$. To complete the proof that $I_{2nT} = o_P(1)$ we need to study $J^0_{rr,nT} \equiv T^{-1}\sum_{t=1}^T n^{-2\rho} r_{nt} r'_{nt}$. Recalling that $r_{nt} = d_{nt} + c_{nt}$, where $d_{nt} = a_{nt} + b_{nt}$, it follows that

$$\begin{aligned} J^0_{rr,nT} &= T^{-1}\sum_{t=1}^T n^{-2\rho}(d_{nt}+c_{nt})(d_{nt}+c_{nt})' \\ &= T^{-1}\sum_{t=1}^T n^{-2\rho} d_{nt}d'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} d_{nt}c'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} c_{nt}d'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} c_{nt}c'_{nt} \\ &\equiv J^0_{rr.1,nT} + J^0_{rr.2,nT} + J^{0\prime}_{rr.2,nT} + J^0_{rr.3,nT}. \end{aligned}$$

Writing $d_{nt} = a_{nt} + b_{nt}$, it follows that

$$\begin{aligned} J^0_{rr.1,nT} &= T^{-1}\sum_{t=1}^T n^{-2\rho} d_{nt}d'_{nt} = T^{-1}\sum_{t=1}^T n^{-2\rho}(a_{nt}+b_{nt})(a_{nt}+b_{nt})' \\ &= T^{-1}\sum_{t=1}^T n^{-2\rho} a_{nt}a'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} a_{nt}b'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} b_{nt}a'_{nt} + T^{-1}\sum_{t=1}^T n^{-2\rho} b_{nt}b'_{nt}. \end{aligned}$$

Take the first term. It follows that

$$\left|T^{-1}\sum_{t=1}^T n^{-2\rho} a_{nt}a'_{nt}\right| \leq \left(T^{-1}\sum_{t=1}^T \left|n^{-\rho}a_{nt}\right|^2\right)^{2/2} = O_P\left(\frac{n^{2(1-\rho)}}{T}\right) \text{ or } O\left(T^{-1}\right),$$

under Assumption 4 or $4''$ (plus Assumptions 1 and $3'$). Similarly,

$$\begin{aligned} \left|T^{-1}\sum_{t=1}^T n^{-2\rho} a_{nt}b'_{nt}\right| &\leq \left(T^{-1}\sum_{t=1}^T \left|n^{-\rho}a_{nt}\right|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^T \left|n^{-\rho}b_{nt}\right|^2\right)^{1/2} \\ &= O_P\left(\frac{n^{1-\rho}}{T^{1/2}}\right) O_P\left(\frac{n^{1-\rho}}{T^{1/2}}\right) = O_P\left(\frac{n^{2(1-\rho)}}{T}\right), \end{aligned}$$

or $O\left(T^{-1}\right)$ if Assumption $4''$ is used instead of Assumption 4. The last term in $J^0_{rr.1,nT}$ can be analyzed similarly concluding the analyzis of $J^0_{rr.1,nT}$. For $J^0_{rr.2,nT}$, replace $c_{nt} \equiv -\sum_{i=1}^n (x_{it}-\bar{x}_i)^2\left(\hat{\beta}-\beta\right) = -W_{nt}\cdot n\left(\hat{\beta}-\beta\right)$, where W_{nt} is defined as above and we let $p=1$ for simplicity. Then we can show that

$$J^0_{rr.2,nT} = T^{-1}\sum_{t=1}^T n^{-2\rho} d_{nt}c'_{nt} = \underbrace{\left[n^{1-\rho}\left(\hat{\beta}-\beta\right)\right]}_{O_P\left(\frac{1}{\sqrt{T}}\right)} \cdot \underbrace{T^{-1}\sum_{t=1}^T \left(n^{-\rho}d_{nt}\right)\cdot W_{nt}}_{O\left(\frac{n^{1-\rho}}{\sqrt{T}}\right) \text{ or } O\left(\frac{1}{\sqrt{T}}\right)},$$

showing that $J^0_{rr.2,nT} = O_P\left(\frac{n^{1-\rho}}{T}\right)$ or $O\left(\frac{1}{T}\right)$. Using the same arguments, we can show that $J^0_{rr.3,nT} = O_P\left(\frac{1}{T}\right)$ under both sets of assumptions. Thus, it follows that $J^0_{rr,nT} = O_P\left(\frac{n^{1-\rho}}{T}\right)$ or $O\left(\frac{1}{T}\right)$, which concludes the proof that $I_{2nT} = o_P(1)$.

Finally we analyze $I_{3,nT}$. By the triangle inequality,

$$|I_{3,nT}| \leq \sum_{\tau=1}^{T-1} \left| k\left(\frac{\tau}{M}\right)\right| \left(\left|J^{\tau}_{sr,nT}\right| + \left|J^{\tau}_{rs,nT}\right| + \left|J^{\tau}_{rr,nT}\right| \right),$$

where for $\tau = 1, \ldots, T$,

$$J^{\tau}_{sr,nT} = T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} s_{nt} r'_{nt+\tau},\ J^{\tau}_{rs,nT} = T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} r_{nt} s'_{nt+\tau},\ \text{and } J^{\tau}_{rr,nT} = T^{-1} \sum_{t=1}^{T-\tau} n^{-2\rho} r_{nt} r'_{nt+\tau}.$$

Using the same arguments as above, we can show that each of these terms is either $O_P\left(\frac{n^{1-\rho}}{\sqrt{T}}\right)$ under Assumptions 1 and 3', or $O_P\left(T^{-1/2}\right)$ if we impose Assumption $4''$, uniformly in τ. Thus, since $\frac{1}{M}\sum_{\tau=-(T-1)}^{T-1} \left|k\left(\frac{\tau}{M}\right)\right| \to \int_{-\infty}^{+\infty} |k(x)|\, dx < \infty$, by Assumption 5, it follows that $|I_{3nT}| = O_P\left(\frac{Mn^{1-\rho}}{\sqrt{T}}\right)$, or $|I_{3,nT}| = O_P\left(\frac{M}{\sqrt{T}}\right)$, if Assumption $4''$ is added.

Proof of Theorem 2.3. The proof follows by Theorems 2.1 and 2.2.

B Appendix B: proofs of the results in Section 3.

First, we state some auxiliary lemmas and their proofs. Then, we prove the results in Section 3.

Lemma B.1 *Under Assumption 1, if $\ell = o(T)$ as $T \to \infty$,*

a) $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x^*_{it}x^{*\prime}_{it} - x_{it}x'_{it}\right) \overset{P^*}{\to} 0$, *in probability.*

b) $n^{-1}\sum_{i=1}^{n}\left(\bar{x}^*_i - \bar{x}_i\right)\left(\bar{x}^*_i - \bar{x}_i\right)' \overset{P^*}{\to} 0$, *in probability.*

c) $\hat{A}^*_{nT} - \hat{A}_{nT} \overset{P^*}{\to} 0$, *in probability, where* $\hat{A}^*_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x^*_{it} - \bar{x}^*_i\right)\left(x^*_{it} - \bar{x}^*_i\right)'$ *and* $\hat{A}_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)'$.

Lemma B.2 *Suppose Assumption 1 holds. For any $\rho \in [1/2, 1]$ such that Assumptions $3'$ and 4 are verified, if $\ell = o\left(\sqrt{T}\right)$ as $T \to \infty$,*

$$B^{-1/2}_{nT,\rho} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{n^{\rho}} \sum_{i=1}^{n} \left(x^*_{it} - \bar{x}^*_i\right) \varepsilon^*_{it} \overset{d^*}{\to} N\left(0, I_p\right),$$

as $n, T \to \infty$, in probability.

Theorem B.1 *Suppose Assumption 1 holds. For any $\rho \in [1/2, 1]$ such that Assumptions $3''$ and 4 are verified, if $\ell = o\left(\sqrt{T}\right)$ as $T \to \infty$, $\hat{B}^*_{nT,\rho} - B_{nT,\rho} \to^{P^*} 0$, in probability.*

Proof of Lemma B.1. For simplicity and without loss of generality, we consider the scalar case with $p = 1$. Proof of a). Let $w_{it} = x^2_{it}$ and write $\bar{w}_i \equiv T^{-1}\sum_{t=1}^{T} w_{it}$. We want to show that $W^*_{nT} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\bar{w}^*_i - \bar{w}_i\right) = o_{P^*}(1)$, in probability. By repeated application of Markov's inequality, it suffices to show that $E\left|E^*\left|W_{nT}^*\right|\right| = o(1)$ as $n, T \to \infty$. Adding and subtracting appropriately implies that

$$W_{nT}^* = \frac{1}{n}\sum_{i=1}^n \left(\bar{w}_i^* - E^*\left(\bar{w}_i^*\right)\right) + \frac{1}{n}\sum_{i=1}^n \left(E^*\left(\bar{w}_i^*\right) - \bar{w}_i\right) \equiv W_{1nT}^* + W_{2nT}^*.$$

Using the properties of the MBB, we can write

$$\bar{w}_i^* - E^*\left(\bar{w}_i^*\right) = k^{-1}\ell^{-1}\sum_{j=1}^k\sum_{t=1}^\ell \left(w_{i,I_j+t} - E^*\left(\bar{w}_i^*\right)\right) \equiv k^{-1}\ell^{-1}\sum_{j=1}^k A_{i,I_j+t},$$

where $A_{i,j+t} \equiv \sum_{t=1}^\ell \left(w_{i,j+t} - E^*\left(\bar{w}_i^*\right)\right)$ and $I_j \sim$ i.i.d. $\{0, 1, \ldots, T-\ell\}$. It follows that for each i,

$$E^*\left|\bar{w}_i^* - E^*\left(\bar{w}_i^*\right)\right| \leq k^{-1}\ell^{-1}\sum_{j=1}^k E^*\left|A_{i,I_j+t}\right| = \ell^{-1}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|A_{i,j+t}\right|.$$

By the triangle inequality,

$$\left|A_{i,j+t}\right| \leq \left|\sum_{t=1}^\ell \left(w_{i,j+t} - E\left(w_{i,j+t}\right)\right)\right| + \left|\sum_{t=1}^\ell \left(E\left(w_{i,j+t}\right) - E^*\left(\bar{w}_i^*\right)\right)\right|,$$

which implies that

$$\begin{aligned} E^*\left|W_{1nT}^*\right| &= \frac{1}{n}\sum_{i=1}^n E^*\left|\bar{w}_i^* - E^*\left(\bar{w}_i^*\right)\right| \leq \frac{1}{n}\sum_{i=1}^n \ell^{-1}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell \left(w_{i,j+t} - E\left(w_{i,j+t}\right)\right)\right| \\ &\quad + \frac{1}{n}\sum_{i=1}^n \ell^{-1}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell \left(E\left(w_{i,j+t}\right) - E^*\left(\bar{w}_i^*\right)\right)\right| \equiv W_{1.1,nT} + W_{1.2,nT}. \end{aligned}$$

Next we show that $E\left(E^*\left|W_{1nT}^*\right|\right) = o(1)$. We start by showing that $E\left|W_{1.1,nT}\right| = o(1)$. Let $\omega_{it} \equiv w_{it} - E\left(w_{it}\right) = w_{it} - \mu_{w,i}$, where $E\left(w_{it}\right) \equiv \mu_{w,i}$ by time stationarity. Under Assumption 1(ii)., $\left\|\omega_{it}\right\|_r \leq \Delta < \infty$, whereas Assumption 1(iii) implies that $\{\omega_{it}\}$ is a zero mean α-mixing process of size $-\frac{2r}{r-2}$, uniformly in $i = 1, \ldots, n$. It follows that

$$E\left|W_{1.1,nT}\right| \leq \frac{1}{n}\sum_{i=1}^n \ell^{-1}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell} E\left|\sum_{t=1}^\ell \omega_{i,t+j}\right| = O\left(\ell^{-1/2}\right) = o(1),$$

provided $\ell \to \infty$, as we assume. In particular, under Assumption 1, a maximal inequality yields $E\left|\sum_{t=1}^\ell \omega_{i,t+j}\right| \leq \left(E\left|\sum_{t=1}^\ell \omega_{i,t+j}\right|^2\right)^{1/2} = O\left(\sqrt{\ell}\right)$, uniformly in i. Next consider $W_{1.2,nT}$. We have that

$$E\left|W_{1.2,nT}\right| = \frac{1}{n}\sum_{i=1}^n E\left|\mu_{wi} - E^*\left(\bar{w}_i^*\right)\right| = \frac{1}{n}\sum_{i=1}^n E\left(E^*\left|\bar{\omega}_i^*\right|\right),$$

where $\bar{\omega}_i^* \equiv T^{-1}\sum_{t=1}^T \omega_{it}^*$. Since $E^*\left|\bar{\omega}_i^*\right| \leq \left(E^*\left|\bar{\omega}_i^*\right|^2\right)^{1/2}$, it follows by Jensen's inequality that

$$E\left|W_{1.2,nT}\right| \leq \frac{1}{n}\sum_{i=1}^n E\left(E^*\left|\bar{\omega}_i^*\right|^2\right)^{1/2} \leq \frac{1}{n}\sum_{i=1}^n \left(E\left(E^*\left|\bar{\omega}_i^*\right|^2\right)\right)^{1/2}.$$

Under Assumption 1, by Lemma A.1 of Gonçalves and White (2005), we can show that

$$E\left(E^*\left|\bar{\omega}_i^*\right|^2\right)=\frac{1}{T^2}\underbrace{E\left(E^*\left|\sum_{t=1}^T\omega_{it}^*\right|^2\right)}_{O(T)+O(\ell^2)}=O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right),$$

uniformly in i. This implies $E\left|W_{1.2,nT}\right|=O\left(\frac{1}{\sqrt{T}}\right)+O\left(\frac{\ell}{T}\right)$ if $\ell=o(T)$, concluding the proof that $W_{1nT}^*=o_{P^*}(1)$ in probability. To end the proof of a), we consider W_{2nT}^*. We have that

$$\left|W_{2nT}^*\right|\leq\frac{1}{n}\sum_{i=1}^n\left|E^*\left(\bar{\omega}_i^*\right)\right|++\frac{1}{n}\sum_{i=1}^n\left|\bar{\omega}_i\right|.$$

The first term is of order $O_P\left(\frac{1}{\sqrt{T}}\right)+O_P\left(\frac{\ell}{T}\right)$, by the same argument as that used to study $W_{1.2,nT}$. The second term is of order $O_P\left(\frac{1}{\sqrt{T}}\right)$ given that $E\left|\sum_{t=1}^T\omega_{it}\right|^2=O(T)$ by a maximal inequality. To prove b), note that

$$n^{-1}\sum_{i=1}^n\left(\bar{x}_i^*-\bar{x}_i\right)^2=n^{-1}\sum_{i=1}^n\left(\bar{x}_i^*-E^*\left(\bar{x}_i^*\right)+E^*\left(\bar{x}_i^*\right)-\bar{x}_i\right)^2\leq 2\left(J_{1,nT}^*+J_{2,nT}^*\right),$$

where $J_{1,nT}^*\equiv n^{-1}\sum_{i=1}^n\left(\bar{x}_i^*-E^*\left(\bar{x}_i^*\right)\right)^2$, and $J_{2,nT}^*\equiv n^{-1}\sum_{i=1}^n\left(E^*\left(\bar{x}_i^*\right)-\bar{x}_i\right)^2$. Using the same arguments as above, we can show that $E\left|E^*\left|J_{l,nT}^*\right|\right|=o(1)$ as $n,T\to\infty$. Consider first $J_{1,nT}^*$. Note that $\bar{x}_i^*-E^*\left(\bar{x}_i^*\right)=T^{-1}\sum_{t=1}^T\left(x_{it}^*-E^*\left(\bar{x}_i^*\right)\right)=k^{-1}\ell^{-1}\sum_{j=1}^k\sum_{t=1}^\ell\left(x_{i,t+I_j}-E^*\left(\bar{x}_i^*\right)\right)\equiv A_{i,I_j}$, where $I_j\sim$ i.i.d. Uniform on $\{0,1,\ldots,T-\ell\}$, and $A_{i,j}=\sum_{t=1}^\ell\left(x_{i,j+t}-E^*\left(\bar{x}_i^*\right)\right)$. We can write

$$J_{1,nT}^*=n^{-1}\sum_{i=1}^n\left(\bar{x}_i^*-E^*\left(\bar{x}_i^*\right)\right)^2=n^{-1}\sum_{i=1}^n\left(k^{-1}\ell^{-1}\sum_{j=1}^k A_{i,I_j}\right)^2,$$

and it follows that

$$\begin{aligned}E^*\left|J_{1,nT}^*\right| &\leq n^{-1}\sum_{i=1}^n k^{-2}\ell^{-2}E^*\left|\left(\sum_{j=1}^k A_{i,I_j}\right)^2\right|\leq n^{-1}\sum_{i=1}^n k^{-2}\ell^{-2}k\sum_{j=1}^k E^*\left|A_{i,I_j}\right|^2\\ &= n^{-1}\sum_{i=1}^n\ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|A_{i,j}\right|^2=n^{-1}\sum_{i=1}^n\ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell z_{i,t+j}+\ell\left(\mu_i-E^*\left(\bar{x}_i^*\right)\right)\right|^2\\ &\leq Cn^{-1}\sum_{i=1}^n\ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell z_{i,t+j}\right|^2+n^{-1}\sum_{i=1}^n\ell^{-2}\left|\ell\left(\mu_i-E^*\left(\bar{x}_i^*\right)\right)\right|^2\equiv F_1+F_2,\end{aligned}$$

where $z_{i,t+j}\equiv x_{i,t+j}-\mu_i$, and where the first inequality holds by the triangle inequality and the second and third hold by the c_r-inequality. We can show that $E\left|F_1\right|=O\left(\ell^{-1}\right)=o(1)$ if $\ell\to\infty$. Specifically, for each i, Assumption 1 implies that $z_{i,t+j}$ is a zero mean α-mixing process with $\alpha_i(k)\leq\alpha(k)$. Thus, by Lemma A.1, we have that $E\left|\sum_{t=1}^\ell z_{i,t+j}\right|^2\leq K\left(\sum_{k=1}^\infty\alpha(k)^{\frac{1}{2}-\frac{1}{r}}\right)^2\sum_{t=1}^\ell\left\|z_{i,t+j}\right\|_r^2$ for some $r>2$. Assumption 1(ii) implies that $\left\|z_{i,t+j}\right\|_r\leq\Delta<\infty$ whereas Assumption 1(iii) implies that $\sum_{k=1}^\infty\alpha(k)^{\frac{1}{2}-\frac{1}{r}}<\infty$, thus proving that $E\left|\sum_{t=1}^\ell z_{i,t+j}\right|^2\leq C\ell$ for some constant C. Thus,

$E\left|F_1\right| \leq n^{-1}\sum_{i=1}^{n}\ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}E\left|\sum_{t=1}^{\ell}z_{i,t+j}\right|^2 \leq Kn^{-1}\sum_{i=1}^{n}\ell^{-2}\ell = O\left(\ell^{-1}\right)$. Next, we show that $E\left|F_2\right| = O\left(T^{-1}\right) + O\left(\left(\frac{\ell}{T}\right)^2\right)$. Since $\mu_i - E^*\left(\bar{x}_i^*\right) = -T^{-1}\sum_{t=1}^{T}E^*\left(x_{it}^* - \mu_i\right) = -T^{-1}\sum_{t=1}^{T}E^*\left(z_{it}^*\right) = -E^*\left(\bar{z}_i^*\right)$, it follows that

$$F_2 = n^{-1}\sum_{i=1}^{n}\ell^{-2}\left|\ell\left(\mu_i - E^*\left(\bar{x}_i^*\right)\right)\right|^2 = n^{-1}\sum_{i=1}^{n}\ell^{-2}\ell^2\left|\mu_i - E^*\left(\bar{x}_i^*\right)\right|^2 \leq n^{-1}\sum_{i=1}^{n}E^*\left(\left|\bar{z}_i^*\right|^2\right).$$

We can show that $E\left|E^*\left(\left|\bar{z}_i^*\right|^2\right)\right| = O\left(\frac{1}{T}\right) + O\left(\frac{\ell^2}{T^2}\right)$ uniformly in i, which implies that $F_2 = o_{P^*}(1)$ in probability. To prove c), note that we can write

$$\begin{aligned}\hat{A}_{nT}^* - \hat{A}_{nT} &= \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}^{*2} - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}x_{it}^2\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}_i^* - \bar{x}_i\right)^2 - 2\frac{1}{n}\sum_{i=1}^{n}\left(\bar{x}_i^* - \bar{x}_i\right)\bar{x}_i \\ &\equiv a_{1,nT}^* - a_{2,nT}^* - a_{3,nT}^*.\end{aligned}$$

By parts a) and b), $a_{1,nT}^* = o_{P^*}(1)$ and $a_{2,nT}^* = o_{P^*}(1)$, in probability, respectively. To show that $a_{3,nT}^* = o_{P^*}(1)$, in probability, it suffices to show that $E\left|E^*\left|a_{3,nT}^*\right|\right| = o(1)$ as $n, T \to \infty$. By the triangle inequality, $E^*\left|a_{3,nT}^*\right| \leq \frac{1}{n}\sum_{i=1}^{n}\bar{x}_iE^*\left|\bar{x}_i^* - \bar{x}_i\right|$, and therefore
$E\left|E^*\left|a_{3,nT}^*\right|\right| \leq \frac{1}{n}\sum_{i=1}^{n}\left(E\left|\bar{x}_i\right|^2\right)^{1/2}\left(E\left(E^*\left|\bar{x}_i^* - \bar{x}_i\right|\right)^2\right)^{1/2}$. We can show that $E\left|\bar{x}_i\right|^2 \leq \Delta > \infty$ whereas $E\left(E^*\left|\bar{x}_i^* - \bar{x}_i\right|\right)^2 = O\left(\frac{1}{\ell}\right) + O\left(\frac{1}{T}\right) = o(1)$, uniformly in i. This completes the proof.

Proof of Lemma B.2. Let $\varepsilon_{it}^{*0} = y_{it}^* - x_{it}^{*\prime}\beta - \alpha_i$ and note that $\varepsilon_{it}^* = \varepsilon_{it}^{*0} - x_{it}^{*\prime}\left(\hat{\beta} - \beta\right) - \left(\hat{\alpha}_i - \alpha_i\right)$. Similarly, $\hat{\varepsilon}_{it} = \varepsilon_{it} - x_{it}'\left(\hat{\beta} - \beta\right) - \left(\hat{\alpha}_i - \alpha_i\right)$. By the FOC for $\hat{\beta}$, $\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\hat{\varepsilon}_{it} = 0$. Thus, adding an subtracting appropriately, we can write

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\varepsilon_{it}^* = \xi_{1,nT}^* + \alpha_{1,nT} - \alpha_{2,nT}^* + \alpha_{3,nT} - \alpha_{4,nT}^*,$$

where

$$\begin{aligned}\xi_{1,nT}^* &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{t=1}^{n}\left(\left(x_{it}^* - \mu_i\right)\varepsilon_{it}^{*0} - \left(x_{it} - \mu_i\right)\varepsilon_{it}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{s_{nt}^*}{n^{\rho}} - \frac{s_{nt}}{n^{\rho}}\right); \\ \alpha_{1,nT} &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(\bar{x}_i - \mu_i\right)\varepsilon_{it}; \\ \alpha_{2,nT}^* &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(\bar{x}_i^* - \mu_i\right)\varepsilon_{it}^{*0}; \\ \alpha_{3,nT} &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(x_{it} - \bar{x}_i\right)\left(x_{it}'\left(\hat{\beta} - \beta\right) + \left(\hat{\alpha}_i - \alpha_i\right)\right); \text{ and} \\ \alpha_{4,nT}^* &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{n^{\rho}}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\left(x_{it}^{*\prime}\left(\hat{\beta} - \beta\right) + \hat{\alpha}_i - \alpha_i\right).\end{aligned}$$

By Theorem Theorem 2.2 of Gonçalves and White (2002) (see also Gonçalves and de Jong (2003) for weaker moment conditions), we can show that $B_{n,T}^{-1/2}\xi_{1,nT}^* \to^{d^*} N\left(0, I_p\right)$, in probability, provided

$\left\{\frac{s_{nt}}{n^{\rho}}\right\}$ satisfies Assumption 3$''$ and $\ell = o\left(\sqrt{T}\right)$. $\alpha_{1,nT}$ is equal to the bias term $R_{nT,\rho}$ and therefore is $o_P(1)$ under Assumption 4. Next we show that $\alpha^*_{2,nT} = o_{P^*}(1)$ in probability. Writing $\bar{z}^*_i = \bar{x}^*_i - \mu_i \equiv T^{-1}\sum_{t=1}^T z^*_{it}$, we have that $\alpha^*_{2,nT} = \frac{1}{T\sqrt{T}}\frac{1}{n^{\rho}}\sum_{i=1}^n\left(\sum_{t=1}^T z^*_{it}\right)\left(\sum_{t=1}^T \varepsilon^{*0}_{it}\right)$. By repeated application of the Cauchy Schwartz inequality, it follows that

$$\begin{aligned} E\left(E^*\left|\alpha^*_{2,nT}\right|\right) &\leq \frac{1}{T\sqrt{T}n^{\rho}}\sum_{i=1}^n E\left(E^*\left|\left(\sum_{t=1}^T z^*_{it}\right)\left(\sum_{t=1}^T \varepsilon^{*0}_{it}\right)\right|\right) \\ &\leq \frac{1}{T\sqrt{T}n^{\rho}}\sum_{i=1}^n\left\{E\left(E^*\left|\sum_{t=1}^T z^*_{it}\right|^2\right)E\left(E^*\left|\sum_{t=1}^T \varepsilon^{*0}_{it}\right|^2\right)\right\}^{1/2}. \end{aligned}$$

Using Assumption 1, we can show that uniformly in i, $E\left(E^*\left|\sum_{t=1}^T z^*_{it}\right|^2\right) = O(T) + O\left(\ell^2\right)$ and similarly for $E\left(E^*\left|\sum_{t=1}^T \varepsilon^{*0}_{it}\right|^2\right)$. This implies that the term in curly brackets is $O(T) + O\left(\ell^2\right) + O\left(\sqrt{T}\ell\right)$. Thus, $E\left(E^*\left|\alpha^*_{2,nT}\right|\right) = O\left(\frac{n^{1-\rho}}{\sqrt{T}}\right) + O\left(\frac{n^{1-\rho}}{\sqrt{T}}\frac{\ell^2}{T}\right) + O\left(\frac{n^{1-\rho}}{\sqrt{T}}\frac{\ell}{\sqrt{T}}\right) = o(1)$ under the assumptions that $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$ and $\frac{\ell}{\sqrt{T}} \to 0$. Finally, we can show that $\alpha_{3,nT} - \alpha^*_{4,nT} = o_{P^*}(1)$. We can write $\alpha_{3,nT} - \alpha^*_{4,nT} = \psi^*_{1,nT} + \psi^*_{2,nT}$, where $\psi^*_{2,nT} = 0$ and $\psi^*_{1,nT} \equiv \left(\hat{A}_{nT} - \hat{A}^*_{nT}\right)\sqrt{T}n^{1-\rho}\left(\hat{\beta} - \beta\right) = o_{P^*}(1) \times O_P(1) = o_{P^*}(1)$, in probability, given Lemma B.1.c) and the fact that $\sqrt{T}n^{1-\rho}\left(\hat{\beta} - \beta\right) = O_P(1)$ by Theorem 2.1, which completes the proof.

Proof of Theorem B.1. Take $p = 1$. We follow the proof of Gonçalves and White (2004), adapting it to the fixed effects estimator context. For any $j = 1, \ldots, k$ and $t = 1, \ldots, \ell$, let $\hat{s}^*_{n,(j-1)\ell+t} = \sum_{i=1}^n\left(x_{i,I_j+t} - \bar{x}^*_i\right)\tilde{\varepsilon}_{i,I_j+t}$, where $\tilde{\varepsilon}_{it} = y_{it} - \bar{x}'_i\hat{\beta}^* - \hat{\alpha}^*_i$, with $\hat{\alpha}^*_i = \bar{y}^*_i - \bar{x}^{*\prime}_i\hat{\beta}^*$, and where I_j are i.i.d Uniform on $\{0, \ldots, T-\ell\}$. Similarly, let $s^*_{n,(j-1)\ell+t} = \sum_{i=1}^n\left(x_{i,I_j+t} - \mu_i\right)\varepsilon_{i,I_j+t}$, where $\varepsilon_{it} = y_{it} - x'_{it}\beta - \alpha_i$. Consider

$$B^{*0}_{nT,\rho} = \frac{1}{k}\sum_{j=1}^k\left(\ell^{-1/2}\sum_{t=1}^{\ell} n^{-\rho}\left(s^*_{n,(j-1)\ell+t} - \bar{s}^*_{nT}\right)\right)^2 = \frac{1}{k}\sum_{j=1}^k\left(\ell^{-1/2}\sum_{t=1}^{\ell} n^{-\rho}s^*_{n,(j-1)\ell+t}\right)^2 - \ell n^{-2\rho}\bar{s}^{*2}_{nT},$$

where $\bar{s}^*_{nT} = T^{-1}\sum_{t=1}^T s^*_{nt}$. We can apply Lemma B.1 of Gonçalves and White (2004) to show that $B^{*0}_{nT,\rho} - B^*_{nT,\rho} = o_{P^*}(1)$ in probability, where $B^*_{nT,\rho} = Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T\frac{s^*_{nt}}{n^{\rho}}\right)$. Since $B^*_{nT,\rho} - B_{nT,\rho} \xrightarrow{P} 0$, it suffices to show that $\hat{B}^*_{nT,\rho} - B^{*0}_{nT,\rho} \xrightarrow{P^*} 0$, in probability. Let $\hat{S}^*_{n,j} \equiv n^{-\rho}\sum_{t=1}^{\ell}\hat{s}^*_{n,(j-1)\ell+t}$ and $S^*_{n,j} \equiv n^{-\rho}\sum_{t=1}^{\ell}s^*_{n,(j-1)\ell+t}$. We have that

$$\hat{B}^*_{nT,\rho} - B^{*0}_{nT,\rho} = \frac{1}{k}\sum_{j=1}^k \ell^{-1}\left(\hat{S}^{*2}_{n,j} - S^{*2}_{n,j}\right) + \ell n^{-2\rho}\bar{s}^{*2}_{nT} \equiv D^*_1 + D^*_2,$$

where $D^*_2 = O_{P^*}\left(\frac{\ell}{T}\right)$ in probability (by an argument similar to that used in Gonçalves and White (2004)). Next we prove that $D^*_1 = o_{P^*}(1)$ in probability. We can write $\hat{s}^*_{nt} = s^*_{nt} + a^*_{nt} + b^*_{nt}$, where

$a^*_{nt} = \sum_{i=1}^n (\mu_i - \bar{x}^*_i)\, \varepsilon^{*0}_{it}$ and

$$b^*_{nt} = -\sum_{i=1}^n (x^*_{it} - \bar{x}^*_i)\, x^*_{it} \left(\hat{\beta}^* - \beta\right) - \sum_{i=1}^n (x^*_{it} - \bar{x}^*_i)(\hat{\alpha}^*_i - \alpha_i) \equiv b^*_{1n,t} + b^*_{2n,t}.$$

It follows that

$$\hat{S}^*_{n,j} = n^{-\rho} \sum_{t=1}^{\ell} s^*_{n,(j-1)\ell+t} + n^{-\rho} \sum_{t=1}^{\ell} a^*_{n,(j-1)\ell+t} + n^{-\rho} \sum_{t=1}^{\ell} b^*_{n,(j-1)\ell+t} \equiv S^*_{n,j} + R^*_{1n,j} + R^*_{2n,j}.$$

and

$$\begin{aligned} |D^*_1| &\leq \frac{2}{k} \sum_{j=1}^k \ell^{-1} \left(\left|R^{*2}_{1,nj}\right| + \left|R^{*2}_{2,nj}\right| + \left|S^*_{n,j} R^*_{1n,j}\right| + \left|S^*_{n,j} R^*_{2n,j}\right| \right) \\ &\leq \frac{2}{k} \sum_{j=1}^k \ell^{-1} \left|R^{*2}_{1,nj}\right| + \frac{2}{k} \sum_{j=1}^k \ell^{-1} \left|R^{*2}_{2,nj}\right| + 2\frac{1}{k} \sum_{j=1}^k \ell^{-1} \left|S^*_{n,j} R^*_{1n,j}\right| + 2\frac{1}{k} \sum_{j=1}^k \ell^{-1} \left|S^*_{n,j} R^*_{2n,j}\right| \\ &\equiv A^* + B^* + C^* + D^*. \end{aligned}$$

We show that each of these terms vanishes in probability. We first prove that $E\left(E^* |A^*|\right) \to 0$. We have that $E\left(E^* |A^*|\right) \leq \frac{2}{k} \sum_{j=1}^k \ell^{-1} E\left(E^* \left|R^{*2}_{1,nj}\right|\right)$. But

$$\begin{aligned} E^* \left|R^{*2}_{1,nj}\right| &= E^* \left| \sum_{t=1}^{\ell} n^{-\rho} \sum_{i=1}^n (\mu_i - \bar{x}^*_i)\, \varepsilon^{*0}_{i,(j-1)\ell+1} \right|^2 = E^* \left| n^{-\rho} \sum_{i=1}^n (\mu_i - \bar{x}^*_i) \sum_{t=1}^{\ell} \varepsilon_{i,I_j+t} \right|^2 \\ &\leq n^{-2\rho} n \sum_{i=1}^n \left(E^* \left((\mu_i - \bar{x}^*_i)^4 \right) \right)^{1/2} \left(E^* \left| \sum_{t=1}^{\ell} \varepsilon_{i,I_j+t} \right|^4 \right)^{1/2}, \end{aligned}$$

implying that

$$E\left(E^* |A^*|\right) \leq 2\ell^{-1} n^{1-2\rho} \sum_{i=1}^n \left[E\left(E^* \left(\bar{z}^{*4}_i\right)\right) \right]^{1/2} \left[E\left(E^* \left| \sum_{t=1}^{\ell} \varepsilon_{i,I_1+t} \right|^4 \right) \right]^{1/2},$$

where $\bar{z}^*_i \equiv \bar{x}^*_i - \mu_i$. By an application of Lemma A.1 of Gonçalves and White (2005), we can show that $E\left(E^*\left(\bar{z}^{*4}_i\right)\right) = O\left(\frac{1}{T^2}\right) + O\left(\frac{\ell^4}{T^4}\right)$ uniformly in i (for this, it suffices that $\|z_{it}\|_{2r} \leq \infty$ and $\{z_{it}\}$ is $\alpha-$mixing of size $-\frac{4r}{r-2}$, for some $r > 2$), whereas

$$E\left(E^* \left| \sum_{t=1}^{\ell} \varepsilon_{i,I_1+t} \right|^4 \right) = \frac{1}{T - \ell + 1} \sum_{j=0}^{T-\ell} E \left| \left| \sum_{t=1}^{\ell} \varepsilon_{i,j+t} \right|^4 \right| = O\left(\ell^2\right),$$

also uniformly in i. Thus $E\left(E^* |A^*|\right) = O\left(\ell^{-1}\right) O\left(n^{2(1-\rho)}\right) \left(O\left(\frac{\ell}{T}\right) + O\left(\frac{\ell^3}{T^2}\right) \right) = O\left(\frac{n^{2(1-\rho)}}{T}\right) + O\left(\frac{n^{2(1-\rho)}}{T} \frac{\ell^2}{T}\right) = o(1)$, provided $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$ and $\frac{\ell^2}{T} \to 0$. Next we show that $E\left(E^* |C^*|\right) = o(1)$. By the Cauchy-Schwartz inequality

$$C^* = \frac{2}{k} \sum_{j=1}^k \ell^{-1} \left|S^*_{n,j} R^*_{1n,j}\right| \leq 2 \left(\frac{1}{k} \sum_{j=1}^k \ell^{-1} \left|S^*_{n,j}\right|^2 \right)^{1/2} \left(\frac{1}{k} \sum_{j=1}^k \ell^{-1} \left|R^*_{1n,j}\right|^2 \right)^{1/2},$$

implying that

$$
\begin{aligned}
E\left(E^{*}\left(\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^{*}R_{1n,j}^{*}\right|\right)\right) &\leq \left(\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}E\left(E^{*}\left|S_{n,j}^{*}\right|^{2}\right)\right)^{1/2}\left(\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}E\left(E^{*}\left|R_{1n,j}^{*}\right|^{2}\right)\right)^{1/2} \\
&= O\left(\frac{n^{2(1-\rho)}}{T}\right)+O\left(\frac{n^{2(1-\rho)}}{T}\frac{\ell^{2}}{T}\right)=o\left(1\right),
\end{aligned}
$$

where we have used the previous result and the fact that we can show that $\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}E\left(E^{*}\left|S_{n,j}^{*}\right|^{2}\right)=O\left(1\right)$ (this relies on an application of a maximal inequality to the array $\left\{\frac{s_{nt}}{n^{\rho}}\right\}$). Next, consider B^{*}:

$$
\begin{aligned}
B^{*} &= \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|R_{2,nj}^{*2}\right|=\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-\rho}b_{n,(j-1)\ell+t}^{*}\right|^{2} \\
&\leq \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-\rho}b_{1n,(j-1)\ell+t}^{*}\right|^{2}+\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-\rho}b_{2n,(j-1)\ell+t}^{*}\right|^{2}\equiv B_{1}^{*}+B_{2}^{*}.
\end{aligned}
$$

We have that

$$
B_{1}^{*}=\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)x_{i,(j-1)\ell+t}^{*}\right|^{2}\left|n^{1-\rho}\left(\hat{\beta}^{*}-\beta\right)\right|^{2}\equiv\frac{1}{T}\Psi^{*}.\left|\sqrt{T}n^{1-\rho}\left(\hat{\beta}^{*}-\beta\right)\right|^{2}.
$$

Because $\sqrt{T}n^{1-\rho}\left(\hat{\beta}^{*}-\beta\right)=O_{P^{*}}\left(1\right)$, it suffices that $\frac{1}{T}\Psi^{*}=o_{P^{*}}\left(1\right)$, in probability. For some constant K,

$$
\frac{1}{T}E\left(E^{*}\left|\Psi^{*}\right|\right)\leq\frac{K}{T}\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\ell\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}E\left(E^{*}\left|\left(x_{i,I_{j}+t}-\bar{x}_{i}^{*}\right)x_{i,I_{j}+t}\right|^{2}\right)=O\left(\frac{\ell}{T}\right)=o\left(1\right),
$$

if $\ell=o\left(T\right)$. This shows that $B_{1}^{*}=o_{P^{*}}\left(1\right)$, in probability. Next consider B_{2}^{*}. Let $\bar{\varepsilon}_{i}^{*0}\equiv T^{-1}\sum_{t=1}^{T}\varepsilon_{it}^{*0}$ and note that $\hat{\alpha}_{i}^{*}-\alpha_{i}=\bar{\varepsilon}_{i}^{*0}-\bar{x}_{i}^{*\prime}\left(\hat{\beta}^{*}-\beta\right)$. It follows that

$$
\begin{aligned}
b_{2n,(j-1)\ell+1}^{*} &= -n^{-\rho}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)\left(\hat{\alpha}_{i}^{*}-\alpha_{i}\right) \\
&= -n^{-\rho}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)\bar{\varepsilon}_{i}^{*0}-n^{-\rho}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)\bar{x}_{i}^{*}\left(\hat{\beta}^{*}-\beta\right),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
B_{2}^{*} &\leq K\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-\rho}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)\bar{\varepsilon}_{i}^{*0}\right|^{2} \\
&\quad +K\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^{*}-\bar{x}_{i}^{*}\right)\bar{x}_{i}^{*}\right|^{2}\left|n^{1-\rho}\left(\hat{\beta}^{*}-\beta\right)\right|^{2} \\
&\equiv M_{1}^{*}+\frac{K}{T}\Psi^{*}\cdot\left|\sqrt{T}n^{1-\rho}\left(\hat{\beta}^{*}-\beta\right)\right|^{2}.
\end{aligned}
$$

The second term is $o_{P^*}(1)$ in probability, as we proved before. For M_1^*, we can argue as for A^* to show that $E\left(E^*\left|M_1^*\right|\right) = O\left(\frac{n^{2(1-\rho)}}{T}\right) + O\left(\frac{\ell^2 n^{2(1-\rho)}}{T^2}\right) = o(1)$ under the assumptions that $\frac{n^{1-\rho}}{\sqrt{T}} \to 0$ and $\frac{\ell^2}{T} \to 0$. Thus $B^* = o_{P^*}(1)$ in probability. Finally, we show that $D^* = o_{P^*}(1)$, in probability. We have that $|D^*|^2 \leq \left[2\left(\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^*\right|^2\right)^{1/2}(B^*)^{1/2}\right]^2 = O_{P^*}(1) \times o_{P^*}(1) = o_{P^*}(1)$, since $\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^*\right|^2 = O_{P^*}(1)$ and $B^* = o_{P^*}(1)$ in probability.

Proof of Theorem 3.1. We can write

$$\begin{aligned}
\sqrt{T}n^{1-\rho}\left(\hat{\beta}^* - \hat{\beta}\right) &= \hat{A}_{nT}^{*-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}n^{-\rho}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\varepsilon_{it}^* = A_{nT}^{-1}B_{nT,\rho}^{1/2}B_{nT}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}n^{-\rho}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\varepsilon_{it}^* \\
&\quad + \left[\left(\hat{A}_{nT}^{-1} - A_{nT}^{-1}\right) + \left(\hat{A}_{nT}^{*-1} - \hat{A}_{nT}^{-1}\right)\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}n^{-\rho}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\varepsilon_{it}^* \\
&\equiv \zeta_{1,nT}^* + \zeta_{2,nT}^*.
\end{aligned}$$

By Lemma B.2, $B_{nT}^{-1/2}A_{nT}\zeta_{1,nT}^* \to^{d^*} N(0, I_p)$ whereas Lemmas A.2.c) and B.1.c), and the fact that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}n^{-\rho}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\varepsilon_{it}^* = O_{P^*}(1)$ imply that $\zeta_{2,nT}^* = o_{P^*}(1)$, in probability.

Proof of Theorem 3.2. The proof follows from Theorems 3.1 and B.1 using standard arguments.

References

[1] Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica, 59, 817-858.

[2] Andrews, D.W.K., 2005. Cross section regression with common shocks, Econometrica, 73, 1551-1585.

[3] Alvarez, J. and M. Arellano, 2003. The time series and cross section asymptotics of dynamic panel data estimators. Econometrica, 71, 1121-1159.

[4] Arellano, M., 1987. Computing robust standard errors for within-groups estimators, Oxford Bulletin of Economics and Statistics, 49, 431-434.

[5] Bai, J., 2009. Panel data models with interactive fixed effects, Econometrica, 77, 1229-1279.

[6] Bester, A., Conley, T. and C. Hansen, 2008. Inference with dependent data using cluster covariance estimators, mimeo, GSB Chicago.

[7] Bester, A., Conley, T., C. Hansen, and T. Vogelsang, 2008. Fixed-b asymptotics for spatially dependent robust nonparametric covariance matrix estimators, mimeo, GSB Chicago.

[8] Billingsley, P., 1995. Probability and Measure. New York: Wiley.

[9] Conley, T.G., 1999. GMM estimation with cross sectional dependence, Journal of Econometrics, 92, 1-45.

[10] Davidson, J., 1994. Stochastic Limit Theory, Oxford: Oxford University Press.

[11] de Jong, R. and J. Davidson, 2000. Consistency of kernel estimators of heteroskedastic and autocorrelated covariance matrices, Econometrica, 68, 235-287.

[12] Driskoll, J.C. and A.C. Kraay, 1998. Consistent covariance matrix estimation with spatially dependent panel data. Review of Economics and Statistics, 80, 549-560.

[13] Gonçalves, S., and R. de Jong, 2003. Consistency of the stationary bootstrap under weak moment conditions, Economics Letters, 81, 273-278.

[14] Gonçalves, S., and T. Vogelsang, 2009. Block bootstrap HAC robust tests: the sophistication of the naive bootstrap. Forthcoming in Econometric Theory.

[15] Gonçalves, S., and H. White, 2002. The bootstrap of the mean for dependent heterogeneous arrays, Econometric Theory, 18, 1367-1384.

[16] Gonçalves, S., and H. White, 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models, Journal of Econometrics, 119, 199-219.

[17] Gonçalves, S., and H. White, 2005. Bootstrap Standard Error Estimates for Linear Regression, Journal of the American Statistical Association,Vol. 100, No. 471, 970-979.

[18] Götze, F., and H.R. Künsch, 1996. Second-order correctness of the blockwise bootstrap for stationary observations, Annals of Statistics, 24, 1914-1933.

[19] Hahn, J. and G. Kuersteiner, 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica, 70, 1639-1657.

[20] Hahn, J. and G. Kuersteiner, 2004. Bias reduction for dynamic nonlinear panel models with fixed effects, manuscript, UC Davis.

[21] Hansen, B.E. 1991. Strong laws for dependent heterogeneous processes. Econometric Theory 7, 213-221.

[22] Hansen, B.E. 1992. Erratum: Strong laws for dependent heterogeneous processes, Econometric Theory 8, 421-422.

[23] Hansen, C.B. 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. Journal of Econometrics, 141, 597-620.

[24] Hounkannounon, B., 2008. Bootstrap for panel regression models with random effects, mimeo, Université de Montréal.

[25] Ibragimov, R. and U. Mueller, 2009. t-statistic Based Correlation and Heterogeneity Robust Inference. Forthcoming in Journal of Business and Economics Statistics.

[26] Kiefer, N.M. and T. J. Vogelsang, 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. Econometric Theory, 21, 1130-1164.

[27] Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations, Annals of Statistics, 17, 1217-1241.

[28] Liu, R.Y., and K. Singh, 1992. Moving blocks jackknife and bootstrap capture weak dependence, in Exploring the Limits of the Bootstrap, ed. by R. LePage and L. Billiard. New York: Wiley.

[29] Moon, R. and M. Weidner, 2009. Likelihood expansion for panel regression models with factors, mimeo, USC.

[30] Vogelsang, 2008. Panel regression inference robust to heteroskedasticity, serial correlation and general forms of spatial correlation, mimeo, Michigan State University.

[31] White, H., 2001. Asymptotic Theory for Econometricians. Academic Press.

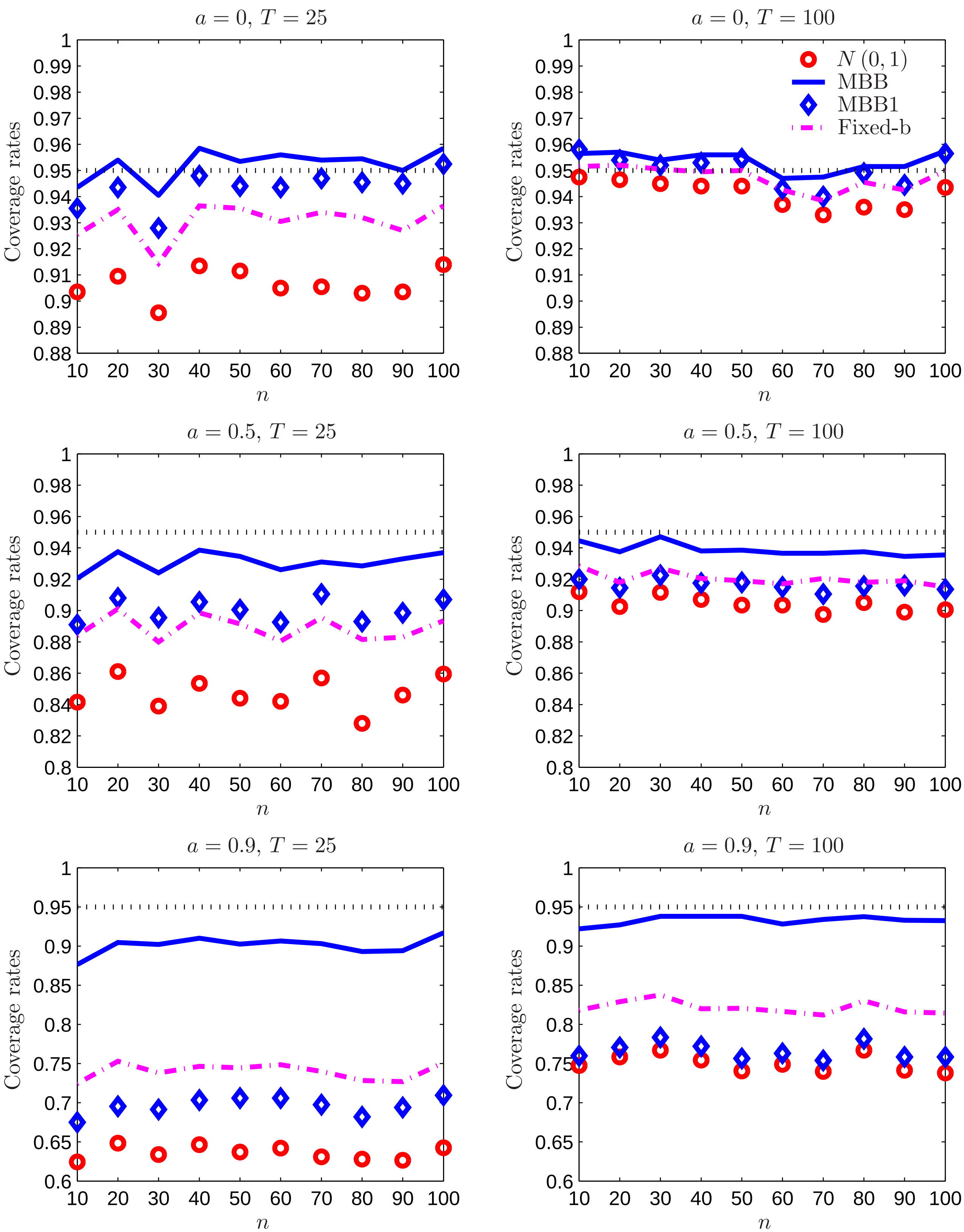

Figure 1: Empirical coverage rates, AR(1) model, $\lambda = \sqrt{0.5}$

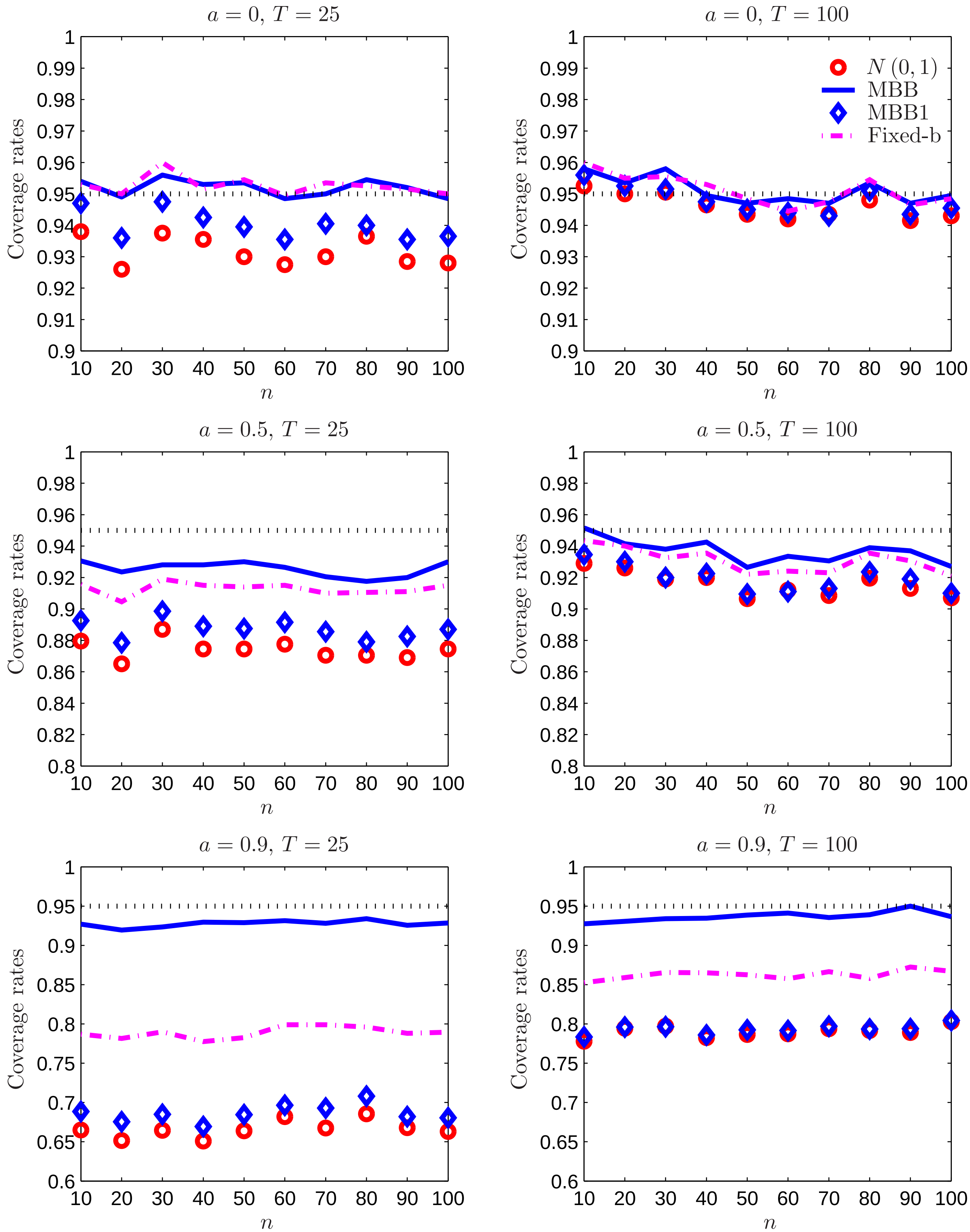

Figure 2: Empirical coverage rates, AR(1) model, $\lambda = 0$