WORKING PAPERS

N° 1754

May 2026

Iterated-Bootstrap Inference For Panel-Data Models

Valérie Heller and Koen Jochmans

Toulouse School of Economics

ITERATED-BOOTSTRAP INFERENCE FOR PANEL-DATA MODELS

Valérie Heller*

Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

Koen Jochmans†

Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France

This version: May 30, 2026

Abstract

Fixed-effect estimators for panel data models suffer from bias. In an $n \times m$ panel the bias is usually of order $1/m$, implying that it is non-negligible unless $n/m \to 0$. Moreover, the limit distribution features a bias term when n and m grow at the same rate. A recent literature has shown that bootstrap inference can correctly account for this asymptotic bias. This implies that inference based on the fixed-effect estimator, when performed by means of the bootstrap, behaves on par with inference based on a bias-corrected estimator. Both procedures are correct provided that $n/m^3 \to 0$. This rate arises because the bootstrap, like bias correction, introduces additional bias of order $1/m^2$. In this paper we argue that, by iterating the bootstrap, one accounts for this higher-order bias, thereby yielding valid inference as long as $n/m^5 \to 0$. The double bootstrap based directly on the (uncorrected) fixed-effect estimator therefore delivers gains equivalent to working with a second-order bias-corrected estimator. To illustrate we provide primitive conditions for iterating a residual bootstrap in the autoregressive model and show by means of a simulation exercise that the gains of iterating the bootstrap are substantial.

JEL Classification: C23

Keywords: bootstrap, higher-order bias correction, panel data

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: valerie.heller@tse-fr.eu.

†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 author 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

Estimators computed from $n \times m$ panel data usually suffer from bias due to the estimation of fixed effects. The bias is of order $^1/_m$, in general, while the variance shrinks like $^1/_{nm}$. This, then, implies that the estimator is asymptotically unbiased only when $^n/_m \to 0$. The literature has investigated several ways of accounting for the bias for the purpose of inference under asymptotics where both n and m grow large. The initial literature proceeded by working with bias-corrected estimators constructed as to remove the leading $^1/_m$ bias (Arellano and Hahn 2007 provide an overview). Such a strategy typically leads to correct inference as long as $^n/_{m^3} \to 0$ (Sartori 2003, Hahn and Newey 2004). Several more recent contributions have shown that the bootstrap, applied directly to the uncorrected estimator, delivers the same inferential gains (Gonçalves and Kaffo 2015, Higgins and Jochmans 2024, 2025).

The intuition behind the requirement that $^n/_{m^3} \to 0$ is that both the bias-correction approach and the bootstrap, while accounting for the leading $^1/_m$ bias, do not correctly capture the remaining $^1/_{m^2}$ bias, as they introduce additional bias terms of order $^1/_{m^2}$. This higher-order bias becomes important in situations where $^n/_{m^3}$ is not close to zero. In this paper we show how simply iterating the bootstrap (Hall 1986, Beran 1987) reduces the bootstrap error to order $^1/_{m^3}$, thereby providing correct inference as long as $^n/_{m^5} \to 0$. So, whereas the (single-layer) bootstrap automatically achieves the gains from a first-order bias correction, the double bootstrap yields inference on par with what a second-order bias correction would achieve. However, while many approaches to first-order bias correction are available, very little is known about second-order corrections. Only the split-panel jackknife corrections of Dhaene and Jochmans (2015) have been justified theoretically. These, however, have the disadvantage that they demand to split the panel into subpanels, which may have a detrimental effect on the small-sample performance of the estimator. The jackknife, at least when applied to higher order, can also be demanding in terms of stationarity requirements.

Our approach to iterating the bootstrap is to be contrasted with using the double bootstrap as a device to prepivot a statistic to account for bias. This route is followed in Kim and Sun (2016). It amounts to a particular way of bootstrapping a (first-order) bias-corrected statistic (as is discussed, for example, in Gonçalves, Cavaliere, Nielsen and Zanelli 2024), and does not improve over a single-layer bootstrap applied directly to the uncorrected estimator.

Our main illustration concerns inference in the autoregressive model with fixed effects, where the within-group estimator is subject to the well-known Nickell (1981) bias. For this problem, Hahn and Kuersteiner (2002) have shown how first-order bias correction yields correct inference when ${}^{n}\!/_{m}$ converges to a positive constant, and Gonçalves and Kaffo (2015) established the validity of a (single-layer) bootstrap scheme under the same asymptotics. We first improve on their results in showing that they continue to yield valid inference as long as ${}^{n}\!/_{m^3} \to 0$. We also show that this condition is sharp, as the bias of order ${}^{1}\!/_{m^2}$ is not correctly captured by either of the two procedures. We then show how the double bootstrap fixes this problem, and correctly accounts for bias up to order ${}^{1}\!/_{m^3}$, thereby leading to size-correct inference as long as ${}^{n}\!/_{m^5} \to 0$. It is also interesting to note that, while the first-order bias in the within-group estimator does not depend on the distribution of the initial conditions, the second-order bias does. As such, a split-sample approach to higher-order bias correction cannot be justified here unless the time-series processes are in steady state at the beginning of the sampling period. To illustrate the improvement of iterating the bootstrap we provide some numerical results on the distribution of p-values in settings where ${}^{n}\!/_{m^3}$ is bounded away from zero (and infinity). The gains from iterating the bootstrap are considerable.

In the next section we describe the general inference problem for panel data. We then show to what extent the bootstrap accounts for bias, and how the double bootstrap improves on this. We initially focus on the distribution of p-values; critical values for tests or confidence intervals are discussed afterwards. We then turn to the autoregressive model, where our results are shown to hold as soon as moments of more than order four exist. All proofs are collected in the Appendix.

2 Incidental-parameter bias in panel data

Consider a statistic T_{nm} constructed from $n \times m$ panel data which, as $n, m \to \infty$, satisfies

$$T_{nm} - B_{nm} \xrightarrow{d} Z,$$

for a random variable Z that follows a mean-zero normal distribution G and a bias term B_{nm} arising from the presence of fixed effects. Typically, $B_{nm} = O(\sqrt{n/m})$, implying that the bias is non-negligible for the purpose of inference unless $n/m \to 0$. To see this, it suffices to focus on the scalar case and consider the distribution of $G(T_{nm})$, the (left-sided) p-value constructed from a naive normal approximation to T_{nm}.[1] From the continuity of G we have

$$G(T_{nm}) \xrightarrow{d} G(Z + B_{nm}) = G(Z) + O(B_{nm}) = V + O(\sqrt{n/m}),$$

with $V \sim \text{Uniform}[0, 1]$. Hence, the distribution of the p-value is asymptotically uniform only when $n/m \to 0$.

2.1 Bootstrap p-values

Performing inference by means of the bootstrap instead of the naive normal approximation mitigates the bias problem. To explain why this is so let T^*_{nm} be the bootstrap version of the statistic T_{nm}, and let $\mathbb{P}^*$ refer to probabilities taken with respect to the bootstrap measure (and, therefore, conditional on the original data). The (left-sided) bootstrap p-value equals

$$\check{P} \coloneqq \mathbb{P}^*(T^*_{nm} \leq T_{nm}).$$

Gonçalves and Kaffo (2015) and Higgins and Jochmans (2024, 2025) have shown results of the form

$$T^*_{nm} - \check{B}_{nm} \xrightarrow{d^*} \check{Z}, \tag{2.1}$$

where $\check{Z} \sim G$ and

$$\check{B}_{nm} - B_{nm} = O_p(\sqrt{n/m^3}), \tag{2.2}$$

[1] As stated here, the p-value implicitly abstracts away from the need to estimate any nuisance parameters on which G would depend in the absence of pivotality. In such a case the p-value would be $\hat{G}(T_{nm})$, where $\hat{G}$ is a some (uniformly) consistent estimator of the function G. The distinction is immaterial for the sequel.

and the notation $A^* \xrightarrow{d^*} A$ means that $\mathbb{P}^*(A^* \leq a) \xrightarrow{p} \mathbb{P}(A \leq a)$ for all continuity points a of the distribution of the random variable A. Then

$$\check{P} \xrightarrow{p} \mathbb{P}^*(\check{Z} \leq Z - (\check{B}_{nm} - B_{nm})) \xrightarrow{p} G(Z - (\check{B}_{nm} - B_{nm})) = V + O_p(\sqrt{n/m^3}),$$

which is asymptotically uniform as $n, m \to \infty$ as long as $n/m^3 \to 0$. This constitutes an improvement of the uniform approximation by an order of magnitude in $1/m$ over the normal approximation.

The bootstrap refinement is of the same order as what can be achieved by working with a bias-corrected statistic of the form $T_{nm} - \hat{B}_{nm}$, where $\hat{B}_{nm}$ is an estimator of B_{nm}. Several such estimators have been proposed. Whichever estimator $\hat{B}_{nm}$ is chosen, we usually have

$$\hat{B}_{nm} - B_{nm} = O_p(\sqrt{n/m^3}),$$

and so $T_{nm} - \hat{B}_{nm} \xrightarrow{d} Z + O_p(\sqrt{n/m^3})$ (see, e.g., Hahn and Newey 2004), from which it then follows that

$$G(T_{nm} - \hat{B}_{nm}) \xrightarrow{p} G(Z - (\hat{B}_{nm} - B_{nm})) = V + O_p(\sqrt{n/m^3}).$$

The bootstrap route is attractive as it by-passes the need to construct an explicit estimator of B_{nm}, which can be a complicated task. It also does not require Z to have a pivotal distribution, as G can depend on nuisance parameters, thereby sidestepping the need for T_{nm} to be a suitably studentized statistic. Perhaps more importantly, though, the bootstrap can be iterated, yielding further refinement to the uniform approximation. We turn to this next.

2.2 Double bootstrap p-values

The reason that the bootstrap and the bias-correction approach control only for first-order bias is that both $\check{B}_{nm}$ and $\hat{B}_{nm}$ are biased estimators of B_{nm}, with bias $O(\sqrt{n/m^3})$, in general. In particular,

$$\check{B}_{nm} = B_{nm} + \frac{D_{nm}}{m} + O_p(\sqrt{n/m^5}),$$

for some $D_{nm} = O(\sqrt{n/m})$. Iterating the bootstrap can yield an improvement. To explain, introduce a second-layer of bootstrap sampling. In complete analogy to $\mathbb{P}^*$, let $\mathbb{P}^{**}$ indicate probabilities conditional on the original data and the first-layer bootstrap data. We will write $A^{**} \overset{d^{**}}{\to} A$ to mean that $\mathbb{P}^{**}(A^{**} \leq a) \overset{p^*}{\to} \mathbb{P}(A \leq a)$ for all continuity points a of the distribution of the random variable A, where $A^* \overset{p^*}{\to} A$ means that $\mathbb{P}^*(\|A^* - A\|_2 > \varepsilon) \overset{p}{\to} 0$ for any $\varepsilon > 0$, and $\|\cdot\|_2$ denotes the Euclidean norm. Later on, we will similarly use $A^{**} \overset{p^{**}}{\to} A$ to indicate that $\mathbb{P}^{**}(\|A^{**} - A\|_2 > \varepsilon) \overset{p^*}{\to} 0$ for any $\varepsilon > 0$.

Let T^{**}_{nm} be the second-layer bootstrap version of T_{nm}. Then it is reasonable to expect

$$T^{**}_{nm} - \check{B}^*_{nm} \overset{d^{**}}{\to} \check{\check{Z}}, \tag{2.3}$$

where $\check{\check{Z}} \sim G$,

$$\check{B}^*_{nm} = \check{B}_{nm} + \frac{\check{D}_{nm}}{m} + O_{p^*}(\sqrt{n/m^5}),$$

and, in the latter expansion, the leading term is once again affected by the presence of first-order bias, that is, it behaves like $\check{D}_{nm} = D_{nm} + O_p(\sqrt{n/m^3})$. But in that case we have

$$(\check{B}^*_{nm} - \check{B}_{nm}) = (\check{B}_{nm} - B_{nm}) + O_{p^*}(\sqrt{n/m^5}). \tag{2.4}$$

To see how this is useful for inference, Let

$$\check{\check{P}} := \mathbb{P}^*(\check{P}^* \leq \check{P}), \qquad \check{P}^* := \mathbb{P}^{**}(T^{**}_{nm} \leq T^*_{nm}).$$

As the notation makes clear, $\check{\check{P}}$ is obtained by bootstrapping the (single-layer) bootstrap p-value $\check{P}$ by means of a second-layer bootstrap. Now, proceeding in the same way as we did before, and using that G is a strictly-increasing function, it can be readily verified that

$$\check{\check{P}} \overset{p}{\to} \mathbb{P}^*(\check{Z} - (\check{B}^*_{nm} - \check{B}_{nm}) \leq Z - (\check{B}_{nm} - B_{nm})) = V + O_p(\sqrt{n/m^5}).$$

Hence, whereas the single bootstrap provides a refinement of order $1/m$ in the uniform approximation to the distribution of p-values, the double bootstrap yields an improvement to $1/m^2$. This refinement will be most noticeable when n/m^3 is not close to zero, where the single-layer bootstrap fails.

The bootstrap can be iterated further, in the manner described in Hall (1986). Each iteration would yield a further improvement of one additional order of magnitude provided, of course, that the bias expansion holds up to a sufficient order; the latter is the case in the autoregressive model that we study below, for example, and has equally been shown in other specific cases (Dhaene and Jochmans 2015, 2017). Of course, the computational cost also increases with the order of iteration, whereas the expected returns will only be substantial in increasingly shorter panels. Whereas performing h iterations has a computational cost that is exponential in h, the remaining bias is non-negligible only when m^{2h+1} is not large relative to n.

2.3 Critical values

Before moving on we explain how the gains from iterating the bootstrap can be harnessed for the construction of critical values. By the test-inversion principle, the argument equally applies to confidence sets.

Consider a test that rejects the null hypothesis when

$$T_{nm} < Q_\alpha$$

for critical value Q_α, chosen with the aim to control the size of the test at α, as $n, m \to \infty$. The bias in T_{nm} invalidates the use of critical values from a zero-mean normal distribution,

$$G^{-1}(\alpha) \coloneqq \inf\{Q : \alpha \leq G(Q)\},$$

unless ${}^{n}/_{m} \to 0$.

The quantile function of the bootstrap distribution is

$$\check{Q}_\alpha \coloneqq \inf\{Q : \alpha \leq \mathbb{P}^*(T^*_{nm} \leq Q)\}.$$

To construct a level-α test using the (single-layer) bootstrap we may use the critical value

$$\check{Q}_\alpha = G^{-1}(\alpha) + \check{B}_{nm} + O_p(\sqrt{{}^{n}/_{m^3}}).$$

Indeed,

$$\mathbb{P}(T_{nm} < \check{Q}_\alpha) = \mathbb{P}(Z < G^{-1}(\alpha) + (\check{B}_{nm} - B_{nm})) + o(1) = \alpha + O(\sqrt{n/m^3})$$

follows.

To improve on the above we can use the double bootstrap. To explain how to do so, let

$$\check{Q}^*_\alpha := \inf\{Q : \alpha \leq \mathbb{P}^{**}(T^{**}_{nm} \leq Q)\}.$$

The double-bootstrap test uses the critical value $\check{Q}_{\check{\alpha}}$, where $\check{\alpha}$ is determined by the equality

$$\mathbb{P}^*(T^*_{nm} < \check{Q}^*_{\check{\alpha}}) = \alpha,$$

that is, by constructing a bootstrap test in the first-layer bootstrap that has actual level equal to α. This calibration step, in tandem with the bootstrap-convergence results, implies

$$\check{Q}^*_{\check{\alpha}} + B_{nm} = G^{-1}(\alpha) + (\check{B}^*_{nm} - \check{B}_{nm}) - (\check{B}_{nm} - B_{nm}) + o_{p^*}(1).$$

The size of the double-bootstrap test then is

$$\mathbb{P}(T_{nm} < \check{Q}_{\check{\alpha}}) \to \mathbb{P}(Z < \check{Q}_{\check{\alpha}} + B_{nm}) = \alpha + O(\sqrt{n/m^5}),$$

which improves over the single-layer bootstrap by one additional order of magnitude in $1/m$.

3 Linear autoregression

The model is

$$y_{it} = \eta_i + \alpha y_{it-1} + v_{it}$$

where the v_{it} are independent and identically distributed with mean zero and variance σ^2. The fixed effect η_i and the initial condition y_{i0} are random draws from some joint distribution which we leave unrestricted. The time-series processes are presumed to be stable, in that $-1 < \alpha < 1$.

3.1 Nickell bias

We will focus on the within-group estimator of α, which is

$$\hat{\alpha} := \frac{\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1} - \bar{y}_{i-})\, y_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1} - \bar{y}_{i-})\, y_{it-1}},$$

where $\bar{y}_{i-} := {}^{1}\!/_{m}\sum_{t=1}^{m} y_{it-1}$. The unit-specific de-meaning arises from treating the η_i as fixed effects and is well-known to introduce bias that is $O(m^{-1})$ in the within-group estimator (Nickell 1981).

Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003) both characterized the asymptotic behavior of the within-group estimator under asymptotics where $n, m \to \infty$ with ${}^{n}\!/_{m}$ converging to a finite constant. For this it suffices to know the bias up to $o(m^{-1})$. They found

$$\sqrt{nm}\left(\hat{\alpha} - \alpha - \frac{c_1(\alpha)}{m}\right) \xrightarrow{d} \mathbf{N}(0, 1-\alpha^2), \qquad c_1(\alpha) := -(1+\alpha).$$

As our interest lies in higher-order corrections, we first generalize this result to higher order.

It is convenient to work with

$$\tau^2 := \frac{\mathbb{E}(y_{i0} - \mu_i)^2}{\gamma^2},$$

where

$$\mu_i := \frac{\eta_i}{1-\alpha} \qquad \text{and} \qquad \gamma^2 := \frac{\sigma^2}{1-\alpha^2}$$

are the long-run steady-state mean and variance of the individual autoregressive processes. As such, τ^2 is a measure of how far away the initial observation y_{i0} is from the relevant stationary distribution (see also Dhaene and Jochmans 2016). To state our results, we let

$$c_2(\alpha, \tau^2) := -(1-\alpha)^{-1}(\alpha\,(1+\alpha) - (\tau^2 - 1))$$

in the following theorem.

Theorem 1. *Suppose that $\mathbb{E}(y_{i0}^4) < \infty$, $\mathbb{E}(\eta_i^4) < \infty$, and $\mathbb{E}(v_{it}^4) < \infty$. Then we have that*

$$\sqrt{nm}\left(\hat{\alpha} - \alpha - \frac{c_1(\alpha)}{m} - \frac{c_2(\alpha, \tau^2)}{m^2}\right) \xrightarrow{d} \mathbf{N}(0, 1-\alpha^2)$$

as $n, m \to \infty$ so that ${}^{n}\!/_{m^5} \to 0$.

The theorem implies a sharpening of the limit result of Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003). Indeed, their approximation remains valid as long as $n/m^3 \to 0$. It further reveals that, while the first-order bias does not depend on the joint distribution of (y_{i0}, η_i), the second-order bias does. One noteworthy implication of this is that a jackknife approach to higher-order bias correction (Dhaene and Jochmans 2015) is not theoretically justified unless the processes are in steady state at the beginning of the sampling period. Finally, although the current statement of the theorem suffices for our purposes here, the bias expansion extends to any higher power of $1/m$ which, by an induction argument applied to our proofs, would allow to justify further iterations of the bootstrap.

We turn to the bootstrap next.

3.2 Bootstrap approximations

In the current context a natural bootstrap scheme is a residual bootstrap. In particular, we consider bootstrap samples $y_{i0}^*, \ldots, y_{im}^*$ generated recursively, starting at $y_{i0}^* = y_{i0}$ followed by iterating on

$$y_{it}^* = \hat{\eta}_i + \hat{\alpha} y_{it-1}^* + v_{it}^*,$$

where $\hat{\eta}_i \coloneqq \bar{y}_i - \hat{\alpha}\bar{y}_{i-}$ and each v_{it}^* is a random draw from the set of residuals $\{\hat{v}_{i1}, \ldots, \hat{v}_{im}\}$, where $\hat{v}_{it} \coloneqq y_{it} - \hat{\eta}_i - \hat{\alpha} y_{it-1}$. Given the bootstrap data we then estimate the autoregressive parameter $\hat{\alpha}$ by the within-group estimator

$$\hat{\alpha}^* \coloneqq \frac{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \bar{y}_{i-}^*)\, y_{it}^*}{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \bar{y}_{i-}^*)\, y_{it-1}^*}.$$

This bootstrap is different from the one considered in Gonçalves and Kaffo (2015) in that they use a wild-bootstrap implementation and set $y_{i0}^* = (1-\hat{\alpha})^{-1}\,\hat{\eta}_i \eqqcolon \hat{\mu}_i$, whereas we maintain $y_{i0}^* = y_{i0}$, which is more natural when initial conditions are left unrestricted. Furthermore, in light of Theorem 1, the latter subtlety will be important for the bootstrap to be higher-order correct.[2]

[2] Gonçalves and Kaffo (2015) allow for the v_{it} to have unit-specific variances σ_i^2. Our bootstrap scheme continues to apply to this case and the results require no modification. In fact, under a full i.i.d. assumption

Theorem 2. *Suppose that* $\mathbb{E}(|y_{i0}|^{2r}) < \infty$, $\mathbb{E}(|\eta_i|^{2r}) < \infty$, *and* $\mathbb{E}(|v_{it}|^{2r}) < \infty$ *for an* $r > 2$. *Then we have that*

$$\sqrt{nm}\left(\hat{\alpha}^* - \hat{\alpha} - \frac{c_1(\alpha)}{m} - \frac{c_2(\alpha, \tau^2)}{m^2} + \frac{c_1(\alpha)}{m^2}\right) \overset{d^*}{\to} \mathbf{N}(0, 1-\alpha^2)$$

as $n, m \to \infty$ *so that* $n/m^5 \to 0$.

This result improves on Gonçalves and Kaffo (2015) in that it implies that the bootstrap correctly reproduces the limit distribution of the (biased) within-group estimator as long as $n/m^3 \to 0$; their Theorem 3.1 presumes that n/m converges to a (finite) positive constant. The theorem also shows how the second-order bias in the bootstrap distribution differs from that in the original within-group estimator, highlighting that the condition that $n/m^3 \to 0$ is sharp.

Theorems 1 and 2, when combined, justify the use of the (reverse-percentile) bootstrap to perform inference, in that critical values computed from it or, equivalently, confidence intervals constructed through it, will lead to asymptotically size-correct inference, and the distribution of p-values will be asymptotically uniform as long as $n/m^3 \to 0$. We further remark that, because the asymptotic variance in Theorem 1 is $1 - \alpha^2$, and $\hat{\alpha} \overset{p}{\to} \alpha$ and $\hat{\alpha}^* \overset{p^*}{\to} \alpha$ as $n, m \to \infty$, the same holds true if these quantities are constructed based on the studentized estimator.

3.3 Iterated-bootstrap approximations

The double bootstrap re-applies the residual bootstrap to the data obtained in each draw of the first-layer bootstrap. Thus, starting at $y_{i0}^{**} = y_{i0}^* = y_{i0}$, we recursively generate data

$$y_{it}^{**} = \hat{\eta}_i^* + \hat{\alpha}^* y_{it-1}^{**} + v_{it}^{**},$$

of the v_{it}, the residual bootstrap could equally be applied by resampling from the full set of residuals across both units and time. In our simulations this had no noticeable impact on performance, however. In either of the two bootstrap designs, it is also possible to rescale the residuals by $\sqrt{m/(m-1)}$; see e.g., MacKinnon (2006). This degrees-of-freedom correction, while theoretically of note, again did not have any observable impact in our numerical experiments. Alternatively, when the v_{it} have time-specific variances σ_t^2 both bias and variance formulae are affected. A wild-bootstrap scheme would correctly capture the bias in that case.

by randomly drawing errors v_{it}^{**} from the first-layer bootstrap residuals $\{\hat{v}_{i1}^*, \ldots, \hat{v}_{im}^*\}$, where $\hat{v}_{it}^* \coloneqq y_{it}^* - \hat{\eta}_i^* - \hat{\alpha}^* y_{it-1}^*$ for $\hat{\eta}_i^* \coloneqq \bar{y}_i^* - \hat{\alpha}^* \bar{y}_{i-}^*$. On these data we then compute the estimator

$$\hat{\alpha}^{**} \coloneqq \frac{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^{**} - \bar{y}_{i-}^{**})\, y_{it}^{**}}{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^{**} - \bar{y}_{i-}^{**})\, y_{it-1}^{**}}.$$

All of this is in complete analogy to the first-layer bootstrap.

The next theorem concerns the limit behavior of the second-layer bootstrap. Notice that the theorem requires no additional conditions relative to the theorem for the single-layer bootstrap.

Theorem 3. *Suppose that* $\mathbb{E}(|y_{i0}|^{2r}) < \infty$, $\mathbb{E}(|\eta_i|^{2r}) < \infty$, *and* $\mathbb{E}(|v_{it}|^{2r}) < \infty$ *for an* $r > 2$. *Then we have that*

$$\sqrt{nm} \left(\hat{\alpha}^{**} - \hat{\alpha}^* - \frac{c_1(\alpha)}{m} - \frac{c_2(\alpha, \tau^2)}{m^2} + 2\, \frac{c_1(\alpha)}{m^2} \right) \overset{d^{**}}{\to} \mathbf{N}(0, 1 - \alpha^2)$$

as $n, m \to \infty$ *so that* ${}^{n}\!/\!{}_{m^5} \to 0$.

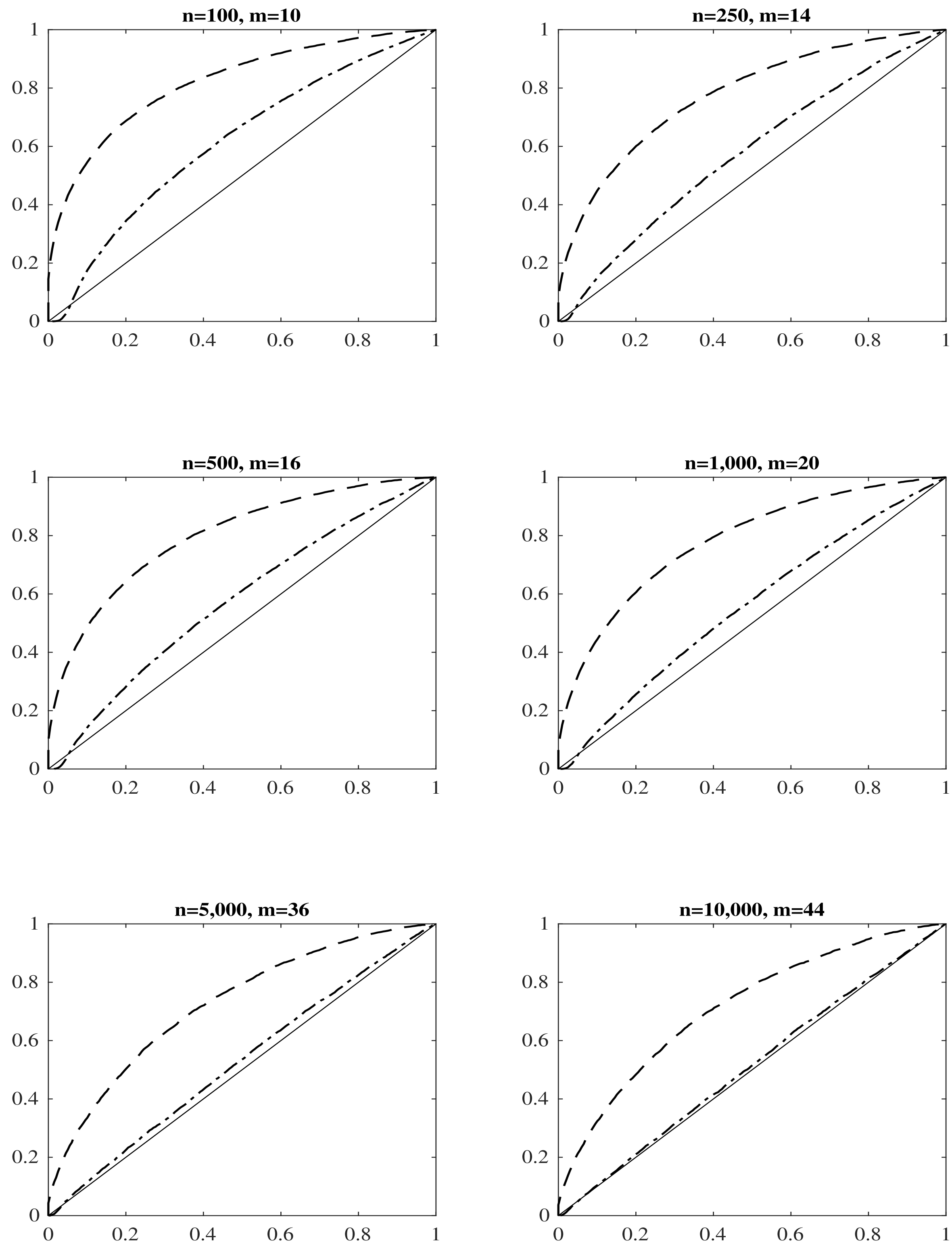
Whereas Theorem 2 validates (2.1) and (2.2) for the autoregressive model, Theorem 3 yields (2.3) and (2.4). Taken together, they then imply that inference based on the double bootstrap correctly accounts for bias in the within-group estimator up to order ${}^{1}\!/\!{}_{m^2}$, instead of only ${}^{1}\!/\!{}_{m}$. Again, $\hat{\alpha}^{**} - \alpha = o_{p^{**}}(1)$, and so the claim equally applies to the studentized estimator.

3.4 Numerical illustrations

We next provide numerical support for our analysis. We simulated data from autoregressive processes with $\alpha = {}^{1}\!/\!{}_{2}$, all fixed effects set to zero, and all initial conditions set to zero, in which case τ^2 is equal to zero. The innovations were drawn from the standard-normal distribution.

Figure 1 contains the cumulative distribution function (over a total of $10,000$ Monte Carlo replications for each design) of p-values obtained via the bootstrap (computed from 199 replications; dashed line) and by the iterated bootstrap (computed through a further round of 199 replications in the inner loop; dashed-dotted line) for different combinations

Figure 1: Empirical distribution of p-values



of n and m. Each plot also contains the 45° line (the uniform distribution; full line) as a reference point.

The data sizes consider have $m = 2\lceil \sqrt[3]{n} \rceil$ for growing values of n. This is to mimic an asymptotic scheme where $n, m \to \infty$ so that n/m^3 is non-vanishing. In such a case, p-values obtained via either a normal or a bootstrap approximation to a bias-corrected estimator (such as, say, Hahn and Kuersteiner 2002 or Dhaene and Jochmans 2015) or through a single-layer bootstrap of the within-group estimator are affected by (second-order) bias and are, therefore, not asymptotically uniform.[3] The p-values computed via the double bootstrap, in contrast, remain valid.

The plots in the figure show a clear improvement of iterating the bootstrap for all sample sizes considered. Furthermore, as predicted by the theory, as the sample size grows, the distribution of the p-values from the (single-run) bootstrap converges to one that lies above the uniform. This is so because it does not correctly replicate the second-order bias in the within-group estimator. Iterating the bootstrap corrects this error in the original bootstrap and so remains correct as long as $n/m^5 \to 0$. Indeed, the plots confirm that, as n and m increase, the empirical distribution of the iterated-bootstrap p-values approaches the uniform distribution.

[3] We note that our derivations for the autoregressive model allow to generalize the bias-correction approach of Hahn and Kuersteiner (2002), $\hat{\hat{\alpha}}_1 \coloneqq \hat{\alpha} - c_1(\hat{\alpha})/m$, to second order. Moreover, the adjusted estimator

$$\hat{\hat{\alpha}}_2 \coloneqq \hat{\alpha} - \frac{c_1(\hat{\alpha})}{m} - \frac{c_2(\hat{\alpha}, \hat{\tau}^2)}{m^2} + \frac{c_1(\hat{\alpha})}{m^2}$$

is free of bias to order $1/m^3$. Consequently, $\sqrt{nm}(\hat{\hat{\alpha}}_2 - \alpha) \xrightarrow{d} \mathbf{N}(0, 1-\alpha^2)$ as $n, m \to \infty$ with $n/m^5 \to 0$, whereas $\sqrt{nm}(\hat{\hat{\alpha}}_1 - \alpha) \xrightarrow{d} \mathbf{N}(0, 1-\alpha^2)$ requires that $n/m^3 \to 0$. In our simulations, inference based on a normal approximation applied to $\hat{\hat{\alpha}}_1$ and $\hat{\hat{\alpha}}_2$ performed similarly to the bootstrap, with distributions somewhat further away from the 45° line. The curves for the naive approach that does not account for bias is degenerate at zero in all our designs.

4 Conclusion

The bootstrap has been found useful as a means to circumvent the incidental-parameter problem in panel-data models, in that it automatically accounts for the first-order bias in the fixed-effect estimator. We show how iterating the bootstrap correctly accounts for bias to second order. Moreover, with $n \times m$ panel data, the (single-layer) bootstrap is correct provided that $n/m^3 \to 0$ whereas the double bootstrap only requires that $n/m^5 \to 0$. As such, iterating the bootstrap on the (uncorrected) fixed-effect estimator yields gains on par with working with a second-order bias-corrected estimator. In the linear autoregressive model these gains were shown to come under no extra conditions, and simulations show them to be substantial.

Appendix

Proof of Theorem 1. Let A_m be the $m \times m$ strictly lower-triangular matrix

$$(A_m)_{t,t'} \coloneqq \begin{cases} \alpha^{t-t'-1} & \text{if } t > t' \\ 0 & \text{if } t \leq t' \end{cases}$$

and let a_m be the column vector of length m with $(a_m)_t \coloneqq \alpha^{t-1}$. Backward substitution towards the initial condition allows to write $w_i \coloneqq (y_{i0}, \ldots, y_{im-1})'$, the vector of lagged outcomes, as

$$w_i = A_m v_i + a_m(y_{i0} - \mu_i) + \iota_m \, \mu_i, \tag{A.1}$$

where $v_i \coloneqq (v_{i1}, \ldots, v_{im})'$ and ι_m is the vector of ones of length m. For the within-group estimator we then have

$$\hat{\alpha} - \alpha = \frac{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1} - \bar{y}_{i-}) \, v_{it}}{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1} - \bar{y}_{i-}) \, y_{it-1}} = \frac{\sum_{i=1}^n w_i' M_m v_i}{\sum_{i=1}^n w_i' M_m w_i},$$

where $M_m \coloneqq I_m - \iota_m(\iota_m' \iota_m)\iota_m'$ is the $m \times m$ demeaning matrix; in particular, for our purposes, $M_m w_i = w_i - \iota_m \bar{y}_{i-}$.

Using (A.1),

$$b_m \coloneqq \frac{\mathbb{E}(w_i' M_m v_i)}{m} = \frac{\mathbb{E}(v_i' M_m A_m v_i)}{m} = \frac{\sigma^2 \operatorname{tr}(M_m A_m)}{m} = -\frac{\sigma^2}{m} \frac{(\iota_m' A_m \iota_m)}{m},$$

because $\mathbb{E}(v_i v_i') = \sigma^2 I_m$ and the matrix A_m has only zeros on the diagonal. Similarly, again using (A.1),

$$\gamma_m^2 := \frac{\mathbb{E}(w_i' M_m w_i)}{m} = \frac{\mathbb{E}(v_i' A_m' M_m A_m v_i)}{m} + \frac{\mathbb{E}(y_{i0}-\mu_i)^2\,(a_m' M_m a_m)}{m}.$$

Here,

$$\frac{\mathbb{E}(v_i' A_m' M_m A_m v_i)}{m} = \frac{\sigma^2\,\mathrm{tr}(A_m' M_m A_m)}{m}, \qquad \frac{\mathbb{E}(y_{i0}-\mu_i)^2\,(a_m' M_m a_m)}{m} = \frac{\sigma^2}{1-\alpha^2}\frac{\tau^2\,(a_m' M_m a_m)}{m},$$

using the definition of τ^2 given in the main text. Now consider the ratio of the expectations,

$$\beta_m := \frac{b_m}{\gamma_m^2} = -\frac{m^{-1}(\iota_m' A_m \iota_m)}{\mathrm{tr}(A_m' M_m A_m) + \tau^2\,(1-\alpha^2)^{-1}\,(a_m' M_m a_m)},$$

which depends only on m, the length of the panel, on α, the autoregressive parameter, and on τ^2, our measure of how much the initial conditions of the individual time series processes deviate from the respective steady-state means. The ratio β_m is an exact expression for the fixed-m probability limit of the within-group estimator. When $\tau^2 = 1$ it corresponds to the expression given in Nickell (1981).

We first consider the behavior of β_m as $m \to \infty$. A few lines of algebra suffice to obtain

$$\iota_m' A_m \iota_m = \frac{1}{1-\alpha}\left(1 - \frac{1}{m}\frac{1-\alpha^m}{1-\alpha}\right), \qquad a_m' M_m a_m = \frac{1-\alpha^{2m}}{1-\alpha^2} - \frac{1}{m}\left(\frac{1-\alpha^m}{1-\alpha}\right)^2,$$

and

$$\mathrm{tr}(A_m' M_m A_m) = \left(\frac{m}{1-\alpha^2} + \frac{1}{(1-\alpha)^2}\right)\left(1 - \frac{1}{m}\frac{1-\alpha^{2m}}{1-\alpha^2}\right) - \frac{2}{(1-\alpha)^2}\left(1 - \frac{1}{m}\frac{1-\alpha^m}{1-\alpha}\right).$$

Using that, as $m \to \infty$, the quantity α^m approaches zero faster than any power of $1/m$, we can proceed as in Dhaene and Jochmans (2015, Online Appendix) and obtain the expansion

$$\beta_m = \frac{c_1(\alpha)}{m} + \frac{c_2(\alpha,\tau^2)}{m^2} + O(m^{-3}),$$

for coefficients

$$c_1(\alpha) := -(1+\alpha), \qquad c_2(\alpha,\tau^2) := -(1-\alpha)^{-1}(\alpha(1+\alpha) - (\tau^2 - 1)).$$

The expansion can be extended up to any order but the current form suffices for our purposes.

Next we turn to the asymptotic behavior of the numerator and denominator of the within-group estimator, when centered around their expectation. Consider the numerator,

$$\frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}\left(\left(y_{it-1}-\bar{y}_{i-}\right)v_{it}-b_m\right)=\frac{1}{nm}\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1}-\mu_i)\,v_{it}-\frac{1}{n}\sum_{i=1}^{n}\left(\left(\bar{y}_{i-}-\mu_i\right)\bar{v}_i+b_m\right).$$

The first term on the right-hand side is a martingale difference sequence. We therefore have

$$\operatorname{var}\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1}-\mu_i)\,v_{it}\right)=\frac{\sigma^2}{m}\,\mathbb{E}\left(\sum_{t=1}^{m}\mathbb{E}((y_{it-1}-\mu_i)^2|y_{i0},\eta_i)\right)$$

by the law of total variance. The conditional variances appearing inside the sum are the diagonal elements of

$$\mathbb{E}((w_i-\mu_i)(w_i-\mu_i)'|\,y_{i0},\eta_i)=\sigma^2\,(A_mA_m')+(y_{i0}-\mu_i)^2\,a_ma_m',$$

where we have once again relied on (A.1). Summing over time then amounts to taking the trace of this matrix. Now,

$$\operatorname{tr}(A_mA_m')=\frac{m}{1-\alpha^2}-\frac{1-\alpha^{2m}}{(1-\alpha^2)^2},\qquad \operatorname{tr}(a_ma_m')=(a_m'a_m)=\frac{1-\alpha^{2m}}{1-\alpha^2},$$

from which

$$\operatorname{var}\left(\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1}-\mu_i)\,v_{it}\right)=\frac{\sigma^4}{1-\alpha^2}+O(m^{-1})=\sigma^2\gamma^2+O(m^{-1}) \tag{A.2}$$

follows readily. Furthermore, by a central limit theorem for martingale difference sequences,

$$\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1}-\mu_i)\,v_{it}\xrightarrow{d}\mathbf{N}(0,\sigma^2\gamma^2)$$

as $n,m\to\infty$. To see this, we verify the conditions of Corollary 5.26 in White (2001) applied to the random variable $z_t:={}^1\!/\!\sqrt{n}\sum_{i=1}^{n}(y_{it-1}-\mu_i)\,v_{it}$. We show that $\mathbb{E}(z_t^4)=O(1)$, that $\operatorname{var}({}^1\!/\!\sqrt{m}\sum_{t=1}^{m}z_t)$ is bounded away from zero for all m sufficiently large, and that ${}^1\!/\!m\sum_{t=1}^{m}z_t^2\xrightarrow{p}\operatorname{var}({}^1\!/\!\sqrt{m}\sum_{t=1}^{m}z_t)$. First, using cross-sectional independence and the fact that $\mathbb{E}(z_t)=0$,

$$\mathbb{E}(z_t^4)=n^{-1}\,\mathbb{E}((y_{it-1}-\mu_i)^4)\,\mathbb{E}(v_{it}^4)+\left(\mathbb{E}((y_{it-1}-\mu_i)^2)\;\mathbb{E}(v_{it}^2)\right)^2=O(1),$$

because all variables involved have finite fourth-order moments. Second, from (A.2), we have

$$\operatorname{var}\left(\frac{1}{\sqrt{m}}\sum_{t=1}^{m} z_t\right) = \frac{\sum_{t=1}^{m}\mathbb{E}(z_t^2)}{m} = \frac{\sum_{t=1}^{m}\mathbb{E}\left((y_{it-1}-\mu_i)^2\, v_{it}^2\right)}{m} = \frac{\sigma^4}{1-\alpha^2} + O(m^{-1}),$$

which is clearly bounded-away from zero for all m sufficiently large. It remains only to show that $1/m\sum_{t=1}^{m}(z_t^2 - \mathbb{E}(z_t^2)) \xrightarrow{p} 0$, which will follow from establishing that its variance,

$$\frac{\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}\mathbb{E}(z_{t_1}^2\, z_{t_2}^2) - \mathbb{E}(z_{t_1}^2)\,\mathbb{E}(z_{t_2}^2)}{m^2}, \tag{A.3}$$

converges to zero. We have

$$\mathbb{E}(z_{t_1}^2\, z_{t_2}^2) = \frac{\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\mathbb{E}(z_{i_1t_1}\, z_{i_2t_1}\, z_{i_3t_2}\, z_{i_4t_2})}{n^2}, \qquad z_{it} \coloneqq (y_{it-1}-\mu_i)\, v_{it}.$$

The expectation in the summand is non-zero only when (i) $i_1 = i_2 = i_3 = i_4$; (ii) $i_1 = i_2$ and $i_3 = i_4$; (iii) $i_1 = i_3$ and $i_2 = i_4$; or (iv) $i_1 = i_4$ and $i_2 = i_3$. The contribution of quadruples of Type (i) to $\mathbb{E}(z_{t_1}^2\, z_{t_2}^2)$ equals

$$\frac{\sum_{i=1}^{n}\mathbb{E}(z_{it_1}^2\, z_{it_2}^2)}{n^2} = O(n^{-1}).$$

Type (ii) terms contribute $\mathbb{E}(z_{it_1}^2)\,\mathbb{E}(z_{it_2}^2)$, which then cancel out with the second term in the summand in (A.3). The contribution of Types (iii) and (iv) are both $\mathbb{E}(z_{it_1} z_{it_2})^2$; note that this expectation is zero unless $t_1 = t_2$, in which case it is $O(1)$. Consequently, the latter two types provide a total contribution to (A.3) that is $O(m^{-1})$. We, therefore, obtain that

$$\frac{\sum_{t_1=1}^{m}\sum_{t_2=1}^{m}\mathbb{E}(z_{t_1}^2\, z_{t_2}^2) - \mathbb{E}(z_{t_1}^2)\,\mathbb{E}(z_{t_2}^2)}{m^2} = O(n^{-1}) + O(m^{-1}),$$

and so our final requirement for the central limit theorem to go through has been verified.

Now turn to variance of the second term in the numerator. After scaling by $\sqrt{nm}$ it is

$$\operatorname{var}\left(\sqrt{\frac{m}{n}}\sum_{i=1}^{n}\left((\bar{y}_{i-}-\mu_i)\,\bar{v}_i + b_m\right)\right) = m\,\mathbb{E}((\bar{y}_{i-}-\mu_i)^2\,\bar{v}_i^2) - m\, b_m^2$$

Here, $m\, b_m^2 = O(m^{-1})$. Further, $m(\bar{y}_{i-}-\mu_i) = \iota_m'(w_i - \mu_i) = \iota_m' A_m v_i + \iota_m' a_m\,(y_{i0}-\mu_i)$, the conditional variance $\mathbb{E}((\bar{y}_{i-}-\mu_i)^2\,\bar{v}_i^2|\, y_{i0}, \eta_i)$ is equal to

$$\frac{\mathbb{E}((\lambda_m' v_i)^2\,(\iota_m' v_i)^2)}{m^4} + (y_{i0}-\mu_i)^2\,\frac{(\iota_m' a_m)^2\,\mathbb{E}((\iota_m' v_i)^2)}{m^4} + 2(y_{i0}-\mu_i)\,\frac{(\iota_m' a_m)\,\mathbb{E}((\lambda_m' v_i)(\iota_m' v_i)^2)}{m^4},$$

where we use the shorthand $\lambda_m = A_m' \iota_m$. Here, the first term is

$$\frac{\mathbb{E}((\lambda_m' v_i)^2 \, (\iota_m' v_i)^2)}{m^4} = \frac{\sum_{t_1=1}^m \sum_{t_2=1}^m \sum_{t_3=1}^m \sum_{t_4=1}^m (\lambda_m)_{t_1} (\lambda_m)_{t_2} \mathbb{E}(v_{it_1} v_{it_2} v_{it_3} v_{it_4})}{m^4} = O(m^{-2}),$$

because $\mathbb{E}(v_{it_1} v_{it_2} v_{it_3} v_{it_4})$ is zero unless all indices are equal (which happens on m occassions) or two pairs of indices are equal (which happens on $3m^2$ occassions). The expectations in the second and third term are easily seen to be of a smaller order of magnitude. Therefore, $m \, \mathbb{E}((\bar{y}_{i-} - \mu_i)^2 \, \bar{v}_i^2) = O(m^{-1})$, and so

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n ((\bar{y}_{i-} - \mu_i) \, \bar{v}_i + b_m) = o_p(1)$$

follows. Therefore, this second term in the numerator is asymptotically negligible for our purposes.

Next consider the recentered denominator,

$$\frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m ((y_{it-1} - \bar{y}_{i-}) \, y_{it-1} - \gamma_m^2).$$

Given the existence of fourth-order moments of all variables involved it is readily seen that

$$\mathbb{E}\left(\left(\frac{1}{m} \sum_{t=1}^m ((y_{it-1} - \bar{y}_{i-}) \, y_{it-1} \right)^2 \right) = \frac{\mathbb{E}((w_i' M_m w_i) \, (w_i' M_m w_i))}{m^2} = O(1),$$

using (A.1) along with independence of the errors over time. Also, $\gamma_m^2 = O(1)$. Therefore,

$$\frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m ((y_{it-1} - \bar{y}_{i-}) \, y_{it-1} - \gamma_m^2) = O_p(n^{-1/2}),$$

which will suffice for our purposes.

Combining the results obtained so far with an expansion yields

$$\sqrt{nm}(\hat{\alpha} - \alpha) = \sqrt{nm} \, \beta_m + \frac{1}{\sqrt{nm}} \sum_{i=1}^n \sum_{t=1}^m \frac{(y_{it-1} - \mu_i) \, v_{it}}{\gamma^2} + o_p(1),$$

noting that $\gamma_m^2 \to \gamma^2$ as $m \to \infty$. Hence, as $n, m \to \infty$, $\sqrt{nm}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathbf{N}(0, 1 - \alpha^2)$ provided that $n/m \to 0$, whereas

$$\sqrt{nm} \left(\hat{\alpha} - \alpha - \frac{c_1(\alpha)}{m} \right) \xrightarrow{d} \mathbf{N}(0, 1 - \alpha^2)$$

provided that ${}^{n}/_{m^3} \to 0$, while

$$\sqrt{nm}\left(\hat{\alpha}-\alpha-\frac{c_1(\alpha)}{m}-\frac{c_2(\alpha,\tau^2)}{m^2}\right)\xrightarrow{d}\mathbf{N}(0,1-\alpha^2)$$

provided that ${}^{n}/_{m^5} \to 0$. □

Proof of Theorem 2. In analogy to the previous proof, let $\hat{A}_m$ be the $m \times m$ matrix with entries

$$(\hat{A}_m)_{t,t'} := \begin{cases} \hat{\alpha}^{t-t'-1} & \text{if } t > t' \\ 0 & \text{if } t \leq t' \end{cases}$$

and let $\hat{a}_m$ be the column vector of length m with $(\hat{a}_m)_t := \hat{\alpha}^{t-1}$. Then we can write $w_i^* := (y_{i0}^*, \ldots, y_{im-1}^*)'$ as

$$w_i^* = \hat{A}_m v_i^* + \hat{a}_m (y_{i0} - \hat{\mu}_i) + \iota_m \hat{\mu}_i, \tag{A.4}$$

where $v_i^* := (v_{i1}^*, \ldots, v_{im}^*)'$ and $\hat{\mu}_i := (1-\hat{\alpha})^{-1}\hat{\eta}_i$, and we recall that the bootstrap is initiated with $y_{i0}^* = y_{i0}$.

We have

$$\hat{\alpha}^* - \hat{\alpha} := \frac{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \bar{y}_{i-}^*)\, v_{it}^*}{\sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \bar{y}_{i-}^*)\, y_{it-1}^*} = \frac{\sum_{i=1}^n w_i^{*\prime} M_m v_i^*}{\sum_{i=1}^n w_i^{*\prime} M_m w_i^*}.$$

Using (A.4),

$$\hat{b}_m := \frac{\sum_{i=1}^n \mathbb{E}^*(w_i^{*\prime} M_m v_i^*)}{nm} = \frac{\sum_{i=1}^n \mathbb{E}^*(v_i^{*\prime} M_m \hat{A}_m v_i^*)}{nm} = \frac{\hat{\sigma}^2 \operatorname{tr}(M_m \hat{A}_m)}{m} = -\frac{\hat{\sigma}^2}{m}\frac{(\iota_m' \hat{A}_m \iota_m)}{m},$$

where $\hat{\sigma}^2 := {}^1/_n \sum_{i=1}^n \hat{\sigma}_i^2 = {}^1/_{nm} \sum_{i=1}^n \sum_{t=1}^m \hat{v}_{it}^2$. This is so because $\mathbb{E}^*(v_i^* v_i^{*\prime}) = \hat{\sigma}_i^2 I_m$ for each $1 \leq i \leq n$. The latter observation follows from noting that $\mathbb{E}^*(v_{it}^{*2}) = {}^1/_m \sum_{t'=1}^m \hat{v}_{it'}^2 = \hat{\sigma}_i^2$ and $\mathbb{E}^*(v_{it_1}^* v_{it_2}^*) = {}^1/_{m^2} \sum_{t_1=1}^m \sum_{t_2=1}^m \hat{v}_{it_1}\hat{v}_{it_2} = ({}^1/_m \sum_{t=1}^m \hat{v}_{it})^2 = 0$ when $t_1 \neq t_2$, because the within-group residuals sum to zero for each $1 \leq i \leq n$ due to the estimation of the fixed effects. In the same way

$$\hat{\gamma}_m^2 := \frac{\sum_{i=1}^n \mathbb{E}^*(w_i^{*\prime} M_m w_i^*)}{nm} = \frac{\sum_{i=1}^n \mathbb{E}^*(v_i^{*\prime} \hat{A}_m' M_m \hat{A}_m v_i^*)}{nm} + \frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^2 \,(\hat{a}_m' M_m \hat{a}_m)}{nm}.$$

Here,

$$\frac{\sum_{i=1}^n \mathbb{E}^*(v_i^{*\prime} \hat{A}_m' M_m \hat{A}_m v_i^*)}{nm} = \frac{\hat{\sigma}^2 \operatorname{tr}(\hat{A}_m' M_m \hat{A}_m)}{m},$$

and

$$\frac{\sum_{i=1}^{n}(y_{i0}-\hat{\mu}_i)^2\,(\hat{a}_m' M_m \hat{a}_m)}{nm} = \frac{\hat{\sigma}^2}{1-\hat{\alpha}^2}\frac{\hat{\tau}^2\,(\hat{a}_m' M_m \hat{a}_m)}{m}, \qquad \hat{\tau}^2 := \frac{{}^1\!/\!_n \sum_{i=1}^{n}(y_{i0}-\hat{\mu}_i)^2}{\hat{\gamma}^2},$$

for $\hat{\gamma}^2 := \hat{\sigma}^2\,(1-\hat{\alpha}^2)^{-1}$. The bootstrap mean of both the numerator and denominator are thus natural plug-in versions of their respective population counterparts. Moreover, from the expansion of β_m in the preceding proof, we have that, conditional on the data, $\hat{\beta}_m$ satisfies the same expansion, with the parameters α and τ^2 replaced by $\hat{\alpha}$ and $\hat{\tau}^2$, respectively. Furthermore, because it is a smooth function of the parameters, as $n, m \to \infty$,

$$\hat{\beta}_m := \frac{\hat{b}_m}{\hat{\gamma}_m^2} = \frac{c_1(\hat{\alpha})}{m} + \frac{c_2(\hat{\alpha},\hat{\tau}^2)}{m^2} + O_p(m^{-3}),$$

because both $\hat{\alpha}$ and $\hat{\tau}^2$ are consistent under such an asymptotic. Moreover, from above we know that $\hat{\alpha}-\alpha = \beta_m + O_p((nm)^{-1/2})$, whereas it is readily deduced that we equally have $\hat{\tau}^2-\tau^2 = O(m^{-1}) + O_p(n^{-1/2})$. A simple expansion of the leading coefficients c_1 and c_2 gives

$$\frac{c_1(\hat{\alpha})}{m} = \frac{c_1(\alpha)}{m} - \frac{(\hat{\alpha}-\alpha)}{m} = \frac{c_1(\alpha)}{m} - \frac{c_1(\alpha)}{m^2} + O(m^{-3}) + O_p\left(\frac{1}{m\sqrt{nm}}\right),$$

and $c_2(\hat{\alpha},\hat{\tau}^2)/m^2 = c_2(\alpha,\tau^2)/m^2 + O(1/m^3) + O_p\left(1/m^2\sqrt{n}\right)$. Consequently, combining results yields

$$\hat{\beta}_m = \frac{c_1(\alpha)}{m} + \frac{c_2(\alpha,\tau^2)-c_1(\alpha)}{m^2} + O_p(m^{-3}) + o_p\left(\frac{1}{\sqrt{nm}}\right), \tag{A.5}$$

and so the expansion for the bootstrap estimator agrees with the expansion for the original estimator to first order, but not beyond.

We now turn to asymptotic normality of the term $1/\sqrt{m}\sum_{t=1}^{m} z_t^*$, for $z_t^* := 1/\sqrt{n}\sum_{i=1}^{n} z_{it}^*$ and $z_{it}^* := (y_{it-1}^*-\hat{\mu}_i)\,v_{it}^*$. To do so we use a conditional version of Corollary 5.26 in White (2001); this approach is similar to the one taken in Gonçalves and Kilian (2004). We first show that, for an $r > 2$, $\mathbb{E}^*(|z_t^*|^r) = O_p(1)$. Note that $\mathbb{E}^*(z_{it}^*) = 0$ and that the z_{it}^* are independent across units conditional on the data. Hence, by Rosenthal's inequality, it suffices to show that

$$\frac{\sum_{i=1}^{n}\mathbb{E}^*(|z_{it}^*|^r)}{n} = O_p(1).$$

By the Cauchy-Schwarz inequality,

$$\frac{\sum_{i=1}^{n}\mathbb{E}^*(|z_{it}^*|^r)}{n} \leq \left(\frac{\sum_{i=1}^{n}\mathbb{E}^*(|y_{it-1}^*-\hat{\mu}_i|^{2r})}{n}\right)^{1/2}\left(\frac{\sum_{i=1}^{n}\mathbb{E}^*(|v_{it}^*|^{2r})}{n}\right)^{1/2}.$$

We handle each term in turn. As $\mathbb{E}^*(v_{it}^{*\,2r}) = {}^1\!/\!_m \sum_{t=1}^m |\hat{v}_{it}|^{2r}$ and the residuals have the form $\hat{v}_{it} = (v_{it} - \bar{v}_i) - (\hat{\alpha} - \alpha)\,(y_{it-1} - \bar{y}_{i-})$, for ${}^1\!/\!_n \sum_{i=1}^n \mathbb{E}^*(|v_{it}^*|^{2r}) = O_p(1)$ to hold it suffices that

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}(|v_{it} - \bar{v}_i|^{2r})}{nm} = O(1), \qquad \frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}(|y_{it-1} - \bar{y}_{i-}|^{2r})}{nm} = O(1).$$

This is the case as $\mathbb{E}(|v_{it}|^{2r}) < \infty$ and $\mathbb{E}(|y_{it-1}|^{2r}) < \infty$. Next, for the second term in the upper bound, using (A.4) and letting $\hat{\chi}_1, \ldots, \hat{\chi}_m$ be the columns of the matrix $\hat{A}_m'$, we have that

$$\mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^{2r}) = \mathbb{E}^*(|\hat{\chi}_t' v_i^* + \hat{\alpha}^{t-1}(y_{i0} - \hat{\mu}_i)|^{2r}) \lesssim \mathbb{E}^*(|\hat{\chi}_t' v_i^*|^{2r}) + |\hat{\alpha}|^{2r(t-1)}\, |y_{i0} - \hat{\mu}_i|^{2r},$$

with the last transition following from the c_r-inequality. Here and later, $A \lesssim B$ means that $A \leq O_p(1)\, B$. Again using Rosenthal's inequality,

$$\mathbb{E}^*(|\hat{\chi}_t' v_i^*|^{2r}) = \mathbb{E}^*\left(\left|\sum_{t_1=1}^m (\hat{\chi}_t)_{t_1} v_{it_1}^*\right|^{2r}\right) \lesssim \sum_{t_1=1}^m |(\hat{\chi}_t)_{t_1}|^{2r}\; \mathbb{E}^*(|v_{it_1}^*|^{2r}) + \left(\sum_{t_1=1}^m (\hat{\chi}_t)_{t_1}^2\; \mathbb{E}^*(v_{it_1}^{*2})\right)^r.$$

Applying Jensen's inequality to the second term on the right-hand side and using that $\mathbb{E}^*(|v_{it}^*|^{2r}) = {}^1\!/\!_m \sum_{t=1}^m |\hat{v}_{it}|^{2r}$ gives

$$\frac{\sum_{i=1}^n \mathbb{E}^*(|\hat{\chi}_t' v_i^*|^{2r})}{n} \lesssim \left(\left(\sum_{t_1=1}^m |(\hat{\chi}_t)_{t_1}|^{2r}\right) + \left(\sum_{t_1=1}^m (\hat{\chi}_t)_{t_1}^2\right)^r\right) \frac{\sum_{i=1}^n \sum_{t_2=1}^m |\hat{v}_{it_2}|^{2r}}{nm} = O_p(1),$$

as $\sum_{t_1=1}^m |(\hat{\chi}_t)_{t_1}|^{2r} \xrightarrow{p} (1 - |\alpha|^{2r})(1 - |\alpha|^{2r(t-1)}) = O(1)$ and ${}^1\!/\!_{nm} \sum_{i=1}^n \sum_{t_2=1}^m |\hat{v}_{it_2}|^{2r} = O_p(1)$ from above. For the remaining term, $\hat{\mu}_i = (1-\hat{\alpha})^{-1}(\eta_i + \bar{v}_i) - (1-\hat{\alpha})^{-1}(\hat{\alpha} - \alpha)\,\bar{y}_{i-}$ gives the decomposition

$$\frac{\sum_{i=1}^n |y_{i0} - \hat{\mu}_i|^{2r}}{n} \lesssim \frac{\sum_{i=1}^n |y_{i0}|^{2r}}{n} + O_p(1) \left(\frac{\sum_{i=1}^n |\bar{y}_{i-}|^{2r}}{n} + \frac{\sum_{i=1}^n |\eta_i|^{2r}}{n} + \frac{\sum_{i=1}^n |\bar{v}_i|^{2r}}{n}\right) = O_p(1),$$

with the last transition again following from the moment requirements in the theorem. We can thus conclude that $\mathbb{E}^*(|z_t^*|^r) = O_p(1)$.

We next turn to the conditional variance of ${}^1\!/\!\sqrt{m} \sum_{t=1}^m z_t^*$ which, by the martingale property, is equal to

$$\frac{\sum_{t=1}^m \mathbb{E}^*(z_t^{*2})}{m} = \left(\frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n}\right) \frac{\operatorname{tr}(\hat{A}_m \hat{A}_m')}{m} + \left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^2\, \hat{\sigma}_i^2}{n}\right) \frac{(\hat{a}_m' \hat{a}_m)}{m}.$$

Here, $m^{-1}\operatorname{tr}(\hat{A}_m\hat{A}_m') = (1-\alpha^2)^{-1} + O_p(m^{-1})$ and $m^{-1}\operatorname{tr}(\hat{a}_m\hat{a}_m') = o_p(1)$. Furthermore, by an expansion,

$$\frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n} = \frac{\sum_{i=1}^n (1/m \sum_{t=1}^m v_{it}^2)^2}{n} + o_p(1).$$

The dominant right-hand side term behaves like

$$\frac{\sum_{i=1}^n (1/m \sum_{t=1}^m v_{it}^2)^2}{n} = \frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2\neq t_1} v_{it_1}^2\, v_{it_2}^2}{nm^2} + \frac{\sum_{i=1}^n \sum_{t=1}^m v_{it}^4}{nm^2} = \sigma^4 + o_p(1),$$

because

$$\frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2\neq t_1} v_{it_1}^2\, v_{it_2}^2}{nm^2} = \sigma^4 - \frac{\sigma^4}{m} + o_p(1), \quad \text{and} \quad \frac{\sum_{i=1}^n \sum_{t=1}^m v_{it}^4}{nm^2} = \frac{\mathbb{E}(v_{it}^4)}{m} + o_p(1).$$

This result, together with the fact that $1/n \sum_{i=1}^n (y_{i0}-\hat{\mu}_i)^4 = O_p(1)$ (as shown above), also implies that

$$\left(\frac{\sum_{i=1}^n (y_{i0}-\hat{\mu}_i)^2\, \hat{\sigma}_i^2}{n}\right) \leq \left(\frac{\sum_{i=1}^n (y_{i0}-\hat{\mu}_i)^4}{n}\right)^{1/2} \left(\frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n}\right)^{1/2} = O_p(1).$$

Taken together,

$$\operatorname{var}^*\left(\frac{\sum_{t=1}^m z_t^*}{\sqrt{m}}\right) \xrightarrow{p} \frac{\sigma^4}{1-\alpha^2} = \sigma^2\,\gamma^2$$

follows.

It remains only to show that $1/m \sum_{t=1}^m (z_t^{*2} - \mathbb{E}(z_t^2)) \xrightarrow{p^*} 0$. As we have already established that $1/m \sum_{t=1}^m (\mathbb{E}^*(z_t^{*2}) - \mathbb{E}(z_t^2)) \xrightarrow{p} 0$, we only need that $1/m \sum_{t=1}^m (z_t^{*2} - \mathbb{E}^*(z_t^{*2})) \xrightarrow{p^*} 0$. Now

$$\frac{\sum_{t=1}^m (z_t^{*2} - \mathbb{E}^*(z_t^{*2}))}{m} = \frac{\sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{t=1}^m (z_{i_1t}^* z_{i_2t}^* - \mathbb{E}^*(z_{i_1t}^* z_{i_2t}^*))}{nm}$$

and, as $\mathbb{E}^*(z_{i_1t}^* z_{i_2t}^*) = 0$ when $i_1 \neq i_2$, we can decompose this as

$$\frac{\sum_{i_1=1}^n \sum_{i_2\neq i_1}^n \sum_{t=1}^m z_{i_1t}^* z_{i_2t}^*}{nm} + \frac{\sum_{i=1}^n \sum_{t=1}^m (z_{it}^{*2} - \mathbb{E}^*(z_{it}^{*2}))}{nm}$$

and analyze both terms separately. The variance of the first term (conditional on the data) is

$$\frac{\sum_{i_1=1}^n \sum_{i_2\neq i_1} \sum_{i_3=1}^n \sum_{i_4\neq i_3} \sum_{t_1=1}^m \sum_{t_2=1}^m \mathbb{E}^*(z_{i_1t_1}^* z_{i_2t_1}^* z_{i_3t_2}^* z_{i_4t_2}^*)}{n^2m^2}.$$

By cross-sectional independence, the expectation in the summand is zero unless either (i) $i_1 = i_3 \neq i_2 = i_4$ or (ii) $i_1 = i_4 \neq i_2 = i_3$ holds. In either of these cases, by the martingale property, the expectation is non-zero only when $t_1 = t_2 = t$ (say), where it equals $\mathbb{E}^*(z_{i_1 t}^{*2})\,\mathbb{E}^*(z_{i_2 t}^{*2})$ in either case. Therefore, the variance of the first term takes the form

$$\frac{\sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{t=1}^m \mathbb{E}^*(z_{i_1 t}^{*2})\,\mathbb{E}^*(z_{i_2 t}^{*2})}{n^2 m^2}.$$

Now,

$$\mathbb{E}^*(z_{it}^{*2}) = \mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^2)\,\hat{\sigma}_i^2 \lesssim \operatorname{tr}(\hat{\chi}_t \hat{\chi}_t')\,\hat{\sigma}_i^4 + \hat{\alpha}^{2(t-1)}\,(y_{i0} - \hat{\mu}_i)^2\,\hat{\sigma}_i^2,$$

and so the variance is bounded from above (up to a multiplicative factor) by the sum of the three terms

$$\frac{\sum_{t=1}^m (\hat{\chi}_t' \hat{\chi}_t)^2}{m^2} \left(\frac{\sum_{i_1=1}^n \hat{\sigma}_i^4}{n}\right)^2, \qquad \frac{\sum_{t=1}^m \hat{\alpha}^{4(t-1)}}{m^2} \left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^2\,\hat{\sigma}_i^2}{n}\right)^2,$$

and

$$2\frac{\sum_{t=1}^m \hat{\alpha}^{2(t-1)}\,(\hat{\chi}_t' \hat{\chi}_t)}{m^2} \frac{\sum_{i_1=1}^n (y_{i_1 0} - \hat{\mu}_{i_1})^2\,\hat{\sigma}_{i_1}^2}{n} \frac{\sum_{i_2=1}^n \hat{\sigma}_{i_2}^4}{n}.$$

Given our results so far it is immediate that each of these terms is of the form $o_p(1)\,O_p(1)$ and, hence, $o_p(1)$. Therefore,

$$\frac{\sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \sum_{t=1}^m z_{i_1 t}^* z_{i_2 t}^*}{nm} \xrightarrow{p^*} 0$$

follows. We then turn our attention to the second term. Its conditional variance is given by

$$\mathbb{E}^*\left(\left(\frac{\sum_{i=1}^n \sum_{t=1}^m (z_{it}^{*2} - \mathbb{E}^*(z_{it}^{*2}))}{nm}\right)^2\right) = \frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2=1}^m \operatorname{cov}^*(z_{it_1}^{*2}, z_{it_2}^{*2})}{n^2 m^2}.$$

Split the sum by cases, with Case (i) $t_1 = t_2$ and Case (ii) $t_1 \neq t_2$. The contribution of Case (i) is $1/n^2m^2 \sum_{i=1}^n \sum_{t=1}^m \mathbb{E}^*(z_{it}^{*4}) - \mathbb{E}^*(z_{it}^{*2})^2$, and we can then focus on establishing that

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}^*(z_{it}^{*4})}{n^2 m^2} = o_p(1).$$

For this we use that $\mathbb{E}^*(z_{it}^{*4}) = \mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^4)\,\mathbb{E}^*(v_{it}^{*4})$ is bounded from above by a multiple of

$$\left(\left(\sum_{t_1=1}^m (\hat{\chi}_t)_{t_1}^4\right) + \left(\sum_{t_1=1}^m (\hat{\chi}_t)_{t_1}^2\right)^2\right) \left(\frac{\sum_{t_2=1}^m \hat{v}_{it_2}^4}{m}\right)^2 + |\hat{\alpha}|^{4(t-1)}\,(y_{i0} - \hat{\mu}_i)^4 \left(\frac{\sum_{t_2=1}^m \hat{v}_{it_2}^4}{m}\right).$$

Here,

$$\frac{1}{n^2}\sum_{i=1}^{n}\left(\frac{\sum_{t_2=1}^{m}\hat{v}_{it_2}^4}{m}\right)^2 \le \left(\frac{\sum_{i=1}^{n}\sum_{t_2=1}^{m}\hat{v}_{it_2}^4}{nm}\right)^2 = O_p(1),$$

whereas

$$\frac{\sum_{i=1}^{n}(y_{i0}-\hat{\mu}_i)^4\ \left({}^1/_m\sum_{t_2=1}^{m}\hat{v}_{it_2}^4\right)}{n^2} \le \frac{\sum_{i=1}^{n}(y_{i0}-\hat{\mu}_i)^4}{n}\ \frac{\max_{1\le i\le n}\left({}^1/_m\sum_{t_2=1}^{m}\hat{v}_{it_2}^4\right)}{n} = O_p(1),$$

because ${}^1/_n\sum_{i=1}^{n}(y_{i0}-\hat{\mu}_i)^4 = O_p(1)$ and

$$\max_{1\le i\le n}\left(\frac{\sum_{t_2=1}^{m}\hat{v}_{it_2}^4}{m}\right) \le \frac{\sum_{i=1}^{n}\sum_{t_2=1}^{m}\hat{v}_{it_2}^4}{m} = O_p(n).$$

Therefore, as $\sum_{t=1}^{m}\hat{\alpha}^{4(t-1)} = O_p(1)$ and $\sum_{t=1}^{m}\sum_{t_1=1}^{m}(\hat{\chi}_t)_{t_1}^4 = O_p(m)$, it follows on combining results that

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}^*(z_{it}^{*4})}{n^2m^2} = O_p(m^{-1}).$$

Next, exploiting symmetry, the Case (ii) term can be written as

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{t_2\neq t_1}\mathrm{cov}^*(z_{it_1}^{*2}, z_{it_2}^{*2})}{n^2m^2} = 2\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{d=1}^{m-t}\mathrm{cov}^*(z_{it}^{*2}, z_{it+d}^{*2})}{n^2m^2}.$$

It is useful to introduce, for given t and $d>0$, the vectors

$$(\acute{\hat{\chi}}_{t+d})_{t'} \coloneqq \begin{cases}(\hat{\chi}_{t+d})_{t'} & \text{if } t'<t\\ 0 & \text{if } t'\ge t\end{cases}, \qquad (\grave{\hat{\chi}}_{t+d})_{t'} \coloneqq \begin{cases}0 & \text{if } t'\le t\\ (\hat{\chi}_{t+d})_{t'} & \text{if } t'>t\end{cases}.$$

Then $\hat{\chi}_{t+d}'\,v_i^* = \acute{\hat{\chi}}_{t+d}'\,v_i^* + \hat{\alpha}^{d-1}v_{it}^* + \grave{\hat{\chi}}_{t+d}'\,v_i^*$ separates the shocks in time periods running up to t from the one occurring at t and from those occurring afterwards. It will not be necessary to carry this additional notation around for long. Moreover, observing that $\acute{\hat{\chi}}_{t+d}'\,v_i^* = \hat{\alpha}^d\hat{\chi}_t'\,v_i^*$, we can combine this decomposition with recursive substitution to obtain

$$z_{it+d}^* = \hat{\alpha}^d\,v_{it+d}^*\,(y_{it-1}^*-\hat{\mu}_i) + \hat{\alpha}^{d-1}v_{it+d}^*\,v_{it}^* + v_{it+d}^*\,(\grave{\hat{\chi}}_{t+d}'v_i^*).$$

With $z_{it}^* = v_{it}^*\,(y_{it-1}^*-\hat{\mu}_i)$, the covariance $\mathrm{cov}^*(z_{it}^{*2}, z_{it+d}^{*2})$ is then composed of three terms. These are

$$\hat{\alpha}^{2(d-1)}\,\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^*-\hat{\mu}_i)^2, v_{it+d}^{*2}\,v_{it}^{*2}), \qquad \hat{\alpha}^{2d}\,\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^*-\hat{\mu}_i)^2, v_{it+d}^{*2}(y_{it-1}^*-\hat{\mu}_i)^2),$$

and

$$2\,\hat{\alpha}^{2d-1}\,\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2, v_{it+d}^{*2} v_{it}^*(y_{it}^* - \hat{\mu}_i)).$$

We note that all terms involving $\grave{\hat{\chi}}'_{t+d} v_i^*$ do not contribute to the covariance as they involve only future shocks. For the first of these three covariances, by the law of total covariance,

$$\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2, v_{it+d}^{*2}\, v_{it}^{*2}) = \hat{\sigma}_i^2\, \mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^2)\, \mathrm{var}^*(v_{it}^{*2}),$$

where $\mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^2) \lesssim \mathbb{E}^*((\hat{\chi}_t' v_i)^2) + \hat{\alpha}^{2(t-1)}\,((y_{i0} - \hat{\mu}_i))^2$ and $\mathbb{E}^*((\hat{\chi}_t' v_i)^2 = (\hat{\chi}_t'\hat{\chi}_t)\,\hat{\sigma}_i^2$, and $\mathrm{var}^*(v_{it}^{*2}) \lesssim \mathbb{E}^*(v_{it}^{*4}) = {}^1\!/_m \sum_{t=1}^m \hat{v}_{it}^4$. Therefore, the total contribution of these terms, that is,

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \sum_{d=1}^{m-t} \hat{\alpha}^{2(d-1)}\,\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2, v_{it+d}^{*2}\, v_{it}^{*2})}{n^2 m^2},$$

is bounded by the sum of

$$\frac{\sum_{t=1}^m (\hat{\chi}_t'\hat{\chi}_t)\,(\sum_{d=1}^{m-t} \hat{\alpha}^{2(d-1)})}{m^2}\,\frac{\sum_{i=1}^n \hat{\sigma}_i^4\,({}^1\!/_m \sum_{t'=1}^m \hat{v}_{it'}^4)}{n^2} = O_p(m^{-1})$$

and

$$\frac{\sum_{t=1}^m (\hat{\alpha}^{2(t-1)})\,(\sum_{d=1}^{m-t} \hat{\alpha}^{2(d-1)})}{m^2}\,\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^2\,\hat{\sigma}_i^2\,({}^1\!/_m \sum_{t'=1}^m \hat{v}_{it'}^4)}{n^2} = O_p(m^{-1}).$$

Here, the orders of magnitude follow from previously-established results. For example, for the first term, we used $\sum_{d=1}^{m-t} \hat{\alpha}^{2(d-1)} \leq \sum_{d=1}^m \hat{\alpha}^{2(d-1)} \xrightarrow{p} (1-\alpha^2)^{-1}(1-\alpha^{2m}) = O(1)$ and $\sum_{t=1}^m (\hat{\chi}_t'\hat{\chi}_t) \xrightarrow{p} (1-\alpha^2)^{-1}\, m + o(m) = O(m)$, along with

$$\frac{\sum_{i=1}^n \hat{\sigma}_i^4\,({}^1\!/_m \sum_{t'=1}^m \hat{v}_{it'}^4)}{n^2} \leq \frac{\max_{1\leq i\leq n} \hat{\sigma}_i^4}{n}\,\frac{\sum_{i=1}^n \sum_{t=1}^m \hat{v}_{it}^4}{nm} = O_p(1).$$

The two other covariance terms can be handled similarly. Moreover, again by the law of total covariance,

$$\mathrm{cov}^*(v_{it}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2, v_{it+d}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2) = \hat{\sigma}_i^4\,\mathrm{var}^*((y_{it-1}^* - \hat{\mu}_i)^2) \lesssim \hat{\sigma}_i^4\,\mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^4).$$

Here $\mathbb{E}^*((y_{it-1}^* - \hat{\mu}_i)^4) \lesssim (\hat{\chi}_t'\hat{\chi}_t)^2\,({}^1\!/_m \sum_{t'=1}^m \hat{v}_{it'}^4) + \hat{\alpha}^{4(t-1)}\,(y_{i0} - \hat{\mu}_i)^4$, and so we have the upper bound

$$\left(\frac{\sum_{d=1}^m \hat{\alpha}^{2d}}{m}\right)\left(\frac{(\sum_{t=1}^m (\hat{\chi}_t'\hat{\chi}_t)^2)}{m}\,\frac{\sum_{i=1}^n \sum_{t=1}^m \hat{\sigma}_i^4\,\hat{v}_{it}^4}{n^2 m} + \frac{(\sum_{t=1}^m \hat{\alpha}^{4(t-1)})}{m}\,\frac{\sum_{i=1}^n \hat{\sigma}_i^4\,(y_{i0} - \hat{\mu}_i)^4}{n^2}\right),$$

which is again $O_p(m^{-1})$. For the final term,

$$|\text{cov}^*(v_{it}^{*2}(y_{it-1}^* - \hat{\mu}_i)^2, v_{it+d}^{*2} v_{it}^*(y_{it-1}^* - \hat{\mu}_i))| \leq \mathbb{E}^*(v_{it+d}^{*2})\, \mathbb{E}^*(|v_{it}^*|^3)\, \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^3).$$

Now, $\mathbb{E}^*(|v_{it}^*|^3)^2 \leq \mathbb{E}^*(|v_{it}^*|^4)\, \mathbb{E}^*(|v_{it}^*|^2)$ and $\mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^3)^2 \leq \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^4)\, \mathbb{E}(|y_{i0} - \hat{\mu}_i|^2)$ by the Cauchy-Schwarz inequality. Then, using the well-known inequality $(a+b) \geq 2\sqrt{ab}$,

$$\mathbb{E}^*(|v_{it}^*|^3)\, \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^3) \lesssim \mathbb{E}^*(v_{it}^{*4})\, \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^2) + \mathbb{E}^*(v_{it}^{*2})\, \mathbb{E}^*(|y_{it-1} - \hat{\mu}_i|^4).$$

Then, exploiting that $\mathbb{E}^*(v_{it+d}^{*2})$ does not depend on d and that $\mathbb{E}^*(v_{it}^{*2})^2 \leq \mathbb{E}^*(v_{it}^{*4})$, these contributions translate into an upper bound on the final term that is composed of the sum of

$$\frac{\left(\sum_{d=1}^m |\hat{\alpha}|^{2d-1}\right)}{m} \frac{\max_{1\leq i\leq n} \left(1/m \sum_{t=1}^m \hat{v}_{it}^4\right)}{n} \frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}^*(v_{it}^{*2})\; \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^2)}{nm} = O_p(m^{-1})$$

and

$$\frac{\left(\sum_{d=1}^m |\hat{\alpha}|^{2d-1}\right)}{m} \frac{\max_{1\leq i\leq n} \left(1/m \sum_{t=1}^m \hat{v}_{it}^4\right)}{n} \frac{\sum_{i=1}^n \sum_{t=1}^m \; \mathbb{E}^*(|y_{it-1}^* - \hat{\mu}_i|^4)}{nm} = O_p(m^{-1}),$$

where the order of magnitude of each term is again immediate given previously-established results. This handles the Case (ii) term, and so

$$\frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2=1}^m \text{cov}^*(z_{it_1}^{*2}, z_{it_2}^{*2})}{n^2 m^2} = O_p(m^{-1})$$

follows. We have thus shown that $1/m \sum_{t=1}^m (z_t^{*2} - \mathbb{E}(z_t^2)) \xrightarrow{p^*} 0$, and so, with all conditions for the central limit theorem verified, we obtain

$$\frac{1}{\sqrt{nm}} \sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \hat{\mu}_i)\, v_{it}^* \xrightarrow{d^*} \mathbf{N}(0, \sigma^2\gamma^2) \tag{A.6}$$

as $n, m \to \infty$.

To complete the analysis of the numerator of the within-group estimator we show that

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \left((\bar{y}_{i-}^* - \hat{\mu}_i)\, \bar{v}_i^* + \hat{b}_m \right) = o_{p^*}(1). \tag{A.7}$$

The summand has (conditional on the data) mean zero by construction. Moreover, from above,

$$\hat{b}_m = -{}^1\!/_n \sum_{i=1}^{n} \frac{\hat{\sigma}_i^2}{m} \frac{(\iota_m' \hat{A}_m \iota_m)}{m},$$

and so the conditional variance is

$$\frac{m}{n} \sum_{i=1}^{n} \mathbb{E}^*((\bar{y}_{i-}^* - \hat{\mu}_i)^2 \, \bar{v}_i^{*2}) - \frac{m}{n} \sum_{i=1}^{n} \left(\frac{\hat{\sigma}_i^2}{m} \frac{(\iota_m' \hat{A}_m \iota_m)}{m} \right)^2.$$

The second term on the right-hand side can be written as

$$\left(\frac{(\iota_m' \hat{A}_m \iota_m)}{m} \right)^2 \frac{{}^1\!/_n \sum_{i=1}^{n} \hat{\sigma}_i^4}{m} \lesssim \frac{{}^1\!/_n \sum_{i=1}^{n} \hat{\sigma}_i^4}{m} = O_p(m^{-1}).$$

For the first right-hand side term, the recursion in (A.4) gives

$$\mathbb{E}^*((\bar{y}_{i-}^* - \hat{\mu}_i)^2 \, \bar{v}_i^{*2}) \lesssim \frac{\mathbb{E}^*((\hat{\lambda}_m' v_i^*)^2 \, (\iota_m' v_i^*)^2)}{m^4} + (y_{i0} - \hat{\mu}_i)^2 \, \frac{(\iota_m' \hat{a}_m)^2 \, \mathbb{E}^*((\iota_m' v_i^*)^2)}{m^4},$$

where, in analogy to λ_m, we let $\hat{\lambda}_m = \hat{A}_m' \iota_m$. Now, using results from above, we easily have

$$\frac{m}{n} \sum_{i=1}^{n} \frac{\mathbb{E}^*((\hat{\lambda}_m' v_i^*)^2 \, (\iota_m' v_i^*)^2)}{m^4} \lesssim O_p(m^{-1}) \left(\frac{\sum_{i=1}^{n} \sum_{t=1}^{m} \hat{v}_{it}^4}{nm} + \frac{\sum_{i=1}^{n} \hat{\sigma}_i^4}{n} \right) = o_p(1).$$

Similarly,

$$\frac{m}{n} \sum_{i=1}^{n} (y_{i0} - \hat{\mu}_i)^2 \, \frac{(\iota_m' \hat{a}_m)^2 \, \mathbb{E}^*((\iota_m' v_i^*)^2)}{m^4} = \frac{(\iota_m' \hat{a}_m)^2}{m^2} \, \frac{\sum_{i=1}^{n} (y_{i0} - \hat{\mu}_i)^2 \, \hat{\sigma}_i^2}{n} = O_p(m^{-2}),$$

because $\iota_m' \hat{a}_m = O_p(1)$ and ${}^1\!/_n \sum_{i=1}^{n} (y_{i0} - \hat{\mu}_i)^2 \, \hat{\sigma}_i^2 = O_p(1)$ again follow from established results together with the Cauchy-Shwarz inequality. This yields (A.7).

For the denominator of the within-group estimator, in turn, it is enough to show that

$$\frac{1}{nm} \sum_{i=1}^{n} \sum_{t=1}^{m} ((y_{it-1}^* - \bar{y}_{i-}^*) \, y_{it-1}^* - \hat{\gamma}_m^2) = O_{p^*}(n^{-1/2}). \tag{A.8}$$

Here, again, the conditional mean is equal to zero and, for the conditional variance, we need only show that

$$\frac{\sum_{i=1}^{n} \mathbb{E}^*((w_i^{*\prime} M_m w_i^*)^2)}{nm^2} = O_p(1).$$

Let $\hat{H}_m := M_m \hat{A}_m$ and $\hat{H}_m^2 := \hat{H}_m' \hat{H}_m = \hat{A}_m' M_m \hat{A}_m$. Then

$$\mathbb{E}^*((w_i^{*\prime} M_m w_i^*)^2) \lesssim \mathbb{E}^*((v_i^{*\prime} \hat{H}_m^2 v_i^*)^2) + (y_{i0} - \hat{\mu}_i)^2 \, \hat{\sigma}_i^2 \, \|\hat{H}_m' \hat{a}_m\|_2^2 + (y_{i0} - \hat{\mu}_i)^4 \, \|M_m \hat{a}_m\|_2^4.$$

From above we have ${}^1\!/\!_n \sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^4 = O_p(1)$ and ${}^1\!/\!_n \sum_{i=1}^n \hat{\sigma}_i^4 = O_p(1)$, and so also ${}^1\!/\!_n \sum_{i=1}^n (y_{i0} - \hat{\mu}_i)^2 \, \hat{\sigma}_i^2 = O_p(1)$, whereas $\|\hat{H}_m' \hat{a}_m\|_2^2 = O_p(1)$ and $\|M_m \hat{a}_m\|_2^2 = O_p(1)$ both follow from the consistency of $\hat{\alpha}$ and the fact that $-1 < \alpha < 1$. Therefore, we readily have

$$\frac{\sum_{i=1}^n \mathbb{E}^*((w_i^{*\prime} M_m w_i^*)^2)}{nm^2} = \frac{\sum_{i=1}^n \mathbb{E}^*((v_i^{*\prime} \hat{H}_m^2 v_i^*)^2)}{nm^2} + O_p(m^{-2}).$$

For the remaining right-hand side term,

$$\frac{\sum_{i=1}^n \mathbb{E}^*((v_i^{*\prime} \hat{H}_m^2 v_i^*)^2)}{nm^2} \lesssim \frac{\operatorname{tr}(\hat{H}_m^2)^2}{m^2} \frac{1}{n} \sum_{i=1}^n \left(\frac{\sum_{t=1}^m \hat{v}_{it}^4}{m} + \left(\frac{\sum_{t=1}^m \hat{v}_{it}^2}{m} \right)^2 \right) = O_p(1)$$

again follows by the same arguments as used before, because $\operatorname{tr}(\hat{H}_m^2) = O_p(m)$. This establishes (A.8).

The results in (A.7) and (A.8), together with the fact that $\hat{\gamma}_m^2 \xrightarrow{p} \gamma^2$, allow us to write

$$\sqrt{nm}(\hat{\alpha}^* - \hat{\alpha}) = \sqrt{nm}\, \hat{\beta}_m + \frac{1}{\sqrt{nm}} \sum_{i=1}^n \sum_{t=1}^m \frac{(y_{it-1}^* - \hat{\mu}_i)\, v_{it}^*}{\gamma^2} + o_{p^*}(1),$$

which may be combined with the expansion of $\hat{\beta}_m$ in (A.5) and the asymptotic-normality result in (A.6) to see that, as $n, m \to \infty$, $\sqrt{nm}(\hat{\alpha}^* - \hat{\alpha}) \xrightarrow{d^*} \mathbf{N}(0, 1 - \alpha^2)$ provided that ${}^n\!/\!_m \to 0$, whereas

$$\sqrt{nm} \left(\hat{\alpha}^* - \hat{\alpha} - \frac{c_1(\alpha)}{m} \right) \xrightarrow{d^*} \mathbf{N}(0, 1 - \alpha^2)$$

provided that ${}^n\!/\!_{m^3} \to 0$, while

$$\sqrt{nm} \left(\hat{\alpha}^* - \hat{\alpha} - \frac{c_1(\alpha)}{m} - \frac{c_2(\alpha, \tau^2) - c_1(\alpha)}{m^2} \right) \xrightarrow{d^*} \mathbf{N}(0, 1 - \alpha^2)$$

provided that ${}^n\!/\!_{m^5} \to 0$. □

Proof of Theorem 3. The proof of this theorem is similar to the proof of Theorem 2. Now,

$$(\hat{A}_m^*)_{t,t'} := \begin{cases} \hat{\alpha}^{*\, t-t'-1} & \text{if } t > t' \\ 0 & \text{if } t \leq t' \end{cases},$$

and $\hat{a}^*_m$ is the column vector of length m with $(\hat{a}^*_m)_t := \hat{\alpha}^{*t-1}$. We can then again write $w^{**}_i := (y^{**}_{i0}, \dots, y^{**}_{im-1})'$ as

$$w^{**}_i = \hat{A}^*_m v^{**}_i + \hat{a}^*_m (y_{i0} - \hat{\mu}^*_i) + \iota_m \, \hat{\mu}^*_i, \tag{A.9}$$

where $v^{**}_i := (v^{**}_{i1}, \dots, v^{**}_{im})'$ and $\hat{\mu}^*_i := (1 - \hat{\alpha}^*)^{-1} \hat{\eta}^*_i$, where we recall that the processes start at $y^{**}_{i0} = y^*_{i0} = y_{i0}$.

We have

$$\hat{\alpha}^{**} - \hat{\alpha}^* := \frac{\sum_{i=1}^n \sum_{t=1}^m (y^{**}_{it-1} - \bar{y}^{**}_{i-}) \, v^{**}_{it}}{\sum_{i=1}^n \sum_{t=1}^m (y^{**}_{it-1} - \bar{y}^{**}_{i-}) \, y^{**}_{it-1}} = \frac{\sum_{i=1}^n w^{**\prime}_i M_m v^{**}_i}{\sum_{i=1}^n w^{**\prime}_i M_m w^{**}_i}.$$

Using (A.9),

$$\hat{b}^*_m := \frac{\sum_{i=1}^n \mathbb{E}^{**}(w^{**\prime}_i M_m v^{**}_i)}{nm} = \frac{\hat{\sigma}^{*2} \, \mathrm{tr}(M_m \hat{A}^*_m)}{m} = -\frac{\hat{\sigma}^{*2}}{m} \frac{(\iota'_m \hat{A}^*_m \iota_m)}{m},$$

where, in analogy to the first-layer bootstrap, $\hat{\sigma}^{*2} := {}^1\!/\!_n \sum_{i=1}^n \hat{\sigma}^{*2}_i = {}^1\!/\!_{nm} \sum_{i=1}^n \sum_{t=1}^m \hat{v}^{*2}_{it}$. Also,

$$\hat{\gamma}^{*2}_m := \frac{\sum_{i=1}^n \mathbb{E}^{**}(w^{**\prime}_i M_m w^{**}_i)}{nm} = \frac{\hat{\sigma}^{*2} \, \mathrm{tr}(\hat{A}^{*\prime}_m M_m \hat{A}^*_m)}{m} + \frac{\hat{\sigma}^{*2}}{1 - \hat{\alpha}^{*2}} \frac{\hat{\tau}^{*2} \, (\hat{a}^{*\prime}_m M_m \hat{a}^*_m)}{m},$$

for

$$\hat{\tau}^{*2} := \frac{{}^1\!/\!_n \sum_{i=1}^n (y_{i0} - \hat{\mu}^*_i)^2}{\hat{\gamma}^{*2}}$$

and $\hat{\gamma}^{*2} := \hat{\sigma}^{*2} \, (1 - \hat{\alpha}^{*2})^{-1}$. Then, in the same way as in the previous proof, we can expand

$$\hat{\beta}^*_m := \frac{\hat{b}^*_m}{\hat{\gamma}^{*2}_m} = \frac{c_1(\hat{\alpha}^*)}{m} + \frac{c_2(\hat{\alpha}^*, \hat{\tau}^{*2})}{m^2} + O_{p^*}(m^{-3}),$$

first around $\hat{\alpha}$ and then around α, and use Theorems 1 and 2 to arrive at the bias expansion

$$\hat{\beta}^*_m = \frac{c_1(\alpha)}{m} + \frac{c_2(\alpha, \tau^2) - 2 \, c_1(\alpha)}{m^2} + O_{p^*}(m^{-3}) + o_{p^*}\left(\frac{1}{\sqrt{nm}}\right). \tag{A.10}$$

Observe that then, indeed,

$$\sqrt{nm}(\hat{\beta}^*_m - \hat{\beta}_m) - \sqrt{nm}(\hat{\beta}_m - \beta^*_m) = O_{p^*}(\sqrt{{}^n\!/\!_{m^5}}),$$

as desired.

We then move on to establish that

$$\frac{1}{\sqrt{nm}}\sum_{i=1}^{n}\sum_{t=1}^{m}(y_{it-1}^{**}-\hat{\mu}_i^{*})\,v_{it}^{**}\overset{d^{**}}{\to}\mathbf{N}(0,\sigma^2\gamma^2) \tag{A.11}$$

as $n, m \to \infty$. To do so we again verify the conditions required for the central limit theorem in Corollary 5.26 in White (2001). Letting $z_{it}^{**} := (y_{it-1}^{**}-\hat{\mu}_i^{*})\,v_{it}^{**}$ and $z_t^{**} := {}^{1}\!/\!\sqrt{n}\sum_{i=1}^{n} z_{it}^{**}$, the first of three conditions that needs to be shown to hold is that, for some $r > 2$, we have $\mathbb{E}^{**}(|z_t^{**}|^r) = O_{p^*}(1)$. Again, by an application of Rosenthal's inequality, this is implied by

$$\frac{\sum_{i=1}^{n}\mathbb{E}^{**}(|z_{it}^{**}|^r)}{n} = O_{p^*}(1),$$

which we now show. By the Cauchy-Schwarz inequality,

$$\frac{\sum_{i=1}^{n}\mathbb{E}^{**}(|z_{it}^{**}|^r)}{n} \le \left(\frac{\sum_{i=1}^{n}\mathbb{E}^{**}(|v_{it}^{**}|^{2r})}{n}\right)^{1/2}\left(\frac{\sum_{i=1}^{n}\mathbb{E}^{**}(|y_{it-1}^{**}-\hat{\mu}_i^{*}|^{2r})}{n}\right)^{1/2}, \tag{A.12}$$

and we take each of the right-hand side terms, in turn. First, as $\mathbb{E}^{**}(|v_{it}^{**}|^{2r}) = {}^{1}\!/\!{}_{m}\sum_{t=1}^{m}|\hat{v}_{it}^{*}|^{2r}$ and $\hat{v}_{it}^{*} = (v_{it}^{*}-\bar{v}_i^{*}) - (\hat{\alpha}^{*}-\hat{\alpha})\,(y_{it-1}^{*}-\bar{y}_{i-}^{*})$, it suffices to show that

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}|v_{it}^{*}-\bar{v}_i^{*}|^{2r}}{nm} = O_{p^*}(1), \qquad \frac{\sum_{i=1}^{n}\sum_{t=1}^{m}|y_{it-1}^{*}-\bar{y}_{i-}^{*}|^{2r}}{nm} = O_{p^*}(1),$$

because we know from Theorem 2 that $|\hat{\alpha}^{*}-\hat{\alpha}| = o_{p^*}(1)$. Now, for these two conditions it is, in turn, sufficient to verify that

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}^{*}(|v_{it}^{*}-\bar{v}_i^{*}|^{2r})}{nm} = O_p(1), \qquad \frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}^{*}(|y_{it-1}^{*}-\bar{y}_{i-}^{*}|^{2r})}{nm} = O_p(1),$$

by appealing to Markov's inequality. The first of these is immediate as we have already established that

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}^{*}(|v_{it}^{*}|^{2r})}{nm} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{m}|\hat{v}_{it}|^{2r}}{nm} = O_p(1).$$

The argument for the second result equally follows from previous results. Moreover, we have the bound

$$\frac{\sum_{i=1}^{n}\sum_{t=1}^{m}\mathbb{E}^{*}(|y_{it-1}^{*}|^{2r})}{nm} \lesssim \frac{\sum_{i=1}^{n}\sum_{t=1}^{m}|v_{it}^{*}|^{2r}}{nm} + \frac{\sum_{t=1}^{m}|\hat{\alpha}|^{2r(t-1)}}{m}\,\frac{\sum_{i=1}^{n}|y_{i0}-\hat{\mu}_i|^{2r}}{n} + \frac{\sum_{i=1}^{n}\hat{\mu}_i^{2r}}{n},$$

where each of the right-hand side terms has already been shown to be $O_p(1)$. This handles the first term in (A.12). For the second term, using (A.9), we have

$$\frac{\sum_{i=1}^n \mathbb{E}^{**}(|y_{it-1}^{**} - \hat{\mu}_i^*|^{2r})}{n} \lesssim \frac{\sum_{i=1}^n \mathbb{E}^{**}(|\hat{\chi}_t^{*\prime} v_i^{**}|^{2r})}{n} + |\hat{\alpha}^*|^{2r(t-1)} \frac{\sum_{i=1}^n |y_{i0} - \hat{\mu}_i^*|^{2r}}{n}.$$

By Rosenthal's inequality and the observation that $|\hat{\alpha}^* - \alpha| < |\hat{\alpha}^* - \hat{\alpha}| + |\hat{\alpha} - \alpha| = o_{p^*}(1)$,

$$\frac{\sum_{i=1}^n \mathbb{E}^{**}(|\hat{\chi}_t^{*\prime} v_i^{**}|^{2r})}{n} \lesssim \frac{\sum_{i=1}^n \sum_{t=1}^m |\hat{v}_{it}^*|^{2r}}{nm} = O_{p^*}(1).$$

Further, $|\hat{\alpha}^*|^{2r(t-1)} = O_{p^*}(1)$ and $1/n \sum_{i=1}^n |y_{i0}|^{2r} = O_p(1)$ because $\mathbb{E}(|y_{i0}|^{2r}) < \infty$, whereas we have

$$\frac{\sum_{i=1}^n \hat{\mu}_i^{*2r}}{n} = \frac{1}{|1 - \hat{\alpha}^*|^{2r}} \frac{\sum_{i=1}^n |\hat{\eta}_i + \bar{v}_i^* - (\hat{\alpha}^* - \hat{\alpha})\, \bar{y}_{i-}^*|^{2r}}{n} = O_{p^*}(1)$$

because $|1 - \hat{\alpha}^*|^{2r} = O_{p^*}(1)$ and the term on the right-hand side is bounded by a multiple of

$$\frac{\sum_{i=1}^n |\hat{\eta}_i|^{2r}}{n} + \frac{\sum_{i=1}^n |\bar{v}_i|^{*2r}}{n} + |\hat{\alpha}^* - \hat{\alpha}| \frac{\sum_{i=1}^n |\bar{y}_{i-}|^{*2r}}{n} = O_{p^*}(1).$$

This concludes the analysis of (A.12) and confirms that the first condition for the central limit theorem holds.

The second requirement for the central limit theorem is that

$$\text{var}^{**}\left(\frac{\sum_{t=1}^m z_t^{**}}{\sqrt{m}}\right) \xrightarrow{p^*} (1 - \alpha^2)^{-1} \sigma^4.$$

In the same way as before, the variance can be written as

$$\left(\frac{\sum_{i=1}^n \hat{\sigma}_i^{*4}}{n}\right) \frac{\text{tr}(\hat{A}^* \hat{A}^{*\prime})}{m} + \left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i^*)^2\, \hat{\sigma}_i^{*2}}{n}\right) \frac{(\hat{a}_m^{*\prime} \hat{a}_m^*)}{m}.$$

Here, it is easy to see that $m^{-1}\,\text{tr}(\hat{A}^* \hat{A}^{*\prime}) \xrightarrow{p^*} (1 - \alpha^2)^{-1}$, using $\hat{\alpha}^* \xrightarrow{p^*} \alpha$. Furthermore, we can see that

$$\frac{\sum_{i=1}^n \hat{\sigma}_i^{*4}}{n} = \frac{\sum_{i=1}^n (1/m \sum_{t=1}^m (v_{it}^* - \bar{v}_i^*)^2)^2}{n} + o_{p^*}(1),$$

by expanding the square and using that $\hat{\alpha}^* \xrightarrow{p^*} \alpha$, $1/nm \sum_{i=1}^n \sum_{t=1}^m (y_{it-1}^* - \bar{y}_{i-}^*)^4 = O_{p^*}(1)$, and $1/nm \sum_{i=1}^n \sum_{t=1}^m (v_{it}^* - \bar{v}_i^*)^4 = O_{p^*}(1)$. We can then focus on the dominant term. Adding and subtracting $\hat{\sigma}_i^2$ to the summand and working out the square we can establish that we have

$$\frac{\sum_{i=1}^n (1/m \sum_{t=1}^m (v_{it}^* - \bar{v}_i^*)^2)^2}{n} = \frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n} + o_{p^*}(1).$$

Indeed,

$$\frac{\sum_{i=1}^n ({}^1\!/\!_m \sum_{t=1}^m (v_{it}^* - \bar{v}_i^*)^2 - \hat{\sigma}_i^2)^2}{n} \lesssim \frac{\sum_{i=1}^n ({}^1\!/\!_m \sum_{t=1}^m v_{it}^{*2} - \hat{\sigma}_i^2)^2}{n} + \frac{\sum_{i=1}^n ({}^1\!/\!_m \sum_{t=1}^m v_{it}^*)^4}{n},$$

where

$$\mathbb{E}^* \left(\frac{\sum_{i=1}^n ({}^1\!/\!_m \sum_{t=1}^m v_{it}^{*2} - \hat{\sigma}_i^2)^2}{n} \right) = \frac{1}{m} \left(\frac{\sum_{i=1}^n \mathbb{E}^*(v_{it}^{*4}) - \hat{\sigma}_i^4}{n} \right) = O_p(m^{-1}),$$

and

$$\mathbb{E}^* \left(\frac{\sum_{i=1}^n ({}^1\!/\!_m \sum_{t=1}^m v_{it}^*)^4}{n} \right) \lesssim \frac{1}{m^3} \left(\frac{\sum_{i=1}^n \mathbb{E}^*(v_{it}^{*4})}{n} \right) + \frac{1}{m^2} \left(\frac{\sum_{i=1}^n \mathbb{E}^*(v_{it}^{*2})^2}{n} \right) = O_p(m^{-2}).$$

Therefore,

$$\frac{\sum_{i=1}^n \hat{\sigma}_i^{*4}}{n} = \frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n} + o_{p^*}(1).$$

Also, $m^{-1} \, (\hat{a}_m^{*\prime} \hat{a}_m^*) \xrightarrow{p^*} 0$ while

$$\left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i^*)^2 \, \hat{\sigma}_i^{*2}}{n} \right) \leq \left(\left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i^*)^4}{n} \right) \right)^{1/2} \left(\left(\frac{\sum_{i=1}^n \hat{\sigma}_i^{*4}}{n} \right) \right)^{1/2} = O_{p^*}(1)$$

is readily deduced. Hence,

$$\text{var}^{**} \left(\frac{\sum_{t=1}^m z_t^{**}}{\sqrt{m}} \right) = \frac{1}{1-\alpha^2} \frac{\sum_{i=1}^n \hat{\sigma}_i^4}{n} + o_{p^*}(1) \xrightarrow{p} \frac{\sigma^4}{1-\alpha^2}$$

follows on recalling that ${}^1\!/\!_n \sum_{i=1}^n \hat{\sigma}_i^4 \xrightarrow{p} \sigma^4$, as was shown previously as part of the proof of Theorem 2. This, then, shows that the second condition for the central limit theorem holds.

The third and final requirement for (A.11) is that

$$\frac{\sum_{t=1}^m (z_t^{**2} - \mathbb{E}^{**}(z_t^{**2}))}{m} = o_{p^{**}}(1).$$

We again proceed by splitting the term into two contributions, as

$$\frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2 \neq t_1} z_{it_1}^{**} \, z_{it_2}^{**}}{nm} + \frac{\sum_{i=1}^n \sum_{t=1}^m z_{it}^{**2} - \mathbb{E}^{**}(z_{it}^{**2})}{nm}.$$

The variance of the first term is

$$\mathbb{E}^{**} \left(\left(\frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2 \neq t_1} z_{it_1}^{**} \, z_{it_2}^{**}}{nm} \right)^2 \right) \lesssim \frac{\sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{t=1}^m \mathbb{E}^{**}(z_{i_1 t}^{**2}) \, \mathbb{E}^{**}(z_{i_2 t}^{**2})}{n^2 m^2}.$$

Denoting by $\hat{\chi}_1^*, \ldots, \hat{\chi}_m^*$ the columns of the matrix $\hat{A}_m^{*\prime}$ we can again use familiar arguments to obtain the decomposition $\mathbb{E}^{**}((y_{it-1}^{**} - \hat{\mu}_i^*)^2) = (\hat{\chi}_t^{*\prime}\hat{\chi}_t^*)\, \hat{\sigma}_i^{*2} + \hat{\alpha}^{*2(t-1)}\, (y_{i0} - \hat{\mu}_i^*)^2$, with which we can readily verify that the bound on the variance itself is bounded by the sum of (multiples of)

$$\frac{\sum_{t=1}^m (\hat{\chi}_t^{*\prime}\hat{\chi}_t^*)^2}{m^2} \left(\frac{\sum_{i_1=1}^n \hat{\sigma}_i^{*4}}{n} \right)^2, \qquad \frac{\sum_{t=1}^m \hat{\alpha}^{*4(t-1)}}{m^2} \left(\frac{\sum_{i=1}^n (y_{i0} - \hat{\mu}_i^*)^2\, \hat{\sigma}_i^{*2}}{n} \right)^2,$$

and

$$2\frac{\sum_{t=1}^m \hat{\alpha}^{*2(t-1)}\, (\hat{\chi}_t^{*\prime}\hat{\chi}_t^*)}{m^2} \frac{\sum_{i_1=1}^n (y_{i_1 0} - \hat{\mu}_{i_1}^*)^2\, \hat{\sigma}_{i_1}^{*2}}{n} \frac{\sum_{i_2=1}^n \hat{\sigma}_{i_2}^{*4}}{n},$$

all of which are $o_{p^*}(1)$ from previous calculations. This deals with the first contribution. The second contribution has variance

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \mathrm{var}^{**}(z_{it}^{**2})}{n^2 m^2} + \frac{\sum_{i=1}^n \sum_{t_1=1}^m \sum_{t_2 \neq t_1} \mathrm{cov}^{**}(z_{it_1}^{**2}, z_{it_2}^{**2})}{n^2 m^2}.$$

For the first term,

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \mathrm{var}^{**}(z_{it}^{**2})}{n^2 m^2} \lesssim \frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}^{**}(z_{it}^{**4})}{n^2 m^2},$$

where, in the same way as in the proof of Theorem 2, $\mathbb{E}^{**}(z_{it}^{**4})$ is bounded by a multiple of

$$\left(\left(\sum_{t_1=1}^m (\hat{\chi}_t^*)_{t_1}^4 \right) + \left(\sum_{t_1=1}^m (\hat{\chi}_t^*)_{t_1}^2 \right)^2 \right) \left(\frac{\sum_{t_2=1}^m \hat{v}_{it_2}^{*4}}{m} \right)^2 + |\hat{\alpha}^*|^{4(t-1)}\, (y_{i0} - \hat{\mu}_i^*)^4 \left(\frac{\sum_{t_2=1}^m \hat{v}_{it_2}^{*4}}{m} \right).$$

We can handle each of these terms by the same arguments as used in the previous proof. We then find that

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \mathbb{E}^{**}(z_{it}^{**4})}{n^2 m^2} = O_{p^*}(m^{-1}),$$

as desired. Finally, for the second term, we again write

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \sum_{t_2 \neq t_1} \mathrm{cov}^{**}(z_{it_1}^{**2}, z_{it_2}^{**2})}{n^2 m^2} = 2\frac{\sum_{i=1}^n \sum_{t=1}^m \sum_{d=1}^{m-t} \mathrm{cov}^{**}(z_{it}^{**2}, z_{it+d}^{**2})}{n^2 m^2},$$

and proceed as before. Moreover, we use $\hat{\chi}_{t+d}^{*\prime}\, v_i^{**} = \acute{\hat{\chi}}_{t+d}^{*\prime}\, v_i^{**} + \hat{\alpha}^{*d-1} v_{it}^{**} + \grave{\hat{\chi}}_{t+d}^{*\prime}\, v_i^{**}$, where

$$(\acute{\hat{\chi}}_{t+d}^*)_{t'} := \begin{cases} (\hat{\chi}_{t+d}^*)_{t'} & \text{if } t' < t \\ 0 & \text{if } t' \geq t \end{cases}, \qquad (\grave{\hat{\chi}}_{t+d}^*)_{t'} := \begin{cases} 0 & \text{if } t' \leq t \\ (\hat{\chi}_{t+d}^*)_{t'} & \text{if } t' > t \end{cases}.$$

Here, $\acute{\hat{\chi}}^{*\prime}_{t+d}\, v_i^{**} = \hat{\alpha}^{*d} \hat{\chi}_t^{*\prime}\, v_i^{**}$. Therefore, the covariance between z_{it}^{**2} and

$$z_{it+d}^{**2} = (\hat{\alpha}^d\, v_{it+d}^{**}\, (y_{it-1}^{**} - \hat{\mu}_i^*) + \hat{\alpha}^{*d-1} v_{it+d}^{**}\, v_{it}^{**} + v_{it+d}^{**}\, (\grave{\hat{\chi}}^{*\prime}_{t+d} v_i^{**}))^2,$$

$\operatorname{cov}^{**}(z_{it}^{**2}, z_{it+d}^{**2})$, has the three components

$$\hat{\alpha}^{*2(d-1)} \operatorname{cov}^{**}(v_{it}^{**2}(y_{it-1}^{**} - \hat{\mu}_i)^{*2}, v_{it+d}^{**2}\, v_{it}^{**2}), \qquad \hat{\alpha}^{*2d} \operatorname{cov}^{*}(v_{it}^{**2}(y_{it-1}^{**} - \hat{\mu}_i^*)^2, v_{it+d}^{**2}(y_{it-1}^{**} - \hat{\mu}_i^*)^2),$$

and

$$2\, \hat{\alpha}^{*2d-1} \operatorname{cov}^{**}(v_{it}^{**2}(y_{it-1}^{**} - \hat{\mu}_i^*)^2, v_{it+d}^{**2} v_{it}^{**}\, (y_{it}^{**} - \hat{\mu}_i^*)).$$

These terms are completely analogous to the terms encountered in the proof of Theorem 2, and they can be handled in the same way. As an example, for the contributions of the second term we have the upper bound

$$\left(\frac{\sum_{d=1}^m \hat{\alpha}^{*2d}}{m}\right)\left(\frac{(\sum_{t=1}^m (\hat{\chi}_t^{*\prime} \hat{\chi}_t^*)^2)}{m} \frac{\sum_{i=1}^n \sum_{t=1}^m \hat{\sigma}_i^{*4}\, \hat{v}_{it}^{*4}}{n^2 m} + \frac{(\sum_{t=1}^m \hat{\alpha}^{*4(t-1)})}{m} \frac{\sum_{i=1}^n \hat{\sigma}_i^{*4}\, (y_{i0} - \hat{\mu}_i^*)^4}{n^2}\right)$$

and we have already analysed all of its components; from these we can obtain that the bound is $O_{p^*}(m^{-1})$. Therefore,

$$\frac{\sum_{i=1}^n \sum_{t=1}^m \sum_{t_2 \neq t_1} \operatorname{cov}^{**}(z_{it_1}^{**2}, z_{it_2}^{**2})}{n^2 m^2} = o_{p^*}(1),$$

and the final condition for the central limit theorem has been verified. Hence, (A.11) follows.

To complete the proof of Theorem 3 it only remains to verify the asymptotic negligibility of two terms. The first of these relates to the numerator of the within-group estimator and is

$$\sqrt{\frac{m}{n}} \sum_{i=1}^n \left((\bar{y}_{i-}^{**} - \hat{\mu}_i^*)\, \bar{v}_i^{**} + \hat{b}_m^*\right) = o_{p^{**}}(1).$$

The second, in turn, is about the denominator and reads

$$\frac{1}{nm} \sum_{i=1}^n \sum_{t=1}^m ((y_{it-1}^{**} - \bar{y}_{i-}^{**})\, y_{it-1}^{**} - \hat{\gamma}_m^{*2}) = O_{p^{**}}(n^{-1/2}).$$

Both conditions are direct double-bootstrap counterparts to the first-layer bootstrap results in (A.7) and (A.8), respectively, and their proofs are essentially the same. For brevity we omit the details here.

Combining all results, we have that, as $n, m \to \infty$, $\sqrt{nm}(\hat{\alpha}^{**} - \hat{\alpha}^*) \overset{d^{**}}{\to} \mathbf{N}(0, 1-\alpha^2)$ provided that $n/m \to 0$, whereas

$$\sqrt{nm}\left(\hat{\alpha}^{**} - \hat{\alpha}^* - \frac{c_1(\alpha)}{m}\right) \overset{d^{**}}{\to} \mathbf{N}(0, 1-\alpha^2)$$

provided that $n/m^3 \to 0$, while

$$\sqrt{nm}\left(\hat{\alpha}^{**} - \hat{\alpha}^* - \frac{c_1(\alpha)}{m} - \frac{c_2(\alpha, \tau^2) - 2c_1(\alpha)}{m^2}\right) \overset{d^{**}}{\to} \mathbf{N}(0, 1-\alpha^2)$$

provided that $n/m^5 \to 0$. □

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. 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.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika 74*, 457–468.

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). Likelihood inference in an autoregression with fixed effects. *Econometric Theory 32*, 1178–1215.

Dhaene, G. and K. Jochmans (2017). Profile-score adjustments for incidental-parameter problems. Mimeo.

Gonçalves, S., G. Cavaliere, M. Nielsen, and E. Zanelli (2024). Bootstrap inference in the presence of bias. *Journal of the American Statistical Association 119*, 2908–2918.

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 L. Kilian (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics 123*, 89–120.

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 W. K. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*, 1295–1319.

Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics 14*, 1431–1452.

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

Higgins, A. and K. Jochmans (2025). Inference in dynamic models for panel data using the moving block bootstrap. Mimeo.

Kim, M. S. and Y. Sun (2016). Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory 32*, 1523–1568.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics. *Economic Record 82*, S2–S18.

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

Sartori, N. (2003). Modified profile likelihood in models with stratum nuisance parameters. *Biometrika 90*, 533–549.

White, H. (2001). *Asymptotic Theory for Econometricians.* Academic Press.