

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

Ayden Higgins*

University of Exeter

Koen Jochmans[†]

Toulouse School of Economics

This version: October 16, 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 consecutive 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.

We are grateful to Anders Kock, Roger Moon, Ulrich Mueller, and conference and seminar participants at various institutions for comments and suggestions.

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. This bias remains important in large samples unless the number of time-series observations, m , is large relative to the cross-sectional sample size, n , in the sense that $n/m \rightarrow 0$ as $n, m \rightarrow \infty$. This is not usually the case. Several ways to correct for the bias have been proposed; see [Arellano and Hahn \(2007\)](#) for an overview. Doing so recenters the estimator’s limit distribution around zero, thereby restoring the validity of inference procedures based on it.

An alternative approach, which has been 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, then, allows to perform inference based on the bootstrap distribution in the usual manner; no adjustment for the presence of bias needs to be made.¹ The difficulty here lies in constructing a bootstrap scheme that correctly reproduces the bias. [Higgins and Jochmans \(2024\)](#) show that the parametric bootstrap does so in quite general nonlinear models. Outside the parametric framework, however, there are currently results only for the linear autoregressive model. In particular, [Gonçalves and Kaffo \(2015\)](#) show that a recursive version of the wild bootstrap permits inference using the within-group estimator. They also demonstrate the failure of several other conventional bootstrap schemes to replicate bias in their setting, illustrating the difficulty in devising a successful procedure.

In this paper we study the linear model where the errors are allowed to be correlated with both past and future regressors. This is a workhorse model in applied economics. Indeed, many regressors react to past outcomes, creating feedback. Microeconomic examples are fertility decisions in wage equations, labor-force participation decisions, and household demand systems (see [Arellano and Carrasco 2003](#) and [Browning and Lechene](#)

¹There is a recent interest in the validity of the bootstrap in the presence of (asymptotic) bias more broadly; see [Cattaneo and Jansson \(2018\)](#) and [Cavaliere, Gonçalves, Nielsen and Zanelli \(2024\)](#).

2003 for discussions). A macroeconomic example are panel local projections (Jordà and Taylor 2025, Mei, Sheng and Shi 2025). As in the autoregressive problem of Gonçalves and Kaffo (2015), which is nested in our setup, the bias in the within-group estimator arises due to the correlation between the time-demeaned regressors and the errors. Its precise form, however, depends on the feedback process. A bootstrap designed around any form of recursion requires the process as an input and will, therefore, not be helpful in this setting. Whether a bootstrap scheme that mimics the bias for feedback processes of an unknown form can be constructed is not known.

We show that a version of the moving block bootstrap of Künsch (1989), where blocks of consecutive cross-sections are resampled with replacement, replicates the distribution of the within-group estimator under asymptotics where $n/m \rightarrow c \in [0, +\infty)$. The potential for this bootstrap scheme to replicate bias in our context has not previously been explored. Gonçalves (2011) has shown that it is capable of yielding correct inference in a version of our model with cross-sectional dependence. However, the assumptions under which this result was established include a condition on n and m to ensure that the estimator is asymptotically unbiased; in our context, this condition amounts to the rate requirement that $n/m \rightarrow 0$.

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 latter estimator centered around the truth—together with its chief implications for estimation and inference. A final section reports on a numerical experiment. 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

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

from $n \times m$ panel data, treating the n intercepts $\alpha_1, \dots, \alpha_n$ 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 for all t .
- (iv) For each i , $\{(x_{it}, \varepsilon_{it})\}$ is stationary mixing with the mixing coefficients, a_i , satisfying

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

for some $s > 4r/r-2$.²

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

Assumption 2. For all (n, m) sufficiently large,

- (i) The covariance matrix $\Sigma_{n,m} := 1/nm \sum_{i=1}^n \sum_{t=1}^m \mathbb{E}(z_{it}z'_{it})$ is positive definite, and

²Here,

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

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

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

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

is positive definite.

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

Assumption 3. $n, m \rightarrow \infty$ with $n/m \rightarrow c \in [0, +\infty)$.

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.³

The within-group estimator of β is

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

where $\bar{y}_i := 1/m \sum_{t=1}^m y_{it}$ and $\bar{x}_i := 1/m \sum_{t=1}^m x_{it}$. Under Assumptions 1-3 this estimator is consistent and asymptotically normally-distributed, but its limit distribution features an asymptotic bias term unless $n/m \rightarrow 0$. The bias arises from the fact that the normal equations factor as

$$\sum_{i=1}^n \sum_{t=1}^m (x_{it} - \bar{x}_i) \varepsilon_{it} = \sum_{i=1}^n \sum_{t=1}^m z_{it} \varepsilon_{it} - \sum_{i=1}^n m \text{cov}(\bar{z}_i, \bar{\varepsilon}_i) + o_P(\sqrt{nm}),$$

where we let $\bar{z}_i := 1/m \sum_{t=1}^m z_{it}$ and $\bar{\varepsilon}_i := 1/m \sum_{t=1}^m \varepsilon_{it}$. While the first right-hand side term in this expression is mean zero and, after scaling, satisfies a central limit theorem, the second right-hand side term will shift the limit distribution away from zero, thereby causing bias.

³In the conventional first-order autoregressive model, for example, $x_{it} = y_{it-1}$ and $\varepsilon_{it} \sim \text{i.i.d.}(0, \sigma^2)$, 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*

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

where $b_{n,m} := -1/n \sum_{i=1}^n \sum_{\tau=1}^{m-1} (m-\tau)/m (\mathbb{E}(z_{it}\varepsilon_{it+\tau}) + \mathbb{E}(z_{it+\tau}\varepsilon_{it}))$ and $\mathcal{V} := \lim_{n,m \rightarrow \infty} \mathcal{V}_{n,m}$ for $\mathcal{V}_{n,m} := \Sigma_{n,m}^{-1} \Omega_{n,m} \Sigma_{n,m}^{-1}$.

Theorem 1 generalizes results available for models with predetermined regressors generated through a specified process such as those derived in [Hahn and Kuersteiner \(2002\)](#), [Alvarez and Arellano \(2003\)](#), [Dhaene and Jochmans \(2016\)](#), and [Chudik, Pesaran and Yang \(2018\)](#) for certain linear processes.

In light of Theorem 1, a way forward would be to estimate the bias term $b_{n,m}$ by means of a conventional HAC estimator, using plug-in estimators of the various covariances appearing in it based on the fixed-effect estimator of the parameters $\alpha_1, \dots, \alpha_n$ and β , and recenter the within-group normal equations around it. [Hahn and Kuersteiner \(2011\)](#) proposed using

$$\hat{b}_{n,m} := -1/n \sum_{i=1}^n \sum_{\tau=-h}^h 1/m \left(\sum_{t=\max\{1, 1+\tau\}}^{\min\{m, m+\tau\}} (x_{it-\tau} - \bar{x}_i) \hat{\varepsilon}_{it} \right),$$

where $\hat{\varepsilon}_{it} := (y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)\hat{\beta}$ are the within-group residuals and h is a chosen bandwidth parameter that grows with m so that $h/m \rightarrow 0$ ([Hahn and Kuersteiner 2011](#), Theorem 2). Under regularity conditions this yields an estimator that is asymptotically unbiased. An alternative would be to use a jackknife estimator as in [Dhaene and Jochmans \(2015\)](#). In either case, asymptotically-valid inference can then be performed by relying on the limit distribution in the theorem. As an alternative, in the next section, we validate the use of a bootstrap procedure on the within-group estimator to directly perform inference based on it.

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 $\varpi_1, \dots, \varpi_p$ be a random sample from the discrete uniform distribution on $\{0, \dots, m - q\}$, the bootstrap time series $\{(y_{it}^*, x_{it}^*)\}$ is generated as

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

for $1 \leq p' \leq p$ and $1 \leq q' \leq q$. Here, the random variables $\varpi_1, \dots, \varpi_p$ select starting points for the different blocks. When the block length, q , is set to one this reduces to [Efron's \(1979\)](#) bootstrap applied to the cross-sections. We will require q to grow with m for the bootstrap to work.

The within-group estimator computed from the sample generated via our bootstrap is

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

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 $q \rightarrow \infty$ 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 $n/m \rightarrow 0$. Because the bias in $\sqrt{nm}(\hat{\beta} - \beta)$ is of the order $\sqrt{n/m}$, this rate condition ensures that the limit distribution of the within-group estimator does not feature a bias term.

The proof of Theorem [2](#) shows that the bootstrap normal equations for $\hat{\beta}$ satisfy,

$$\sum_{i=1}^n \sum_{t=1}^m (x_{it}^* - \bar{x}_i^*) \hat{\varepsilon}_{it}^* = \sum_{i=1}^n \sum_{t=1}^m (z_{it}^* \varepsilon_{it}^* - z_{it} \varepsilon_{it}) - \sum_{i=1}^n m \text{cov}^*(z_i^*, \bar{\varepsilon}_i^*) + o_{P^*}(\sqrt{nm}),$$

where $\hat{\varepsilon}_{i(p'-1)q+q'}^* := \hat{\varepsilon}_{i\varpi_{p'}+q'}$, for $1 \leq p' \leq p$ and $1 \leq q' \leq q$ and we let $\bar{z}_i^* := 1/m \sum_{t=1}^m z_{it}^*$ and $\bar{\varepsilon}_i^* := 1/m \sum_{t=1}^m \varepsilon_{it}^*$. Here, the $mcov^*(\bar{z}_i^*, \bar{\varepsilon}_i^*)$ are unit-specific moving-block bootstrap estimators of the covariances $mcov(\bar{z}_i, \bar{\varepsilon}_i)$ which cause the bias in Theorem 1 (see, e.g., [Gonçalves and White 2002](#)). Indeed, here, after scaling, the first term yields the limit distribution that features in Theorem 1 while the second term is a consistent estimator of the bias term $b_{n,m}$, ensuring that the bootstrap distribution is correctly centered.

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

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

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

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

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,

$$\check{\theta} := \hat{\theta} - \hat{Q}_{1/2}$$

is a bootstrap-based bias-corrected estimator of θ with the same properties as the corrected estimators discussed above.

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

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

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

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

where $\hat{\epsilon}_{it}^* := (y_{it}^* - \bar{y}_i^*) - (x_{it}^* - \bar{x}_i^*)\hat{\beta}$. The latter, in turn, is an estimator of $\Omega_{n,m}$. If we further introduce the shorthands

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

we can define

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

By the same arguments as in [Gonçalves \(2011, Theorem B.1 and Lemma B.1.\(iii\)\)](#) we can verify that $\hat{\mathcal{R}}^* - \mathcal{R} \xrightarrow{P^*} 0$ and $\hat{\mathcal{R}} - \mathcal{R} \xrightarrow{P} 0$.

Inference can thus be performed using the normal approximation to the estimator $\check{\theta}$, as

$$\sqrt{nm} \hat{\sigma}^{-1} (\check{\theta} - \theta) \xrightarrow{L} N(0, 1),$$

where $\hat{\sigma} := (c' \hat{\mathcal{R}} c)^{1/2}$. This distributional approximation also goes through with $\hat{\mathcal{R}}$ replaced by any other consistent estimator of \mathcal{R} . Furthermore, the variance estimator can equally be used along with other bias-corrected estimators, such as the one proposed by [Hahn and Kuersteiner \(2011\)](#).

Together with [Theorem 2](#), the bootstrap consistency of the variance estimator implies that the reverse-percentile method equally applies to studentized statistics. If we redefine

$$\hat{Q}_\alpha := \inf \{ Q : \alpha \leq \mathbb{P}^*(\hat{\sigma}^{*-1}(\hat{\theta}^* - \hat{\theta}) \leq Q) \},$$

where $\hat{\sigma}^* := (c' \hat{\mathcal{R}}^* c)^{1/2}$, then

$$\mathbb{P}(\hat{\theta} - \hat{\sigma} \hat{Q}_\alpha \leq \theta) \rightarrow \alpha$$

under the assumptions of [Theorem 2](#). This ensures that decision rules for null hypotheses on θ involving critical values obtained from the bootstrap quantiles \hat{Q}_α deliver asymptotic size control.

4 Simulations

In this section we provide simulation results for three different designs. Each design has the structure

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

with $\varepsilon_{it} \sim \text{i.i.d. } N(0, 1)$, $\beta = 1/2$, and $(n, m) = (50, 50)$. 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. The three designs differ in how the regressor is generated. The first design has $x_{it} = y_{it-1}$ and thus corresponds to the conventional first-order autoregressive model. The second design has $x_{it} = \{y_{it-1} > 0\}$ and is, thus, nonlinear. The third design has $x_{it} = 0.5x_{it-1} + 0.5y_{it-1} + \eta_{it}$ for $\eta_{it} \sim \text{i.i.d. } N(0, 1)$ and corresponds to a bivariate vector autoregression.

The designs correspond to Tables 1–3. Each table contains the bias and standard deviation of three estimators. The naive estimator is the standard within-group estimator, $\hat{\beta}$. The bootstrap estimator is the estimator $\check{\beta}$, where the bias is corrected for using the median of the bootstrap distribution; we provide results for block sizes $q \in \{2, 5, 10\}$. The final estimator is the analytical plug-in correction of [Hahn and Kuersteiner \(2011\)](#); we give results for bandwidth choices $h \in \{1, 5, 10\}$. The tables also provide the empirical coverage rate of 95% confidence intervals constructed in different ways. First, for each of the estimators just discussed, we provide rates for the confidence interval constructed using the large-sample normal approximation. For this we used the bootstrap to estimate the variance for all estimators; we again report the coverage for $q \in \{2, 5, 10\}$. We then also give coverage rates based on the reverse-percentile and studentized reverse-percentile bootstrap (applied directly to the within-group estimator) in the last two rows of each table. All bootstrap calculations were based on 1999 bootstrap replications. The simulation results are calculated over 2500 Monte Carlo repetitions.

Table 1 illustrates the downward [Nickell \(1981\)](#) bias in the within-group estimator and the corresponding undercoverage in confidence intervals that ignore this bias in their construction. Both the bootstrap and analytical approach reduce the bias, and both are

Table 1: Simulation results for the autoregressive model

	Naive	Moving block bootstrap			Analytical correction		
		q			h		
		2	5	10	1	5	10
Bias	-0.0303	-0.0221	-0.0109	-0.0070	-0.0151	-0.0049	-0.0106
Std. dev.	0.0179	0.0178	0.0183	0.0193	0.0176	0.0182	0.0184
Coverage (0.95)							
Limit distribution:							
$q = 2$	0.6052	0.7696	— — —	— — —	0.8712	0.9360	0.8972
$q = 5$	0.6032	— — —	0.8956	— — —	0.8660	0.9324	0.8944
$q = 10$	0.6044	— — —	— — —	0.9052	0.8644	0.9344	0.8944
Bootstrap:							
Reverse-percentile	— — —	0.7969	0.9496	0.9516	— — —	— — —	— — —
Studentized	— — —	0.7896	0.9444	0.9484	— — —	— — —	— — —

The data were generated as $y_{it} = x_{it}\beta + \varepsilon_{it}$ with $x_{it} = y_{it-1}$, $\varepsilon_{it} \sim \text{i.i.d } N(0, 1)$, y_{i0} drawn from the steady-state distribution, and $\beta = 1/2$. The bootstrap was implemented using 1999 replications and the different choices for q in the table. The analytical correction is the one of [Hahn and Kuersteiner \(2011\)](#), based on the different bandwidth choices h in the table. In all cases, the variance of the limit distribution was estimated using the moving block bootstrap. The sample size considered was $(n, m) = (50, 50)$. Results are based on 2500 Monte Carlo replications.

Table 2: Simulation results for the binary-regressor model

	Naive	Moving block bootstrap			Analytical correction		
		q			h		
		2	5	10	1	5	10
Bias	-0.0387	-0.0226	-0.0103	-0.0087	-0.0078	-0.0066	-0.0146
Std. dev.	0.0412	0.0413	0.0423	0.0439	0.0411	0.0418	0.0418
Coverage (0.95)							
Limit distribution:							
$q = 2$	0.8456	0.9160	— — —	— — —	0.9416	0.9404	0.9272
$q = 5$	0.8376	— — —	0.9296	— — —	0.9408	0.9372	0.9268
$q = 10$	0.8416	— — —	— — —	0.9220	0.9420	0.9376	0.9276
Bootstrap:							
Reverse-percentile	— — —	0.9136	0.9240	0.8948	— — —	— — —	— — —
Studentized	— — —	0.9136	0.9236	0.8972	— — —	— — —	— — —

The data were generated as $y_{it} = x_{it}\beta + \varepsilon_{it}$ with $x_{it} = \{y_{it-1} > 0\}$, $\varepsilon_{it} \sim \text{i.i.d } N(0, 1)$, y_{i0} drawn from the steady-state distribution, and $\beta = 1/2$. The bootstrap was implemented using 1999 replications and the different choices for q in the table. The analytical correction is the one of [Hahn and Kuersteiner \(2011\)](#), based on the different bandwidth choices h in the table. In all cases, the variance of the limit distribution was estimated using the moving block bootstrap. The sample size considered was $(n, m) = (50, 50)$. Results are based on 2500 Monte Carlo replications.

more successful in doing so when additional autocovariances are taken into account. Indeed, their respective biases are decreasing in q and h . In all cases inference based on the limit distribution improves as a consequence of bias correction. For the bootstrap the coverage rates improve with q while for the analytical plug-in correction their behavior is not monotone in h . In any event, the percentile-based bootstrap confidence intervals based on the within-group estimator outperform those based on the asymptotic approximation in all cases.

In [Table 2](#) the undercoverage of the naive approach is somewhat less severe. This is so because the standard deviation of the estimator is larger while the size of the bias is roughly unaltered. Both the bootstrap and analytical approach to bias correction are effective, and

Table 3: Simulation results for the VAR model

	Naive	Moving block bootstrap			Analytical correction		
		q			h		
		2	5	10	1	5	10
Bias	-0.0149	-0.0127	-0.0080	-0.0046	-0.0111	-0.0047	-0.0050
Std. dev.	0.0130	0.0130	0.0132	0.0138	0.0129	0.0130	0.0132
Coverage (0.95)							
Limit distribution:							
$q = 2$	0.7780	0.8248	— — —	— — —	0.8536	0.9232	0.9216
$q = 5$	0.7792	— — —	0.8916	— — —	0.8528	0.9252	0.9204
$q = 10$	0.7860	— — —	— — —	0.9160	0.8620	0.9292	0.9248
Bootstrap:							
Reverse-percentile	— — —	0.8304	0.9100	0.9356	— — —	— — —	— — —
Studentized	— — —	0.8256	0.9040	0.9252	— — —	— — —	— — —

The data were generated as $y_{it} = x_{it}\beta + \varepsilon_{it}$ with $x_{it} = 0.5y_{it-1} + 0.5x_{it-1} + u_{it}$ for $\varepsilon_{it} \sim \text{i.i.d } N(0, 1)$ and $u_{it} \sim \text{i.i.d } N(0, 1)$. The (y_{it}, x_{it}) processes were initialized with draws from their steady-state distributions, and $\beta = 1/2$. The bootstrap was implemented using 1999 replications and the different choices for q in the table. The analytical correction is the one of [Hahn and Kuersteiner \(2011\)](#), based on the different bandwidth choices h in the table. In all cases, the variance of the limit distribution was estimated using the moving block bootstrap. The sample size considered was $(n, m) = (50, 50)$. Results are based on 2500 Monte Carlo replications.

more so for larger values of q and h . This, then, largely rectifies the undercoverage. The coverage rates for the various techniques also show considerably less sensitivity to the choice of the tuning parameters q and h .

Table 3 finally, contains the results related to the VAR setting. It confirms the overall picture observed in the pure autoregressive model.

5 Concluding remarks

The within-group estimator of linear models computed from $n \times m$ panel data suffers from asymptotic bias under asymptotics where $n/m \rightarrow c \in [0, +\infty)$ when regressors are not strictly exogenous. We have shown that, in spite of the presence of this bias, the moving block bootstrap permits valid inference based on the within-group estimator. This finding presents an encouraging addition to the literature where, until now, it was not known whether models with unspecified feedback were amenable to bootstrapping in the presence of incidental-parameter bias.

We conjecture that the validity of our bootstrap scheme extends to a large class of nonlinear problems, at the level of generality of [Hahn and Kuersteiner \(2011\)](#). Moreover, while nonlinearity introduces an additional bias term of order $1/m$, this bias is equally captured by the moving block bootstrap as described here. We have established this to be indeed the case for the variance estimator in a semiparametric version of the many normal means problem of [Neyman and Scott \(1948\)](#), where the distribution of the data is unspecified and observations are stationary over time. A rigorous theory for the nonlinear case is left for future work.

In the linear setting of the current paper we have, as most of the literature that is concerned with bias correction, proceeded under the assumption of independence in the cross-sectional dimension. [Gonçalves \(2011\)](#) provided limit theory for the within-group estimator under various degrees of cross-sectional dependence. In her setting, the normal equations are assumed to converge at the rate $1/\sqrt{n^\rho m}$ for some $\rho \in [0, 1]$, and asymptotic bias is assumed away by imposing that $n^\rho/m \rightarrow 0$. The form of the bias—and with it the ability of the moving block bootstrap to replicate it—is, however, not affected by the presence of cross-sectional dependence. On the other hand, as ρ decreases and the amount of dependence increases, the convergence rate of the estimator goes down, and the less pertinent the presence of bias at order $1/m$ becomes. In the most severe case we have $\rho = 0$; the cross-sectional sample size becomes irrelevant in the asymptotic analysis and no bias issue arises.

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 $\hat{\beta}$,

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

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

$$|\hat{\Sigma}_{n,m} - \Sigma_{n,m}| \leq |\hat{\Sigma}_{n,m} - \check{\Sigma}_{n,m}| + |\check{\Sigma}_{n,m} - \Sigma_{n,m}|, \quad (\text{A.1})$$

for

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

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

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

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

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

by an application of Chebychev's inequality and the fact that, by another application of Hansen (1991, Corollary 3), $\sup_{1 \leq i \leq n} \mathbb{E}((\sum_{t=1}^m (z_{it}^2 - \mathbb{E}(z_{it}^2)))^2) = O(m)$ 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

$$1/\sqrt{nm} \sum_{i=1}^n \sum_{t=1}^m z_{it} \varepsilon_{it} - (\sqrt{n/m}) 1/n \sum_{i=1}^n \left(1/\sqrt{m} \sum_{t=1}^m z_{it} \right) \left(1/\sqrt{m} \sum_{t=1}^m \varepsilon_{it} \right), \quad (\text{A.2})$$

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 $\lim_{n,m \rightarrow \infty} \Omega_{n,m}$.

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

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

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

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

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

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

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

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

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),

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

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 $\sup_{1 \leq i \leq n} \text{var}(\chi_i) = O(1)$. Therefore (A.3) holds.

Combined all results obtained reveals that

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

from which the theorem follows readily. \square

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

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

where

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

and $\hat{\varepsilon}_{i(p'-1)q+q'}^* := \hat{\varepsilon}_{i\varpi_{p'}+q'}$, 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, $\hat{\varepsilon}_{it} := (y_{it} - \bar{y}_i) - (x_{it} - \bar{x}_i)\hat{\beta}$. 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

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

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

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

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

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

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

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

Now, by definition of the moving block bootstrap scheme in the first step and the fact that $\varpi_1, \dots, \varpi_p$ are i.i.d. from the uniform distribution on the set $\{0, 1, \dots, m-q\}$ —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

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

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

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

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

$$\begin{aligned} \mathbb{E} \left(\left| \frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m (\mathbb{E}(x_{it}^2) - \mathbb{E}^*(x_{it}^{*2})) \right| \right) &\leq \frac{1}{nm} \sum_{i=1}^n \mathbb{E} \left(\left| \mathbb{E}^* \left(\sum_{t=1}^m (x_{it}^{*2} - \mathbb{E}(x_{it}^2)) \right) \right| \right) \\ &\leq \frac{1}{nm} \sum_{i=1}^n \sqrt{\mathbb{E} \left(\mathbb{E}^* \left(\left| \sum_{t=1}^m (x_{it}^{*2} - \mathbb{E}(x_{it}^2)) \right|^2 \right) \right)}. \end{aligned}$$

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,

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

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

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

Here,

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

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

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

and we have already shown that $\frac{1}{n} \sum_{i=1}^n \bar{z}_i^2 = o_P(1)$ 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

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

we obtain

$$(\bar{x}_i^* - \mathbb{E}(\bar{x}_i))^2 = \frac{1}{p^2 q^2} \sum_{p'_1=1}^p \sum_{q'_1=1}^q \sum_{p'_2=1}^p \sum_{q'_2=1}^q z_{i\varpi_{p'_1}+q'_1} z_{i\varpi_{p'_2}+q'_2}.$$

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,

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

Here, the last term involves the bootstrap mean $\mathbb{E}^*(\bar{z}_i^*) = \frac{1}{q(m-q+1)} \sum_{t=0}^{m-q} \sum_{q'=1}^q z_{it+q'}$. 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

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

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

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

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

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

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

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

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),

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

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

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

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

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

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

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

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

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

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

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

where χ_i was defined in the proof of Theorem 1 and

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

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

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

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

$$\chi_i^* = \frac{1}{pq} \sum_{p'_1=1}^p \sum_{q'_1=1}^q \sum_{p'_2=1}^p \sum_{q'_2=1}^q z_{i\varpi_{p'_1}+q'_1} \varepsilon_{i\varpi_{p'_2}+q'_2}.$$

Note that

$$\chi_i^* = \frac{1}{pq} \sum_{p'=1}^p \sum_{q'_1=1}^q \sum_{q'_2=1}^q z_{i\varpi_{p'}+q'_1} \varepsilon_{i\varpi_{p'}+q'_2} + \frac{1}{pq} \sum_{p'_1=1}^p \sum_{q'_1=1}^q \sum_{p'_2 \neq p'_1}^p \sum_{q'_2=1}^q z_{i\varpi_{p'_1}+q'_1} \varepsilon_{i\varpi_{p'_2}+q'_2} \quad (\text{A.9})$$

holds.

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

$$\frac{1}{pq} \sum_{p'=1}^p \sum_{q'_1=1}^q \sum_{q'_2=1}^q z_{i\varpi_{p'}+q'_1} \varepsilon_{i\varpi_{p'}+q'_2} = \hat{\gamma}_i^* + q \bar{z}_i^* \bar{\varepsilon}_i^*,$$

where

$$\hat{\gamma}_i^* := \frac{1}{p} \sum_{p'=1}^p \left(\frac{1}{\sqrt{q}} \sum_{q'_1=1}^q (z_{i\varpi_{p'}+q'_1} - \bar{z}_i^*) \right) \left(\frac{1}{\sqrt{q}} \sum_{q'_2=1}^q (\varepsilon_{i\varpi_{p'}+q'_2} - \bar{\varepsilon}_i^*) \right).$$

We will first show that

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

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

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

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

Let

$$\tilde{\gamma}_i := \frac{1}{p} \sum_{p'=1}^p \left(\frac{1}{\sqrt{q}} \sum_{q'_1=1}^q (z_{i\varpi_{p'}+q'_1} - \mathbb{E}^*(\bar{z}_i^*)) \right) \left(\frac{1}{\sqrt{q}} \sum_{q'_2=1}^q (\varepsilon_{i\varpi_{p'}+q'_2} - \mathbb{E}^*(\bar{\varepsilon}_i^*)) \right).$$

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

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

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

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

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,

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

Next, exploiting the fact that the $\varpi_1, \dots, \varpi_p$ 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

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

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

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

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

$$\frac{1}{pq} \sum_{p'_1=1}^p \sum_{q'_1=1}^q \sum_{p'_2 \neq p'_1}^p \sum_{q'_2=1}^q z_{i\varpi_{p'_1}+q'_1} \varepsilon_{i\varpi_{p'_2}+q'_2} = \frac{1}{p} \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p \tilde{X}_{i\varpi_{p'_1}} \tilde{Y}_{i\varpi_{p'_2}}.$$

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 $\varpi_1, \dots, \varpi_p$ are independent, and so

$$\mathbb{E}^* \left(\frac{1}{p} \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p \tilde{X}_{i\varpi_{p'_1}} \tilde{Y}_{i\varpi_{p'_2}} \right) = (p-1)q \mathbb{E}^*(\tilde{z}_i^*) \mathbb{E}^*(\tilde{\varepsilon}_i^*).$$

Define

$$r_i^* := \frac{1}{p} \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p (\tilde{X}_{i\varpi_{p'_1}} \tilde{Y}_{i\varpi_{p'_2}} - \mathbb{E}^*(\tilde{X}_{i\varpi_{p'_1}} \tilde{Y}_{i\varpi_{p'_2}})).$$

Then, collecting terms,

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

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

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

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

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

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

$$r_i^* = (p-1)/p \sum_{p'=1}^p X_{i\varpi_{p'}} \mathbb{E}^*(\tilde{Y}_{i\varpi}) + (p-1)/p \sum_{p'=1}^p Y_{i\varpi_{p'}} \mathbb{E}^*(\tilde{X}_{i\varpi}) + 1/p \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p X_{i\varpi_{p'_1}} Y_{i\varpi_{p'_2}}.$$

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

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

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

$$\mathbb{E}^* \left(1/p^2 \sum_{p'_1=1}^p \sum_{p'_2=1}^p X_{i\varpi_{p'_1}} X_{j\varpi_{p'_2}} \right) = \mathbb{E}^* \left(1/p^2 \sum_{p'=1}^p X_{i\varpi_{p'}} X_{j\varpi_{p'}} \right) = 1/p \mathbb{E}^*(X_{i\varpi} X_{j\varpi}),$$

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

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

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

$$1/np \sum_{i=1}^n \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p X_{i\varpi_{p'_1}} Y_{i\varpi_{p'_2}} = o_{P^*}(1), \quad (\text{A.12})$$

we begin by calculating its conditional variance,

$$1/n^2 \sum_{i=1}^n \sum_{j=1}^n 1/p^2 \sum_{p'_1=1}^p \sum_{p'_2 \neq p'_1}^p \sum_{p''_1=1}^p \sum_{p''_2 \neq p''_1}^p \mathbb{E}^*(X_{i\varpi_{p'_1}} Y_{i\varpi_{p'_2}} X_{j\varpi_{p''_1}} Y_{j\varpi_{p''_2}}).$$

The conditional mean inside the summation is zero unless either (i) $p'_1 = p''_1$ and $p'_2 = p''_2$ or (ii) $p'_1 = p''_2$ and $p'_2 = p''_1$ because both $X_{i\varpi}$ and $Y_{i\varpi}$ are mean zero and the $\varpi_1, \dots, \varpi_p$ are independent. The contribution of the Case (i) terms to the conditional variance is given by

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

while the contribution of the Case (ii) terms equals

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

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

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

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. \square

References

- Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.
- 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 R. Carrasco (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics* 115, 125–157.
- 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.
- Browning, M. and V. Lechene (2003). Children and demand: Direct and non-direct effects. *Review of Economics of the Household* 1, 9–31.
- Cattaneo, M. D. and M. Jansson (2018). Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica* 86, 955–995.
- 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 esti-

- mation 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.
- Dhaene, G. and K. Jochmans (2016). Bias-corrected estimation of panel vector autoregressions with fixed effects. *Economics Letters* 145, 98–103.
- 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.
- Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.
- 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. *Econo-*

- metrica* 92, 411–427.
- Jordà, O. and A. M. Taylor (2025). Local projections. *Journal of Economic Literature* 63, 59–110.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
- Mei, Z., L. Sheng, and Z. Shi (2025). Nickell bias in panel local projections: Financial crises are worse than you think. Mimeo.
- 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 r th 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.