INFERENCE IN DYNAMIC MODELS FOR PANEL DATA USING THE MOVING BLOCK BOOTSTRAP

Ayden Higgins* University of Exeter

Koen Jochmans† Toulouse School of Economics

This preliminary version: February 4, 2025

Abstract

Inference in linear panel data models is complicated by the presence of fixed effects when (some of) the regressors are not strictly exogenous. Under asymptotics where the number of cross-sectional observations and time periods grow at the same rate, the within-group estimator is consistent but its limit distribution features a bias term. In this paper we show that a panel version of the moving block bootstrap, where blocks of adjacent cross-sections are resampled with replacement, replicates the limit distribution of the within-group estimator. Confidence ellipsoids and hypothesis tests based on the reverse-percentile bootstrap are thus asymptotically valid without the need to take the presence of bias into account.

JEL Classification: C23

Keywords: asymptotic bias, bootstrap, dynamic model, fixed effects, inference

*Address: University of Exeter Business School, Rennes Drive, Exeter EX4 4PU, United Kingdom. E-mail: a.higgins@exeter.ac.uk.

†Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.

Funded by the European Union (ERC-NETWORK-101044319) and by the French Government and the French National Research Agency under the Investissements d' Avenir program (ANR-17-EURE-0010). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

1 Introduction

Econometric models for panel data almost invariably feature fixed effects. Their presence complicates estimation and inference as they introduce bias in the fixed-effect estimator of the parameter of interest, in general. A leading case is the estimation of the regression slopes in a linear model when the regressors are not strictly exogenous. The canonical example is an autoregressive model, where the bias of the within-group estimator was first derived in influential work by Nickell (1981), but correlation between contemporaneous errors and both past and future regressors is the norm rather than the exception. The resulting bias remains important in large samples unless the number of time-series observations, m, is large relative to the cross-sectional sample size, n. This is, however, not a situation regularly encountered in practice.

Under asymptotics where n and m grow at the same rate the fixed-effect estimator is generally consistent but suffers from asymptotic bias. Several ways to estimate this bias to restore the validity of inference based on the limit distribution have been proposed; see Arellano and Hahn (2007) and Dhaene and Jochmans (2015). An alternative recently explored by Gonçalves and Kaffo (2015) and Higgins and Jochmans (2024) is to use the bootstrap to replicate the distribution of the fixed-effect estimator, including its bias. This allows to perform inference based on the bootstrap distribution in the usual manner, i.e., no adjustment for the presence of bias needs to be made. The difficulty here lies in devising a bootstrap scheme that correctly reproduces the bias. Higgins and Jochmans (2024) show that the parametric bootstrap does so in quite general nonlinear models. Gonçalves and Kaffo (2015) focus on the linear autoregressive model and show that the wild bootstrap permits inference using the within-group estimator. They also demonstrate the failure of several other conventional bootstrap schemes to replicate bias, illustrating the difficulty in devising a successful procedure. An important limitation of the methods available to date is that they cannot handle unspecified feedback processes, leaving open the question of how to do so and, indeed, whether or not this is possible.

In this paper we study the linear model where the errors are allowed to be correlated

with both past and future regressors. This is the workhorse model in applied economics. We show that a version of the moving block bootstrap of Künsch (1989), where blocks of adjacent cross-sections are resampled with replacement, correctly replicates the distribution of the within-group estimator under asymptotics where $n/m \to c \in [0, +\infty)$. This bootstrap scheme was not considered in Gonçalves and Kaffo (2015). Gonçalves (2011) did show that this procedure is capable of yielding correct inference in our model even under the additional complication of cross-sectional dependence. However, the assumptions under which she established this result include a condition on n and m that renders the bias of the estimator small relative to its standard deviation; in our context, this amounts to the rate condition $n/m \to 0$. Here, instead, we focus on the ability of the moving block bootstrap to replicate the bias when m need not be small relative to n and we abstract from cross-sectional dependence. Nevertheless, as the expression of the bias would remain unchanged, our main findings should generalize to such a situation.

Our contribution should be interpreted against the backdrop of a recent interest in the validity of the bootstrap in the presence of (asymptotic) bias; see also Cavaliere, Gonçalves, Nielsen and Zanelli (2024) in addition to the work already mentioned above. We conjecture that the validity of our bootstrap scheme extends to a large class of nonlinear problems. We have established this to be the case for the variance estimator in the classical many normal means problem of Neyman and Scott (1948). A more general theory is left for future work.

The paper is structured as follows. In the next section we first formally state our model and assumptions, and derive the limit distribution of the within-group estimator. This yields a somewhat more general expression of its asymptotic bias than is available in the literature. The following section introduces the moving block bootstrap and states our main result—that is, the distribution of the bootstrap within-group estimator, centered around the within-group estimator and conditional on the data, is consistent for the limit distribution of the within-group estimator centered around the truth—together with its chief implications for estimation and inference. A final section reports on a numerical experiment that was conducted to support our claims. All proofs are collected in the

Appendix.

2 Linear regression with fixed effects

We are interested in estimation of and inference on the slope vector β in the linear model

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it}, \qquad \mathbb{E}(\varepsilon_{it}) = 0, \qquad \mathbb{E}(x_{it}\varepsilon_{it}) = 0,$$

from $n \times m$ panel data, treating the n intercepts $\alpha_1, \ldots, \alpha_n$ as unknown parameters to be estimated.

We will work under a set of three assumptions which we state next. These assumptions are standard in the literature. The first assumption contains moment requirements and mixing conditions.

Assumption 1.

(i) The variables x_{it} and ε_{it} have uniformly bounded moments of order $2r$ for some $r > 2$.

(ii) The variables $x_{it}\varepsilon_{it}$ have uniformly bounded moments of order $3r$.

(iii) The variables x_{it} and ε_{it} are independent across i.

(iv) The processes $\{(x_{it}, \varepsilon_{it})\}$ are stationary mixing and their mixing coefficients, a_i, satisfy

$$\sup_{1 \le i \le n} a_i(h) = O(h^{-s})$$

for some $s > {}^{4r}/_{r-2}$.[1]

The second assumption states conventional rank conditions on certain covariance matrices. Here and in the sequel we let $z_{it} \coloneqq x_{it} - \mathbb{E}(x_{it})$.

[1] We recall that

$$a_i(h) \coloneqq \sup_{1 \le t \le m} \sup_{A \in \mathcal{A}_{it}} \sup_{B \in \mathcal{B}_{it+h}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B)|,$$

for $\mathcal{A}_{it}$ and $\mathcal{B}_{it}$ the sigma algebras generated by the sequences $(x_{it}, \varepsilon_{it}), (x_{it-1}, \varepsilon_{it-1}), \ldots$ and $(x_{it}, \varepsilon_{it}), (x_{it+1}, \varepsilon_{it+1}), \ldots$, respectively.

Assumption 2.

(i) The covariance matrix $\Sigma_{n,m} \coloneqq {}^{1}\!/_{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} \mathbb{E}(z_{it} z_{it}')$ *is uniformly positive definite.*

(ii) The variance-covariance matrix of ${}^{1}\!/_{\sqrt{nm}} \sum_{i=1}^{n} \sum_{t=1}^{m} z_{it} \varepsilon_{it}$,

$$\Omega_{n,m} \coloneqq {}^{1}\!/_{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} \left(\mathbb{E}(z_{it} z_{it}' \varepsilon_{it}^2) + \sum_{\tau=1}^{m-1} {}^{(m-\tau)}\!/_{m} \, \mathbb{E}((z_{it} z_{it+\tau}' + z_{it+\tau} z_{it}') \, \varepsilon_{it} \varepsilon_{it+\tau}) \right),$$

is uniformly positive definite.

The third assumption states the asymptotic approximation under which we proceed.

Assumption 3. $n, m \to \infty$ *with* ${}^{n}\!/_{m} \to c \in [0, +\infty)$.

Because we allow for $\mathbb{E}(x_{it} \varepsilon_{it+\tau})$ and $\mathbb{E}(x_{it+\tau} \varepsilon_{it})$ to be non-zero when $\tau \neq 0$, both the within-group least-squares estimator and generalized method-of-moment estimators as in Arellano and Bond (1991) will be inconsistent, in general, under asymptotics where m is held fixed.[2] In fact, it would appear that β is not point identified in such a framework without further conditions. Reversely, asymptotics where n does not diverge are unsuitable for typical microeconometric applications.

The within-group estimator of β is

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

where $\bar{y}_i \coloneqq {}^{1}\!/_{m} \sum_{t=1}^{m} y_{it}$ and $\bar{x}_i \coloneqq {}^{1}\!/_{m} \sum_{t=1}^{m} x_{it}$. Under Assumptions 1-3 this estimator is consistent and asymptotically normally-distributed, but its limit distribution features an asymptotic bias term unless ${}^{n}\!/_{m} \to 0$.

In the following theorem, we let $\Upsilon \coloneqq \lim_{n,m \to \infty} \Upsilon_{n,m}$ for $\Upsilon_{n,m} \coloneqq \Sigma_{n,m}^{-1} \Omega_{n,m} \Sigma_{n,m}^{-1}$.

[2] In the conventional first-order autoregressive model, for example, $x_{it} = y_{it-1}$ and $\varepsilon_{it} \sim$ i.i.d.$(0, \sigma^2)$, and so $\mathbb{E}(x_{it} \varepsilon_{it+\tau}) = 0$ but $\mathbb{E}(x_{it+\tau} \varepsilon_{it}) = \beta^{\tau-1} \sigma^2 \neq 0$ for all $\tau > 0$. This, then, leads to the well-known Nickell (1981) bias in the within-group estimator. More generally, while moment-based strategies are available that can handle situations where $\mathbb{E}(x_{it+\tau} \varepsilon_{it}) \neq 0$, these approaches require that $\mathbb{E}(x_{it} \varepsilon_{it+\tau}) = 0$ at least for two known values of $\tau \leq m - t$ to yield valid moment conditions that can be exploited to construct an estimator; see Arellano (2003, Chapter 8).

Theorem 1. *Let Assumptions 1-3 hold. Then*

$$\sqrt{nm}\,(\hat{\beta} - \beta - \Sigma_{n,m}^{-1} b_{n,m}/m) \xrightarrow{L} N(0, \Upsilon),$$

where $b_{n,m} := -1/n \sum_{i=1}^{n} \sum_{\tau=1}^{m-1} (m-\tau)/m \left(\mathbb{E}(z_{it}\varepsilon_{it+\tau}) + \mathbb{E}(z_{it+\tau}\varepsilon_{it})\right).$

Theorem 1 generalizes results available for models with predetermined regressors generated through a specified process such as those derived in Hahn and Kuersteiner (2002) and Chudik, Pesaran and Yang (2018).

3 Moving block bootstrap

Our bootstrap scheme consists of applying the moving block bootstrap of Künsch (1989) in the time-series dimension of the data, jointly for all cross-sectional units. Moreover, for integers p and q with $m = p \times q$, we randomly select p blocks of q consecutive cross-sections from the original data; the blocks may overlap. Our bootstrap sample is then obtained on concatenating these p blocks. If we let $\varpi_1, \ldots \varpi_p$ be a random sample from the discrete uniform distribution on $\{0, \ldots, m-q\}$, the bootstrap time series $\{(y_{it}^*, x_{it}^*)\}$ is generated as

$$y_{i\,(p'-1)q+q'}^* := y_{i\,\varpi_{p'}+q'}, \qquad x_{i\,(p'-1)q+q'}^* := x_{i\,\varpi_{p'}+q'},$$

for $1 \leq p' \leq p$ and $1 \leq q' \leq q$. Here, the random variables $\varpi_1, \ldots, \varpi_p$ select starting points for the different blocks. When the block length, q, is set to one we recover the bootstrap as originally introduced by Efron (1979) although, in our setting, we will require q to grow with m.

The within-group estimator computed from the bootstrap sample so constructed equals

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

where $\bar{y}_i^* := 1/m \sum_{t=1}^{m} y_{it}^*$ and $\bar{x}_i^* := 1/m \sum_{t=1}^{m} x_{it}^*$.

The following theorem is our main result. In it, as usual, we let $\mathbb{P}^*$ denote a probability computed with respect to the bootstrap measure, that is, conditional on the original data.

Theorem 2. *Let Assumptions 1-3 hold and suppose that $q \to \infty$ with $q = o(\sqrt{m})$. Then*

$$\mathbb{P}\left(\sup_a \left|\mathbb{P}^*(\sqrt{nm}(\hat{\beta}^* - \hat{\beta}) \leq a) - \mathbb{P}(\sqrt{nm}(\hat{\beta} - \beta) \leq a)\right| > \epsilon\right) = o(1)$$

for any $\epsilon > 0$.

This result generalizes the findings in Gonçalves (2011), where Assumption 3 is replaced by the stronger requirement that $n/m \to 0$. Because the bias in $\sqrt{nm}(\hat{\beta} - \beta)$ is of the order $\sqrt{n/m}$, her rate condition ensures that the limit distribution of the within-group estimator does not feature a bias term.

Theorem 2 has a number of useful implications. A first is that inference based on the bootstrap remains valid in the presence of asymptotic bias. To see this, say we wish to perform inference on linear contrasts of the form $\theta \coloneqq c'\beta$, where c is a chosen vector of conformable dimension. For $\alpha \in (0, 1)$, let

$$\hat{Q}_\alpha \coloneqq \inf\{Q : \alpha \leq \mathbb{P}^*(\hat{\theta}^* - \hat{\theta} \leq Q)\},$$

with $\hat{\theta} \coloneqq c'\hat{\beta}$ and $\hat{\theta}^* \coloneqq c'\hat{\beta}^*$. By van der Vaart (2000, Lemma 23.3) we have that, under the conditions of Theorem 2,

$$\lim_{n,m\to\infty} \mathbb{P}(\hat{\theta} - \hat{Q}_\alpha \leq \theta) \to \alpha,$$

which allows the construction of confidence intervals and decision rules to conduct inference on θ.

Theorem 2 also implies that the median of the bootstrap distribution converges to the median of the limit distribution which, by Theorem 1, equals the asymptotic bias. Hence,

$$\check{\beta} \coloneqq \hat{\beta} - \text{median}^*(\hat{\beta}^* - \hat{\beta}).$$

is a bootstrap-based bias-corrected estimator of β.

The variance of the limit distribution in Theorem 1 can be estimated using conventional HAC methods or by means of resampling via the moving block bootstrap. To describe the latter way, consider

$$\hat{\Omega}^*_{n,m} \coloneqq 1/np \sum_{i=1}^{n} \sum_{p'=1}^{p} V_{\varpi_{p'}} V'_{\varpi_{p'}}, \qquad V_\varpi \coloneqq \left(1/\sqrt{q} \sum_{q'=1}^{q} (x_{i\varpi+q'} - \bar{x}^*_i)\, \hat{\epsilon}^*_{i\varpi+q'}\right),$$

where $\hat{\epsilon}^*_{it} := (y_{it} - \bar{y}^*_i) - (x_{it} - \bar{x}^*_i)\hat{\beta}^*$ are bootstrap residuals. This is an estimator of the (conditional) bootstrap variance

$$\hat{\Omega}_{n,m} := \text{var}^* \left({}^1\!/\!_{\sqrt{nm}} \sum_{i=1}^{n} \sum_{t=1}^{m} (x^*_{it} - \bar{x}^*_i)\, \hat{\varepsilon}^*_{it} \right),$$

where $\hat{\varepsilon}^*_{it} := (y^*_{it} - \bar{y}^*_i) - (x^*_{it} - \bar{x}^*_i)\hat{\beta}$. The latter, in turn, is an estimator of $\Omega_{n,m}$; it is well-known that it can be computed without resampling the data (see Künsch 1989). If we further introduce the shorthands

$$\hat{\Sigma}^*_{n,m} := {}^1\!/\!_{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x^*_{it} - \bar{x}^*_i)(x^*_{it} - \bar{x}^*_i)', \quad \hat{\Sigma}_{n,m} := {}^1\!/\!_{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)',$$

we can define

$$\hat{\Upsilon}^* := (\hat{\Sigma}^*_{n,m})^{-1} \hat{\Omega}^*_{n,m} (\hat{\Sigma}^*_{n,m})^{-1}, \qquad \hat{\Upsilon} := \hat{\Sigma}^{-1}_{n,m} \hat{\Omega}_{n,m} \hat{\Sigma}^{-1}_{n,m}.$$

By the same arguments as in Gonçalves (2011, Theorem B.1 and its proof) we can verify that $\hat{\Upsilon}^* - \Upsilon \xrightarrow{P^*} 0$ and $\hat{\Upsilon} - \Upsilon \xrightarrow{P} 0$.

Combined with our theorems, this yields two additional results. The first of these is that inference can be performed using the normal approximation to our bias-corrected estimator, using $\hat{\Upsilon}$. The second is that the reverse-percentile bootstrap can be applied to studentized quantities to perform inference in the same way as before. In the context of linear contrasts of the form $\theta = c'\beta$, for example, if we let $\hat{\sigma}^{*2} := c'\,\hat{\Upsilon}^* c$ and $\hat{\sigma}^2 := c'\,\hat{\Upsilon} c$, and redefine

$$\hat{Q}_\alpha := \inf\{Q : \alpha \leq \mathbb{P}^*({}^{(\hat{\theta}^* - \hat{\theta})}\!/\!_{\hat{\sigma}^*} \leq Q)\},$$

then $\mathbb{P}(\hat{\theta} - \hat{\sigma}\,\hat{Q}_\alpha \leq \theta) \to \alpha$ under the assumptions of Theorem 2. This implies that decision rules for null hypotheses on θ involving critical values obtained from the bootstrap distribution of ${}^{(\hat{\theta}^* - \hat{\theta})}\!/\!_{\hat{\sigma}^*}$ deliver asymptotic size control without the need to correct for bias.

Table 1: Quantile estimates

p	q	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
$(n,m) = (200,200)$										
40	5	0.0726	0.1497	0.2327	0.3214	0.4161	0.5168	0.6239	0.7384	0.8617
20	10	0.0880	0.1710	0.2581	0.3499	0.4466	0.5481	0.6548	0.7663	0.8822
10	20	0.0963	0.1780	0.2632	0.3536	0.4497	0.5520	0.6601	0.7735	0.8903
$(n,m) = (500,500)$										
50	10	0.0859	0.1727	0.2631	0.3572	0.4547	0.5558	0.6605	0.7689	0.8816
25	20	0.0940	0.1825	0.2737	0.3682	0.4664	0.5677	0.6724	0.7802	0.8903
20	25	0.0963	0.1843	0.2750	0.3693	0.4672	0.5685	0.6734	0.7815	0.8917

4 Simulations

To support Theorem 2 we report some numerical results. Consider a stationary first-order autoregression with standard-normal innovations, that is

$$y_{it} = x_{it}\beta + \varepsilon_{it}$$

with $\varepsilon_{it} \sim$ i.i.d. $N(0,1)$ for all $1 \leq t \leq m$, $x_{i1} \sim$ i.i.d. $N(0, 1/(1-\beta^2))$, and $x_{it} = y_{it-1}$ for $2 \leq t \leq m$. The within-group estimator is invariant to the distribution of the fixed effects, so setting $\alpha_i = 0$ for all $1 \leq i \leq n$ is without loss of generality. It is well-known that, here,

$$\sqrt{nm}(\hat{\beta} - \beta) = -\sqrt{n/m}(1+\beta) + \sqrt{1/1-\beta^2}\, N(0,1) + o_P(1)$$

as $n, m \to \infty$. In Table 1 we report the average (over 10,000 Monte Carlo replications) of the bootstrap distribution of $\sqrt{nm}(\hat{\beta}^* - \hat{\beta})$ (as computed using 1,999 bootstrap replications) evaluated at the percentiles of the normal distribution in the above limit statement (which is not centered at zero). This effectively yields a QQ-plot, although we provide it in tabular form. We do this for two different sample sizes (combinations of n and m) and for various choices of the number and length of the bootstrap blocks (p and q). The table concerns simulations under a data generating process where $\beta = 0$. It shows that the quantiles of the

bootstrap distribution are, on average, close to those of the limit distribution. The same phenomenon was equally observed in a wider set of simulation designs and also carries over to the bootstrap distribution of the studentized estimator using the variance estimators given above.

Appendix

For notational simplicity, and without loss of generality, we take x_{it} to be univariate throughout the appendix.

Proof of Theorem 1. From the definition of $\hat{\beta}$,

$$\sqrt{nm}(\hat{\beta}-\beta) = \hat{\Sigma}_{n,m}^{-1}\left(1/\sqrt{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}-\bar{x}_i)\,\varepsilon_{it}\right),$$

where, recall, $\hat{\Sigma}_{n,m} = 1/nm\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}-\bar{x}_i)^2$.

We first show that $|\hat{\Sigma}_{n,m}-\Sigma_{n,m}| = o_P(1)$. By using the triangle inequality we can write

$$|\hat{\Sigma}_{n,m}-\Sigma_{n,m}| \leq |\hat{\Sigma}_{n,m}-\check{\Sigma}_{n,m}| + |\check{\Sigma}_{n,m}-\Sigma_{n,m}|, \tag{A.1}$$

for

$$\check{\Sigma}_{n,m} \coloneqq 1/nm\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}-\mathbb{E}(x_i))^2 = 1/nm\sum_{i=1}^{n}\sum_{t=1}^{m}z_{it}^2.$$

We proceed by showing that each of the terms on the right-hand side converges to zero in probability.

For the first term in (A.1), on working out the square and re-arranging we observe that

$$\hat{\Sigma}_{n,m}-\check{\Sigma}_{n,m} = -1/n\sum_{i=1}^{n}\bar{z}_i^2,$$

with $\bar{z}_i \coloneqq 1/m\sum_{t=1}^{m}z_{it}$. Therefore, for any $\epsilon>0$,

$$\mathbb{P}(|\hat{\Sigma}_{n,m}-\check{\Sigma}_{n,m}|>\epsilon) \leq \frac{\mathbb{E}(|\hat{\Sigma}_{n,m}-\check{\Sigma}_{n,m}|)}{\epsilon} \leq \frac{1/n\sum_{i=1}^{n}\mathbb{E}(\bar{z}_i^2)}{\epsilon} = O(m^{-1})$$

by an application of Markov's inequality in the first step, the Cauchy-Schwarz inequality in the second step, and the fact that $\sup_{1\leq i\leq n}\mathbb{E}(\bar{z}_i^2) = O(m^{-1})$. The latter follows from

Hansen (1991, Corollary 3). Moreover, by Assumption 1 the time-series processes $\{z_{it}\}$ are strong mixing with mixing coefficients a_i for which $\sup_{1\leq i\leq n} a_i(h) =: \bar{a}(h)$ satisfies the summability condition $\sum_{h=1}^{+\infty} \bar{a}(h)^{1/2-1/r} < +\infty$ for some $r > 2$ for which $\sup_{1\leq i\leq n} \mathbb{E}(|z_{it}|^r)$ is bounded. Therefore, by an application of Hansen's (1991) corollary, we can deduce that

$$\sup_{1\leq i\leq n} \mathbb{E}\left(\left(\sum_{t=1}^{m} z_{it}\right)^2\right) \lesssim \sum_{t=1}^{m} \sup_{1\leq i\leq n} \mathbb{E}(|z_{it}|^r)^{2/r} = O(m),$$

where, here and later, we use $A \lesssim B$ to indicate that there exists a finite constant C such that $A \leq C\,B$. From this it then follows that $\sup_{1\leq i\leq n} \mathbb{E}(\bar{z}_i^2) = O(m^{-1})$, as claimed. This handles the first term in (A.1).

For the second term in (A.1), we have

$$\check{\Sigma}_{n,m} - \Sigma_{n,m} = {}^1\!/_{nm} \sum_{i=1}^{n}\sum_{t=1}^{m} (z_{it}^2 - \mathbb{E}(z_{it}^2)).$$

Therefore, for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\check{\Sigma}_{n,m} - \Sigma_{n,m}\right| > \epsilon\right) \leq \frac{{}^1\!/_{n^2} \sum_{i=1}^{n} \mathbb{E}\left(\left({}^1\!/_{m} \sum_{t=1}^{m} (z_{it}^2 - \mathbb{E}(z_{it}^2))\right)^2\right)}{\epsilon^2} = O(n^{-1}m^{-1}),$$

by an application of Chebychev's inequality and the fact that, by another application of Hansen (1991, Corollary 3), $\sup_{1\leq i\leq n} \mathbb{E}\left(\left(\sum_{t=1}^{m}(z_{it}^2 - \mathbb{E}(z_{it}^2))\right)^2\right) = O(m)$ follows in the same way as before. With both right-hand side terms of (A.1) $o_P(1)$ we have thus shown the desired result that $|\hat{\Sigma}_{n,m} - \Sigma_{n,m}| = o_P(1)$.

We next derive the limit distribution of the within-group least-squares normal equations,

$${}^1\!/_{\sqrt{nm}} \sum_{i=1}^{n}\sum_{t=1}^{m} (x_{it} - \bar{x}_i)\,\varepsilon_{it}.$$

Adding and subtracting $\mathbb{E}(x_{it})(\varepsilon_{it} - \bar{\varepsilon}_i)$ to the summands and re-arranging allows us to write this as

$${}^1\!/_{\sqrt{nm}} \sum_{i=1}^{n}\sum_{t=1}^{m} z_{it}\varepsilon_{it} - \left(\sqrt{{}^n\!/_m}\right) {}^1\!/_n \sum_{i=1}^{n} \left({}^1\!/_{\sqrt{m}} \sum_{t=1}^{m} z_{it}\right)\left({}^1\!/_{\sqrt{m}} \sum_{t=1}^{m} \varepsilon_{it}\right), \tag{A.2}$$

and we analyse each of the terms in turn.

The first term in (A.2) is a scaled sample average of zero-mean random variables with variance covariance matrix $\Omega_{n,m}$ that satisfies the conditions of White (2000, Theorem 5.20)

(with the required order of mixing following from Assumption 1 together with Theorem 3.49 of White (2000)). Therefore, it converges in law to a normal random variable with mean zero and variance $\lim_{n,m\to\infty} \Omega_{n,m}$.

The second term in (A.2), on the other hand, will generate bias. To see this we introduce

$$\chi_i := \left(1/\sqrt{m} \sum_{t=1}^{m} z_{it}\right)\left(1/\sqrt{m} \sum_{t=1}^{m} \varepsilon_{it}\right) = 1/m \sum_{t_1=1}^{m}\sum_{t_2=1}^{m} z_{it_1}\varepsilon_{it_2}.$$

Recalling that $\mathbb{E}(z_{it}\varepsilon_{it}) = 0$ we have

$$\mathbb{E}(\chi_i) = \sum_{\tau=1}^{m-1} (m-\tau/m)\left(\mathbb{E}(z_{it}\varepsilon_{it+\tau}) + \mathbb{E}(z_{it+\tau}\varepsilon_{it})\right),$$

and so $-1/n \sum_{i=1}^{n} \mathbb{E}(\chi_i) = b_{n,m}$. We now show that

$$1/n \sum_{i=1}^{n} (\chi_i - \mathbb{E}(\chi_i)) = o_P(1). \tag{A.3}$$

For any $\epsilon > 0$, by Chebychev's inequality and independence of the data in the cross-section,

$$\mathbb{P}\left(\left|1/n \sum_{i=1}^{n} (\chi_i - \mathbb{E}(\chi_i))\right| > \epsilon\right) \leq \frac{1}{n} \frac{1/n \sum_{i=1}^{n} \mathbb{E}((\chi_i - \mathbb{E}(\chi_i))^2)}{\epsilon^2} \leq \frac{1}{n} \frac{\sup_{1\leq i\leq n} \operatorname{var}(\chi_i)}{\epsilon^2},$$

so that it suffices to show that $\sup_{1\leq i\leq n} \operatorname{var}(\chi_i) = O(1)$. Note that the variance of $m\chi_i$ equals

$$\mathbb{E}\left(\left(\sum_{t_1=1}^{m} z_{it_1}\right)^2 \left(\sum_{t_2=1}^{m} \varepsilon_{it_2}\right)^2\right) - \mathbb{E}\left(\sum_{t_1=1}^{m}\sum_{t_2=1}^{m} z_{it_1}\varepsilon_{it_2}\right) \mathbb{E}\left(\sum_{t_1=1}^{m}\sum_{t_2=1}^{m} z_{it_1}\varepsilon_{it_2}\right). \tag{A.4}$$

For the first term in (A.4), the Cauchy-Schwarz inequality yields

$$\mathbb{E}\left(\left(\sum_{t_1=1}^{m} z_{it_1}\right)^2 \left(\sum_{t_2=1}^{m} \varepsilon_{it_2}\right)^2\right) \leq \sqrt{\mathbb{E}\left(\left(\sum_{t_1=1}^{m} z_{it_1}\right)^4\right)} \sqrt{\mathbb{E}\left(\left(\sum_{t_2=1}^{m} \varepsilon_{it_2}\right)^4\right)},$$

while, by an application Hansen (1991, Corollary 3),

$$\sup_{1\leq i\leq n} \mathbb{E}\left(\left(\sum_{t=1}^{m} z_{it}\right)^4\right)^{1/4} \lesssim \sqrt{\sum_{t=1}^{m} \sup_{1\leq i\leq n} \mathbb{E}(|z_{it}|^r)^2} = O(m^{1/2}),$$

$$\sup_{1\leq i\leq n} \mathbb{E}\left(\left(\sum_{t=1}^{m} \varepsilon_{it}\right)^4\right)^{1/4} \lesssim \sqrt{\sum_{t=1}^{m} \sup_{1\leq i\leq n} \mathbb{E}(|\varepsilon_{it}|^r)^2} = O(m^{1/2}),$$

so that $\sup_{1\leq i\leq n}\mathbb{E}\left(\left(\sum_{t_1=1}^{m} z_{it_1}\right)^2\left(\sum_{t_2=1}^{m}\varepsilon_{it_2}\right)^2\right)=O(m^2)$. For the second term in (A.4), we have

$$\sup_{1\leq i\leq n}\mathbb{E}\left(\sum_{t_1=1}^{m}\sum_{t_2=1}^{m} z_{it_1}\varepsilon_{it_2}\right)\leq\sqrt{\sup_{1\leq i\leq n}\mathbb{E}\left(\left|\sum_{t=1}^{m} z_{it}\right|^2\right)}\sqrt{\sup_{1\leq i\leq n}\mathbb{E}\left(\left|\sum_{t=1}^{m}\varepsilon_{it}\right|^2\right)}=O(m),$$

with the order of magnitude already having been established above. It thus follows that the term in (A.4) is $O(m^2)$ uniformly, and so that $\sup_{1\leq i\leq n}\operatorname{var}(\chi_i)=O(1)$. Therefore (A.3) holds.

Combined all results obtained reveals that

$$\sqrt{nm}(\hat{\beta}-\beta)={}^{1}\!/_{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}\Sigma_{n,m}^{-1}z_{it}\varepsilon_{it}+\left(\sqrt{{}^{n}\!/_{m}}\right)\Sigma_{n,m}^{-1}b_{n,m}+o_P(1),$$

from which the theorem follows readily. □

Proof of Theorem 2 From the definition of $\hat{\beta}^*$,

$$\sqrt{nm}(\hat{\beta}^*-\hat{\beta})=(\hat{\Sigma}_{n,m}^*)^{-1}\left({}^{1}\!/_{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}^*-\bar{x}_i^*)\,\hat{\varepsilon}_{it}^*\right),\tag{A.5}$$

where

$$\hat{\Sigma}_{n,m}^*={}^{1}\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}^*-\bar{x}_i^*)^2$$

and $\hat{\varepsilon}_{i\,(p'-1)q+q'}^*\coloneqq\hat{\varepsilon}_{i\,\varpi_{p'}+q'}$, for $1\leq p'\leq p$ and $1\leq q'\leq q$, are the resampled residuals associated with the within-group estimator computed from the original data; moreover, $\hat{\varepsilon}_{it}\coloneqq(y_{it}-\bar{y}_i)-(x_{it}-\bar{x}_i)\hat{\beta}$. As in the proof of Theorem 1 we will proceed in multiple steps.

We first show that $|\hat{\Sigma}_{n,m}^*-\hat{\Sigma}_{n,m}|=o_{P^*}(1)$, that is, that

$$\mathbb{P}(\mathbb{P}^*(|\hat{\Sigma}_{n,m}^*-\hat{\Sigma}_{n,m}|>\epsilon^*)>\epsilon)=o(1)$$

for all $\epsilon^*>0$ and $\epsilon>0$. Because

$$\hat{\Sigma}_{n,m}^*-\hat{\Sigma}_{n,m}={}^{1}\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}({x_{it}^*}^2-x_{it}^2)-{}^{1}\!/_{n}\sum_{i=1}^{n}({\bar{x}_i^*}^2-{\bar{x}_i}^2),\tag{A.6}$$

it suffices to show that each of the terms on the right-hand side is $o_{P^*}(1)$; we handle them in turn next.

For the first term in (A.6) we first add and subtract $\mathbb{E}^*({x_{it}^*}^2)$ to the summand in a first step and then add and subtract $\mathbb{E}({x_{it}}^2)$ in a second step. We then obtain the decomposition

$$\begin{aligned}{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}({x_{it}^*}^2-x_{it}^2) &= {}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}\ ({x_{it}}^2)\right)+{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\mathbb{E}(x_{it}^2)-\mathbb{E}^*({x_{it}^*}^2)\right)\\ &\quad-{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2\ -\mathbb{E}^*({x_{it}^*}^2)\right)\end{aligned}$$

on re-arranging terms. For the first of these three terms, by iterated application of Markov's inequality,

$$\mathbb{P}\left(\mathbb{P}^*\left(\left|{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}(x_{it}^2)\right)\right|>\epsilon^*\right)>\epsilon\right)\lesssim\mathbb{E}\left(\mathbb{E}^*\left(\left|{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}(x_{it}^2)\right)\right|\right)\right)$$

for all ϵ^* and $\epsilon>0$. Furthermore,

$$\mathbb{E}^*\left(\left|{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}(x_{it}^2)\right)\right|\right)\leq{}^1\!/_{n}\sum_{i=1}^{n}\mathbb{E}^*\left(\left|{}^1\!/_{m}\sum_{t=1}^{m}{x_{it}^*}^2-{}^1\!/_{m}\sum_{t=1}^{m}\mathbb{E}(x_{it}^2)\right|\right).$$

Now, by definition of the moving block bootstrap scheme in the first step and the fact that $\varpi_1,\ldots,\varpi_p$ are i.i.d. from the uniform distribution on the set $\{0,1,\ldots,m-q\}$—so that the probability that $\varpi_{p'}=q'$ is equal to ${}^1\!/_{m-q+1}$ for all $1\leq p'\leq p$ and any $0\leq q'\leq m-q$, independent of p' and q'—together with stationarity of the data in the second step, we have

$$\begin{aligned}\mathbb{E}^*\left(\left|\left({}^1\!/_{m}\sum_{t=1}^{m}{x_{it}^*}^2-{}^1\!/_{m}\sum_{t=1}^{m}\mathbb{E}(x_{it}^2)\right)\right|\right)&=\mathbb{E}^*\left(\left|\left({}^1\!/_{pq}\sum_{p'=1}^{p}\sum_{q'=1}^{q}x^2_{i\varpi_{p'}+q'}-{}^1\!/_{m}\sum_{t=1}^{m}\mathbb{E}(x_{it}^2)\right)\right|\right)\\ &={}^1\!/_{(m-q+1)}\sum_{t=0}^{m-q}\left(\left|\left({}^1\!/_{q}\sum_{q'=1}^{q}(x^2_{it+q'}-\mathbb{E}(x^2_{it+q'}))\right)\right|\right).\end{aligned}$$

Hence, $\mathbb{E}\left(\mathbb{E}^*\left(\left|{}^1\!/_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}(x_{it}^2)\right)\right|\right)\right)$ is bounded from above by

$${}^1\!/_{nq(m-q+1)}\sum_{i=1}^{n}\sum_{t=0}^{m-q}\mathbb{E}\left(\left|\left(\sum_{q'=1}^{q}(x^2_{it+q'}-\mathbb{E}(x^2_{it+q'}))\right)\right|\right)=O(q^{-1/2})=o(1),$$

where the order of magnitude follows from the, by now familiar, arguments from Hansen (1991). This handles the first of the three right-hand side terms in the decomposition at

the start of this paragraph. For the second term, by the triangle inequality in a first step and iterating the Cauchy-Schwarz inequality in the second step

$$\mathbb{E}\left(\left|{}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\mathbb{E}(x_{it}^2)-\mathbb{E}^*({x_{it}^*}^2)\right)\right|\right) \leq {}^1\!/\!_{nm}\sum_{i=1}^{n}\mathbb{E}\left(\left|\mathbb{E}^*\left(\sum_{t=1}^{m}\left({x_{it}^*}^2-\mathbb{E}(x_{it}^2)\right)\right)\right|\right)$$
$$\leq {}^1\!/\!_{nm}\sum_{i=1}^{n}\sqrt{\mathbb{E}\left(\mathbb{E}^*\left(\left|\sum_{t=1}^{m}({x_{it}^*}^2-\mathbb{E}(x_{it}^2))\right|^2\right)\right)}.$$

Using Gonçalves and White (2005, Lemma A.1) we have

$$\sup_{1\leq i\leq n}\mathbb{E}\left(\mathbb{E}^*\left(\left|\sum_{t=1}^{m}({x_{it}^*}^2-\mathbb{E}(x_{it}^2))\right|^2\right)\right)=O(m)+O(q^2).$$

Therefore, by Markov's inequality,

$$\mathbb{P}\left(\left|{}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\mathbb{E}(x_{it}^2)-\mathbb{E}^*({x_{it}^*}^2)\right)\right|>\epsilon\right)=O({}^{\sqrt{m}}\!/\!_{m}+{}^{q}\!/\!_{m})=o(1)$$

for all $\epsilon>0$. This handles the second term. Finally, for the third term in the decomposition, ${}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}^*({x_{it}^*}^2)\right)$, we add and subtract $\mathbb{E}(x_{it}^2)$ to each of the summands to see that it equals ${}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)+{}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\mathbb{E}(x_{it}^2)-\mathbb{E}^*({x_{it}^*}^2)\right)$. The second of these terms has already been shown to be $o_{P^*}(1)$. Thus,

$${}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}^*({x_{it}^*}^2)\right)={}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)+o_{P^*}(1).$$

Here,

$$\mathbb{P}\left(\left|{}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)>\epsilon\right|\right)\leq\frac{\mathbb{E}\left(\left|{}^1\!/\!_{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)\right|\right)}{\epsilon}$$
$$\leq\frac{{}^1\!/\!_{nm}\sum_{i=1}^{n}\mathbb{E}\left(\left|\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)\right|\right)}{\epsilon}$$
$$\leq\frac{{}^1\!/\!_{m}\sqrt{\sup_{1\leq i\leq n}\mathbb{E}\left(\left|\sum_{t=1}^{m}\left(x_{it}^2-\mathbb{E}(x_{it}^2)\right)\right|^2\right)}}{\epsilon}$$

is $o(m^{-1/2})$ for any $\epsilon>0$ by Hansen (1991, Corollary 3). We may then conclude that the first right-hand side term in (A.6) is $o_{P^*}(1)$.

For the second term in (A.6), we can add and subtract terms to write

$$1/n \sum_{i=1}^{n} (\bar{x}_i^{*2} - \bar{x}_i^{\,2}) = 1/n \sum_{i=1}^{n} (\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2 + 2/n \sum_{i=1}^{n} (\bar{x}_i^* - \bar{x}_i)\, \mathbb{E}(\bar{x}_i) - 1/n \sum_{i=1}^{n} \bar{z}_i^2,$$

and we have already shown that $1/n \sum_{i=1}^{n} \bar{z}_i^2 = o_P(1)$ in the proof of Theorem 1. We next look at the remaining two terms on the right-hand side. For the first one, because we have

$$\bar{x}_i^* - \mathbb{E}(\bar{x}_i) = 1/m \sum_{t=1}^{m} (x_{it}^* - \mathbb{E}(x_{it})) = 1/pq \sum_{p'=1}^{p} \sum_{q'=1}^{q} (x_{i\varpi_{p'}+q'} - \mathbb{E}(x_{it})) = 1/pq \sum_{p'=1}^{p} \sum_{q'=1}^{q} z_{i\varpi_{p'}+q'},$$

we obtain

$$(\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2 = 1/p^2q^2 \sum_{p_1'=1}^{p} \sum_{q_1'=1}^{q} \sum_{p_2'=1}^{p} \sum_{q_2'=1}^{q} z_{i\varpi_{p_1'}+q_1'}\, z_{i\varpi_{p_2'}+q_2'}.$$

Its expectation conditional on the data factors as the sum of two terms; the first corresponds to contributions where $p_1' = p_2'$ and the second collects the terms where $p_1' \neq p_2'$. Moreover,

$$\mathbb{E}^*\left((\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2\right) = 1/pq^2\, 1/(m-q+1) \sum_{t=0}^{m-q} \left(\sum_{q'=1}^{q} z_{it+q'} \right)^2 + \left(1/q(m-q+1) \sum_{t=0}^{m-q} \sum_{q'=1}^{q} z_{it+q'} \right)^2.$$

Here, the last term involves the bootstrap mean $\mathbb{E}^*(\bar{z}_i^*) = 1/q(m-q+1) \sum_{t=0}^{m-q} \sum_{q'=1}^{q} z_{it+q'}$. By the Cauchy-Schwarz inequality, $\mathbb{E}(|\mathbb{E}^*(\bar{z}_i^*)|^2) \leq \mathbb{E}(\mathbb{E}^*(|\bar{z}_i^*|^2))$. Therefore, taking expectations yields

$$\mathbb{E}\left(\mathbb{E}^*\left((\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2\right)\right) \leq 1/pq^2\, 1/(m-q+1) \sum_{t=0}^{m-q} \mathbb{E}\left(\left| \sum_{q'=1}^{q} z_{it+q'} \right|^2 \right) + \mathbb{E}(\mathbb{E}^*(|\bar{z}_i^*|^2)).$$

By an application of Hansen (1991, Corollary 3) and Gonçalves and White (2005, Lemma A.1) to the first and the second term, respectively, and recalling that $m = pq$, we arrive at

$$\sup_{1 \leq i \leq n} \mathbb{E}\left(\mathbb{E}^*\left((\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2\right)\right) = O(1/m) + O(m/m^2 + q^2/m^2) = o(1),$$

from which $1/n \sum_{i=1}^{n} (\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2 = o_{P^*}(1)$ follows. Next,

$$1/n \sum_{i=1}^{n} (\bar{x}_i^* - \bar{x}_i)\, \mathbb{E}(\bar{x}_i) \leq \sqrt{1/n \sum_{i=1}^{n} (\bar{x}_i^* - \bar{x}_i)^2} \sqrt{1/n \sum_{i=1}^{n} \mathbb{E}(\bar{x}_i)^2}.$$

Here, the second right-hand side term is $O(1)$ uniformly in $1 \leq i \leq n$ because x_{it} has uniformly-bounded moments of sufficient order while

$$1/n \sum_{i=1}^{n} (\bar{x}_i^* - \bar{x}_i)^2 \leq 2/n \sum_{i=1}^{n} (\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2 + 2/n \sum_{i=1}^{n} (\bar{x}_i - \mathbb{E}(\bar{x}_i))^2 = o_{P^*}(1)$$

uniformly in $1 \leq i \leq n$ because both terms on the right have already been shown to be $o_{P^*}(1)$ uniformly in $1 \leq i \leq n$. Through (A.6) we have thus shown that

$$|\hat{\Sigma}_{n,m}^* - \hat{\Sigma}_{n,m}| = o_{P^*}(1)$$

holds.

We now turn to the numerator in (A.5),

$$1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x_{it}^* - \bar{x}_i^*)\, \hat{\varepsilon}_{it}^*.$$

First, by the first-order condition of the within-group estimator it holds that

$$1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x_{it} - \bar{x}_i)\, \hat{\varepsilon}_{it} = 0,$$

where we recall that $1/m \sum_{t=1}^{m} \hat{\varepsilon}_{it} = 0$ for all $1 \leq i \leq n$ by definition of the within-group estimator. Therefore,

$$1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x_{it}^* - \bar{x}_i^*) \hat{\varepsilon}_{it}^* = 1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} \left((x_{it}^* - \bar{x}_i^*)\, \hat{\varepsilon}_{it}^* - (x_{it} - \bar{x}_i)\, \hat{\varepsilon}_{it} \right).$$

Next, $\hat{\varepsilon}_{it} = (y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)\hat{\beta} = -(x_{it} - \bar{x}_i)(\hat{\beta} - \beta) + (\varepsilon_{it} - \bar{\varepsilon}_i)$ by definition of the within-group residuals. Furthermore, by definition of the moving block bootstrap, equally,

$$\hat{\varepsilon}_{it}^* = -(x_{it}^* - \bar{x}_i^*)(\hat{\beta} - \beta) + (\varepsilon_{it}^* - \bar{\varepsilon}_i^*),$$

where $\varepsilon_{i\,(p'-1)q+q'}^* \coloneqq \varepsilon_{i\,\varpi_{p'}+q'}$, for $1 \leq p' \leq p$ and $1 \leq q' \leq q$. Substituting these expressions into the summands above yields

$$1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} (x_{it}^* - \bar{x}_i^*)\, \hat{\varepsilon}_{it}^* = 1/\sqrt{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} \left((x_{it}^* - \bar{x}_i^*)\, \varepsilon_{it}^* - (x_{it} - \bar{x}_i)\, \varepsilon_{it} \right) + o_{P^*}(1), \quad \text{(A.7)}$$

where the $o_{P^*}(1)$ term is equal to

$$-\left(1/nm\sum_{i=1}^{n}\sum_{t=1}^{m}(x_{it}^{*}-\bar{x}_{i}^{*})^{2}-(x_{it}-\bar{x}_{i})^{2}\right)\sqrt{nm}(\hat{\beta}-\beta);$$

the term in brackets is equal to $\hat{\Sigma}_{n,m}^{*}-\hat{\Sigma}_{n,m}$, which had already been shown to be $o_{P^*}(1)$, and $\sqrt{nm}(\hat{\beta}-\beta)=O_P(1)$ from Theorem 1. It thus remains only to analyse the leading term in (A.7). Re-centering both x_{it}^{*} and x_{it} around $\mathbb{E}(x_i)$ and re-arranging terms allows to write this leading term as

$$1/\sqrt{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}(z_{it}^{*}\varepsilon_{it}^{*}-z_{it}\varepsilon_{it})-(\sqrt{n/m})\,1/n\sum_{i=1}^{n}(\chi_{i}^{*}-\chi_{i}), \tag{A.8}$$

where χ_i was defined in the proof of Theorem 1 and

$$\chi_{i}^{*}:=\left(1/\sqrt{m}\sum_{t=1}^{m}z_{it}^{*}\right)\left(1/\sqrt{m}\sum_{t=1}^{m}\varepsilon_{it}^{*}\right)=1/m\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}z_{it_1}^{*}\varepsilon_{it_2}^{*}.$$

is its natural bootstrap counterpart. We again proceed term by term.

First, from Gonçalves and White (2002, Theorem 2.2), the first term on the right-hand side of (A.8) satisfies

$$1/\sqrt{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\Omega_{n,m}^{-1/2}\,(z_{it}^{*}\varepsilon_{it}^{*}-z_{it}\varepsilon_{it})\xrightarrow{L^*}N(0,I),$$

where $\xrightarrow{L^*}$ means convergence in law, conditional on the data. The remaining part in (A.8) will contribute bias. Because $1/\sqrt{m}\sum_{t=1}^{m}z_{it}^{*}=1/\sqrt{pq}\sum_{p'=1}^{p}\sum_{q'=1}^{q}z_{i\varpi_{p'}+q'}$ and, equally, $1/\sqrt{m}\sum_{t=1}^{m}\varepsilon_{it}^{*}=1/\sqrt{pq}\sum_{p'=1}^{p}\sum_{q'=1}^{q}\varepsilon_{i\varpi_{p'}+q'}$ we have

$$\chi_{i}^{*}=1/pq\sum_{p_1'=1}^{p}\sum_{q_1'=1}^{q}\sum_{p_2'=1}^{p}\sum_{q_2'=1}^{q}z_{i\varpi_{p_1'}+q_1'}\varepsilon_{i\varpi_{p_2'}+q_2'}.$$

Note that

$$\chi_{i}^{*}=1/pq\sum_{p'=1}^{p}\sum_{q_1'=1}^{q}\sum_{q_2'=1}^{q}z_{i\varpi_{p'}+q_1'}\varepsilon_{i\varpi_{p'}+q_2'}+1/pq\sum_{p_1'=1}^{p}\sum_{q_1'=1}^{q}\sum_{p_2'\neq p_1'}\sum_{q_2'=1}^{q}z_{i\varpi_{p_1'}+q_1'}\varepsilon_{i\varpi_{p_2'}+q_2'} \tag{A.9}$$

holds.

The first term in (A.9) is close to a bootstrap-based covariance estimator in the sense of Götze and Künsch (1996). Moreover,

$$1/pq \sum_{p'=1}^{p} \sum_{q_1'=1}^{q} \sum_{q_2'=1}^{q} z_{i\varpi_{p'}+q_1'} \varepsilon_{i\varpi_{p'}+q_2'} = \hat{\gamma}_i^* + q\, \bar{z}_i^* \bar{\varepsilon}_i^*,$$

where

$$\hat{\gamma}_i^* := 1/p \sum_{p'=1}^{p} \left(1/\sqrt{q} \sum_{q_1'=1}^{q} (z_{i\varpi_{p'}+q_1'} - \bar{z}_i^*) \right) \left(1/\sqrt{q} \sum_{q_2'=1}^{q} (\varepsilon_{i\varpi_{p'}+q_2'} - \bar{\varepsilon}_i^*) \right).$$

We will first show that

$$\left| 1/n \sum_{i=1}^{n} (\hat{\gamma}_i^* - \mathbb{E}(\chi_i)) \right| = o_{P^*}(1). \tag{A.10}$$

To do so we first use arguments similar to those used in Gonçalves and White (2004, Proof of Lemma B.1) to show that

$$\sup_{1 \le i \le n} |\hat{\gamma}_i^* - \hat{\gamma}_i| = o_{P^*}(1),$$

for $\hat{\gamma}_i := m \operatorname{cov}^*(\bar{z}_i^*, \bar{\varepsilon}_i^*)$, an estimator of the long-run covariance. We then use Gonçalves and White (2002, Corollary 2.1) to claim that $\sup_{1 \le i \le n} |\hat{\gamma}_i - \mathbb{E}(\chi_i)| = o_P(1)$, from which the desired result will follow.

Let

$$\tilde{\gamma}_i := 1/p \sum_{p'=1}^{p} \left(1/\sqrt{q} \sum_{q_1'=1}^{q} (z_{i\varpi_{p'}+q_1'} - \mathbb{E}^*(\bar{z}_i^*)) \right) \left(1/\sqrt{q} \sum_{q_2'=1}^{q} (\varepsilon_{i\varpi_{p'}+q_2'} - \mathbb{E}^*(\bar{\varepsilon}_i^*)) \right).$$

With $X_{i\varpi} := 1/\sqrt{q} \sum_{q'=1}^{q} (z_{i\varpi+q'} - \mathbb{E}^*(\bar{z}_i^*))$ and $Y_{i\varpi} := 1/\sqrt{q} \sum_{q'=1}^{q} (\varepsilon_{i\varpi+q'} - \mathbb{E}^*(\bar{\varepsilon}_i^*))$ we can write

$$\tilde{\gamma}_i = 1/p \sum_{p'=1}^{p} X_{i\varpi_{p'}} Y_{i\varpi_{p'}}, \qquad \hat{\gamma}_i = \mathbb{E}^*(X_{i\varpi} Y_{i\varpi}).$$

For any $1 < a \le 2$, we have

$$\mathbb{E}^*(|\tilde{\gamma}_i - \hat{\gamma}_i|^a) = 1/p^a\, \mathbb{E}^* \left(\left| \sum_{p'=1}^{p} (X_{i\varpi_{p'}} Y_{i\varpi_{p'}} - \mathbb{E}^*(X_{i\varpi_{p'}} Y_{i\varpi_{p'}})) \right|^a \right).$$

Here the summands inside the expectation on the right-hand side are independent and identically distributed zero-mean random variables, conditional on the data, and so, by an

application of Burkholder's inequality,

$$\mathbb{E}^*(|\tilde{\gamma}_i - \hat{\gamma}_i|^a) \lesssim {}^1\!/\!_{p^a}\, \mathbb{E}^* \left(\left| \sum_{p'=1}^{p} (X_{i\varpi_{p'}} Y_{i\varpi_{p'}} - \mathbb{E}^*(X_{i\varpi_{p'}} Y_{i\varpi_{p'}}))^2 \right|^{a/2} \right).$$

Next, exploiting the fact that the $\varpi_1, \ldots, \varpi_p$ are independent and identically distributed, together with well-known inequalities on the expectation of powers of a sum of random variables (as given in, e.g., von Bahr and Esseen 1965) we may bound the right-hand side by

$${}^1\!/\!_{p^a}\, \mathbb{E}^* \left(\sum_{p'=1}^{p} \left| (X_{i\varpi_{p'}} Y_{i\varpi_{p'}} - \mathbb{E}^*(X_{i\varpi_{p'}} Y_{i\varpi_{p'}}))^a \right| \right) = {}^1\!/\!_{p^{a-1}}\, \mathbb{E}^* \left(|(X_{i\varpi} Y_{i\varpi} - \mathbb{E}^*(X_{i\varpi} Y_{i\varpi}))^a| \right)$$

in a first step, and then by ${}^{2^a}\!/\!_{p^{a-1}}\, \mathbb{E}^*(|X_{i\varpi} Y_{i\varpi}|^a)$ in a second step. By the Cauchy-Schwarz inequality,

$${}^{2^a}\!/\!_{p^{a-1}}\, \mathbb{E}^*(|X_{i\varpi} Y_{i\varpi}|^a) \leq {}^{2^a}\!/\!_{p^{a-1}} \sqrt{\mathbb{E}^*(|X_{i\varpi}|^{2a})} \sqrt{\mathbb{E}^*(|Y_{i\varpi}|^{2a})}.$$

To prove that $\sup_{1\leq i\leq n} |\tilde{\gamma}_i - \hat{\gamma}_i| = o_{P^*}(1)$ we show that $\sup_{1\leq i\leq n} \mathbb{E}(\mathbb{E}^*(|\tilde{\gamma}_i - \hat{\gamma}_i|^a)) = o(1)$ by noting that

$${}^{2^a}\!/\!_{p^{a-1}}\, \mathbb{E} \left(\sqrt{\mathbb{E}^*(|X_{i\varpi}|^{2a})} \sqrt{\mathbb{E}^*(|Y_{i\varpi}|^{2a})} \right) \leq {}^{2^a}\!/\!_{p^{a-1}} \sqrt{\mathbb{E}(\mathbb{E}^*(|X_{i\varpi}|^{2a}))\, \mathbb{E}(\mathbb{E}^*(|Y_{i\varpi}|^{2a}))},$$

and that

$$\mathbb{E}(\mathbb{E}^*(|X_{i\varpi}|^{2a})) = {}^1\!/\!_{(m-q+1)\, q^a} \sum_{t=0}^{m-q} \mathbb{E} \left(\left| \sum_{q'=1}^{q} z_{it+q'} \right|^{2a} \right) = O(1),$$

$$\mathbb{E}(\mathbb{E}^*(|Y_{i\varpi}|^{2a})) = {}^1\!/\!_{(m-q+1)\, q^a} \sum_{t=0}^{m-q} \mathbb{E} \left(\left| \sum_{q'=1}^{q} \varepsilon_{it+q'} \right|^{2a} \right) = O(1),$$

uniformly by Hansen (1991, Corollary 3). Indeed, putting everything together reveals that

$$\sup_{1\leq i\leq n} |\tilde{\gamma}_i - \hat{\gamma}_i| = O_{P^*}({}^1\!/\!_{p^{a-1}}) = O_{P^*}(({}^q\!/\!_m)^{a-1}) = o_{P^*}(1),$$

using that $m = pq$ and recalling that $a > 1$.

Next, we observe that $\tilde{\gamma}_i - \hat{\gamma}_i^* = {}^1\!/\!_{p^2} \sum_{p_1'=1}^{p} \sum_{p_2'=1}^{p} X_{i\varpi_{p_1'}} Y_{i\varpi_{p_2'}}$. Then

$$\sup_{1\leq i\leq n} |\tilde{\gamma}_i - \hat{\gamma}_i^*| \leq \sup_{1\leq i\leq n} \left| {}^1\!/\!_p \sum_{p'=1}^{p} X_{i\varpi_{p'}} \right| \sup_{1\leq i\leq n} \left| {}^1\!/\!_p \sum_{p'=1}^{p} Y_{i\varpi_{p'}} \right| = O_{P^*}({}^q\!/\!_m) = o_{P^*}(1),$$

using Gonçalves and White (2002, Theorem 2.2). By the triangle inequality, we have thus shown that $\sup_{1\leq i\leq n}|\hat{\gamma}_i^* - \hat{\gamma}_i| = o_{P^*}(1)$. Finally, from Gonçalves and White (2002, Corollary 2.1), we have that

$$\sup_{1\leq i\leq n} |\hat{\gamma}_i - \mathbb{E}(\chi_i)| = o_P(1).$$

From this, (A.10) follows.

To proceed, let $\tilde{X}_{i\varpi} := {}^1\!/\!_{\sqrt{q}} \sum_{q'=1}^{q} z_{i\varpi+q'}$ and $\tilde{Y}_{i\varpi} := {}^1\!/\!_{\sqrt{q}} \sum_{q'=1}^{q} \varepsilon_{i\varpi+q'}$. These are the uncentered versions of $X_{i\varpi}$ and $Y_{i\varpi}$, respectively. The second term in (A.9) then writes as

$${}^1\!/\!_{pq} \sum_{p_1'=1}^{p} \sum_{q_1'=1}^{q} \sum_{p_2'\neq p_1'} \sum_{q_2'=1}^{q} z_{i\varpi_{p_1'}+q_1'} \varepsilon_{i\varpi_{p_2'}+q_2'} = {}^1\!/\!_{p} \sum_{p_1'=1}^{p} \sum_{p_2'\neq p_1'} \tilde{X}_{i\varpi_{p_1'}} \tilde{Y}_{i\varpi_{p_2'}}.$$

We have $\mathbb{E}^*(\tilde{X}_{i\varpi_{p_1'}} \tilde{Y}_{i\varpi_{p_2'}}) = \mathbb{E}^*(\tilde{X}_{i\varpi_{p_1'}})\, \mathbb{E}^*(\tilde{Y}_{i\varpi_{p_2'}})$ because the $\varpi_1, \ldots, \varpi_p$ are independent, and so

$$\mathbb{E}^* \left({}^1\!/\!_{p} \sum_{p_1'=1}^{p} \sum_{p_2'\neq p_1'} \tilde{X}_{i\varpi_{p_1'}} \tilde{Y}_{i\varpi_{p_2'}} \right) = (p-1)q\, \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\bar{\varepsilon}_i^*).$$

Define

$$r_i^* := {}^1\!/\!_{p} \sum_{p_1'=1}^{p} \sum_{p_2'\neq p_1'} (\tilde{X}_{i\varpi_{p_1'}} \tilde{Y}_{i\varpi_{p_2'}} - \mathbb{E}^*(\tilde{X}_{i\varpi_{p_1'}} \tilde{Y}_{i\varpi_{p_2'}})).$$

Then, collecting terms,

$$\chi_i^* - \chi_i = \mathbb{E}(\chi_i) + r_i^* + q\{\bar{z}_i^*\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\bar{\varepsilon}_i^*)\} - m\{\bar{z}_i\, \bar{\varepsilon}_i - \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\bar{\varepsilon}_i^*)\} + o_{P^*}(1).$$

By adding and subtracting terms to the each of the components making up the difference, we can write $\bar{z}_i^*\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\varepsilon_i^*)$ as

$$\big(\bar{z}_i^* - \mathbb{E}^*(\bar{z}_i^*)\big)\big(\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{\varepsilon}_i^*)\big) + \big(\bar{z}_i^* - \mathbb{E}^*(\bar{z}_i^*)\big)\, \mathbb{E}^*(\bar{\varepsilon}_i^*) + \big(\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{\varepsilon}_i^*)\big)\, \mathbb{E}^*(\bar{z}_i^*).$$

From Fitzenberger (1997, Lemma A.1) and Gonçalves and White (2002, Theorem 2.2) we have $\sup_{1\leq i\leq n}|\bar{z}_i^* - \mathbb{E}^*(\bar{z}_i^*)| = O_{P^*}({}^1\!/\!_{\sqrt{m}})$ and $\sup_{1\leq i\leq n}|\mathbb{E}^*(\bar{z}_i^*)| = O_P({}^1\!/\!_{\sqrt{m}} + {}^q\!/\!_{m})$, and, similarly, $\sup_{1\leq i\leq n}|\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{\varepsilon}_i^*)| = O_{P^*}({}^1\!/\!_{\sqrt{m}})$ and $\sup_{1\leq i\leq n}|\mathbb{E}^*(\bar{\varepsilon}_i^*)| = O_P({}^1\!/\!_{\sqrt{m}} + {}^q\!/\!_{m})$. We thus deduce that $q \sup_{1\leq i\leq n}|\bar{z}_i^*\bar{\varepsilon}_i^* - \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\bar{\varepsilon}_i^*)| = o_{P^*}(1)$. Proceeding in the same way gives $m \sup_{1\leq i\leq n}|\bar{z}_i\, \bar{\varepsilon}_i - \mathbb{E}^*(\bar{z}_i^*)\, \mathbb{E}^*(\bar{\varepsilon}_i^*)| = o_P(1)$. Therefore,

$${}^1\!/\!_{n} \sum_{i=1}^{n} ((\chi_i^* - \chi_i) - \mathbb{E}(\chi_i)) = {}^1\!/\!_{n} \sum_{i=1}^{n} r_i^* + o_{P^*}(1),$$

and we are left only with showing that $1/n \sum_{i=1}^{n} r_i^* = o_{P^*}(1)$ to complete our derivation of the asymptotic bias.

To do so, first, repeated adding and subtracting of $\mathbb{E}^*(\tilde{X}_{i\varpi})$ and $\mathbb{E}^*(\tilde{Y}_{i\varpi})$ allows to write

$$r_i^* = (p-1)/p \sum_{p'=1}^{p} X_{i\varpi_{p'}} \mathbb{E}^*(\tilde{Y}_{i\varpi}) + (p-1)/p \sum_{p'=1}^{p} Y_{i\varpi_{p'}} \mathbb{E}^*(\tilde{X}_{i\varpi}) + 1/p \sum_{p_1'=1}^{p} \sum_{p_2' \neq p_1'} X_{i\varpi_{p_1'}} Y_{i\varpi_{p_2'}}.$$

This corresponds to a Hoeffding decomposition (conditional on the data) into a Hájek projection, which constitutes the first two right-hand side terms, and a remainder term. We now proceed by looking at each of these three terms, in turn. We begin by showing that

$$(p-1)/np \sum_{i=1}^{n} \sum_{p'=1}^{p} X_{i\varpi_{p'}} \mathbb{E}^*(\tilde{Y}_{i\varpi}) = o_{P^*}(1). \tag{A.11}$$

As $\mathbb{E}^*(X_{i\varpi_{p'}} \mathbb{E}^*(\tilde{Y}_{i\varpi})) = 0$, the conditional variance of the left-hand side of (A.11) equals

$$\mathbb{E}^* \left(\left((p-1)/np \sum_{i=1}^{n} \sum_{p'=1}^{p} X_{i\varpi_{p'}} \mathbb{E}^*(\tilde{Y}_{i\varpi}) \right)^2 \right) = (p-1)^2/pn^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}^*(X_{i\varpi} X_{j\varpi}) \mathbb{E}^*(\tilde{Y}_{i\varpi}) \mathbb{E}^*(\tilde{Y}_{j\varpi}),$$

where we have used that

$$\mathbb{E}^* \left(1/p^2 \sum_{p_1'=1}^{p} \sum_{p_2'=1}^{p} X_{i\varpi_{p_1'}} X_{j\varpi_{p_2'}} \right) = \mathbb{E}^* \left(1/p^2 \sum_{p'=1}^{p} X_{i\varpi_{p'}} X_{j\varpi_{p'}} \right) = 1/p \, \mathbb{E}^* \left(X_{i\varpi} X_{j\varpi} \right),$$

which holds because $\mathbb{E}^*(X_{i\varpi} X_{j\varpi'}) = \mathbb{E}^*(X_{i\varpi}) \mathbb{E}^*(X_{j\varpi'}) = 0$ whenever $\varpi \neq \varpi'$. We can expand the conditional variance as

$$(p-1)^2/pn^2 \sum_{i=1}^{n} \mathbb{E}^*(X_{i\varpi}^2) \mathbb{E}^*(\tilde{Y}_{i\varpi})^2 + (p-1)^2/pn^2 \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}^*(X_{i\varpi} X_{j\varpi}) \mathbb{E}^*(\tilde{Y}_{i\varpi}) \mathbb{E}^*(\tilde{Y}_{j\varpi}).$$

Here, $\mathbb{E}^*(X_{i\varpi}^2)$ is a bootstrap estimator of the long-run variance of the scaled sample mean $\sqrt{m}\, \bar{z}_i$; this variance is uniformly bounded under our assumptions. Furthermore, from Gonçalves and White (2002, Corollary 2.1), $\mathbb{E}^*(X_{i\varpi}^2)$ is uniformly consistent. Also, using Fitzenberger (1997, Lemma A.1), $\sup_{1 \leq i \leq n} \mathbb{E}^*(\tilde{Y}_{i\varpi})^2 = O_P(q/m + q^2/m^{3/2} + q^3/m^2)$. Therefore,

$$(p-1)^2/pn^2 \sum_{i=1}^{n} \mathbb{E}^*(X_{i\varpi}^2) \mathbb{E}^*(\tilde{Y}_{i\varpi})^2 = O(1/n) \, O_P(1 + q/m^{1/2} + q^2/m) = o_P(1).$$

Similarly, $\mathbb{E}^*(X_{i\varpi}X_{j\varpi})$ is a consistent estimator of the long-run covariance between the scaled means $\sqrt{m}\,\bar{z}_i$ and $\sqrt{m}\,\bar{z}_j$. Because the cross-sectional observations are independent, $|\mathbb{E}^*(X_{i\varpi}X_{j\varpi})| = o_P(1)$ when $i \neq j$. We thus obtain

$$(p-1)^2/pn^2 \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}^*(X_{i\varpi}X_{j\varpi})\, \mathbb{E}^*(\tilde{Y}_{i\varpi})\, \mathbb{E}^*(\tilde{Y}_{j\varpi}) = o_P(p) \left(\sup_{1 \leq i \leq n} \mathbb{E}^*(\tilde{Y}_{i\varpi}) \right)^2 = o_P(1).$$

Equation (A.11) follows by Markov's inequality. By symmetry,

$$(p-1)/np \sum_{i=1}^{n} \sum_{p'=1}^{p} Y_{i\varpi_{p'}}\, \mathbb{E}^*(\tilde{X}_{i\varpi}) = o_{P^*}(1)$$

can be shown in the same way.

Now turn to the remainder term. To show that

$$1/np \sum_{i=1}^{n} \sum_{p_1'=1}^{p} \sum_{p_2' \neq p_1'} X_{i\varpi_{p_1'}} Y_{i\varpi_{p_2'}} = o_{P^*}(1), \tag{A.12}$$

we begin by calculating its conditional variance,

$$1/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} 1/p^2 \sum_{p_1'=1}^{p} \sum_{p_2' \neq p_1'} \sum_{p_1''=1}^{p} \sum_{p_2'' \neq p_1''} \mathbb{E}^*(X_{i\varpi_{p_1'}} Y_{i\varpi_{p_2'}} X_{j\varpi_{p_1''}} Y_{j\varpi_{p_2''}}).$$

The conditional mean inside the summation is zero unless either (i) $p_1' = p_1''$ and $p_2' = p_2''$ or $p_1' = p_2''$ and $p_2' = p_1''$ because both $X_{i\varpi}$ and $Y_{i\varpi}$ are mean zero and the $\varpi_1, \ldots, \varpi_p$ are independent. The contribution of the Case (i) terms to the conditional variance is given by

$$(p-1)/p\ 1/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}^*(X_{i\varpi}\, X_{j\varpi})\, \mathbb{E}^*(Y_{i\varpi}\, Y_{j\varpi})$$

while the contribution of the Case (ii) terms equals

$$(p-1)/p\ 1/n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}^*(X_{i\varpi}Y_{j\varpi})\, \mathbb{E}^*(Y_{i\varpi}\, X_{j\varpi}).$$

Each of these two terms can be handled in the same way as before. For example, the first term can be decomposed as

$$(p-1)/p\ 1/n^2 \sum_{i=1}^{n} \mathbb{E}^*(X_{i\varpi}^2)\, \mathbb{E}^*(Y_{i\varpi}^2) + (p-1)/p\ 1/n^2 \sum_{i=1}^{n} \sum_{j \neq i} \mathbb{E}^*(X_{i\varpi}\, X_{j\varpi})\, \mathbb{E}^*(Y_{i\varpi}\, Y_{j\varpi}).$$

Here, each contribution is again $o_P(1)$ because $\sup_{1\leq i\leq n}\mathbb{E}^*(X_{i\varpi}^2)$ and $\sup_{1\leq i\leq n}\mathbb{E}^*(Y_{i\varpi}^2)$ are $O_P(1)$ and $\sup_{1\leq i\neq j\leq n}\mathbb{E}^*(X_{i\varpi}X_{j\varpi})$ and $\sup_{1\leq i\neq j\leq n}\mathbb{E}^*(Y_{i\varpi}Y_{j\varpi})$ are $o_P(1)$. (A.12) follows and, with it,

$$-1/n\sum_{i=1}^{n}(\chi_i^*-\chi_i)=b_{n,m}+o_{P^*}(1).$$

The proof of the theorem is then complete. □

References

Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–297.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. In R. Blundell, W. K. Newey, and T. Persson (Eds.), *Advances In Economics and Econometrics*, Volume III. Econometric Society: Cambridge University Press.

Cavaliere, G., S. Gonçalves, M. Ø. Nielsen, and E. Zanelli (2024). Bootstrap inference in the presence of bias. Forthcoming in *Journal of the American Statistical Association*.

Chudik, A., M. H. Pesaran, and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*, 816–836.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*, 991–1030.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics 7*, 1–26.

Fitzenberger, B. (1997). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics 82*, 235–287.

Gonçalves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual effects. *Econometric Theory 27*, 1048–1082.

Gonçalves, S. and M. Kaffo (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics 186*, 407–426.

Gonçalves, S. and H. White (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory 18*, 1367–1384.

Gonçalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics 119*, 199–219.

Gonçalves, S. and H. White (2005). Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association 100*, 970–979.

Götze, G. and H. R. Künsch (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics 24*, 1914–1933.

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.

Hansen, B. E. (1991). Strong laws for dependent heterogeneous processes. *Econometric Theory 7*, 213–221.

Higgins, A. and K. Jochmans (2024). Bootstrap inference for fixed-effect models. *Econometrica 92*, 411–427.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics 17*, 1217–1241.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–1426.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

von Bahr, B. and C.-G. Esseen (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics 36*, 299–303.

White, H. (2000). *Asymptotic Theory for Econometricians, Revised Edition*. Emerald.