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

Sílvia Gonçalves†
Université de Montréal

December 2008

Abstract

The main contribution of this paper is to show the first order asymptotic validity of the moving blocks bootstrap for fixed effects OLS estimators of panel linear regression models with individual fixed effects. We show that this bootstrap method is robust to serial and cross sectional dependence of unknown form under the assumption that n is an arbitrary nondecreasing function of T, thus allowing for the possibility that both n and T diverge to infinity. 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

A common approach to handle serial dependence in panel data linear regression models is to construct the so-called clustered standard errors, as first proposed by Arellano (1987), and later analyzed by Kezdi (2002), Bertrand, Duflo and Mullainathan (2004), and more recently by Hansen (2007) and Vogelsang (2008). As these papers show, the clustered standard errors are robust to arbitrary levels of autocorrelation, and their finite sample performance is extremely good across a variety of values of n (the cross sectional dimension) and T (the time series dimension). The clustered standard errors essentially eliminate the size distortions that are associated with the standard OLS variance estimators derived under the assumption of serial and cross sectional independence.

A major drawback of the clustered standard errors is that their validity depends on the assumption of cross sectional independence. Nevertheless, many panel data sets are characterized by dependencies among individuals due for instance to the presence of common shocks. An example is a panel data set where countries respond to a common macroeconomic or political shock. Another example is when individual financial assets respond to a common market shock. In such cases, the assumption of cross

*I would like to thank Barbara Rossi for very stimulating discussions during my visit to Duke University in April 2008, which eventually led to the idea of this paper. I would also like to thank Serena Ng for the reference to the paper by Driskoll and Kraay (1998). Finally, I would like to thank Tim Vogelsang and all the participants at the conference "Bootstrap and Numerical Methods in Time Series", Granger Center, University of Nottingham, September 2008 for useful comments on a preliminary version of this paper.

†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

sectional independence is clearly not satisfied, violating the crucial condition under which the clustered standard errors are consistent, resulting in severe size distortions in finite samples.

Driskoll and Kraay (1998) (henceforth D&K) propose standard errors for panel data estimators that are robust to serial and cross sectional correlation of unknown forms. Their approach consists of applying a standard nonparametric heteroskedasticity and autocorrelation consistent (HAC) variance estimator to the cross sectional average of the moment conditions identifying the parameter of interest. The consistency of the D&K standard errors is established under the assumption that T goes to infinity, independently of the behavior of n. In particular, n can either be fixed as T grows to infinity or grow to infinity with T at any arbitrary rate. The simulation results of D&K show that their approach dominates the standard approach based on standard errors that assume independence across individuals and over time when this assumption is not satisfied. Nevertheless, the simulation results also show that there are important finite sample distortions associated with the D&K approach, especially when the degree of serial correlation in the panel is large. Tests based on the D&K standard errors tend to over reject in finite samples and confidence intervals tend to undercover.

Recently, Vogelsang (2008) proposes a new asymptotic theory for test statistics studentized with HAC variance estimators of cross sectional averages 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.

Here 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 the n individual observations at each point in time. Because it does not resample the individual observations directly, the panel MBB is 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.

The main contribution of this paper is to establish the first order asymptotic validity of the panel MBB in the context of a panel linear regression model with individual fixed effects subject to heteroskedasticity, and serial and cross sectional dependence of unknown forms. We first derive the asymptotic distribution of the fixed effects estimator and show the consistency of the D&K standard errors in this context under the assumption that n is an arbitrary nondecreasing function of T. Although quite general, the D&K setup does not cover this case 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). Building on these results, we then prove the consistency of the MBB distribution of the fixed effects estimator sampling distribution. We also show the first order asymptotic validity of MBB Wald tests, which implies the first order asymptotic validity of MBB percentile-t confidence intervals.

We follow D&K and assume that the panel is the realization of a mixing random field where the mixing condition is imposed only in the time dimension, i.e. we require that the dependence between any two observations decreases to zero as the time distance between them increases, without imposing a priori any restriction on the amount of cross sectional dependence. Nevertheless, as it turns out, a crucial condition for our results is that the long run variance of the cross sectional averages is positive for all n, T sufficiently large. This essentially precludes weak dependence in the cross section dimension. For instance, it is not satisfied when individuals are independent. A leading example where it is satisfied is when the cross sectional correlation between any two individuals does not decay to zero as the "distance" between them increases. Under these conditions, we show that the fixed effects estimator is $\sqrt{T}$-convergent (as opposed to $\sqrt{nT}$-convergent) despite the fact that T and n are large. Because individuals are allowed to be arbitrarily dependent, our results only exploit the time series variation, which explains the slower rate of convergence.

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 and cross sectional correlation. In particular, it outperforms the standard normal approximation based on robust D&K standard errors and the fixed-b asymptotic approximation of Vogelsang (2008).

The rest of this paper is organized as follows. In Section 2, we derive the asymptotic distribution of the fixed effects estimator and state the consistency result of an appropriate version of the D&K standard errors. Section 3 contains the bootstrap results for the fixed effects estimator. Section 4 reports the Monte Carlo simulation results and Section 5 concludes. Three mathematical appendices are included. Appendix A contains results for the panel sample mean. In particular, we prove the consistency of the MBB for the sample mean of a panel assumed to be the realization of a mixing random field. These results are auxiliary in proving the results for the MBB of the fixed effects estimator, but are of interest in their own right. Appendix B contains the proofs of the results in Section 2 whereas Appendix C contains the proofs of the results in Section 3.

2 Asymptotic theory for 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 (1). In particular, we suppose that the data are a realization of a random field.

To characterize the dependence structure of the panel over time and across individuals, we follow D&K, and adopt the following definition of a mixing random field. For each t, let $Z_{t,\infty} = \{z_{1t}, z_{2t}, \ldots, z_{nt}, \ldots\}$ denote the set containing the t^{th} observation on all cross sectional units ($i = 1, 2, \ldots, n, \ldots$) for a given random vector z_{it}.

Definition 1 *The random field $\{z_{it}\}$ is α-mixing of size $-a$ if $\alpha(k) = O\left(k^{-\lambda}\right)$ for some $\lambda > a$, where*

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

and $\mathcal{F}_{-\infty}^{t} = \sigma(\ldots, Z_{t-1,\infty}, Z_{t,\infty})$ and $\mathcal{F}_{t+k}^{+\infty} = \sigma(Z_{t+k,\infty}, Z_{t+k+1,\infty}, \ldots)$ denote the sigma fields generated by the corresponding set of random variables.

This definition of mixing does not impose any restriction on the cross sectional dependence, only requiring that for any (i, j) pair, z_{it} and $z_{j,t+k}$ become asymptotically independent as $k \to \infty$. In particular, no mixing condition is imposed on the cross section dimension. The cost of such generality is that we will only be able to obtain $\sqrt{T}$ convergence results (as opposed to $\sqrt{nT}$ convergence). In what follows, we let $\|z_{it}\|_p \equiv (E|z_{it}|^p)^{1/p}$ denote the L_p norm of a random vector, where $|z_{it}|$ denotes its Euclidean norm.

We make the following assumptions.

Assumption 1

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

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

1c. $\{(x'_{it}, \varepsilon_{it}) : i = 1, \ldots, n;\ t = 1, \ldots, T\}$ are the realization of a time stationary mixing random field of size $-\frac{2r}{r-2}$, for some $r > 2$.

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

1e. $\left\{B_{nT} \equiv Var\left(T^{-1/2} \sum_{t=1}^{T} n^{-1} \sum_{i=1}^{n} (x_{it} - E(x_{it}))\varepsilon_{it}\right)\right\}$ is uniformly positive definite, i.e. B_{nT} is positive definite for all n, T and $\det B_{n,T} > \epsilon > 0$ for all n and T sufficiently large.

1f. n is a nondecreasing function of T.

Assumption 1a. requires the regressors to be weakly exogenous, thus allowing for dynamic panel models. Under Assumption 1c. the regressors and the regression errors are the realization of a strong mixing random field as defined above. The crucial part of this assumption is that mixing occurs in the time dimension, without any dependence constraints in the cross section dimension. The time stationarity assumption is mainly imposed to simplify the proofs. It could be relaxed provided extra regularity conditions controlling the degree of time heterogeneity in the data were imposed. Assumption 1e. imposes a restriction on the cross sectional dependence. In particular, it is not satisfied if cross sectional units are weakly dependent, i.e. if the dependence between any two individual observations decreases as the distance between them increases. A leading example where Assumption 1e. is satisfied is when the data generating process of $(x_{it}', \varepsilon_{it})'$ contains a time varying factor that is common to all individuals. In this case, the cross section correlations do not decrease to zero as the distance between individuals increases. A leading example where Assumption 1e. is not satisfied is when observations are cross sectionally independent. In this case, $\det B_{nT} \to 0$ as $n, T \to \infty$ and Assumption 1e. is violated. Weak cross section dependence allows for a faster rate of convergence ($\sqrt{nT}$ as opposed to $\sqrt{T}$, as we derive here), in which case B_{nT} is replaced with $V_{nT} \equiv Var\left(T^{-1/2} \sum_{t=1}^{T} n^{-1/2} \sum_{i=1}^{n} (x_{it} - E(x_{it})) \varepsilon_{it}\right)$. The requirement that V_{nT} be uniformly positive definite is compatible with weak cross sectional dependence whereas it is not when applied to B_{nT}. Assumption 1f. allows two cases: (i) $n \to \infty$ as $T \to \infty$, and (ii) n fixed as $T \to \infty$. We will write $n, T \to \infty$ whenever Assumption 1f. holds.

Our first result is as follows.

Theorem 2.1 *Under Assumption 1, as $n, T \to \infty$,* $B_{nT}^{-1/2} A_{nT} \sqrt{T} \left(\hat{\beta}_{nT} - \beta\right) \to^{d} N(0, I_p)$.

The proof of Theorem 2.1 and of all the results in this Section are in Appendix B. Theorem 2.1 shows that $\hat{\beta}_{nT}$ is asymptotically distributed as a standard normal distribution with an asymptotic covariance matrix given by $C_{nT} \equiv A_{nT}^{-1} B_{nT} A_{nT}^{-1}$. Despite the fact that both n and T are large, the convergence rate of $\hat{\beta}$ is only $\sqrt{T}$ and not $\sqrt{nT}$. The reason for this slower rate of convergence is that we allow for strong cross sectional dependence, which effectively means that most of the variation is coming from the time series dimension and not from the cross sectional dependence. Under our Assumption 1, $\hat{\beta}_{nT}$ is consistent and asymptotically unbiased even for dynamic panel models, where the regressors are not strictly exogenous. This contrasts with the results in Hahn and Kursteiner (2002) (see also Alvarez and Arellano (2003)) where an asymptotic bias in $\hat{\beta}_{nT}$ appears when n and T grow at the same rate. Because Hahn and Kuersteiner (2002) assume cross sectional independence, the convergence rate of $\hat{\beta}_{nT}$ is $\sqrt{nT}$. The asymptotic bias of $\hat{\beta}_{nT}$ disappears as $T \to \infty$, when cross sectional dependence is of the strong type (as assumed in Assumption 1) and only $\sqrt{T}$ convergence is achieved.

To use the normal approximation given in Theorem 2.1, we need a consistent estimator of $C_{nT} \equiv A_{nT}^{-1} B_{nT} A_{nT}^{-1}$. A consistent estimator of A_{nT} under Assumption 1 is $\hat{A}_{nT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'$. See Lemma B.1c) in Appendix B. Next we propose a consistent estimator of B_{nT}. Let $s_{nt} \equiv n^{-1} \sum_{i=1}^{n} (x_{it} - E(x_{it})) \varepsilon_{it}$ denote the cross sectional average of the individual scores for β (after concentrating out α_i). Then

$$B_{nT} = Var\left(T^{-1/2}\sum_{t=1}^{T} s_{nt}\right) = \Gamma_{nT}(0) + \sum_{\tau=1}^{T-1}\left(\Gamma_{nT}(\tau) + \Gamma'_{nT}(\tau)\right),$$

where $\Gamma_{nT}(\tau) \equiv T^{-1}\sum_{t=1}^{T-\tau} E\left(s_{nt}s'_{nt+\tau}\right)$. We propose the following kernel estimator of B_{nT},

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

where $k(\cdot)$ is a kernel function, M is a bandwidth parameter, and for any $\tau \geq 0$, $\hat{\Gamma}_{nT}(\tau) = T^{-1}\sum_{t=1}^{T-\tau} \hat{s}_{nt}\hat{s}'_{nt+\tau}$, with $\hat{s}_{nt} = n^{-1}\sum_{i=1}^{n}(x_{it} - \bar{x}_i)\hat{\varepsilon}_{it}$, and $\hat{\varepsilon}_{it} = y_{it} - \bar{y}_i - (x_{it} - \bar{x}_i)'\hat{\beta}$ the fixed effects OLS residuals of model (1).

$\hat{B}_{nT}$ is a standard HAC estimator of the long run variance of the cross sectional averages $s_{nt} \equiv n^{-1}\sum_{i=1}^{n}(x_{it} - E(x_{it}))\varepsilon_{it}$. Because $E(x_{it})$ and ε_{it} are unknown, we replace these with $\bar{x}_i$ (a consistent estimator of $E(x_{it})$ under our time stationarity assumption) and $\hat{\varepsilon}_{it}$, the fixed effects OLS residuals.

In the context of GMM estimators with panel data, D&K propose 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. Our proposed estimator $\hat{B}_{nT}$ is an extension of the D&K approach to the case of linear panel regression models with individual fixed effects.

To prove the consistency of $\hat{B}_{nT}$ for B_{nT} we strengthen Assumption 1 as follows.

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

1c′. $\{(x'_{it}, \varepsilon_{it}) : i = 1, \ldots, n;\ t = 1, \ldots, T\}$ are the realization of a time stationary mixing random field of size $-\frac{4r}{r-2}$, for some $r > 2$.

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

Assumption 2 $k(\cdot) \in \mathcal{K}$, 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 2 corresponds to Assumption 2 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 3 $M \to \infty$ as $T \to \infty$ and $\frac{M^2}{T} \to 0$.

Theorem 2.2 *Suppose Assumption 1 strengthened by Assumptions 1b′. and 1c′ holds. If Assumptions 2 and 3 also hold,* $\hat{B}_{nT} - B_{nT} \to^{P} 0$ *as* $n, T \to \infty$.

Theorems 2.1 and 2.2 justify using the standard normal approximation for computing critical values of studentized statistics based on $\hat{C}_{nT}$.

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 $\left\{z_{it}^* = (y_{it}^*, x_{it}^{*\prime})' : i=1,\dots,n, t=1,\dots,T\right\}$ obtained with the MBB as follows. Let $Z_{t,n} \equiv (z_{1t}', z_{2t}', \dots, z_{nt}')'$ denote the $n(p+1)\times 1$ vector containing the n cross sectional observations on z_{it}. Let $\ell = \ell_T \in \mathbb{N}$ $(1 \le \ell < T)$ denote the length of the blocks and let $B_{t,\ell} = \{Z_{t,n}, Z_{t+1,n}, \dots, Z_{t+\ell-1,n}\}$ 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 $\{B_{1,\ell}, \dots, B_{T-\ell+1,\ell}\}$. Thus, if we let $I_1, \dots, I_k$ be i.i.d. random variables uniformly distributed on $\{0, \dots, T-\ell\}$, the MBB pseudo-data $\left\{Z_{t,n}^*, t=1,\dots,T\right\}$ is the result of arranging the elements of the k resampled blocks $B_{I_1+1,\ell}, \dots, B_{I_k+1,\ell}$ in a sequence: $Z_{1,n}^* = Z_{I_1+1,n}, Z_{2,n}^* = Z_{I_1+2,n}, \dots, Z_{\ell,n}^* = Z_{I_1+\ell,n}, Z_{\ell+1,n}^* = Z_{I_2+1,n}, \dots, Z_{kl,n}^* = Z_{I_k+\ell,n}$. Thus the 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 $Z_{nT}^* \to^{P^*} 0$, 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 $Z_{nT}^* \to^{d_*} Z$ as $h \to 0$, 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.

We strengthen Assumption 1b$'$. as follows.

1b$''$. For some $r > 2$ and $\delta > 0$, $\|x_{it}\|_{2(r+\delta)} \le \Delta < \infty$ and $\|\varepsilon_{it}\|_{2(r+\delta)} \le \Delta < \infty$ for all $i = 1, \dots, n$ and $t = 1, \dots, T$.

Theorem 3.1 *Suppose Assumption 1 strengthened by Assumptions 1b$''$ and 1c$'$ holds. If $\ell_T \to \infty$ and $\ell_T = o\left(\sqrt{T}\right)$, then for any $\varepsilon > 0$,* $P\left(\sup_{x\in\mathbb{R}^p}\left|P^*\left(\sqrt{T}\left(\hat{\beta}^*-\hat{\beta}\right)\le x\right)-P\left(\sqrt{T}\left(\hat{\beta}-\beta\right)\le x\right)\right|>\varepsilon\right)\to 0.$

The proofs of Theorem 3.1 and of all the results in this section are in Appendix C. Theorem 3.1 justifies using the order statistics of the bootstrap distribution of $\sqrt{T}\left(\hat{\beta}^* - \hat{\beta}\right)$ to approximate the quantiles of the distribution of $\sqrt{T}\left(\hat{\beta} - \beta\right)$. This result is useful for constructing bootstrap percentile confidence intervals for β with asymptotically correct coverage probabilities. Although it does not immediately justify the use of the bootstrap for constructing bootstrap percentile t intervals or testing hypotheses about β based on studentized statistics, it is an important first step in that direction, as we now show.

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. The Wald statistic for testing H_0 is

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

where $\hat{A}_{nT}$ is a consistent estimator of A_{nT} and $\hat{B}_{nT}$ is a consistent estimator of B_{nT}, as we showed in Section 2. Our bootstrap Wald statistic 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

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

Lemma C.1.c) shows that under our assumptions $\hat{A}_{nT}^* - \hat{A}_{nT} \to^{P^*} 0$ in probability, which together with Lemma B.1.c) implies that $\hat{A}_{nT}^* - A_{nT} \to^{P^*} 0$ in probability. To define $\hat{B}_{nT}^*$, let $\hat{s}_{nt}^* = n^{-1}\sum_{i=1}^{n}\left(x_{it}^* - \bar{x}_i^*\right)\hat{\varepsilon}_{it}^*$, 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 $t = 1, \ldots, \ell$, $\hat{s}_{n,(j-1)\ell+t}^* = n^{-1}\sum_{i=1}^{n}\left(x_{i,I_j+t} - \bar{x}_i^*\right)\tilde{\varepsilon}_{i,I_j+t}$, 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,

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

$\hat{B}_{nT}^*$ 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 C.1 in Appendix C proves that $\hat{B}_{nT}^* - B_{nT} \to^{P^*} 0$ in probability. Thus, we can state the following result.

Theorem 3.2 *Suppose Assumption 1 strengthened by Assumptions 1b$''$ and 1c$'$ holds. If $\ell_T \to \infty$ and $\ell_T = o\left(\sqrt{T}\right)$, then under H_0 for any $\varepsilon > 0$, $P\left(\sup_{x\in\mathbb{R}}\left|P^*\left(\mathcal{W}_{nT}^* \leq x\right) - P\left(\mathcal{W}_{nT} \leq x\right)\right| > \varepsilon\right) \to 0$.*

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. By the same arguments, we can show the consistency of a MBB t-statistic studentized with $\hat{C}_{nT}^* \equiv \hat{A}_{nT}^{*-1}\hat{B}_{nT}^*\hat{A}_{nT}^{*-1}$, justifying the construction of bootstrap percentile-t intervals for the elements of β.

4 Monte Carlo results

This section provides simulation evidence of the finite sample performance of the MBB in the context of a panel linear regression model with fixed effects. Specifically, we consider the following model

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

where ε_{it} and $x_{it} = (x_{1,it}, x_{2,it}, x_{3,it})'$ are serially and cross sectionally correlated, and are mutually independent. Without loss of generality, we set $\alpha_i = \beta = 0$. To introduce cross sectional dependence we assume a factor structure for the errors and the regressors. In particular,

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

where

$$\begin{aligned} f_{\varepsilon,t} &= \rho f_{\varepsilon,t} + u_{\varepsilon,t}, \quad u_{\varepsilon,t} \sim N\left(0, 1-\rho^2\right) \\ e_{\varepsilon,it} &= \rho e_{\varepsilon,it} + v_{\varepsilon,it},\ v_{\varepsilon,it} \sim N\left(0, \left(1-\rho^2\right)\left(1-\lambda^2\right)\right), \end{aligned}$$

where the innovations $u_{\varepsilon,t}$ and $v_{\varepsilon,it}$ are mutually independent and uncorrelated over time and across units. We can show that for any τ,

$$E\left(\varepsilon_{it}\varepsilon_{j,t+\tau}\right) = \begin{cases} \rho^\tau & \text{if } i = j, \\ \lambda^2\rho^\tau & \text{if } i \neq j, \end{cases}$$

where we use the convention that $0^0 = 1$. Thus, the error term for each individual is correlated over time according to ρ^τ whereas the error terms of any two individuals (i, j) are equicorrelated according to $\lambda^2\rho^\tau$. A similar factor structure is assumed for each regressor, i.e. for $l = 1, 2, 3$, and independently of each other, we let

$$\begin{aligned} x_{l,it} &= \lambda f_{l,t} + e_{l,it}, \\ f_{l,t} &= \rho f_{l,t-1} + u_{l,t},\ u_{l,t} \sim N\left(0, 1-\rho^2\right) \\ e_{l,it} &= \rho e_{l,it} + v_{l,it},\ v_{l,it} \sim N\left(0, \left(1-\rho^2\right)\left(1-\lambda^2\right)\right), \end{aligned}$$

with $u_{l,t}$ and $v_{l,it}$ mutually independent and uncorrelated over time and across units. As previously, this implies a cross sectional contemporaneous correlation equal to λ^2 and a lagged cross sectional correlation equal to $\lambda^2\rho^\tau$.

We can easily verify that this model satisfies our regularity conditions provided $\lambda \neq 0$ and $|\rho| < 1$. In particular, the moment conditions on ε_{it} and x_{it} are automatically satisfied given that the innovations driving the factor processes and the idiosyncratic errors have normal distributions. Because each factor and idiosyncratic error are generated by stationary AR(1) models, their sum forms an α-mixing random field with exponentially decaying α coefficients. Finally, we can show that $A_{nT} = I_3$, where

I_3 is the 3×3 identity matrix , which is nonsingular, and

$$\lim_{n,T\to\infty} B_{nT} = \lambda^4 \left(1 + 2\sum_{\tau=1}^{\infty} \rho^{2\tau}\right) I_3,$$

where $\lambda^4 \left(1 + 2\sum_{\tau=1}^{\infty} \rho^{2\tau}\right) > 0$ provided $|\rho| < 1$ and $\lambda \neq 0$. Although $\lambda = 0$ is not covered by the regularity conditions in this paper, we also consider this case in the simulations. Specifically, 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 $\lambda^4 = 0.25$).

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 $\hat{\beta}_{nT}$, the fixed-effects estimator of β, 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 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 $\hat{s}_{nt} \equiv n^{-1}\sum_{i=1}^{n}\hat{s}_{it}$. We also ran results with the QS kernel. To conserve space and because these results follow the same patterns as those for the Bartlett kernel, we only present results for the Bartlett kernel.

We consider intervals based on the normal approximation (AT), on the new fixed-b asymptotic theory of Vogelsang (2008) (Fixed-b), and on the bootstrap. The AT 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 $b = \frac{M}{T}$ with 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. An alternative approach 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 (2008) 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. We also consider the naive bootstrap approach in this study. It is implemented with the Bartlett kernel and a data-driven bandwidth chosen by Andrews'

(1991) procedure applied to the bootstrap scores $\hat{s}_{nt}^* \equiv n^{-1}\sum_{i=1}^n \hat{s}_{it}^*$.

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 $\ell = 1$.

The results are in Figures 1-6. Each figure contains results for a particular (λ, ρ) combination, where $\lambda \in \{0, \sqrt{0.5}\}$ and $\rho \in \{0, 0.5, 0.9\}$. We find three panels in each figure, corresponding to three different values of $T \in \{25, 50, 100\}$. Each panel depicts the actual coverage rates of each interval as a function of $n \in \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. 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 six types of intervals: the asymptotic theory (AT) intervals based on the normal approximation, the fixed-b (Fixed-b) intervals based on the Vogelsang (2008) approach, the MBB intervals based on $\hat{B}_{nT}^*$ given in Section 3, implemented with a data-driven block size (MBB) and a block size equal to 1 (MBB1), and the naive MBB intervals implemented either with a data-driven block size (N-MBB) or a block size of 1 (N-MBB1).

Figures 1-3 consider the case when there is cross sectional dependence, i.e. $\lambda = \sqrt{0.5}$, and $\rho \in \{0, 0.5, 0.9\}$, respectively. Figure 1 shows that when there is no serial correlation ($\rho = 0$) but individuals are cross sectionally correlated, some finite sample distortions arise for $T = 25$, especially for the AT intervals (whose rates are in the range 89.5%-91.5%). The Fixed-b intervals outperform the AT intervals by a small margin, with rates between 91.5%-93.5% for $T = 25$. Among all bootstrap methods, the best is the MBB, which uses the Götze and Künsch (1996) variance estimator and a data-driven block size (the selected ℓ was on average 1.60 across all values of n and T). The N-MBB with $\ell = 1$ has rates that are very close to the MBB intervals and it dominates MBB1. For $T = 100$, the differences between all methods disappears and they all perform very well. When we increase ρ to 0.5, Figure 2 shows that the performance of all methods deteriorates, but this is more pronounced for the AT intervals (with rates around 85% when $T = 25$). The Fixed-b intervals outperform the AT intervals, displaying rates between 88% and 90% when $T = 25$, followed by the MBB1 and the N-MBB1. The N-MBB1 tends to dominate the MBB1, but both are worse than the MBB and the N-MBB implemented with a data driven ℓ (with an average value of 2.00 across all values of n when $T = 25$, of 2.7 when $T = 50$, and of 3.5 when $T = 100$). The MBB tends to slightly dominate N-MBB when $T = 25$, but the differences disappear for $T = 50$ and $T = 100$. Figure 3 shows that when $\rho = 0.9$ and $\lambda = \sqrt{0.5}$, the degree of undercoverage increases significantly for all methods. Of all methods, the AT intervals are the most distorted, with coverage rates between 62% and 65% for $T = 25$ (these rates increase to about 68% for $T = 50$ and to 75% for $T = 100$, across all values of n). Next come the block bootstrap methods with $\ell = 1$, with the N-MBB1 slightly better than the MBB1. The Fixed-b intervals outperform the i.i.d. bootstrap methods and the AT intervals, for all values of T and n. In particular, the Fixed-b intervals dominate the N-MBB1. This finding appears to be at odds with the evidence in Gonçalves and Vogelsang (2008). Nevertheless, the difference of results can be explained by the fact that here we implement the fixed-b intervals and the naive i.i.d. bootstrap with an automatic bandwidth whereas Gonçalves and Vogelsang (2008) compare the two methods across a range of fixed bandwidths. In particular, the bandwidth used to studentize the fixed-effects estimator $\hat{\beta}_{1,nT}$ is not necessarily the same as the bandwidth used to studentize $\hat{\beta}^*_{1,nT}$. Instead, the same (fixed) bandwidth is used for the two methods in Gonçalves and Vogelsang (2008). This could explain the difference of results. Overall, the best method is MBB, followed by the N-MBB (both use a data-driven block size equal on average to 4.4 when $T = 25$, 7.9 when $T = 50$, 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%. The difference between the MBB and the N-MBB intervals tends to decrease when T increases.

Figures 4-6 contain the results for $\lambda = 0$ and $\rho \in \{0, 0.5, 0.9\}$, respectively. A comparison between these figures and Figures 1-3 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 ρ, T, and n.

Although our regularity conditions effectively rule out the cross sectional independent case, Figures 4-6 suggest that the D&K approach and the MBB approach proposed here continue to be valid in this case. 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 $\sqrt{T}$-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). We conjecture that a similar result holds in our context. In particular, the same test statistics based on the D&K standard errors can be computed and are asymptotically valid in the two cases (mixing and no-mixing in the cross sectional dimension), a feature also shared by the panel MBB. By exploiting the cross sectional mixing conditions, $\hat{\beta}_{nT}$ is $\sqrt{nT}$-convergent, and not $\sqrt{T}$-convergent as we proved in Section 2, with an asymptotic variance equal to $B^0_{nT} \equiv Var\left(\sqrt{nT}\hat{\beta}_{nT}\right) = nB_{nT}$. The analogue of Assumption 1e. in this case requires $\left\{B^0_{nT}\right\}$ to be uniformly positive definite, which is a weaker requirement than Assumption 1e. For our specific DGP, when $\lambda = 0$, $B^0_{nT} = \left(1 + 2\sum_{\tau=1}^{\infty} \rho^{2\tau}\right) I_3$, which is positive definite provided $|\rho| < 1$. The appropriate t-statistic is $t^0_{\hat{\beta}_1} = \frac{\sqrt{nT}\left(\hat{\beta}_{1,nT} - \beta_1\right)}{\sqrt{\hat{B}^0_{nT}}}$, where $\hat{B}^0_{nT} = n\hat{B}_{nT}$, which implies that $t^0_{\hat{\beta}_1} = t_{\hat{\beta}_1}$. Thus, although the fixed effects estimator has a different rate of convergence according to degree of cross sectional dependence, the same t-statistics and Wald statistics can be used. Similarly, the same bootstrap statistics apply and are asymptotically valid independently of the degree of cross sectional dependence in the panel. Providing a set of primitive conditions under which the theoretical results derived here cover the case of weak cross sectional dependence is an important extension of our results, which we will consider elsewhere.

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 (thus allowing for the possibility that both n and T diverge to infinity). 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.

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 D&K 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)). 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: The panel sample mean

In this Appendix we study the panel sample mean of $\{z_{it} : i = 1, \ldots, n, t = 1, \ldots, T\}$, the realization of a random field defined on a given probability space $(\Omega, \mathcal{F}, P)$. The Appendix is divided in three parts. The first part contains some auxiliary results that will be used throughout the proofs. The second part contains the asymptotic theory results for the sample mean. The third part contains the bootstrap results for the sample mean.

Auxiliary results

The first auxiliary result is 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*

$\alpha(k) \equiv \sup_t \sup_{\left\{A\in\mathcal{G}_{-\infty}^{nt}, B\in\mathcal{G}_{t+k}^{n,+\infty}\right\}} |P(A\cap B) - P(A)P(B)|$, *where* $\mathcal{G}_{-\infty}^{nt} = \sigma(\ldots, \mathcal{X}_{nt})$ *and* $\mathcal{G}_{t+k}^{n,+\infty} = \sigma(\mathcal{X}_{n,t+k}, \ldots)$. *Then for any* $1 \leq p < r$,

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

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

Lemma A.2 *Suppose* $\{z_{it} : i = 1, \ldots, n, \ldots; t = 1, \ldots, T, \ldots\}$ *is an* α*-mixing random field of size* $-a$, *as defined in Definition 1. Let* n *be a nondecreasing function of* T. *Then* $\bar{z}_{t,n} \equiv n^{-1}\sum_{i=1}^{n} z_{it}$ *is a double mixing array of size* $-a$ *with mixing coefficients bounded above by those of* $\{z_{it}\}$.

For each $i = 1, \ldots, n$, let $\mathcal{G}_{-\infty}^{i,t} = \sigma(\ldots, z_{i,t-1}, z_{it})$ and $\mathcal{G}_{t+k}^{i,+\infty} = \sigma(z_{i,t+k}, z_{i,t+k+1}, \ldots)$ denote the σ-fields generated by the corresponding set of random variables. For each individual i, we can define the individual mixing coefficients as $\alpha_i(k) \equiv \sup_t \sup_{\left\{A\in\mathcal{G}_{-\infty}^{i,t}, B\in\mathcal{G}_{t+k}^{i,+\infty}\right\}} |P(A\cap B) - P(A)P(B)|$.

Lemma A.3 *Suppose* $\{z_{it} : i = 1, \ldots, n, \ldots; t = 1, \ldots, T, \ldots\}$ *is an* α*-mixing random field of size* $-a$, *as defined in Definition 1. Then, for each* i, *the stochastic process given by* $\{z_{it} : t = 1, \ldots, T, \ldots\}$ *is* $\alpha-$*mixing of size* $-a$ *with* $\sup_i \alpha_i(k) \leq \alpha(k)$ *for all* k.

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 mixing 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. See D&K, 1998, proof of Result 1.

Proof of Lemma A.3. The results follows because $\mathcal{G}_{-\infty}^{i,t} \subseteq \mathcal{F}_{-\infty}^{t}$ and $\mathcal{G}_{t+k}^{i,+\infty} \subseteq \mathcal{F}_{t+k}^{+\infty}$, which implies that $\alpha_i(k) \leq \alpha(k)$ for all i and k.

Asymptotic theory for the panel sample mean

Let z_{it} de a $p \times 1$ vector and let $\mu_{it} \equiv E(z_{it})$ for $i = 1, \ldots, n$, $t = 1, \ldots, T$. The parameter of interest is the time average of the cross sectional average of individual means $\overline{\overline{\mu}}_{nT} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mu_{it}$. For simplicity, we will assume that there is no time mean heterogeneity, i.e. $\mu_{it} = \mu_i$ for all $t = 1, 2, \ldots, T$, $i = 1, 2, \ldots, n$, in which case $\overline{\overline{\mu}}_{nT} = n^{-1}\sum_{i=1}^{n}\mu_i$. If in addition there is no individual mean heterogeneity, $\mu_{it} = \mu$ for all (i, t), and $\overline{\overline{\mu}}_{nT} = \mu$. We estimate $\overline{\overline{\mu}}_{n,T}$ with the panel sample mean, $\overline{\overline{z}}_{nT} \equiv \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} z_{it}$. We make the following assumptions.

Assumption A.

A1. For some $r > 2$, $\|z_{it}\|_r \leq \Delta < \infty$ for all (i, t).

A2. $\{z_{it} : i = 1, \ldots, n;\ t = 1, \ldots, T\}$ is the realization of an α-mixing random field of size $-\frac{r}{r-2}$ for some $r > 2$, as defined in Definition 1.

A3. $\Sigma_{nT} \equiv Var\left(T^{-1/2}\sum_{t=1}^{T} n^{-1}\sum_{i=1}^{n} z_{it}\right)$ is positive definite uniformly in n and T, i.e. $\Sigma_{n,T}$ is definite positive for each (n,T) and $\det(\Sigma_{n,T}) \geq \kappa > 0$ for all n,T sufficiently large.

A4. n is a nondecreasing function of T.

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

a) $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - E(z_{it})) \to^{P} 0$.

b) $\Sigma_{nT}^{-1/2}\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - E(z_{it})) \to^{d} N(0, I_d)$.

Next we provide a consistent estimator of Σ_{nT}. Let $\Gamma_{nT}(\tau) \equiv T^{-1}\sum_{t=1}^{T-\tau} E\left(\bar{z}_{nt}\bar{z}'_{nt+\tau}\right)$ for any $\tau \geq 0$. We can write

$$\Sigma_{nT} \equiv Var\left(T^{-1/2}\sum_{t=1}^{T}\bar{z}_{nt}\right) = \Gamma_{nT}(0) + \sum_{\tau=1}^{T-1}\left(\Gamma_{nT}(\tau) + \Gamma'_{nT}(\tau)\right).$$

The HAC estimator of Σ_{nT} is given by

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

where $\hat{\Gamma}_{nT}(\tau) = T^{-1}\sum_{t=1}^{T-\tau}\bar{z}_{nt}\bar{z}'_{nt+\tau}$ for any $\tau \geq 0$, and M is the bandwidth parameter. The following result shows that $\hat{\Sigma}_{nT}$ is a consistent estimator for Σ_{nT} provided the following additional assumptions hold.

Assumption C $M \to \infty$ as $T \to \infty$ and $\frac{M}{T} \to 0$.

Theorem A.2 *Under Assumption A, Assumption 2 in the main text, and Assumption C, $\hat{\Sigma}_{nT} - \Sigma_{nT} \to^{P} 0$ as $n,T \to \infty$.*

Proof of Theorem A.1. a) follows automatically given b) and given that $\Sigma_{n,T}^{-1/2} = O(1)$. To prove b), note that we can write $\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}(z_{it} - E(z_{it})) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ddot{z}_{nt}$, where $\ddot{z}_{nt} \equiv n^{-1}\sum_{t=1}^{n}(z_{it} - E(z_{it})) \equiv \bar{z}_{nt} - \bar{\mu}_n$. By Lemma A.2, $\{\ddot{z}_{nt}\}$ is a zero mixing double array of the same size as $\{z_{it}\}$. Hence, the result follows by Theorem 5.20 of White (2001) under Assumption A.

Proof of Theorem A.2. The proof follows immediately from Theorem 2.1 of de Jong and Davidson (2000) upon noting that $\{\ddot{z}_{nt} = \bar{z}_{nt} - \bar{\mu}_n\}$ is an L_r-integrable zero mean mixing array of size $-\frac{r}{r-2}$ for some $r > 2$ and therefore satisfies Assumption 1 of that Theorem. Assumption B and C correspond to their Assumptions 2 and 3, respectively.

Bootstrap results for the sample mean

Given a bootstrap resample $\{z_{it}^*\}$ obtained with the PMBB, we can compute the resampled version of the panel sample mean as $\bar{\bar{z}}_{nT}^* = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} z_{it}^* = \frac{1}{T}\sum_{t=1}^{T}\bar{z}_{nt}^*$, where $\bar{z}_{nt}^* = n^{-1}\sum_{i=1}^{n} z_{it}^*$ is the MBB resample cross sectional average for observation t. To prove the consistency of the MBB we need to strengthen Assumption A as follows.

A1′. $\|z_{it}\|_{r+\delta} \leq \Delta < \infty$ for some $r > 2$ and some small $\delta > 0$, for all (i, t).

A2′. $\{z_{it}\}$ is an α-mixing random field of size $-(2+\delta)(r+\delta)/(r-2)$.

Theorem A.3 *Assume $\{z_{it}\}$ satisfies Assumption A strengthening by A1′ and A2′. If $\ell_T \to \infty$ with $\ell_T = o\left(T^{1/2}\right)$, then*

a) $n^{-1}T^{-1}\sum_{i=1}^{T}\sum_{t=1}^{T}\left(z_{it}^* - z_{it}\right) \to^{P^*} 0$, *in probability.*

b) $\Sigma_{n,T}^{-1/2}n^{-1}T^{-1/2}\sum_{i=1}^{T}\sum_{t=1}^{T}\left(z_{it}^* - z_{it}\right) \to^{d^*} N\left(0, I_d\right)$ *under P^* with probability P approaching one as $n, T \to \infty$.*

Our next result shows the consistency of

$$\hat{\Sigma}_{nT}^* = k^{-1}\sum_{i=1}^{k}\left(\ell^{-1/2}\sum_{t=1}^{\ell}\left(\bar{z}_{I_i+t,n} - \bar{\bar{z}}_{nT}^*\right)\right)^2$$

for the bootstrap variance $\Sigma_{nT}^* = Var^*\left(\sqrt{T}\bar{\bar{z}}_{nT}^*\right)$ as n and $T \to \infty$ jointly. This estimator was proposed by Götze and Künsch (1996) in the pure time series context.

Lemma A.4 *Assume $\{z_{it}\}$ satisfies Assumption A strengthened by A1′ and A2′. If $\ell_T \to \infty$ with $\ell_T = o\left(T^{1/2}\right)$ we have that for any $\varepsilon > 0$ and $\delta > 0$, $\lim_{n,T\to\infty} P\left(P^*\left(\left|\hat{\Sigma}_{nT}^* - \Sigma_{nT}^*\right| > \varepsilon\right) > \delta\right) = 0$.*

Proof of Theorem A.3. By Lemma A.2, $\bar{z}_{nt}$ is a double array that satisfies the moment and memory assumptions of Gonçalves and White (2002) (Assumption 2.1 (a) and (b). See also Assumption 1 of Gonçalves and de Jong (2003) for weaker moment conditions). In particular, $\bar{z}_{nt}$ is trivially NED on a mixing process because it is itself a mixing array (as proven in D&K). Thus, the result follows by Theorem 2.2 of Gonçalves and White (2002).

Proof of Lemma A.4. Apply Lemma A.2 and Lemma B.1 of Gonçalves and White (2004).

B Appendix B: 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).

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

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

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) $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)' - A_{nT} \to^{P} 0$.

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

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

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

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

Proof of Lemma B.1. a) We apply Theorem A.1.a) with $z_{it}=\left(x_{it,k}-\mu_{i,k}\right)\left(x_{it,l}-\mu_{i,l}\right)$, a typical (k,l) element of $\left(x_{it}-\mu_i\right)\left(x_{it}-\mu_i\right)'$. Under Assumption 1b., $\|z_{it}\|_r\leq\Delta<\infty$, whereas Assumption 1c. implies that $\{z_{it}\}$ is α-mixing of size $-\frac{2r}{r-2}$ (hence, of size $-\frac{r}{r-2}$), for some $r>2$. b) Since $\bar{x}_i=T^{-1}\sum_{t=1}^{T}x_{it}$, we can write

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

where we let $z_{it}\equiv x_{it}-\mu_i$. We show that $E\left|R_{1,nT}\right|\rightarrow 0$ and consequently $R_{1,nT}\rightarrow^P 0$ 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 $E\left|R_{1,nT}\right|\leq\frac{1}{nT^2}\sum_{i=1}^{n}E\left|\xi_{iT}\xi_{iT}'\right|\leq\frac{1}{nT^2}\sum_{i=1}^{n}\|\xi_{iT}\|_2^2$. 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 $E\left|R_{1,nT}\right|=O\left(\frac{1}{T}\right)=o(1)$ as $T\rightarrow\infty$. Specifically, for each i, Lemma A.3 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 1b. implies that $\|z_{it}\|_r\leq\Delta<\infty$ whereas Assumption 1c. 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

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\bar{x}_i\right)\left(x_{it}-\bar{x}_i\right)'-A_{nT}=I_{1,nT}+I_{2,nT}+I_{3,nT}+I_{4,nT}-A_{nT},$$

where $I_{1,nT}\equiv\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\mu_i\right)\left(x_{it}-\mu_i\right)'$, $I_{2,nT}\equiv\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}-\mu_i\right)\left(\mu_i-\bar{x}_i\right)'$, $I_{3,nT}=I_{2,nT}'$, and $I_{4,nT}\equiv\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mu_i-\bar{x}_i\right)\left(\mu_i-\bar{x}_i\right)'$. By part a) of this Lemma, $I_{1,nT}-A_{nT}\rightarrow 0$. Using the assumption of time stationarity and noting that $\bar{x}_i=T^{-1}\sum_{t=1}^{T}x_{it}$, we can show that $I_{2,nT}$ goes to zero as $n,T\rightarrow\infty$ given part b). The same holds for $I_{3,nT}$ and $I_{4,nT}$, thus completing the proof.

Proof of Lemma B.2. a) This follows from an application of Theorem A.1.b) with $w_{it}=\left(x_{it}-\mu_i\right)\varepsilon_{it}$. In particular, Assumptions 1a. and 1c. imply that w_{it} is a zero mean random field of size $-\frac{2r}{r-2}$ (hence of size $-\frac{r}{r-2}$, as required by Assumption A2) whereas Assumption 1b. implies that $\|w_{it}\|_r\leq\Delta<\infty$ for all (i,t), thus satisfying Assumption A1. Assumption 1e. ensures that Assumption A3 is satisfied.

For b), note that

$$\begin{aligned} R_{2,nT} &\equiv \frac{1}{n\sqrt{T}} \sum_{i=1}^{n} \sum_{t=1}^{T} (\mu_i - \bar{x}_i) \varepsilon_{it} = \frac{1}{n\sqrt{T}} \sum_{i=1}^{n} (\mu_i - \bar{x}_i) \sum_{t=1}^{T} \varepsilon_{it} \\ &= -\frac{1}{nT\sqrt{T}} \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \mu_i) \sum_{s=1}^{T} \varepsilon_{is} = -\frac{1}{nT\sqrt{T}} \sum_{i=1}^{n} \xi_{iT} \eta_{iT}, \end{aligned}$$

where $\xi_{iT} \equiv \sum_{t=1}^{T} (x_{it} - \mu_i) \equiv \sum_{t=1}^{T} z_{it}$ and $\eta_{iT} \equiv \sum_{s=1}^{T} \varepsilon_{is}$. We can show that

$$E\left|R_{2,nT}\right| \leq \frac{1}{nT\sqrt{T}} \sum_{i=1}^{n} E\left|\xi_{iT}\eta_{iT}\right| \leq \frac{1}{nT\sqrt{T}} \sum_{i=1}^{n} \left\|\xi_{iT}\right\|_2 \left\|\eta_{iT}\right\|_2 = O\left(\frac{1}{\sqrt{T}}\right) = o(1),$$

as $T \to \infty$. Consequently, by Markov's inequality, it follows that $R_{2,nT} \to^{P} 0$. In particular, Assumption 1 ensures that $\left\|\xi_{iT}\right\|_2 \leq CT^{1/2}$ (see the proof of Lemma B.1.a)). A similar argument can be used to show that $\left\|\eta_{iT}\right\|_2 \leq CT^{1/2}$ for some constant C independent of T and i. Thus, $E\left|R_{2,nT}\right| = O\left(\frac{1}{\sqrt{T}}\right) = o(1)$ as $T \to \infty$.

Proof of Theorem 2.1. The proof follows standard arguments (see e.g. Gallant and White, 1988, p. 82). In particular, note that $B_{nT}^{-1/2} = O(1)$ and $A_{nT} = O(1)$ (this follows by Assumption 1b) and apply Lemmas B.1 and B.2.

Proof of Theorem 2.2. By definition,

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

where $\hat{s}_{nt} = n^{-1} \sum_{i=1}^{n} (x_{it} - \bar{x}_i) \hat{\varepsilon}_{it}$ and $\hat{\varepsilon}_{it} = \varepsilon_{it} - (x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right) - (\hat{\alpha}_i - \alpha_i)$. Thus,

$$\begin{aligned} \hat{s}_{nt} &= n^{-1} \sum_{i=1}^{n} (x_{it} - \bar{x}_i) \hat{\varepsilon}_{it} = \hat{s}_{nt} = n^{-1} \sum_{i=1}^{n} (x_{it} - \bar{x}_i) \left(\varepsilon_{it} - (x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right) - (\hat{\alpha}_i - \alpha_i)\right) \\ &= n^{-1} \sum_{i=1}^{n} (x_{it} - \mu_i) \varepsilon_{it} + (\mu_i - \bar{x}_i) \varepsilon_{it} - (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right) - (x_{it} - \bar{x}_i)(\hat{\alpha}_i - \alpha_i) \\ &= s_{nt} + a_{nt} + b_{nt} \end{aligned}$$

where $s_{nt} \equiv n^{-1} \sum_{i=1}^{n} (x_{it} - \mu_i) \varepsilon_{it}$, $a_{nt} \equiv n^{-1} \sum_{i=1}^{n} (\mu_i - \bar{x}_i) \varepsilon_{it}$ and
$b_{nt} \equiv -n^{-1} \sum_{i=1}^{n} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \left(\hat{\beta} - \beta\right) - n^{-1} \sum_{i=1}^{n} (x_{it} - \bar{x}_i)(\hat{\alpha}_i - \alpha_i)$. Substituting $\hat{s}_{nt}$ in $\hat{B}_{nT} - B_{nT}$ yields

$$
\begin{aligned}
\hat{B}_{nT}-B_{nT} &= T^{-1}\sum_{t=1}^{T} s_{nt}s_{nt}' + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau} s_{nt}s_{nt+\tau}' + T^{-1}\sum_{t=1}^{T-\tau} s_{nt+\tau}s_{nt}'\right) - B_{nT} \\
&\quad + \left(T^{-1}\sum_{t=1}^{T} s_{nt}\left(a_{nt}+b_{nt}\right)' + T^{-1}\sum_{t=1}^{T} a_{nt}\left(s_{nt}+a_{nt}+b_{nt}\right)' + T^{-1}\sum_{t=1}^{T} b_{nt}\left(s_{nt}+a_{nt}+b_{nt}\right)'\right) \\
&\quad + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau} s_{nt}\left(a_{nt+\tau}+b_{nt+\tau}\right)' + T^{-1}\sum_{t=1}^{T-\tau}\left(a_{nt}+b_{nt}\right)\left(s_{nt+\tau}+a_{nt+\tau}+b_{nt+\tau}\right)'\right) \\
&\quad + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau}\left(a_{nt+\tau}+b_{nt+\tau}\right)s_{nt}' + T^{-1}\sum_{t=1}^{T-\tau}\left(s_{nt+\tau}+a_{nt+\tau}+b_{nt+\tau}\right)\left(a_{nt}+b_{nt}\right)'\right) \\
&\equiv I_{1,nT} + I_{2,nT} + I_{3,nT} + I_{3,nT}'.
\end{aligned}
$$

By Theorem A.2 in Appendix A, under Assumption 1, 2 and 3, $I_{1nT} \to^{P} 0$. Thus, it suffices to show that each of the remaining terms is $o_P(1)$. We start with $I_{2,nT}$, which we write as

$$I_{2,nT} = J_{1,nT} + J_{2,nT} + J_{3,nT},$$

where $J_{1,nT} = T^{-1}\sum_{t=1}^{T} s_{nt}\left(a_{nt}+b_{nt}\right)'$, $J_{2,nT} = T^{-1}\sum_{t=1}^{T} a_{nt}\left(s_{nt}+a_{nt}+b_{nt}\right)'$, and $J_{3,nT} = T^{-1}\sum_{t=1}^{T} b_{nt}\left(s_{nt}+a_{nt}+b_{nt}\right)'$. Next we show that $J_{k,nT} = o_P(1)$ for all $k = 1, 2, 3$, which implies $I_{2,nT} = o_P(1)$. We can write

$$J_{1,nT} = T^{-1}\sum_{t=1}^{T} s_{nt}\left(a_{nt}+b_{nt}\right)' = T^{-1}\sum_{t=1}^{T} s_{nt}a_{nt}' + T^{-1}\sum_{t=1}^{T} s_{nt}b_{nt}' \equiv \zeta_{1,nT} + \zeta_{2,nT}.$$

$\zeta_{1,nT} \to^{P} 0$ if $\left\|\zeta_{1,nT}\right\|_1 \le T^{-1}\sum_{t=1}^{T}\left\|s_{nt}\right\|_2\left\|a_{nt}\right\|_2 \to 0$. By definition of s_{nt}, $\left\|s_{nt}\right\|_2 = \left\|n^{-1}\sum_{i=1}^{n} z_{it}\varepsilon_{it}\right\|_2 \le n^{-1}\sum_{i=1}^{n}\left\|z_{it}\varepsilon_{it}\right\|_2 \le \Delta < \infty$ provided $\left\|z_{it}\varepsilon_{it}\right\|_2 \le \Delta < \infty$. A sufficient condition is that $\left\|z_{it}\right\|_4 \le \Delta < \infty$ and $\left\|\varepsilon_{it}\right\|_4 \le \Delta < \infty$ (which holds under our Assumption 1b′). By definition of a_{nt}, recalling that $\mu_i - \bar{x}_i = -T^{-1}\sum_{t=1}^{T} z_{it}$, and letting $\xi_{iT} = \sum_{s=1}^{T} z_{is}$, we have that

$$a_{nt} = n^{-1}\sum_{i=1}^{n}\left(\mu_i - \bar{x}_i\right)\varepsilon_{it} = -T^{-1}n^{-1}\sum_{i=1}^{n}\left(\sum_{s=1}^{T} z_{is}\right)\varepsilon_{it} \equiv -T^{-1}n^{-1}\sum_{i=1}^{n}\xi_{iT}\varepsilon_{it}.$$

Thus, by first using the Minkowski inequality and then the Cauchy-Schwartz inequality,

$$\left\|a_{nt}\right\|_2 \le T^{-1}n^{-1}\sum_{i=1}^{n}\left\|\xi_{iT}\varepsilon_{it}\right\|_2 \le T^{-1}n^{-1}\sum_{i=1}^{n}\left\|\xi_{iT}\right\|_4\left\|\varepsilon_{it}\right\|_4 \le \Delta T^{-1}n^{-1}\sum_{i=1}^{n}\left\|\xi_{iT}\right\|_4.$$

By definition of the L_4- and the Euclidean norms, $\left\|\xi_{iT}\right\|_4 = \left(E\left|\xi_{iT}\right|^4\right)^{1/4} = \left(E\left|\sum_{k=1}^{p}\xi_{iT,k}^2\right|^2\right)^{1/4} \le \left(\sum_{k=1}^{p}E\left|\xi_{iT,k}\right|^4\right)^{1/4} \le \sum_{k=1}^{p}\left\|\xi_{iT,k}\right\|_4$. For each $k = 1, \ldots, p$, we can show that $\left\|\xi_{iT,k}\right\|_4 \le CT^{1/2}$ for some constant C independent of i. In particular, Lemma A.3 with $p = 4$ together with Lemma A.1 imply that $\left\|\xi_{iT,k}\right\|_4 \le K\sum_{j=1}^{\infty}\alpha(j)^{\frac{1}{4}-\frac{1}{r'}}\left(\sum_{t=1}^{T}\left\|z_{it}\right\|_{r'}^2\right)^{1/2}$ for some $r' > 4$. Setting $r' = 2r$ and using the size condition in Assumption 1c. and the moment condition in Assumption 1b. show that $\left\|\xi_{iT,k}\right\|_4 = O\left(T^{1/2}\right)$ uniformly in i. Thus, $\left\|a_{nt}\right\|_2 \le CT^{-1/2}$ uniformly in n, t. It follows that

$\zeta_{1,nT} = O\left(\frac{1}{\sqrt{T}}\right)$. Next we analyze $\zeta_{2,nT}$. Noting that

$$\hat{\alpha}_i - \alpha_i = T^{-1}\sum_{t=1}^{T}\varepsilon_{it} - T^{-1}\sum_{t=1}^{T}x_{it}'\left(\hat{\beta}-\beta\right) \equiv T^{-1}\eta_{iT} - \bar{x}_i'\left(\hat{\beta}-\beta\right),$$

with $\eta_{iT} = \sum_{t=1}^{T}\varepsilon_{it}$, it follows that

$$\begin{aligned}
b_{nt} &\equiv -n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\left(\hat{\beta}-\beta\right) - n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)(\hat{\alpha}_i-\alpha_i) \\
&= -n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)(x_{it}-\bar{x}_i)'\left(\hat{\beta}-\beta\right) - n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)\left(T^{-1}\eta_{iT}-\bar{x}_i'\left(\hat{\beta}-\beta\right)\right), \\
&= -n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)(x_{it}-2\bar{x}_i)'\left(\hat{\beta}-\beta\right) - T^{-1}n^{-1}\sum_{i=1}^{n}(x_{it}-\bar{x}_i)\eta_{iT} \equiv b_{1,nt}+b_{2,nt}.
\end{aligned}$$

Thus

$$\zeta_{2,nT} = T^{-1}\sum_{t=1}^{T}s_{nt}b_{nt}' = T^{-1}\sum_{t=1}^{T}s_{nt}b_{1,nt}' + T^{-1}\sum_{t=1}^{T}s_{nt}b_{2,nt}' \equiv S_{1,nT}+S_{2,nT}.$$

For the first term, consider

$$\begin{aligned}
Vec\left(S_{1,nT}\right) &= Vec\left(T^{-1}\sum_{t=1}^{T}s_{nt}b_{1,nt}'\right) = Vec\left(T^{-1}\sum_{t=1}^{T}s_{nt}\left(\hat{\beta}-\beta\right)'(x_{it}-2\bar{x}_i)(x_{it}-\bar{x}_i)'\right) \\
&= T^{-1}\sum_{t=1}^{T}Vec\left(s_{nt}\left(\hat{\beta}-\beta\right)'(x_{it}-2\bar{x}_i)(x_{it}-\bar{x}_i)'\right) \\
&= T^{-1}\sum_{t=1}^{T}\left((x_{it}-\bar{x}_i)(x_{it}-2\bar{x}_i)'\otimes s_{nt}\right)Vec\left(\hat{\beta}-\beta\right) \equiv R_{nT}\cdot\left(\hat{\beta}-\beta\right)
\end{aligned}$$

where we have used the fact that $Vec(ABC) = (C'\otimes A)\,Vec(B)$. Since $\hat{\beta}-\beta = O_P\left(\frac{1}{\sqrt{T}}\right) = o_P(1)$, it suffices to show that $R_{nT} = O_P(1)$, or that $E\left|R_{nT}\right| = O(1)$. Using our Assumption 1b′., we can show that there exist a finite matrix Δ such that $E\left|R_{nT}\right| \leq T^{-1}\sum_{t=1}^{T}E\left|(x_{it}-\bar{x}_i)(x_{it}-2\bar{x}_i)'\otimes s_{nt}\right| \leq \Delta$, which shows that $R_{nT} = O_P(1)$ and therefore $S_{1,nT} = o_P(1)$. For $S_{2,nT}$, we have that $\left\|S_{2,nT}\right\|_1 \leq T^{-1}\sum_{t=1}^{T}\left\|s_{nt}\right\|_2\left\|b_{2,nt}\right\|_2$, where $\left\|s_{nt}\right\|_2 \leq \Delta < \infty$ and $\left\|b_{2,nt}\right\|_2 \leq T^{-1}n^{-1}\sum_{i=1}^{n}\left\|x_{it}-\bar{x}_i\right\|_4\left\|\eta_{iT}\right\|_4 = O\left(\frac{1}{\sqrt{T}}\right)$, given that $\left\|x_{it}-\bar{x}_i\right\|_4 \leq 2\left\|x_{it}\right\|_4 \leq \Delta < \infty$ and $\left\|\eta_{iT}\right\|_4 \leq C\sqrt{T}$ uniformly in i. Thus, $\left\|S_{2,nT}\right\|_1 \leq CT^{-1/2}$, showing that $S_{2,nT} = O_P\left(\frac{1}{\sqrt{T}}\right)$. This completes the proof that $J_{1,nT} = o_P(1)$. Next,

$$J_{2,nT} = T^{-1}\sum_{t=1}^{T}a_{nt}s_{nt}' + T^{-1}\sum_{t=1}^{T}a_{nt}a_{nt}' + T^{-1}\sum_{t=1}^{T}a_{nt}b_{nt}' \equiv \psi_{1,nT}+\psi_{2,nT}+\psi_{3,nT}.$$

$\psi_{1,nT} = \zeta_{1,nT}'$ and therefore it follows immediately that $\psi_{1,nT} = o_P(1)$. For $\psi_{2,nT}$, note that $E\left|\psi_{2,nT}\right| \leq T^{-1}\sum_{t=1}^{T}E\left|a_{nt}a_{nt}'\right| \leq T^{-1}\sum_{t=1}^{T}\left\|a_{nt}\right\|_2^2 = O\left(T^{-1}\right)$, since we showed before that $\left\|a_{nt}\right\|_2 \leq CT^{-1/2}$.

Thus, $\psi_{2,nT} = O_P\left(T^{-1}\right) = o_P(1)$. To show that $\psi_{3,nT} = o_P(1)$, we can proceed as for $\zeta_{2,nT}$. Finally,

$$J_{3,nT} = T^{-1}\sum_{t=1}^{T} b_{nt}s'_{nt} + T^{-1}\sum_{t=1}^{T} b_{nt}a_{nt} + T^{-1}\sum_{t=1}^{T} b_{nt}b'_{nt} \equiv \omega_{1,nT} + \omega_{2,nT} + \omega_{3,nT},$$

where $\omega_{1,nT} = \zeta'_{2,nT} = o_P(1)$, $\omega_{2,nT} = \psi'_{3,nT} = o_P(1)$, and the term $\omega_{3,nT}$ is analyzed next. In particular,

$$\begin{aligned}
\omega_{3,nT} &= T^{-1}\sum_{t=1}^{T} b_{nt}b'_{nt} = T^{-1}\sum_{t=1}^{T}\left(b_{1,nt} + b_{2,nt}\right)\left(b_{1,nt} + b_{2,nT}\right)' \\
&= T^{-1}\sum_{t=1}^{T} b_{1,nt}b'_{1,nt} + T^{-1}\sum_{t=1}^{T} b_{1,nt}b'_{2,nt} + T^{-1}\sum_{t=1}^{T} b_{2,nt}b'_{1,nt} + T^{-1}\sum_{t=1}^{T} b_{2,nt}b'_{2,nt}.
\end{aligned}$$

We can show that each of the above terms vanishes in probability. Since the arguments are similar, we consider only the first term in detail,

$$T^{-1}\sum_{t=1}^{T} b_{1,nt}b'_{1,nt} = T^{-1}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(x_{it} - 2\bar{x}_i\right)'\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\left(x_{it} - 2\bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)',$$

with

$$Vec\left(T^{-1}\sum_{t=1}^{T} b_{1,nt}b'_{1,nt}\right) = T^{-1}\sum_{t=1}^{T}\left(\left(x_{it} - \bar{x}_i\right)\left(x_{it} - 2\bar{x}_i\right)' \otimes \left(x_{it} - \bar{x}_i\right)\left(x_{it} - 2\bar{x}_i\right)'\right) Vec\left(\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right).$$

Under our Assumption 1b′., $T^{-1}\sum_{t=1}^{T}\left(\left(x_{it} - \bar{x}_i\right)\left(x_{it} - 2\bar{x}_i\right)' \otimes \left(x_{it} - \bar{x}_i\right)\left(x_{it} - 2\bar{x}_i\right)'\right) = O_P(1)$, whereas $Vec\left(\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right) = O_P\left(\frac{1}{T}\right)$, thus showing that $T^{-1}\sum_{t=1}^{T} b_{1,nt}b'_{1,nt} = o_P(1)$. Next we analyze $I_{3,nT}$. We consider a typical element of $I_{3,nT}$, say the (k,l) element, where $k, l = 1, 2, \ldots, p$,

$$\begin{aligned}
I_{3,nT}^{(k,l)} &= \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}\left(a_{nt+\tau,l} + b_{nt+\tau,l}\right) + T^{-1}\sum_{t=1}^{T-\tau}\left(a_{nt,k} + b_{nt,k}\right)\left(s_{nt+\tau,l} + a_{nt+\tau,l} + b_{nt+\tau,l}\right)\right) \\
&= \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(\begin{array}{c}
T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}a_{nt+\tau,l} + T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}b_{nt+\tau,l} \\
+T^{-1}\sum_{t=1}^{T-\tau} a_{nt,k}s_{nt+\tau,l} + T^{-1}\sum_{t=1}^{T-\tau} a_{nt,k}a_{nt+\tau,l} + T^{-1}\sum_{t=1}^{T-\tau} a_{nt,k}b_{nt+\tau,l} \\
+T^{-1}\sum_{t=1}^{T-\tau} b_{nt,k}s_{nt+\tau,l} + T^{-1}\sum_{t=1}^{T-\tau} b_{nt,k}a_{nt+\tau,l} + T^{-1}\sum_{t=1}^{T-\tau} b_{nt,k}b_{nt+\tau,l}
\end{array}\right) \\
&\equiv M_{1,nT} + M_{2,nT} + M_{3,nT} + M_{4,nT} + M_{5,nT} + M_{6,nT} + M_{7,nT} + M_{8,nT}.
\end{aligned}$$

By time stationarity, $\mu_i - \bar{x}_i = -T^{-1}\sum_{t=1}^{T}\left(x_{it} - \mu_i\right)$, and hence
$a_{nt+\tau} = n^{-1}\sum_{i=1}^{n}\left(\mu_i - \bar{x}_i\right)\varepsilon_{it+\tau} = -T^{-1}n^{-1}\sum_{i=1}^{n}\xi_{iT}\varepsilon_{it+\tau}$, where $\xi_{iT} \equiv \sum_{s=1}^{T} z_{is}$. Thus

$$\begin{aligned}
\left\|M_{1,nT}\right\|_1 &\le \sum_{\tau=1}^{T-1}\left|k\left(\frac{\tau}{M}\right)\right| T^{-1}\sum_{t=1}^{T-\tau}\left\|s_{nt,k}a_{nt+\tau,l}\right\|_1 \\
&\le \sum_{\tau=1}^{T-1}\left|k\left(\frac{\tau}{M}\right)\right| T^{-1}\sum_{t=1}^{T-\tau}\left\|s_{nt,k}\right\|_2\left\|a_{nt+\tau,l}\right\|_2 \le \sum_{\tau=1}^{T-1}\left|k\left(\frac{\tau}{M}\right)\right|\Delta T^{-1/2} \\
&\le C\frac{1}{M}\sum_{\tau=-(T-1)}^{T-1}\left|k\left(\frac{\tau}{M}\right)\right| \cdot \frac{M}{\sqrt{T}}
\end{aligned}$$

given that $\|s_{nt,k}\|_2 \leq \Delta$ and $\|a_{nt+\tau,l}\|_2 \leq CT^{-1/2}$. 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 2, it follows that $\|M_{1,nT}\|_1 = O_P\left(\frac{M}{\sqrt{T}}\right) = o_P(1)$ provided $\frac{M}{\sqrt{T}} \to 0$, as we assume in Assumption 3. For $M_{2,nT}$,

$$M_{2,nT} = \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}b_{1,nt+\tau,l}\right) + \sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)\left(T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}b_{2,nt+\tau,l}\right) \equiv \upsilon_{1,nT} + \upsilon_{2,nT}.$$

$\upsilon_{2,nT}$ may be analyzed in the same way as $M_{1,nT}$ since in particular $\|b_{2,nt+\tau,l}\|_2 \leq CT^{-1/2}$. For $\upsilon_{1,nT}$, substituting $b_{1,nt+\tau,l} = -n^{-1}\sum_{i=1}^n (x_{it+\tau,l} - \bar{x}_{i,l})(x_{it+\tau,l} - 2\bar{x}_i)\left(\hat{\beta} - \beta\right)$ yields

$$\upsilon_{1,nT} = -\sum_{\tau=1}^{T-1} k\left(\frac{\tau}{M}\right)T^{-1}\sum_{t=1}^{T-\tau} s_{nt,k}\left(n^{-1}\sum_{i=1}^{n}(x_{it+\tau,l} - \bar{x}_{i,l})(x_{it+\tau,l} - 2\bar{x}_i)\right)\left(\hat{\beta} - \beta\right) \equiv \pi_{nT}\cdot\left(\hat{\beta} - \beta\right)$$

We can show that $\pi_{nT} = O_P(M)$ whereas $\hat{\beta} - \beta = O_P\left(\frac{1}{\sqrt{T}}\right)$, thus implying that $\upsilon_{1,nT} = O_P\left(\frac{M}{\sqrt{T}}\right) = o_P(1)$. The remaining terms can be analyzed using similar arguments and therefore we omit the details.

C Appendix C: 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 C.1 *Under Assumption 1 strengthened by Assumption 1b$''$ and 1c$'$, if $\ell_T \to \infty$ such that $\ell = o(T)$ as $T \to \infty$,*

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

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

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

Lemma C.2 *Under Assumption 1 strengthened by Assumption 1b$''$ and 1c$'$, $\ell_T \to \infty$ such that $\ell = o(T)$ as $T \to \infty$, $B_{nT}^{-1/2}\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T (x_{it}^* - \bar{x}_i^*)\varepsilon_{it}^* \to^{d^*} N(0, I_p)$, in probability.*

Theorem C.1 *Under Assumption strengthened by Assumption 1b$''$ and 1c$'$, $\ell_T \to \infty$ such that $\ell = o(T)$ as $T \to \infty$, $\hat{B}_{nT}^* - B_{nT} \to^{P^*} 0$, in probability.*

Proof of Lemma C.1. Without loss of generality, we consider the scalar case with $p = 1$. a) follows from an application of Theorem A.3.a) in Appendix A with $w_{it} = x_{it}^2$. Under Assumption 1b$''$, $\|w_{it}\|_{r+\delta} \leq \Delta < \infty$, whereas Assumption 1c$'$. implies that $\{w_{it}\}$ is α-mixing of size $-\frac{4r}{r-2}$ (hence, of size $-\frac{(2+\delta)(r+\delta)}{r-2}$), for some $r > 2$ and some small $\delta > 0$ (in particular, it suffices that $0 < \delta < 1$). To prove b), note that

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

where $J_{1,nT}^* \equiv n^{-1}\sum_{i=1}^n \left(\bar{x}_i^* - E^*\left(\bar{x}_i^*\right)\right)^2$, and $J_{2,nT}^* \equiv n^{-1}\sum_{i=1}^n \left(E^*\left(\bar{x}_i^*\right) - \bar{x}_i\right)^2$. We show that $J_{l,nT}^* = o_{P^*}(1)$ in probability, for $l = 1$ and 2. By repeated application of Markov's inequality, it suffices to 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

$$\begin{aligned} E^*\left|J_{1,nT}^*\right| &\leq n^{-1}\sum_{i=1}^n k^{-2}\ell^{-2}E^*\left|\left(\sum_{j=1}^k A_{i,I_j}\right)^2\right| \leq n^{-1}\sum_{i=1}^n k^{-2}\ell^{-2}k\sum_{j=1}^k E^*\left|A_{i,I_j}\right|^2 \\ &= n^{-1}\sum_{i=1}^n \ell^{-2}E^*\left|A_{i,I_1}\right|^2 = n^{-1}\sum_{i=1}^n \ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|A_{i,j}\right|^2 \\ &= n^{-1}\sum_{i=1}^n \ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell z_{i,t+j} + \ell\left(\mu_i - E^*\left(\bar{x}_i^*\right)\right)\right|^2 \\ &\leq Cn^{-1}\sum_{i=1}^n \ell^{-2}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}\left|\sum_{t=1}^\ell z_{i,t+j}\right|^2 + n^{-1}\sum_{i=1}^n \ell^{-2}\left|\ell\left(\mu_i - E^*\left(\bar{x}_i^*\right)\right)\right|^2 \equiv F_1 + F_2, \end{aligned}$$

where $z_{i,t+j} \equiv x_{i,t+j} - \mu_i$, with $\mu_i \equiv E\left(x_{i,t+j}\right)$, 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, Lemma A.3 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 1b$''$. implies that $\left\|z_{i,t+j}\right\|_r \leq \Delta < \infty$ whereas Assumption 1c$'$. 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 $\mu_i - E^*\left(\bar{x}_i^*\right) = -T^{-1}\sum_{t=1}^T E^*\left(x_{it}^* - \mu_i\right) = -T^{-1}\sum_{t=1}^T E^*\left(z_{it}^*\right) \equiv -E^*\left(\bar{z}_i^*\right)$, it follows that

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

Using Lemma A.1 of Gonçalves and White (2005), we can show that $E\left|E^*\left(\left|\bar{z}_i^*\right|^2\right)\right| = O\left(\frac{1}{T}\right) + O\left(\frac{\ell^2}{T^2}\right) = o(1)$ uniformly in i, which implies that $F_2 = o_{P^*}(1)$ in probability. To prove c), note that we can write

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

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 $E\left(E^*\left|\bar{x}_i^*-\bar{x}_i\right|\right)^2=O\left(\frac{1}{\ell}\right)+O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right)=o(1)$, uniformly in i.

Proof of Lemma C.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 $\hat{\beta}$, $\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(x_{it}-\bar{x}_i\right)\hat{\varepsilon}_{it}=0$. Thus, adding an subtracting appropriately, we have that

$$\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(x_{it}^*-\bar{x}_i^*\right)\varepsilon_{it}^* = \frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T x_{it}^*\left(\varepsilon_{it}^{*0}-x_{it}^{*\prime}\left(\hat{\beta}-\beta\right)-\hat{\alpha}_i-\alpha_i\right)-\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\bar{x}_i^*\varepsilon_{it}^*$$
$$= \frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(\left(x_{it}^*-\mu_i\right)\varepsilon_{it}^{*0}-\left(x_{it}-\mu_i\right)\varepsilon_{it}\right)+\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(\bar{x}_i-\mu_i\right)\varepsilon_{it}$$
$$-\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(\bar{x}_i^*-\mu_i\right)\varepsilon_{it}^{*0}+\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(x_{it}-\bar{x}_i\right)\left(x_{it}'\left(\hat{\beta}-\beta\right)+\left(\hat{\alpha}_i-\alpha_i\right)\right)$$
$$-\frac{1}{n\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T\left(x_{it}^*-\bar{x}_i^*\right)\left(x_{it}^{*\prime}\left(\hat{\beta}-\beta\right)+\hat{\alpha}_i-\alpha_i\right)$$
$$\equiv \xi^*_{1,nT}+\omega_{1,nT}-\omega^*_{2,nT}+\omega_{3,nT}-\omega^*_{4,nT}$$

By Theorem A.3.b), $B_{n,T}^{-1/2}\xi^*_{1,nT}\to^{d^*}N(0,I_p)$, in probability, provided $w_{it}\equiv\left(x_{it}-\mu_i\right)\varepsilon_{it}$ satisfies Assumption A strengthened by A1$'$ and A2$'$ and $\ell=o\left(\sqrt{T}\right)$. Assumption 1.b$''$ suffices for A1$'$ whereas Assumption 1.c$'$ suffices for A2$'$. Lemma B.2.b) shows that $\omega_{1,nT}=o_P(1)$. Next we show that $\omega^*_{2,nT}=o_{P^*}(1)$ in probability. Let $\bar{z}_i^*\equiv\bar{x}_i^*-\mu_i$ and $\bar{\varepsilon}_i^{*0}\equiv T^{-1}\sum_{t=1}^T\varepsilon_{it}^{*0}$. By repeated application of the Cauchy Schwartz inequality, we have that

$$E\left|E^*\left|\omega^*_{2,nT}\right|\right| \leq \frac{\sqrt{T}}{n}\sum_{i=1}^n E\left(E^*\left|\bar{z}_i^*\bar{\varepsilon}_i^{*0}\right|\right)\leq\frac{\sqrt{T}}{n}\sum_{i=1}^n\left[E\left(E^*\left|\bar{z}_i^*\right|^2\right)\right]^{1/2}\left[E\left(E^*\left|\bar{\varepsilon}_i^{*0}\right|^2\right)\right]^{1/2}$$
$$= O\left(\frac{1}{\sqrt{T}}\right)+O\left(\frac{\ell^2}{T}\frac{1}{\sqrt{T}}\right),$$

where we have used the fact that $E\left(E^*\left|\bar{z}_i^*\right|^2\right)=O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right)$ and $E\left(E^*\left|\bar{\varepsilon}_i^{*0}\right|^2\right)=O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right)$ uniformly in i. Finally, we can show that $\omega_{3,nT}^*-\omega_{4,nT}^*=o_{P^*}(1)$. We can write

$$\begin{aligned}\omega_{3,nT}^*-\omega_{4,nT}^* &= \left(\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left(x_{it}-\bar{x}_i\right)\left(x_{it}-\bar{x}_i\right)'-\left(x_{it}^*-\bar{x}_i^*\right)\left(x_{it}^*-\bar{x}_i^*\right)'\right)\sqrt{T}\left(\hat{\beta}-\beta\right)\\ &\quad+\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left(\left(x_{it}-\bar{x}_i\right)-\left(x_{it}^*-\bar{x}_i^*\right)\right)\sqrt{T}\left(\hat{\alpha}_i-\alpha_i\right)\\ &\equiv \psi_{1,nT}^*+\psi_{2,nT}^*,\end{aligned}$$

where $\psi_{2,nT}^*=0$. Since $\psi_{1,nT}^*\equiv\left(\hat{A}_{nT}-\hat{A}_{nT}^*\right)\sqrt{T}\left(\hat{\beta}-\beta\right)=o_{P^*}(1)\times O_P(1)=o_{P^*}(1)$, in probability, given Lemma C.1.c) and the fact that $\sqrt{T}\left(\hat{\beta}-\beta\right)=O_P(1)$ by Theorem 2.1, this completes the proof.

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

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

where $\bar{s}_{nT}^*=T^{-1}\sum_{t=1}^T s_{nt}^*$. We can apply Lemma A.4 to show that $B_{nT}^{*0}-B_{nT}^*\to^{P^*}0$ in probability, where $B_{nT}^*=Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T s_{nt}^*\right)$. For this, it suffices that $s_{nt}\equiv n^{-1}\sum_{i=1}^n\left(x_{it}-\mu_i\right)\varepsilon_{it}$ is such that $\|s_{nt}\|_{r+\delta}\leq\Delta<\infty$ and s_{nt} is $\alpha-$mixing of size $-(2+\delta)(r+\delta)/(r-2)$, which follows under our assumptions. Since $B_{nT}^*-B_{nT}\to^P 0$, it suffices to show that $\hat{B}_{nT}^*-B_{nT}^{*0}\to^{P^*}0$, in probability. Let $\hat{S}_{n,j}^*\equiv\sum_{t=1}^\ell\hat{s}_{n,(j-1)\ell+t}^*$ and $S_{n,j}^*\equiv\sum_{t=1}^\ell s_{n,(j-1)\ell+t}^*$. We have that

$$\hat{B}_{nT}^*-B_{nT}^{*0}=\frac{1}{k}\sum_{j=1}^k\ell^{-1}\left(\hat{S}_{n,j}^{*2}-S_{n,j}^{*2}\right)-\ell\bar{s}_{nT}^{*2}\equiv D_1^*+D_2^*,$$

where $D_2^*=o_{P^*}(1)$ in probability (by an argument similar to that used in GW(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 $a_{nt}^*=n^{-1}\sum_{i=1}^n\left(\mu_i-\bar{x}_i^*\right)\varepsilon_{it}^{*0}$ and

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

It follows that

$$\hat{S}_{n,j}^*=\sum_{t=1}^\ell s_{n,(j-1)\ell+t}^*+\sum_{t=1}^\ell a_{n,(j-1)\ell+t}^*+\sum_{t=1}^\ell b_{n,(j-1)\ell+t}^*\equiv S_{n,j}^*+R_{1n,j}^*+R_{2n,j}^*.$$

and

$$\begin{aligned}
|D_1^*| &\leq \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left(\left|R_{1,nj}^{*2}\right|+\left|R_{2,nj}^{*2}\right|+\left|S_{n,j}^{*}R_{1n,j}^{*}\right|+\left|S_{n,j}^{*}R_{2n,j}^{*}\right|\right)\\
&\leq \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|R_{1,nj}^{*2}\right|+\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|R_{2,nj}^{*2}\right|+2\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^{*}R_{1n,j}^{*}\right|+2\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^{*}R_{2n,j}^{*}\right|\\
&\equiv A^*+B^*+C^*+D^*.
\end{aligned}$$

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

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

implying that

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

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 $E\left(E^*\left(\bar{z}_i^{*4}\right)\right)=O\left(\frac{1}{T^2}\right)+O\left(\frac{\ell^4}{T^4}\right)$ 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 $E\left(E^*\left|A^*\right|\right)=O\left(\ell^{-1}\right)\left(O\left(\frac{\ell}{T}\right)+O\left(\frac{\ell^3}{T^2}\right)\right)=O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right)=o\left(1\right)$. Next we show that $E\left(E^*\left|C^*\right|\right)=o\left(1\right)$. By the well known inequality

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

and by repeated application of the Cauchy-Schwartz inequality,

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

where we have used the previous result and the fact that we can show that $\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}E\left(E^*\left|S_{n,j}^*\right|^2\right)=O(1)$. Next, consider B^*:

$$\begin{aligned}
B^* &= \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|R_{2,nj}^{*2}\right| = \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}b_{n,(j-1)\ell+t}^*\right|^2 \\
&\leq \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}b_{1n,(j-1)\ell+t}^*\right|^2+\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}b_{2n,(j-1)\ell+t}^*\right|^2 \equiv B_1^*+B_2^*,
\end{aligned}$$

where

$$B_1^*=\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^*-\bar{x}_i^*\right)^2\right|^2\left|\hat{\beta}^*-\beta\right|^2 \equiv \frac{1}{T}\Psi^*.\left|\sqrt{T}\left(\hat{\beta}^*-\beta\right)\right|^2.$$

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

$$\begin{aligned}
\frac{1}{T}E\left(E^*\left|\Psi^*\right|\right) &\leq \frac{K}{T}\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\ell\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}E\left(E^*\left|\left(x_{i,I_j+t}-\bar{x}_i^*\right)\right|^4\right) \\
&\leq \frac{2K}{T}n^{-1}\sum_{i=1}^{n}\left(\sum_{t=1}^{\ell}E\left(E^*\left|z_{i,I_j+t}\right|^4\right)+\ell E\left(E^*\left|\bar{z}_i^*\right|^4\right)\right) \\
&= \frac{2K}{T}n^{-1}\sum_{i=1}^{n}\left(\sum_{t=1}^{\ell}\frac{1}{T-\ell+1}\sum_{j=0}^{T-\ell}E\left|z_{i,j+t}\right|^4+\ell E\left(E^*\left|\bar{z}_i^*\right|^4\right)\right) \\
&= O\left(\frac{\ell}{T}\right)+O\left(\frac{\ell}{T^3}\right)+O\left(\frac{\ell^5}{T^5}\right)=o(1),
\end{aligned}$$

if $\ell=o(T)$. This shows that $B_1^*=o_{P^*}(1)$, 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

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

which implies that

$$\begin{aligned}
B_2^* &\leq K\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^*-\bar{x}_i^*\right)\bar{\varepsilon}_i^{*0}\right|^2 \\
&\quad +K\frac{2}{k}\sum_{j=1}^{k}\ell^{-1}\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^*-\bar{x}_i^*\right)\bar{x}_i^*\right|^2\left|\hat{\beta}^*-\beta\right|^2 \\
&\equiv M_1^*+\frac{1}{T}M_2^*\cdot\left|\sqrt{T}\left(\hat{\beta}^*-\beta\right)\right|^2.
\end{aligned}$$

We show that M_1^* and $\frac{1}{T}M_2^*$ are $o_{P^*}(1)$ in probability. We can show that $\frac{1}{T}E\left(E^*\left|M_2^*\right|\right)$ is bounded by $\frac{1}{T}\frac{2}{k}\sum_{j=1}^{k}\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(E\left(E^*\left|z_{i,I_j+t}-\bar{z}_i^*\right|^4\right)\right)^{1/2}$, which is of order $O\left(\frac{\ell}{T}\right)+O\left(\frac{\ell}{T^3}\right)+O\left(\frac{\ell^5}{T^5}\right)=o(1)$, as we just showed. By a similar argument,

$$\begin{aligned} E\left(E^*\left|M_1^*\right|\right) &= \frac{2}{k}\sum_{j=1}^{k}\ell^{-1}E\left(E^*\left|\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(x_{i,(j-1)\ell+t}^*-\bar{x}_i^*\right)\bar{\varepsilon}_i^{*0}\right|^2\right) \\ &\le \frac{2}{k}\sum_{j=1}^{k}\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}E\left[\left(E^*\left|z_{i,I_j+t}-\bar{z}_i^*\right|^4\right)^{1/2}\left(E^*\left|\bar{\varepsilon}_i^{*0}\right|^4\right)^{1/2}\right] \\ &\le \underbrace{\frac{2}{k}\sum_{j=1}^{k}\sum_{t=1}^{\ell}n^{-1}\sum_{i=1}^{n}\left(E\left(E^*\left|z_{i,I_j+t}-\bar{z}_i^*\right|^4\right)\right)^{1/2}}_{O(\ell)+O\left(\frac{\ell}{T^2}\right)+O\left(\frac{\ell^5}{T^4}\right)}\underbrace{\left(E\left(E^*\left|\bar{\varepsilon}_i^{*0}\right|^4\right)\right)}_{O\left(\frac{1}{T}\right)+O\left(\frac{\ell^2}{T^2}\right)}^{1/2}=o(1), \end{aligned}$$

if $\frac{\ell^2}{T}=o(1)$. Thus $B^*=o_{P^*}(1)$ in probability. Finally, we show that $D^*=o_{P^*}(1)$, in probability. We have that $\left|D^*\right|^2\le\left[2\left(\frac{1}{k}\sum_{j=1}^{k}\ell^{-1}\left|S_{n,j}^*\right|^2\right)^{1/2}\left(B^*\right)^{1/2}\right]^2=O_{P^*}(1)\times o_{P^*}(1)=o_{P^*}(1)$, 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

$$\sqrt{T}\left(\hat{\beta}^*-\hat{\beta}\right)=\hat{A}_{nT}^{*-1}\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it}^*-\bar{x}_i^*\right)\varepsilon_{it}^*=\zeta_{1,nT}^*+\zeta_{2,nT}^*,$$

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

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

References

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

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

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

[4] Bertrand, M., Duflo, E. and S. Mullainathan, 2004. How much should we trust difference-in-differences estimates?, Quarterly Journal of Economics, 119, 249-275.

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

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

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

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

[9] 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.

[10] Gallant, A.R., and H. White, 1988. A unified theory of estimation and inference for nonlinear dynamic models. New York: Basil Blackwell.

[11] Gonçalves, S., and R. de Jong, 2003. Consistency of the Stationary Bootstrap under Weak Moment Conditions, Economics Letters, 81, 273-278.

[12] Gonçalves, S., and T. Vogelsang, 2008. Block bootstrap HAC robust tests: the sophistication of the naive bootstrap, manuscript, Université de Montréal.

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

[14] Gonçalves, S., and H. White, 2005. Bootstrap Standard Error Estimates for Linear Regression, Université de Montréal,Vol. 100, No. 471, 970-979.

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

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

[17] 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.

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

[19] Hansen, B.E. 1992. Erratum, Econometric Theory 8, 421-422.

[20] 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.

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

[22] Kezdi, G., 2002. Robust standard error estimation in fixed-effects panel models, mimeo.

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

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

[25] 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.

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

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

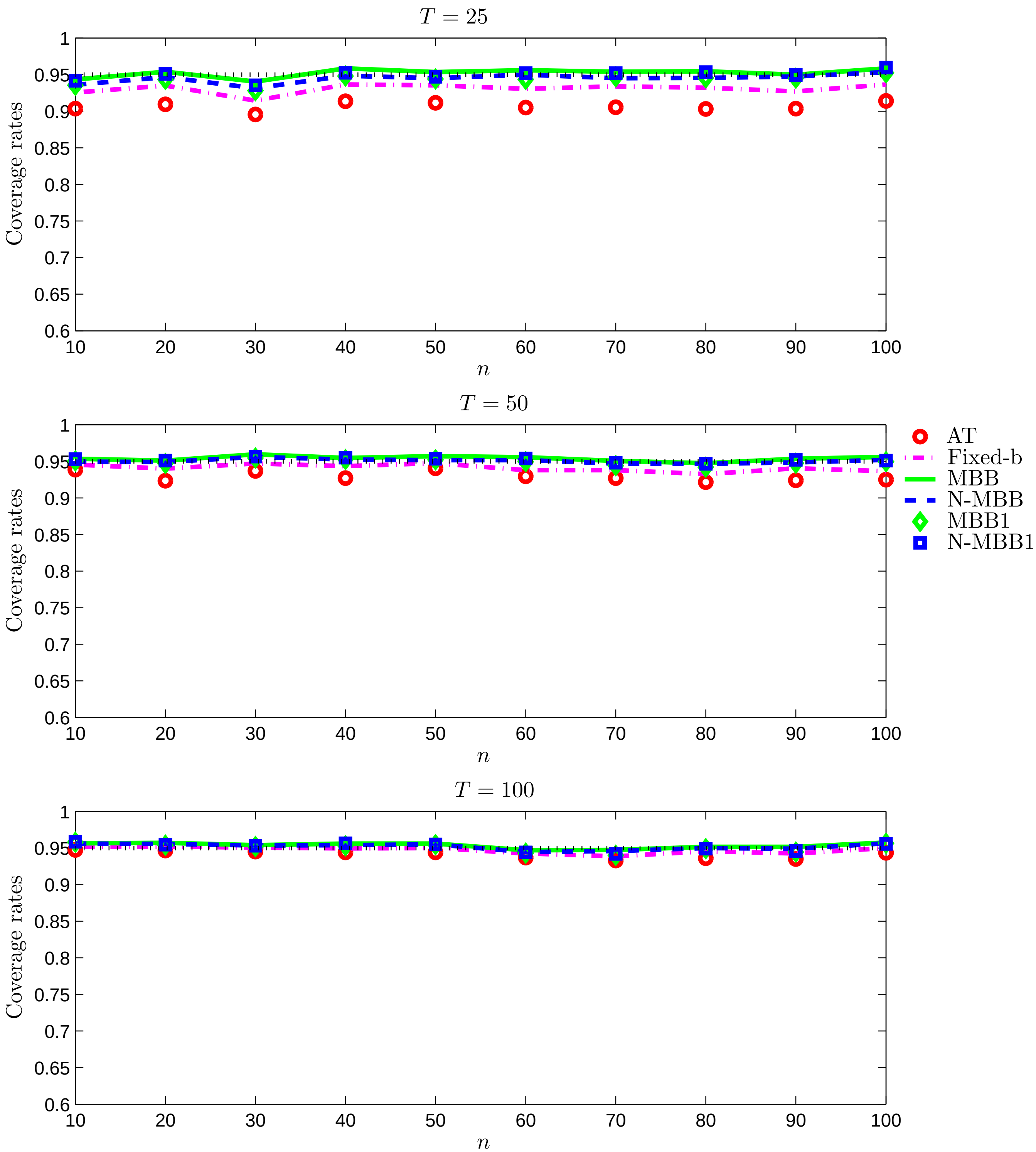

Figure 1: Empirical coverage rates, $\rho = 0.0$ and $\lambda = \sqrt{0.5}$

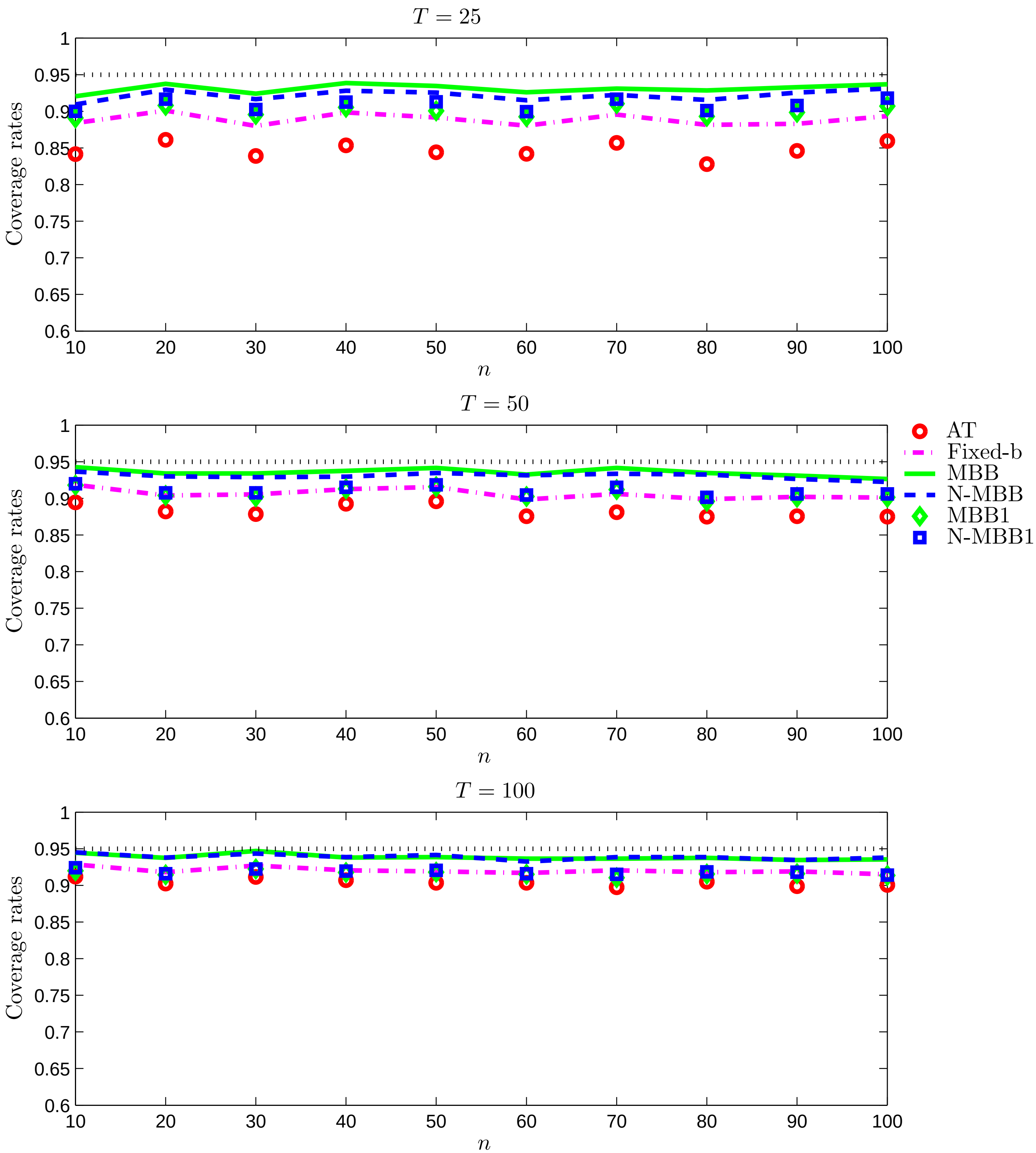

Figure 2: Empirical coverage rates, $\rho = 0.5$ and $\lambda = \sqrt{0.5}$

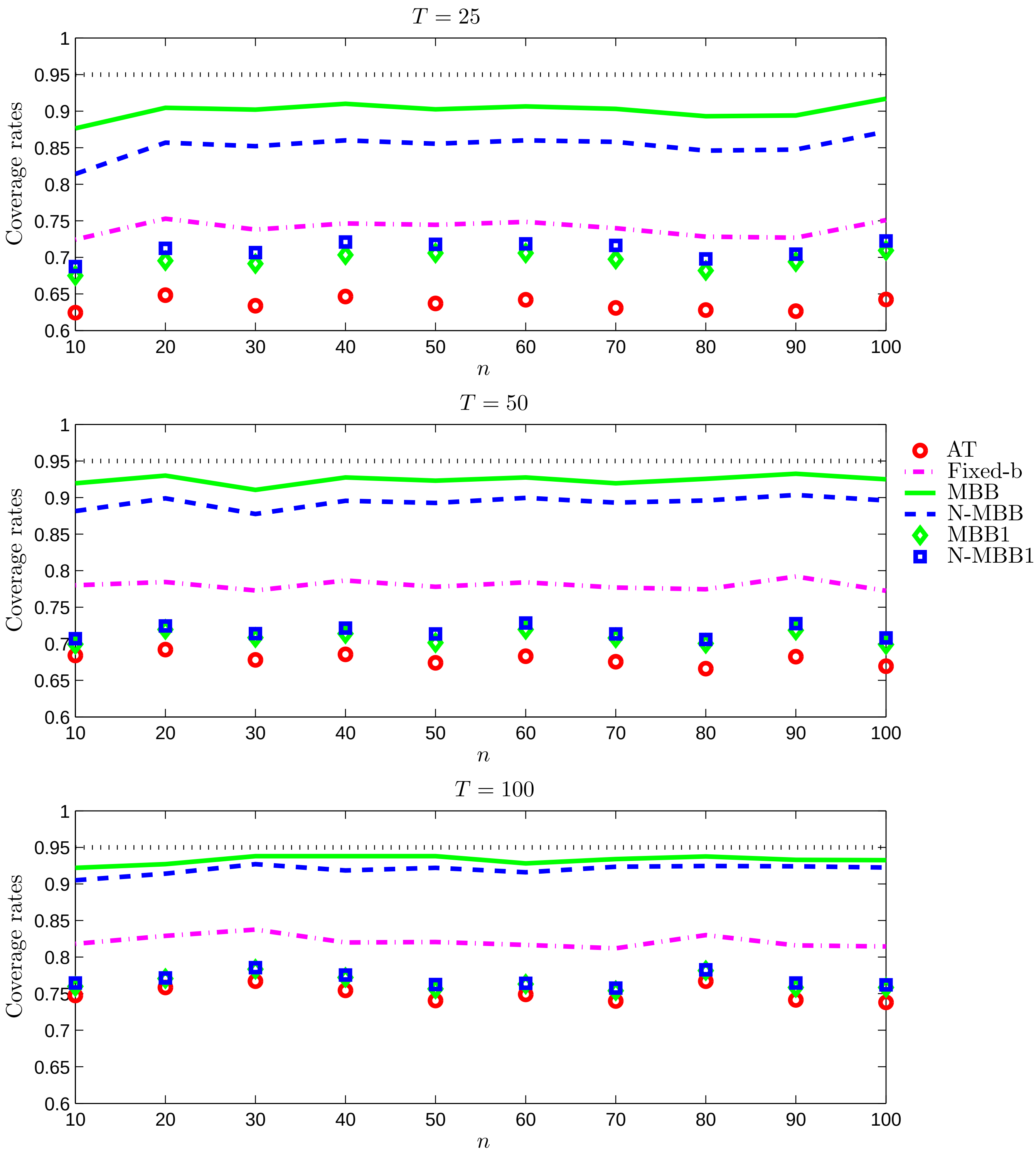

Figure 3: Empirical coverage rates, $\rho = 0.9$ and $\lambda = \sqrt{0.5}$

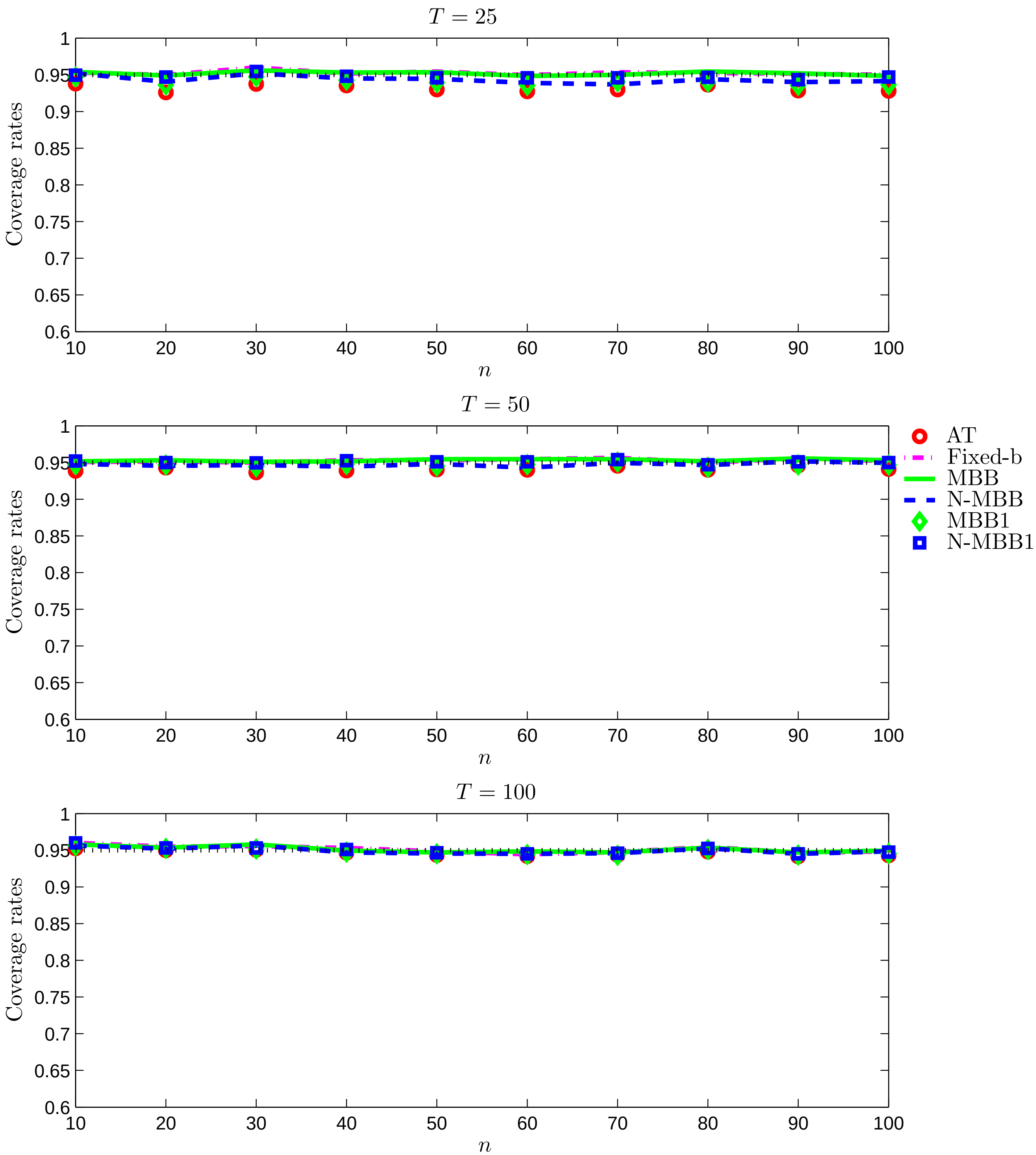

Figure 4: Empirical coverage rates, $\rho = 0.0$ and $\lambda = 0.0$

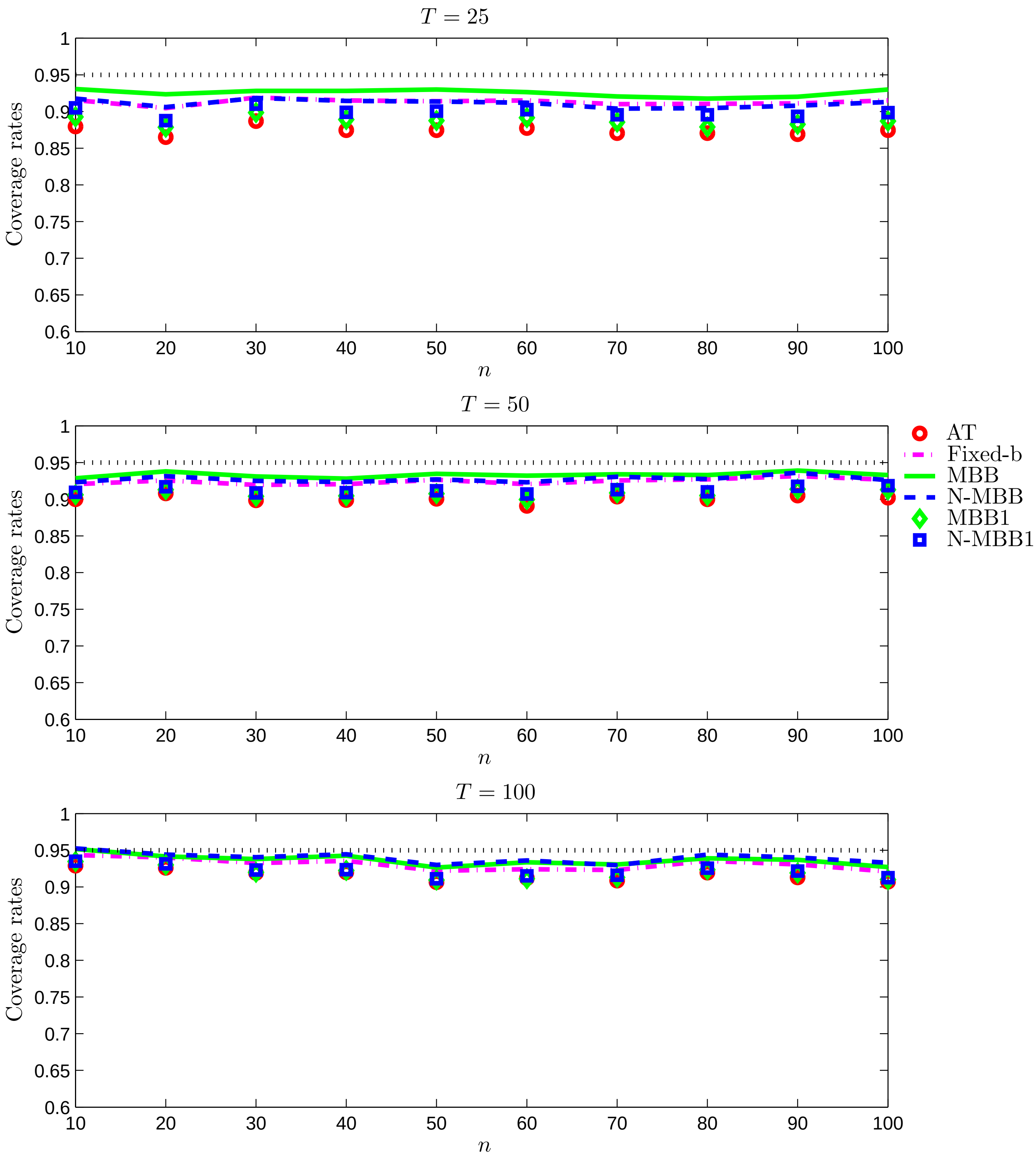

Figure 5: Empirical coverage rates, $\rho = 0.5$ and $\lambda = 0.0$

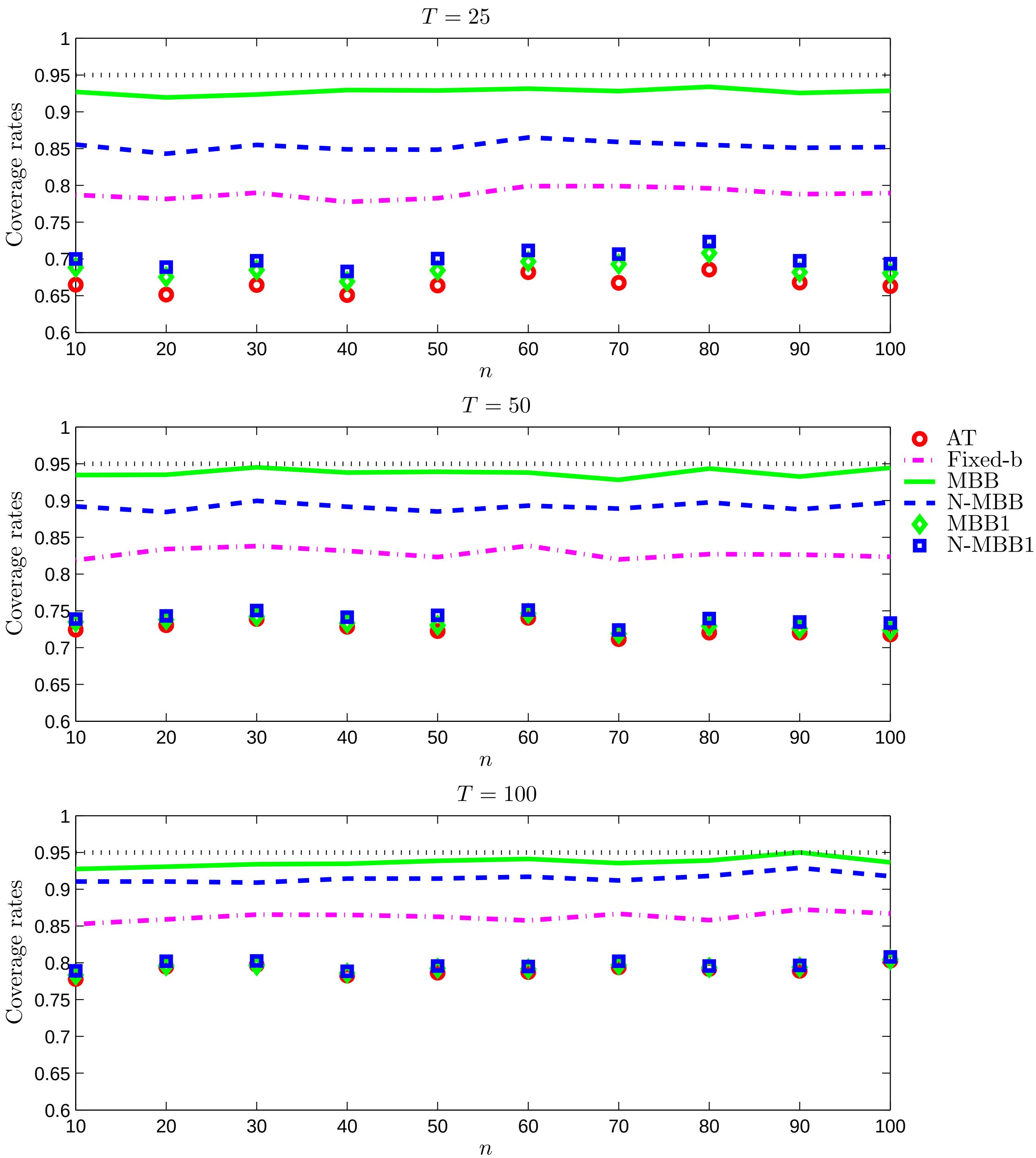

Figure 6: Empirical coverage rates, $\rho = 0.9$ and $\lambda = 0.0$