Bootstrap inference for linear dynamic panel data models with individual fixed effects

Sílvia Gonçalves and Maximilien Kaffo*

CIREQ and Département de Sciences Économiques, Université de Montréal

April 29, 2014

Abstract

This paper's main contribution is to propose and theoretically justify the application of bootstrap methods for inference in autoregressive panel data models with fixed effects. Whereas the focus of the existing literature has been on bias correcting the standard fixed effects OLS estimator (due to the well known incidental parameter bias), our focus here is on improving the quality of inference by relying on the bootstrap instead of the standard normal distribution when computing critical values for test statistics. In particular, we show by simulation that confidence intervals based on the normal distribution can be very distorted in finite samples. Instead, a bootstrap that resamples the residuals and generates the bootstrap observations recursively using the estimated autoregressive panel data model greatly reduces these distortions. We show that this recursive-design residual-based bootstrap fixed effects OLS estimator contains a built-in bias correction term that mimics the incidental parameter bias. Thus, this method can be used to approximate the bias (as well as the entire distribution) of the (biased) fixed effects OLS estimator. This is in contrast with two other methods we consider (a fixed-design residual-based bootstrap and a pairs bootstrap) whose distributions are incorrectly centered at zero. As it turns out, both the recursive-design and the pairs bootstrap are asymptotically valid when applied to the bias-corrected estimator, but the fixed-design bootstrap is not. In the simulations, the recursive-design bootstrap is the method that does best overall.

JEL classification: C15, C22.

Keywords: bootstrap, panel data autoregression, fixed effects, incidental parameter bias.

1 Introduction

Estimation and inference in the context of linear dynamic panel data models is complicated by the presence of fixed effects. Indeed, as noted by Neyman and Scott (1948) and Nickell (1981), estimation of the fixed effects creates an incidental parameter bias in the standard fixed effects OLS estimator that persists even as $n \to \infty$ (and T is fixed). Although this inconsistency disappears when both n and T diverge to infinity, an asymptotic bias appears in the limiting distribution of the fixed effects estimator when n and T grow at the same rate, as shown by Hahn and Kuersteiner (2002). The

*We would like to thank participants at the CEA conference in Calgary (June 2012), the NBER/NSF Time Series conference at Texas A&M (October 2012), and the CESG conference at Queen's University (October 2012). A special thanks to Russell Davidson and James MacKinnon for useful comments and discussions. Gonçalves acknowledges financial support from the NSERC. Kaffo acknowledges financial support from the FQRSC.

existence of the incidental parameter bias has motivated the proposal of many bias reduction methods for panel autoregressive models with fixed effects, including Kiviet (1995), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Bun and Carree (2005), Phillips and Sul (2007), Everaert and Pozzi (2007), Gouriroux, Phillips, and Yu (2010), Fernandez-Val and Weidner (2013) and Lee (2012), among others.

Our focus in this paper is on inference rather than bias correction. In particular, our main goal is to propose bootstrap methods whose finite sample properties improve upon those of the asymptotic normal approximation when computing critical values for test statistics based on bias-corrected estimators. Although this asymptotic approach is justified by the existing literature, our simulations show that asymptotic theory-based confidence intervals for the common autoregressive parameter of an AR(1) model with fixed effects can be severely distorted in finite samples. This provides motivation for the use of the bootstrap.

A natural bootstrap scheme in this context is a recursive-design residual-based bootstrap which resamples the residuals and recursively generates bootstrap observations for the dependent variable using the estimated autoregressive panel data model. The choice of how to generate the bootstrap residuals depends on the assumptions we make on the idiosyncratic error term. Here, we follow most of the existing panel data literature and maintain throughout the assumption of cross sectional independence. In contrast, we allow for time series dependence in the error term by assuming that it satisfies a martingale difference sequence assumption for each individual. This rules out serial correlation but is compatible with time series and cross sectional heteroskedasticity in the error term. To capture both forms of heteroskedasticity, we implement the residual-based bootstrap using the wild bootstrap, where bootstrap residuals are obtained by multiplying the estimated residuals by an external random variable that is i.i.d.$(0, 1)$ across both the time series and the cross sectional dimensions. A version of the recursive-design wild bootstrap method has been applied by Everaert and Pozzi (2007) for bias correction without theoretical justification.

We consider two other bootstrap methods in this paper. One is a version of the residual-based bootstrap that fixes the regressors when generating the bootstrap observations on the dependent variable (i.e. we simply add the wild bootstrap residuals to the estimated conditional mean). We call this method the fixed-design residual-based bootstrap. The other method is a pairs bootstrap which resamples the pairs formed by the dependent and the lagged dependent variables (this amounts to the standard nonparametric bootstrap applied to the pairs). Given the cross sectional independence assumption, our proposal is to resample only in the cross sectional dimension. The main reason why we also consider these two methods is that they have been applied very successfully in the pure time series literature by Gonçalves and Kilian (2004), who showed that they are robust to more general forms of conditional heteroskedasticity (in the form of leverage effects) than the recursive-design residual-based bootstrap. As we will show, even though the three methods we analyze here can be viewed as panel extensions of the bootstrap methods studied by Gonçalves and Kilian (2004), the results we obtain are not a straightforward extension of the results obtained in the pure time series autoregression model due to the presence of the incidental parameter bias.

Our first finding is that only the recursive-design residual-based bootstrap is able to capture the incidental parameter bias term inherent in the fixed effects OLS estimates. The fixed-design residual bootstrap and the pairs bootstrap fail to do so as their bootstrap distributions are incorrectly centered at zero. Thus, although these bootstrap methods are more generally applicable (in that they allow for leverage effects), they do not consistently estimate the distribution of the standard fixed effects estimator in a linear dynamic panel data model with individual specific fixed effects. This is in contrast with the recursive-design bootstrap, which can be used to approximate the whole distribution of the fixed effects OLS estimator, including its bias. We formally prove the consistency of this bootstrap bias, thus providing a theoretical justification for a bootstrap based bias correction as used for instance in Everaert and Pozzi (2007).

Although our results for the recursive-design bootstrap justify bootstrap inference based on the (uncorrected and biased) fixed-effects OLS estimator without the need for an explicit bias correction, further finite sample improvements of the bootstrap approximation can be obtained if we base our inference on a bias-corrected estimator. Bootstrapping a bias-corrected fixed effects estimator essentially removes the incidental parameter bias from the asymptotic distribution, resulting in a t-statistic that is asymptotically pivotal.

Building on the theory of the bootstrap for the standard (biased) fixed effects OLS estimator, we show that the recursive-design bootstrap is asymptotically valid when applied to the bias-corrected estimator of Hahn and Kuersteiner (2002). The asymptotic invalidity of the fixed-design bootstrap for the standard fixed effects estimator extends to the bias-corrected estimator. However, as it turns out, the pairs bootstrap distribution of the bootstrap bias-corrected fixed effects estimator is consistent provided we center the bootstrap bias-corrected estimator around the bias-corrected estimator evaluated on the original sample (instead of its biased version). In the simulations, the recursive-design bootstrap is the method that does best overall, essentially removing the finite sample distortions associated with the confidence intervals that rely on the asymptotic normal distribution.

The existing literature on bootstrapping linear panel data models with fixed effects is surprisingly rather limited. One important exception is Kapetanios (2008), who proposed and studied the pairs bootstrap in the context of panel regression models with strictly exogenous regressors and fixed effects, for which the incidental parameter bias does not exist. More recently, Gonçalves (2011) proved the asymptotic validity of the moving blocks bootstrap under general forms of cross sectional and time series dependence in the regression error of a panel linear regression model. Although the regularity conditions of Gonçalves (2011) allow in principle dynamic regressors, the impact of the incidental parameter bias on inference was ruled out by assuming that $n/T \to 0$. Contrary to these papers, here we establish the consistency of the bootstrap for fixed-effects estimators when the incidental parameter bias is present. A few other papers have recently studied the validity of the bootstrap for panel data models with fixed effects and incidental parameter bias. In particular, Galvão and Kato (2013) study the asymptotic properties of the pairs bootstrap in the context of linear dynamic panel data models with possible misspecification. They find that the pairs bootstrap is asymptotically valid when applied to a bias corrected estimator and that it is robust to misspecification. Similarly, Kaffo (2013) also applies the pairs bootstrap to a bias corrected estimator in the context of nonlinear dynamic panel data models with fixed effects. In both cases, the bootstrap is not able to capture the incidental parameter bias and is only valid when used for inference on a bias corrected estimator. These results (although more general than ours) are entirely parallel to what we find here for the simpler AR(1) panel data model. However, contrary to these papers, here we are able to go a step further and propose a bootstrap method that is also able to capture the bias (the recursive-design bootstrap).

The remainder of the paper is organized as followed. Section 2 introduces the model and the assumptions, and provides a summary of the asymptotic theory for the fixed effects estimator. These results are a restatement of Hahn and Kuersteiner's (2002) results under our set of assumptions (which are slightly different from theirs). Section 3 provides the bootstrap results for the standard fixed effects OLS estimator for the three bootstrap schemes described above. We show that only the recursive-design bootstrap is able to capture the asymptotic bias term. Section 4 relies on the results of Section 3 to prove the consistency of this bootstrap method for estimating the distribution of the biased-corrected fixed effects estimator of Hahn and Kuersteiner (2002). Section 5 contains Monte Carlo results while Section 6 concludes. All proofs are relegated to the Appendix.

2 Assumptions and asymptotic theory for the fixed effects estimator when $n, T \to \infty$

Following Hahn and Kuersteiner (2002), we consider estimation of the autoregressive parameter θ_0 in a stationary linear dynamic panel model with fixed effects[1]

$$y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}, \qquad i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{1}$$

where $|\theta_0| < 1$ and α_i are individual specific fixed effects that capture the unobserved individual heterogeneity. We assume that the initial observation y_{i0} is available. Given the stability condition that $|\theta_0| < 1$ and the assumption that the panel is stationary, the impact of initial conditions does not matter asymptotically when T is large.

The standard fixed effects OLS estimator of θ_0 is given by

$$\hat{\theta} = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right)^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right) \left(y_{it} - \bar{y}_i \right),$$

[1] Our results could be generalized to higher order dynamics at the cost of complicating the notation. Since this would not add any additional insights, we prefer to follow Hahn and Kuersteiner (2002) and focus on this simple AR(1) panel model.

where $\bar{y}_i \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i-} \equiv \frac{1}{T}\sum_{t=1}^{T} y_{it-1}$ are the individual time averages.

The main goal of this section is to provide a set of assumptions under which we can prove the bootstrap results that will follow and at the same time present the asymptotic theory of the fixed effects estimator under these assumptions.

Assumption A1 describes formally our set of assumptions. Note that for a given time series $\{w_t\}$ and for $j \in \mathbb{N}$, we let $cum\left(w_0, w_{t_1}, \ldots, w_{t_j-1}\right)$ denote the j^{th} order joint cumulant of $\left(w_0, w_{t_1}, \ldots, w_{t_{j-1}}\right)$ (see Brillinger, 1981, p. 19), where $t_1, \ldots, t_{j-1}$ are integers[2].

Assumption A1

(i) $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ are independent across i.

(ii) For each i, $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ is a strictly stationary martingale difference sequence, i.e. $E\left(\varepsilon_{it}|\mathcal{F}_i^{t-1}\right) = 0$, a.s., where $\mathcal{F}_i^{t-1} = \sigma\left(\varepsilon_{it-1}, \varepsilon_{it-2}, \ldots\right)$.

(iii) $E\left|\varepsilon_{it}\right|^{4r}$ is uniformly bounded in i and t, for some $r \geq 2$.

(iv) $E\left(\varepsilon_{it}^2\right) = \sigma_i^2$, where $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \sigma^2 < \infty$.

(v) $E(\varepsilon_{it}^2\varepsilon_{it-l}\varepsilon_{it-p}) = \tau_{ilp}$ is uniformly bounded for all i, t, $l \geq 1$, $p \geq 1$; $\tau_{ill} > 0$ for all l, and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tau_{ilp} = \tau_{lp}$, for fixed l, $p \in \mathbb{N}$.

(vi) $\sum_{t_1,t_2,t_3=-\infty}^{+\infty}\left|cum\left(\varepsilon_{it_1}, \varepsilon_{it_2}, \varepsilon_{it_3}, \varepsilon_{i0}\right)\right| < \Delta < \infty$ uniformly in i.

(vii) $\sum_{t_1,t_2,t_3=-\infty}^{+\infty}\left|cum\left(z_{it_1}^{l_1}, z_{it_2}^{l_2}, z_{it_3}^{l_3}, z_{i0}^{l_4}\right)\right| < \Delta < \infty$ uniformly in i, l_1, l_2, l_3 and l_4, where $z_{it}^{l} = \varepsilon_{it}\varepsilon_{it-l}$ and $l_1,\ldots,l_4$ are positive integers.

(viii) $\frac{1}{n}\sum_{i=1}^{n}\left|\alpha_i\right|^2 = O(1)$.

(ix) $n, T \to \infty$ such that $n/T \to \rho < \infty$.

In this paper, we follow the fixed effects approach and treat α_i as parameters to be estimated. Accordingly, Assumption A1 implicitly treats α_i as being constant. Alternatively, our analysis can be

[2] In particular, $cum\left(w_0\right) = E\left(w_0\right)$ and $cum\left(w_0, w_{t_1}\right) = Cov\left(w_0, w_{t_1}\right)$. For a zero random variable, $cum\left(w_0, w_{t_1}, w_{t_2}\right) = E\left(w_0 w_{t_1} w_{t_2}\right)$ and $cum\left(w_0, w_{t_1}, w_{t_2}, w_{t_3}\right) = E\left(w_0 w_{t_1} w_{t_2} w_{t_3}\right) - E\left(w_0 w_{t_1}\right)E\left(w_{t_2} w_{t_3}\right) - E\left(w_0 w_{t_2}\right)E\left(w_{t_1} w_{t_3}\right) - E\left(w_0 w_{t_3}\right)E\left(w_{t_1} w_{t_2}\right)$.

interpreted as being conditional on a random realization of the fixed effects α_i as long as we modify our assumptions by conditioning on α_i.[3]

Assumption A1(i) assumes cross sectional independence. Although we do not impose homogeneity along the cross sectional dimension, we nevertheless require this heterogeneity to disappear asymptotically. Assumption A1(ii) imposes a martingale difference sequence restriction on $\{\varepsilon_{it} : t = 1, 2, \ldots\}$ for each $i = 1, \ldots, n$; time stationarity is also assumed for simplicity. The m.d.s. assumption implies that the model for the conditional mean of y_{it} given $\mathcal{F}_i^{t-1}$ is correctly specified. This is a strong assumption that has been recently relaxed by Galvão and Kato (2013) in the context of possibly misspecified linear dynamic panel data models with fixed effects. Specifically, their results show that the pairs bootstrap is asymptotically valid for inference on a pseudo-true parameter when applied to a bias-corrected estimator. Here, we assume the model is corrected specified for the conditional mean, which allows us to obtain results for the recursive-design bootstrap based on the wild bootstrap. The motivation for this method relies on the fact that the m.d.s assumption restricts the dependence in the time dimension, ruling out serial correlation in ε_{it}, but allows for time series dependence in the form of conditional heteroskedasticity. Allowing for conditional heteroskedasticity over time is important in order to capture GARCH effects, as documented by the increasing literature on estimating large dimensional GARCH panels (see e.g. Engle, Shephard, and Sheppard (2008) and Pakel, Shephard, and Sheppard (2011)). Assumption A1(vi) restricts the fourth order cumulants of ε_{it} whereas Assumption A1(vii) is an additional eighth order restriction on the distribution of the innovations needed to establish a central limit theorem and justify covariance matrix estimation. Given that $|\theta_0| < 1$, it implies Condition 3 of Hahn and Kuersteiner (2002). Assumption A1(ix) assumes that n and T diverge to infinity at the same rate and is standard in this literature.

Under Assumption A1, we can prove the following result. See Appendix A for the proof.

Theorem 2.1 *Let $\{y_{it}\}$ be generated by (1). Under Assumption A1, we have*

$$\sqrt{nT}\left(\hat{\theta} - \theta_0\right) \to^d N\left(D, C\right),$$

where $D = -\sqrt{\rho}(1 + \theta_0)$; and $C = A^{-1}BA^{-1}$, with $A = \sigma^2\left(1 - \theta_0^2\right)^{-1}$ and $B = \sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\tau_{lp}$.

Theorem 2.1 is a restatement of Hahn and Kuesteiner's (2002) Theorem 1 under our Assumption A1. The method of proof follows closely that of Gonçalves and Kilian (2004), adapted to the panel context considered here. In particular, the cross sectional independence assumption A1(i) allows us to use results by Hansen (2007) (see also Moon and Phillips (2004)) to derive the joint asymptotic theory of $\hat{\theta}$ as $n, T \to \infty$ under Assumption A1.

[3] For instance, A1(ii) should read "For each i, $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ is a strictly stationary martingale difference sequence conditional on α_i, i.e. $E\left(\varepsilon_{it}|\mathcal{F}_i^{t-1}, \alpha_i\right) = 0$, where $\mathcal{F}_i^{t-1} = \sigma\left(\varepsilon_{it-1}, \varepsilon_{it-2}, \ldots\right)$.". Similarly, all expectations should be conditional on α_i and the limits in parts (iv) and (v) should be replaced with probability limits. See Remark 1 of Hahn and Kuersteiner (2011) for more details on the appropriate modifications.

Presenting this result and its heuristic derivation is helpful in understanding the reasons for the (in)validity of the different bootstrap methods we consider in the next section. The fixed effects OLS estimator can be represented as

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right)=A_{nT}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right),$$

where $A_{nT}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)^2$. Under Assumption A1, we show in the Appendix that $A_{nT}\to^P A$. Moreover, adding and subtracting $\mu_i\equiv E\left(y_{it-1}\right)=\alpha_i/\left(1-\theta_0\right)$ to the term $\left(y_{it-1}-\bar{y}_{i-}\right)$ and using the fact that the average over t of $\left(\varepsilon_{it}-\bar{\varepsilon}_i\right)$ is zero implies that

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right)=A^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right)+o_P\left(1\right).$$

The following decomposition holds for the normalized score,

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right)=\underbrace{\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\varepsilon_{it}}_{\to^d N(0,B)}-\underbrace{\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\bar{\varepsilon}_i}_{\to^P -A\cdot D},$$

where the stochastic behavior of each of the two terms above is discussed in Lemma A.4 in Appendix A.

This result has two implications for the validity of the bootstrap. First, the bootstrap needs to mimic the asymptotic variance of $\hat{\theta}$ given by $C=A^{-1}BA^{-1}$. This variance has the usual sandwich form under conditional heteroskedasticity. In particular, it depends on the long run variance of the score process (after concentrating out α_i) defined as

$$B=\lim_{n,T\to\infty}Var\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\varepsilon_{it}\right).$$

Theorem 2.1 shows that B depends on[4] τ_{lp}, the limiting value of the cross sectional average of the fourth order cumulants of ε_{it}. When ε_{it} are i.i.d. $\left(0,\sigma^2\right)$, we have that $\tau_{lp}=\sigma^4$ for $l=p$ and $\tau_{lp}=0$ for $l\neq p$, implying that $B=\sigma^4/\left(1-\theta_0^2\right)$. In this case, $B=\sigma^2 A$ and $C=1-\theta_0^2$. But when ε_{it} are heteroskedastic (in either dimension), the fourth order cumulants of ε_{it} do not simplify and the sandwich form for C is obtained. As discussed by Gonçalves and Kilian (2004) in the pure time series context, bootstrap validity depends on replicating the properties of τ_{lp} and this is also true in the panel context.

[4] Note that Hahn and Kuersteiner (2002) obtain a different but equivalent expression for B, given by $\frac{\sigma^4}{1-\theta_0^2}+\chi$, where $\chi\equiv\sum_{t=-\infty}^{\infty}\chi(t,0)$ and $\chi(t_1,t_2)\equiv E[u_{it_1-1}u_{it_2-1}\varepsilon_{it_1}\varepsilon_{it_2}]-E[\varepsilon_{it_1}\varepsilon_{it_2}]E[u_{it_1-1}u_{it_2-1}]$, $u_{it-1}=y_{it-1}-E\left(y_{it-1}\right)$. The constant χ reflects higher order moments of the error term when conditional heteroskedasticity is allowed for and it becomes zero when ε_{it} is i.i.d. $\left(0,\sigma^2\right)$, implying the same value for B. Our expression makes the comparison of our results with Gonçalves and Kilian (2004) easier.

Second, the bootstrap needs to capture the asymptotic bias term D created by the estimation of the fixed effects. As the decomposition above shows (and as was discussed already by Hahn and Kuersteiner (2002)), this noncentrality parameter results from the correlation between the averaged error terms $\bar{\varepsilon}_i$ and the demeaned regressors $y_{it-1} - \mu_i$ and is non zero when $\rho = \lim \frac{n}{T} \neq 0$. As we will see next, the presence of this incidental parameter asymptotic bias is the crucial difference between the application of the bootstrap in the pure time series context considered in Gonçalves and Kilian (2004) and in the panel context considered here.

3 Bootstrap results for the fixed effects estimator

In this section, we study the asymptotic validity of the bootstrap when applied to the fixed effects OLS estimator $\hat{\theta}$. Following Gonçalves and Kilian (2004), we consider three bootstrap methods adapted to the panel AR(1) model considered here. Two of these are residual-based wild bootstrap (WB) methods whereas the third one is a pairs bootstrap that resamples only in the cross sectional dimension (which is justified under our cross sectional independence assumption).

We use the following notation for the bootstrap asymptotics (see Chang and Park (2003) for similar notation and for several useful bootstrap asymptotic properties): Let Z_{nT}^* be a sequence of bootstrap statistics. We write $Z_{nT}^* = o_{P^*}(1)$ in probability, or $Z_{nT}^* \to^{P^*} 0$ in probability, if for any $\varepsilon > 0$, $\delta > 0$, $\lim_{n,T\to\infty} P\left[P^*\left(|Z_{nT}^*| > \delta\right) > \varepsilon\right] = 0$. Similarly, we write $Z_{nT}^* = O_{P^*}(1)$ in probability if for all $\varepsilon > 0$ there exists a $M_\varepsilon < \infty$ such that $\lim_{n,T\to\infty} P\left[P^*\left(|Z_{nT}^*| > M_\varepsilon\right) > \varepsilon\right] = 0$. Finally, we write $Z_{nT}^* \to^{d^*} Z$ in probability if, conditional on the sample, Z_{nT}^* weakly converges to Z under P^*, for all samples contained in a set with probability converging to one. Specifically, we write $Z_{nT}^* \to^{d^*} Z$ in probability if and only if $E^*\left(f\left(Z_{nT}^*\right)\right) \to E\left(f\left(Z\right)\right)$ in probability for any bounded and uniformly continuous function f.

3.1 Recursive-design wild bootstrap

The recursive-design bootstrap generates a panel of pseudo observations $\{y_{it}^*,\ i = 1,\ldots,n;\ t = 1,\ldots,T\}$ recursively from the panel AR(1) model with estimated parameters,

$$y_{it}^* = \hat{\alpha}_i + \hat{\theta} y_{it-1}^* + \varepsilon_{it}^*,\ i = 1,\ldots,n;\ t = 1,\ldots,T,$$

where $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\theta} y_{it-1})$, $i = 1,\ldots,n$ and $\hat{\theta}$ is the fixed effects OLS estimator defined in the previous section (the method remains valid if $\hat{\theta}$ is replaced with any consistent estimator $\tilde{\theta}$ of θ_0). The initial condition is $y_{i0}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}}$, $i = 1,\ldots,n$, which is equivalent to setting y_{i0}^* equal to the stationary mean in the bootstrap world. The bootstrap residuals are obtained with the wild bootstrap $\varepsilon_{it}^* = \hat{\varepsilon}_{it}\eta_{it}$, where $\eta_{it} \sim$ i.i.d.$(0,1)$ over (i,t) with $E^*|\eta_{it}|^4 \leq \Delta < \infty$, and $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta} y_{it-1}$ are the estimated residuals. The wild bootstrap was originally proposed by Wu (1986) and Liu (1988) in the context of cross section regressions with unconditional heteroskedasticity. Its application to the time series autoregressive context was considered by Gonçalves and Kilian (2004) (see also Kreiss (1997)). Here we extend its application to the panel autoregressive context with individual fixed effects (see Gonçalves and Perron (2014) for a recent application to panel factor models).

Letting η_{it} be i.i.d.$(0,1)$ along the two dimensions is appropriate since by Assumption A1 ε_{it} is independent across i and uncorrelated over t (due to the m.d.s. assumption), but we allow for heteroskedasticity in the two dimensions.

The bootstrap analogue of $\hat{\theta}$ is $\hat{\theta}^*_{rd}$, the recursive-design wild bootstrap OLS estimator,

$$\hat{\theta}^*_{rd} = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \bar{y}^*_{i-} \right)^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \bar{y}^*_{i-} \right) \left(y^*_{it} - \bar{y}^*_{i} \right), \tag{2}$$

where $\bar{y}^*_i$ and $\bar{y}^*_{i-}$ are defined analogously to $\bar{y}_i$ and $\bar{y}_{i-}$.

As in Gonçalves and Kilian (2004), we require a strengthening of Assumption A1 to establish the validity of the recursive-design wild bootstrap for the fixed effects OLS estimator.

A1. (v′) $\tau_{ilp} \equiv E\left(\varepsilon^2_{it}\varepsilon_{it-l}\varepsilon_{it-p}\right) = 0$ for all $l \neq p$, for all i, and t, $l \geq 1$, $p \geq 1$.

A1 (v′) is the panel analogue of Assumption A′(iv′) in Gonçalves and Kilian (2004). As they remark, this assumption further restricts the class of conditionally heteroskedastic autoregressive models that are covered by excluding certain asymmetric GARCH and ARCH models (e.g. the popular EGARCH model). This is crucial to prove that the bootstrap variance of $\hat{\theta}^*_{rd}$ is consistent for C.

Theorem 3.1 *Under Assumption A1 strengthened by Assumption A1(v′), it follows that*

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta}) \leq x) - P(\sqrt{nT}(\hat{\theta} - \theta_0) \leq x) \right| \to^{P} 0.$$

The proof of Theorem 3.1 is in Appendix B. The crucial difference compared to the proof of Theorem 3.2 of Gonçalves and Kilian (2004) is the need to account for the incidental parameter bias generated by the estimation of the fixed effects. In particular, Lemma B.4 in Appendix B shows that the incidental parameter bias in the bootstrap world is such that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \hat{\mu}_i \right) \bar{\varepsilon}^*_i \to^{P^*} -A \cdot D,$$

in probability, where $\hat{\mu}_i = \hat{\alpha}_i / \left(1 - \hat{\theta}\right) = E^*\left(y^*_{it-1}\right)$. This, together with the fact that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \hat{\mu}_i \right) \varepsilon^*_{it} \to^{d^*} N\left(0, \tilde{B}\right),$$

in probability, where $\tilde{B} = \sum_{l=1}^{\infty} \theta_0^{2(l-1)} \tau_{ll}$, implies that $\sqrt{nT}\left(\hat{\theta}^*_{rd} - \hat{\theta}\right) \to^{d^*} N\left(D, A^{-1}\tilde{B}A^{-1}\right)$, in probability. Since $\tilde{B} = B$ whenever $\tau_{i,lp} = 0$ for $l \neq p$ (i.e. under A1(v′)), the recursive-design wild bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}^*_{rd}-\hat{\theta}\right)$ is consistent for the distribution of the biased fixed effects OLS estimator $\sqrt{nT}\left(\hat{\theta}-\theta\right)$. In particular, the recursive-design bootstrap contains a built-in bias correction term that mimics the incidental parameter bias induced by the individual fixed effects.

Theorem 3.1 justifies the construction of bootstrap percentile-type confidence intervals for θ_0 without the need for an explicit bias correction method. It does not however justify the use of the bootstrap to consistently estimate the bias of $\hat{\theta}$ without further conditions, for instance that the sequence $\left\{\sqrt{nT}\left(\hat{\theta}^*_{rd}-\hat{\theta}\right)\right\}$ is uniformly integrable (see e.g. Billingsley (1995), Theorem 25.12).

Although our focus in this paper is on using the bootstrap for constructing confidence intervals for θ_0, we now provide a result that theoretically justifies the use of the bootstrap for bias correction. The bootstrap has been used for this purpose in Everaert and Pozzi (2007) without a theoretical justification. Compared to the analytical bias correction method of Hahn and Kuersteiner (2002) (and of many others since then), the bootstrap approach is easy to generalize to more complex models without requiring the need for different analytical formulae.

Following Liu and Singh (1992) and Gonçalves and White (2005), we focus on the following bootstrap fixed effects estimator

$$\tilde{\theta}^* = \begin{cases} \hat{\theta}^*_{rd} & \text{if } \dfrac{1}{nT}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{T}(y^*_{it-1}-\bar{y}^*_{i-})^2 \geq \dfrac{\delta}{2} \\ \hat{\theta} & \text{otherwise,} \end{cases}$$

for some $\delta > 0$. Thus, $\tilde{\theta}^*$ is equal to $\hat{\theta}^*_{rd}$ whenever $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)^2$ is bounded away from zero. Since $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)^2 \rightarrow^{P^*} A > 0$, in probability, it follows that for any $\varepsilon > 0$ and sufficiently large n and T, there exists $\delta > 0$ such that

$$P\left[P^*\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)^2 \geq \frac{\delta}{2}\right) > 1-\varepsilon\right] > 1-\varepsilon. \tag{3}$$

Thus, this modification does not have adverse practical consequences but at the same time it greatly simplifies the theoretical study of the bootstrap bias estimator $D^* = E^*\left(\sqrt{nT}\left(\tilde{\theta}^*-\hat{\theta}\right)\right)$.

Theorem 3.2 *Under the same assumptions as in Theorem 3.1, $D^* \rightarrow^P D$, where $D^* = E^*\left(\sqrt{nT}(\tilde{\theta}^*-\hat{\theta})\right)$ and $D = -\sqrt{\rho}(1+\theta_0)$.*

The proof of Theorem 3.2 is in Appendix B. We show that under Assumption A1 strengthened by A1(v′), $E^*\left(\left|\sqrt{nT}(\tilde{\theta}^*-\hat{\theta})\right|^{1+\delta}\right) = O_P(1)$ for some $\delta > 0$, which is a sufficient condition for the uniform integrability of the sequence $\left\{\left|\sqrt{nT}(\tilde{\theta}^*-\hat{\theta})\right|\right\}$, in probability. This together with Theorem 3.1 implies Theorem 3.2.

To end this section, we discuss bootstrap percentile-t intervals based on the following t-statistic

$$t_{\hat{\theta}^*_{rd}} = \frac{\sqrt{nT}\left(\hat{\theta}^*_{rd} - \hat{\theta}\right)}{\sqrt{\hat{C}^*_{rd}}},$$

where $\hat{C}^*_{rd} = \hat{A}^{*-1}_{rd}\hat{B}^*_{rd}\hat{A}^{*-1}_{rd}$, with

$$\hat{A}^*_{rd} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \text{ and } \hat{B}^*_{rd} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \tilde{\varepsilon}^{*2}_{it}, \tag{4}$$

and $\tilde{\varepsilon}^*_{it} = y^*_{it} - \bar{y}^*_i - \hat{\theta}^*\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$. The statistic $t_{\hat{\theta}^*_{rd}}$ is the bootstrap analogue of $t_{\hat{\theta}} = \sqrt{nT}\left(\hat{\theta} - \theta_0\right)/\sqrt{\hat{C}}$, where $\hat{C}$ is defined as $\hat{C}^*$ using the original data.

Given Theorems 2.1 and 3.1, the asymptotic validity of a bootstrap percentile-t interval based on $t_{\hat{\theta}^*_{rd}}$ follows from the following lemma. It shows the consistency of $\hat{C}^*_{rd}$ towards $C = A^{-1}BA^{-1}$, where $B = \tilde{B}$ under Assumption A1 (v′).

Lemma 3.1 *Under the same assumptions as in Theorem 3.1,* $\hat{C}^*_{rd} \to^{P^*} C = A^{-1}\tilde{B}A^{-1}$, *in probability.*

3.2 Fixed-design wild bootstrap

The fixed-design wild bootstrap generates $\{y^*_{it},\ i = 1, \ldots, n;\ t = 1, \ldots, T\}$ according to

$$y^*_{it} = \hat{\alpha}_i + \hat{\theta} y_{it-1} + \varepsilon^*_{it},\ i = 1, \ldots, n;\ t = 1, \ldots, T, \tag{5}$$

where $\varepsilon^*_{it} = \hat{\varepsilon}_{it}\eta_{it}$, with $\eta_{it} \sim$ i.i.d.$(0,1)$ across (i,t) such that $E^*\left|\eta_{it}\right|^4 \leq \Delta < \infty$. As for the recursive-design wild bootstrap, $\hat{\theta}$ can be replaced by any consistent estimator $\tilde{\theta}$ of θ_0 and $\hat{\alpha}_i$ by $\tilde{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \tilde{\theta}y_{it-1}\right)$.

The fixed-design wild bootstrap estimator is

$$\hat{\theta}^*_{fd} = \left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i-}\right)^2\right)^{-1}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i-}\right)\left(y^*_{it} - \bar{y}^*_i\right). \tag{6}$$

Gonçalves and Kilian (2004) consider this method in the context of a pure time series autoregression and show that it is asymptotically valid for estimating the distribution of the autoregressive parameter under conditional heteroskedasticity of unknown form. In particular, and in contrast to the recursive-design wild bootstrap, the fixed-design wild bootstrap is more generally applicable because it does not require Assumption A1(v′), thus allowing for leverage effects in the form of an asymmetric response of volatility to positive and negative shocks of the same absolute magnitude. It is therefore interesting to know whether this method is valid in the context of a panel autoregression model with individual fixed effects.

Theorem 3.3 *Under Assumption A1, it follows that* $\sqrt{nT}\left(\hat{\theta}^*_{fd} - \hat{\theta}\right) \to^{d^*} N\left(0, C\right)$, *in probability, where* $C = A^{-1}BA^{-1}$, *with* A *and* B *defined as in Theorem 2.1.*

The proof of Theorem 3.3 is in Appendix B. In contrast to the recursive-design wild bootstrap, the fixed-design wild bootstrap is not able to reproduce the noncentrality parameter of the limiting distribution of the fixed effects OLS estimator. The bootstrap distribution of $\sqrt{nT}(\hat{\theta}^*_{fd} - \hat{\theta})$ is incorrectly centered at zero, as $n, T \to \infty$.

The reason for the failure of the fixed-design wild bootstrap to capture the incidental parameter bias is that it destroys the correlation between the average bootstrap residuals $\bar{\varepsilon}^*_i$ and the bootstrap regressors $y^*_{it-1} - \hat{\mu}_i$ because it fixes these at the sample values, i.e. $y^*_{it-1} - \hat{\mu}_i = y_{it-1} - \hat{\alpha}_i / \left(1 - \hat{\theta}\right)$. This implies that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \hat{\mu}_i\right) \bar{\varepsilon}^*_i = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \hat{\mu}_i\right) \bar{\varepsilon}^*_i \to^{P^*} 0,$$

since $E^* \left(\bar{\varepsilon}^*_i\right) = 0$.

Two implications follow from this negative result. First, the fixed-design wild bootstrap cannot be used to approximate the distribution (nor the bias) of the biased fixed effects OLS estimator $\hat{\theta}$. As our simulations show, this method does not replicate the incidental parameter bias of $\hat{\theta}$ and therefore fails when used to construct percentile (or percentile-t) bootstrap confidence intervals for θ_0 based on this estimator. The second implication is that its invalidity extends to bootstrap confidence intervals for θ_0 based on the bias-corrected estimator that relies on the analytical bias correction method of Hahn and Kuersteiner (2002). We will discuss the application of the bootstrap to the bias-corrected estimator of Hahn and Kuersteiner (2002) in Section 4.

3.3 Pairs bootstrap

A third method that is robust to conditional heteroskedasticity of unknown form in the error term of a pure time series autoregressive model is the pairs bootstrap, where one resamples with replacement the vector that collects the dependent variable and its lagged values. This method was also studied by Gonçalves and Kilian (2004), who proved its asymptotic validity under the same assumptions as those underlying the validity of the fixed-design wild bootstrap.

The goal of this section is to study the applicability of a panel version of this bootstrap method in the context of a panel AR(1) model with individual specific fixed effects. Specifically, we consider resampling only in the cross-sectional dimension, by resampling the "pairs" (y_i, y_{i-}), where $y_i = \begin{pmatrix} y_{i1} & \dots & y_{iT} \end{pmatrix}'$ and $y_{i-} = \begin{pmatrix} y_{i0} & \dots & y_{iT-1} \end{pmatrix}'$. This method was proposed by Kapetanios (2008) in the context of a panel regression model with strictly exogeneous regressors and fixed effects, in which case no incidental parameter bias exists[5]. Our contribution here is to analyze the properties of this method for linear dynamic panel models where the incidental parameter bias is present. Note that there are other ways of resampling the pairs (y_{it}, y_{it-1}) in the panel context. For instance, one alternative bootstrap method is to resample only in the time dimension, by resampling the "pairs"

[5]See also Hounkannounon (2010) for the applicability of this method in the context of panel regression models with random effects.

(y_t, y_{t-1}), where $y_t = \begin{pmatrix} y_{1t} & \dots & y_{nt} \end{pmatrix}'$ and $y_{t-1} = \begin{pmatrix} y_{1t-1} & \dots & y_{nt-1} \end{pmatrix}'$. This method was also considered in Kapetanios (2008) and more recently in Gonçalves (2011), who showed the asymptotic validity of the moving blocks bootstrap under general forms of cross sectional dependence and time series dependence in the regression error of a panel linear regression model. Although the regularity conditions of Gonçalves (2011) allow in principle dynamic regressors, the impact of the incidental parameter bias on inference is ruled out by assuming that $n/T \to \rho = 0$. We do not consider this bootstrap method here because we assume cross sectional independence, in which case resampling in the cross sectional dimension is more appropriate.

More specifically, we generate $\left(y_i^*, y_{i-}^*\right) \sim$ i.i.d. $\{(y_i, y_{i-1}) : i = 1, \dots, n\}$, i.e. letting $I_1, \dots, I_n$ be i.i.d. Uniform on $\{1, \dots, n\}$, we have that

$$\left(y_i^*, y_{i-}^*\right) = \begin{pmatrix} y_{I_i,1} & y_{I_i,0} \\ \vdots & \vdots \\ y_{I_i,T} & y_{I_i,T-1} \end{pmatrix}.$$

The pairs bootstrap fixed effects estimator is then defined as the original fixed effects OLS estimator but with $\{(y_{it}, y_{it-1})\}$ replaced with $\left\{\left(y_{it}^*, y_{it-1}^*\right)\right\}$. Let $\hat{\theta}_{pb}^*$ denote this estimator.

Theorem 3.4 *Under Assumption A1, it follows that* $\sqrt{nT}\left(\hat{\theta}_{pb}^* - \hat{\theta}\right) \to^{d^*} N(0, C)$, *in probability, where* $C = A^{-1}BA^{-1}$, *with* A *and* B *defined as in Theorem 2.1.*

Similarly to the fixed-design wild bootstrap, the pairs bootstrap distribution of the bootstrap fixed effects OLS estimator is incorrectly centered at zero.

To understand the reason why the pairs bootstrap fails in capturing the bias, note that the pairs bootstrap fixed effects OLS estimator has the following representation

$$\sqrt{nT}\left(\hat{\theta}_{pb}^* - \hat{\theta}\right) = A_{nT}^{*-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\hat{\varepsilon}_{it}^* - \overline{\hat{\varepsilon}_i^*}\right),$$

where $A_{nT}^* = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2$ is the bootstrap analogue of A_{nT} and $\hat{\varepsilon}_{it}^*$ is the bootstrap version of the error term ε_{it}, i.e. $\hat{\varepsilon}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\theta} y_{it-1}^* \equiv \hat{\varepsilon}_{I_i,t}$. Since ε_{it} depends on α_i (which is a function of i), its bootstrap analogue when resampling in the cross sectional dimension involves resampling $\hat{\alpha}_i$, i.e. $\hat{\varepsilon}_{it}^*$ depends on $\hat{\alpha}_i^* = \hat{\alpha}_{I_i}$, a resampled version of $\hat{\alpha}_i$. Given that resampling only occurs in the cross sectional dimension, we can define

$$s_i^* \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\hat{\varepsilon}_{it}^* - \overline{\hat{\varepsilon}_i^*}\right)$$

as being the bootstrap version of $s_i \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-}\right)\left(\hat{\varepsilon}_{it} - \overline{\hat{\varepsilon}_i}\right)$, i.e. $s_i^* = s_{I_i}$ for all $i = 1, \dots, n$.

It follows that

$$\sqrt{nT}\left(\hat{\theta}_{pb}^{*}-\hat{\theta}\right)=A_{nT}^{*-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_{i}^{*}=A^{-1}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_{i}^{*}}_{\to^{d^*}N(0,B)}+o_{P^*}(1),$$

given that $A_{nT}^{*}\to^{P^*}A$, in probability. Since $I_1,\ldots,I_n$ are i.i.d. uniformly distributed on $\{1,\ldots,n\}$, $\{s_i^*: i=1,\ldots,n\}$ is i.i.d. (conditional on the original observations) and a bootstrap CLT holds for $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^*$, yielding an asymptotic normal distribution for $\sqrt{nT}\left(\hat{\theta}_{pb}^{*}-\hat{\theta}\right)$. Nevertheless, the asymptotic bootstrap population mean turns out to be zero because

$$E^*\left(s_i^*\right)=\frac{1}{n}\sum_{i=1}^{n}s_i=\frac{1}{n}\frac{1}{\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)\left(\hat{\varepsilon}_{it}-\overline{\hat{\varepsilon}_{i}}\right)=0,$$

by the first order condition for the fixed effects OLS estimator. Thus, the limiting bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}_{pb}^{*}-\hat{\theta}\right)$ is (incorrectly) centered at zero.

4 Bootstrapping the bias-corrected estimator

The results of Section 3 justify bootstrap inference on θ_0 based on the recursive-design bootstrap fixed effects OLS estimator $\hat{\theta}_{rd}^{*}$. In particular, Theorem 3.1 justifies the construction of bootstrap percentile intervals for θ_0 whereas Theorem 3.1 together with Lemma 3.1 justify bootstrap percentile-t intervals. Although these approaches are valid and have the advantage of avoiding the need for an explicit bias correction of $\hat{\theta}$, further finite sample improvements of the bootstrap approximation can be obtained if we base our inference on a bias-corrected estimator. Bootstrapping a bias-corrected fixed effects estimator removes the incidental parameter bias from the asymptotic distribution, resulting in a t-statistic that is asymptotically pivotal.

For the particular panel AR(1) model with individual fixed effects that we consider here, a simple analytical formula for the bias of $\hat{\theta}$ has been derived by Hahn and Kuersteiner (2002). Specifically, their bias-corrected fixed effects estimator is given by

$$\hat{\hat{\theta}}=\hat{\theta}+\frac{1}{T}\left(1+\hat{\theta}\right), \tag{7}$$

where $\hat{\theta}$ is the standard biased fixed effects OLS estimator. The intuition for this bias correction is simple: by Theorem 2.1, $\hat{\theta}-\theta_0$ is approximately distributed as $N\left(-\frac{1}{T}\left(1+\theta_0\right),\frac{1}{nT}C\right)$. Therefore, $\Delta=-\frac{1}{T}\left(1+\theta_0\right)$ is the bias of $\hat{\theta}$ of order $O\left(1/T\right)$. The bias-corrected estimator given in (7) is the feasible version of the infeasible bias-corrected estimator given by $\hat{\theta}-\Delta=\hat{\theta}+\frac{1}{T}\left(1+\theta_0\right)$.

The main contribution of this section is to prove the asymptotic validity of the recursive-design bootstrap when applied to $\hat{\hat{\theta}}$. As our simulations in Section 5 show, bootstrap intervals based on $\hat{\hat{\theta}}$ have coverage probabilities that are closer to the desired nominal level than the bootstrap intervals based on $\hat{\theta}$. We also consider the application of the fixed-design and the pairs bootstrap to $\hat{\hat{\theta}}$. Our results show that whereas the asymptotic invalidity of the fixed-design bootstrap to estimate the distribution of $\hat{\theta}$ extends to $\hat{\hat{\theta}}$, this is not the case for the pairs bootstrap, which becomes a valid method of inference when used to estimate the distribution of $\hat{\hat{\theta}}$.

We start by considering the recursive-design wild bootstrap, which we now implement using only bias-corrected estimates. More specifically, the bootstrap panel observations are generated recursively from the estimated panel AR(1) model using the bias-corrected estimates, i.e. we let

$$y_{it}^* = \hat{\hat{\alpha}}_i + \hat{\hat{\theta}} y_{it-1}^* + \varepsilon_{it}^*, \; i = 1, \ldots, n; \; t = 1, \ldots, T, \tag{8}$$

where $\hat{\hat{\alpha}}_i = \frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \hat{\hat{\theta}} y_{it-1} \right)$, $i = 1, \ldots, n$, and $\hat{\hat{\theta}}$ is the bias-corrected fixed effects OLS estimator defined in (7). The initial condition is $y_{i0}^* = \hat{\hat{\alpha}}_i \left(1 - \hat{\hat{\theta}} \right)^{-1}$, $i = 1, \ldots, n$.

Let $\hat{\hat{\theta}}_{rd}^*$ denote the bootstrap version of the bias-corrected fixed effects estimator (7), i.e.

$$\hat{\hat{\theta}}_{rd}^* = \hat{\theta}_{rd}^* + \frac{1}{T} \left(1 + \hat{\theta}_{rd}^* \right), \tag{9}$$

where $\hat{\theta}_{rd}^*$ is as defined in (2) but using bootstrap observations generated as in (8).

Our goal is to show the consistency of the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right)$ for the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$. An immediate consequence of Theorem 2.1 is that $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right) \to^d N(0, C)$ (see Theorem 2 of Hahn and Kuersteiner (2002)). Therefore, it suffices to show that $\sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right) \to^{d^*} N(0, C)$, in probability. This is an immediate consequence of the proof of Theorem 3.1. Heuristically, by replacing $\hat{\hat{\theta}}_{rd}^*$ with (9) we have that

$$\sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right) = \underbrace{\sqrt{nT}\left(\hat{\theta}_{rd}^* - \hat{\hat{\theta}}\right)}_{\to^{d^*} N(D,C)} + \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}_{rd}^*\right)}_{\to^{P^*} \sqrt{\rho}(1+\theta_0) \equiv -D} \quad \longrightarrow^{d^*} N(0, C),$$

where the first term converges in distribution to $N(D, C)$ by Theorem 3.1 (note that we center $\hat{\theta}_{rd}^*$ around $\hat{\hat{\theta}}$ because the bootstrap DGP (8) depends on $\hat{\hat{\theta}}$; using $\hat{\hat{\theta}}$ instead of $\hat{\theta}$ does not change the consistency result of Theorem 3.1 as long as we center $\hat{\hat{\theta}}_{rd}^*$ around $\hat{\hat{\theta}}$ because $\hat{\hat{\theta}}$ is a consistent estimator of θ_0). The second term converges in probability to $-D$ because $\hat{\theta}_{rd}^*$ is a consistent estimator of θ_0 (albeit biased) and $n/T \to \rho$ under Assumption A1.

Theorem 4.1 below states this result formally.

Theorem 4.1 *Under the same assumptions as in Theorem 3.1, we have that*

$$\sup_{x \in \mathbb{R}} \left| P^*\left(\sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right) \leq x\right) - P\left(\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right) \leq x\right) \right| \to^P 0.$$

Theorem 4.1 justifies using the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right)$ to consistently estimate the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$. The consistency of the distribution of the bootstrap t-statistic $t_{\hat{\hat{\theta}}_{rd}^*} = \sqrt{nT}\left(\hat{\hat{\theta}}_{rd}^* - \hat{\hat{\theta}}\right) / \sqrt{\hat{C}^*}$ follows whenever $\hat{C}^*$ is a consistent estimator of C, as in Lemma 3.1. In particular, our proposal is to choose $\tilde{C}^*_{rd} = \tilde{A}^{*-1}_{rd} \tilde{B}^*_{rd} \tilde{A}^{*-1}_{rd}$, where $\tilde{A}^*_{rd}$ and $\tilde{B}^*_{rd}$ are exactly as defined in (4) with the difference that $\{y^*_{it}\}$ is generated as in (8) and $\tilde{\varepsilon}^*_{it}$ is a function of $\hat{\hat{\theta}}^*_{rd}$ instead of $\hat{\theta}^*_{rd}$. To conserve space, we do not provide the formal result but note that the same exact arguments used to prove Lemma 3.1 can be applied to show the consistency of $\tilde{C}^*_{rd}$ towards C. The Monte Carlo simulation results of the next section show that the finite sample properties of this approach are superior to the asymptotic normal approximation.

Next, we explain why the fixed-design bootstrap method is not asymptotically valid when applied to $\hat{\hat{\theta}}$. Let $\hat{\hat{\theta}}^*_{fd}$ denote the bootstrap version of $\hat{\hat{\theta}}$ where $\hat{\theta}^*_{fd}$ is computed as (6) with $\{y^*_{it}\}$ generated using equation (5) with $\hat{\theta}$ (and $\hat{\alpha}_i$) replaced with $\hat{\hat{\theta}}$ (and $\hat{\hat{\alpha}}_i$). Proceeding as for the recursive-design bootstrap, the following decomposition holds

$$\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right) = \underbrace{\sqrt{nT}\left(\hat{\theta}^*_{fd} - \hat{\hat{\theta}}\right)}_{\to^{d^*} N(0,C)} + \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}^*_{fd}\right)}_{\to^{P^*} \sqrt{\rho}(1+\theta_0) \equiv -D} \quad \longrightarrow^{d^*} N\left(-D, C\right),$$

where in particular Theorem 3.3 justifies the convergence of the the first term. This shows that the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ is incorrectly centered at $-D$ (the correct mean should be zero since the asymptotic distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$ is centered at 0).

In contrast, the pairs bootstrap is asymptotically valid when applied to $\hat{\hat{\theta}}$. In this case, letting $\hat{\hat{\theta}}^*_{pb}$ denote the bootstrap version of $\hat{\hat{\theta}}$ based on the biased fixed effects estimator $\hat{\theta}^*_{pb}$, we have that

$$\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right) = \underbrace{\sqrt{nT}\left(\hat{\theta}^*_{pb} - \hat{\theta}\right)}_{\to^{d^*} N(0,C)} - \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}\right)}_{\to^{P} -\sqrt{\rho}(1+\theta_0) \equiv D} + \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}^*_{pb}\right)}_{\to^{P^*} \sqrt{\rho}(1+\theta_0) \equiv -D} \quad \longrightarrow^{d^*} N\left(0, C\right).$$

Thus, although the pairs bootstrap does not provide a consistent estimator of the distribution of $\sqrt{nT}\left(\hat{\theta} - \theta_0\right)$ (because its asymptotic distribution is incorrectly centered at zero), the pairs bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right)$ is consistent for the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$. The formal result is stated in the following theorem.

Theorem 4.2 *Under the same assumptions as in Theorem 3.4, we have that*

$$\sup_{x \in \mathbb{R}} \left| P^*(\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right) \leq x) - P(\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right) \leq x) \right| \to^P 0.$$

For bootstrap percentile-t intervals based on the pairs bootstrap, we consider $t^*_{pb} = \sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right) / \sqrt{\tilde{C}^*_{pb}}$, with $\tilde{C}^*_{pb} = \tilde{A}^{*-1}_{pd} \tilde{B}^*_{pd} \tilde{A}^{*-1}_{pd}$, where $\tilde{A}^*_{pb}$ and $\tilde{B}^*_{pb}$ are defined as in (4) evaluated on the pairs bootstrap data and bias-corrected estimator. The analogue of Lemma 3.1 is as follows.

Lemma 4.1 *Under the same assumptions as in Theorem 3.4 ,* $\tilde{C}^*_{pb} \to^{P^*} C = A^{-1} B A^{-1}$*, in probability.*

5 Simulations

The goal of this section is to evaluate the finite sample performance of the three bootstrap methods studied in the previous section. We generate a panel of AR(1) processes with GARCH errors using the following equation

$$y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}, i = 1, \ldots, n; t = 1, \ldots, T, \tag{10}$$

where ε_{it} is such that $E\left(\varepsilon_{it}|\mathcal{F}_{it-1}\right) = 0$ and $Var\left(\varepsilon_{it}|\mathcal{F}_{it-1}\right) = \sigma_{it}^2$, with

$$\sigma_{it}^2 = \gamma_i\left(1 - \omega - \beta\right) + \omega\varepsilon_{it-1}^2 + \beta\sigma_{it-1}^2, \tag{11}$$

where $\gamma_i > 0$, $\omega, \beta \in [0, 1)$, and $\omega + \beta < 1$. See Pakel, Shephard, and Sheppard (2011) for more details on this particular GARCH specification. Because $\omega + \beta < 1$, these GARCH$(1, 1)$ processes are stationary but heterogeneous. In particular, the unconditional variance is given by γ_i. In the simulations, we set $\varepsilon_{it} \sim N\left(0, \sigma_{it}^2\right)$ where σ_{it} is given by (11) with $\sigma_{i0}^2 = \gamma_i$, the unconditional variance. Following Pakel, Shephard, and Sheppard (2011), we let $\gamma_i \sim$ i.i.d. $U[0.02, 0.05]$, which matches the range of annual volatility of most stock returns. The initial observations are drawn from the stationary distribution, $y_{i0} \mid \alpha_i, \gamma_i \sim N\left(\frac{\alpha_i}{1-\theta_0}, \frac{\gamma_i}{1-\theta_0^2}\right)$ and we set ω and β equal to 0.30 and 0.65, respectively. Since the fixed-effects estimator is invariant to α_i, we let $\alpha_i = 0$; in addition, we let $\theta_0 \in \{0.3, 0.6, 0.9, 0.99\}$, and consider $n \in \{20,\ 40,\ 60,\ 80,\ 100\}$ and $T \in \{10,\ 20,\ 30\}$.

Tables 1 and Figures 1-4 summarize our results, which are based on 2500 Monte Carlo simulations with 999 bootstrap replications each.

Table 1 reports the bias properties of the different methods. The first column corresponds to the true finite sample bias $E\left(\hat{\theta} - \theta_0\right)$ whereas the second column reports the estimated bias using the analytical correction of Hahn and Kuersteiner (2002) (i.e. $-\frac{1}{T}\left(1 + \hat{\theta}\right)$). The remaining three columns pertain to the bootstrap bias estimators based on the recursive-design wild bootstrap (RD), the fixed-design wild bootstrap (FD) and the pairs bootstrap (PB). To implement the residual-based wild bootstrap methods, we let η_{it} follow the Rademacher distribution (i.e. $\eta_{it} = 1$ with probability 0.5 and -1 with probability 0.5). We also used $\eta_{it} \sim N(0, 1)$ and η_{it} chosen according to Mammen (1993) but these choices were dominated by the Rademacher distribution, confirming the results by Davidson and Flachaire (2008) who advocate the use of the Rademacher distribution.

The simulation results in Table 1 confirm our theory. The FD and and the PB do not capture the incidental parameter bias whereas the RD does. An interesting result is that the RD outperforms the analytical bias correction of Hahn and Kuersteiner (2002), especially as θ_0 approaches 1.

Figures 1-4 report coverage rates of nominal 95% intervals for θ_0 based on the different bootstrap methods and the asymptotic normal distribution. We consider intervals based on $\hat{\theta}$ (Figures 1 and 2) and intervals based on its bias-corrected version $\hat{\hat{\theta}}$ (Figures 3 and 4). The bootstrap can yield both equal-tailed and symmetric intervals whereas the normal distribution generates symmetric intervals

Table 1: Performance of the bootstrap for bias-correction

			Bias				
T	n	θ_0	True	AT	RD	FD	PB
10	20	0.3	-0.098	-0.090	-0.098	0.000	0.000
		0.6	-0.138	-0.116	-0.135	0.000	-0.003
		0.9	-0.185	-0.141	-0.178	0.000	-0.005
		0.99	-0.256	-0.164	-0.232	0.000	-0.008
	60	0.3	-0.098	-0.09	-0.099	0.000	0.000
		0.6	-0.135	-0.117	-0.133	0.000	-0.001
		0.9	-0.180	-0.142	-0.174	0.000	-0.002
		0.99	-0.248	-0.165	-0.227	0.000	-0.004
	100	0.3	-0.099	-0.09	-0.099	0.000	0.000
		0.6	-0.135	-0.116	-0.133	0.000	-0.001
		0.9	-0.179	-0.142	-0.173	0.000	-0.002
		0.99	-0.245	-0.165	-0.225	0.000	-0.003
20	20	0.3	-0.049	-0.048	-0.049	0.000	0.000
		0.6	-0.070	-0.061	-0.069	0.000	-0.002
		0.9	-0.093	-0.075	-0.091	0.000	-0.004
		0.99	-0.130	-0.089	-0.124	0.000	-0.005
	60	0.3	-0.050	-0.047	-0.050	0.000	0.000
		0.6	-0.069	-0.062	-0.067	0.000	-0.001
		0.9	-0.089	-0.076	-0.088	0.000	-0.002
		0.99	-0.124	-0.089	-0.119	0.000	-0.002
	100	0.3	-0.050	-0.048	-0.050	0.000	0.000
		0.6	-0.067	-0.062	-0.067	0.000	-0.001
		0.9	-0.087	-0.076	-0.087	0.000	-0.001
		0.99	-0.122	-0.089	-0.118	0.000	-0.002

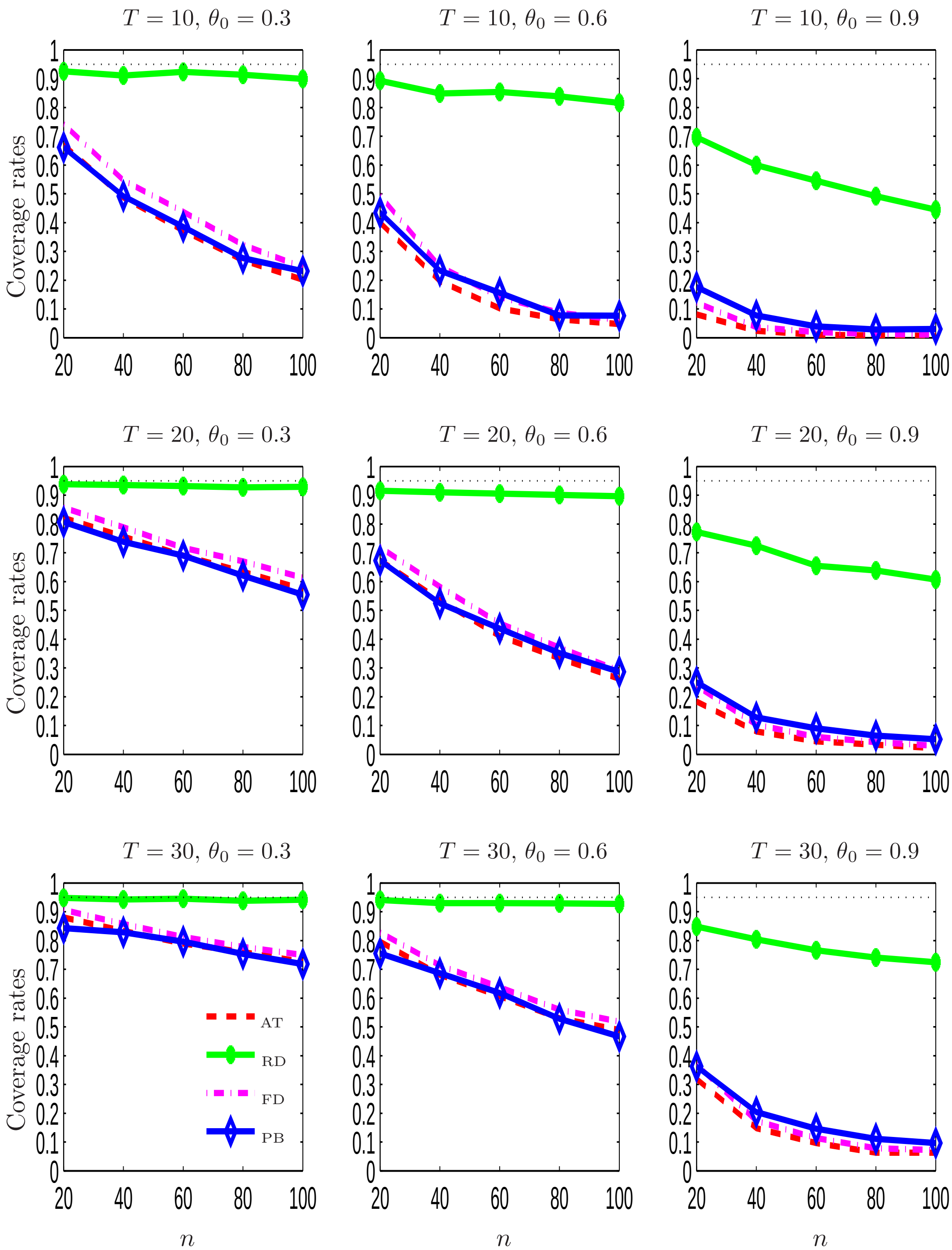

Figure 1: Coverage rates of nominal 95% symmetric intervals based on $\hat{\theta}$

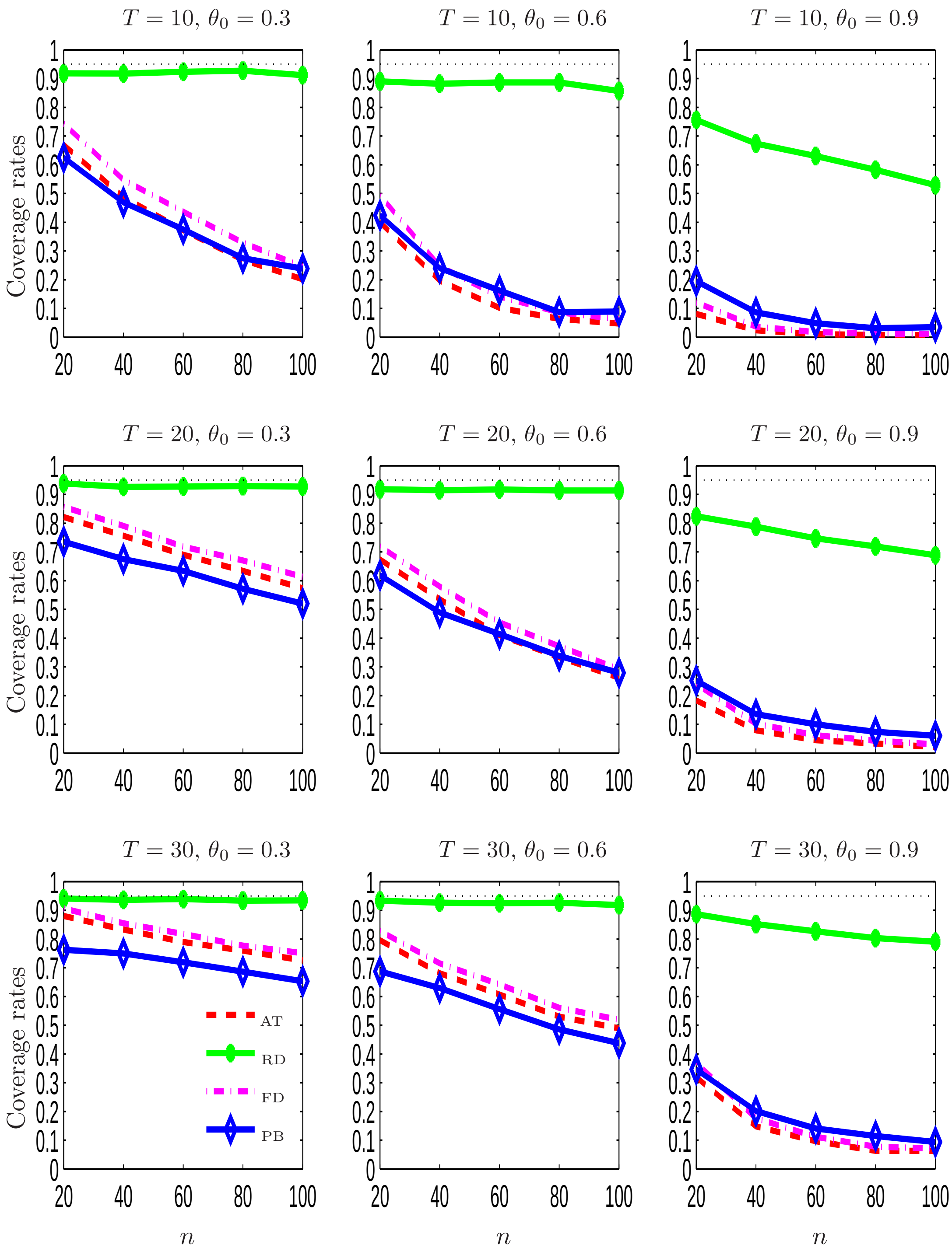

Figure 2: Coverage rates of nominal 95% equal-tailed confidence intervals based on $\hat{\theta}$

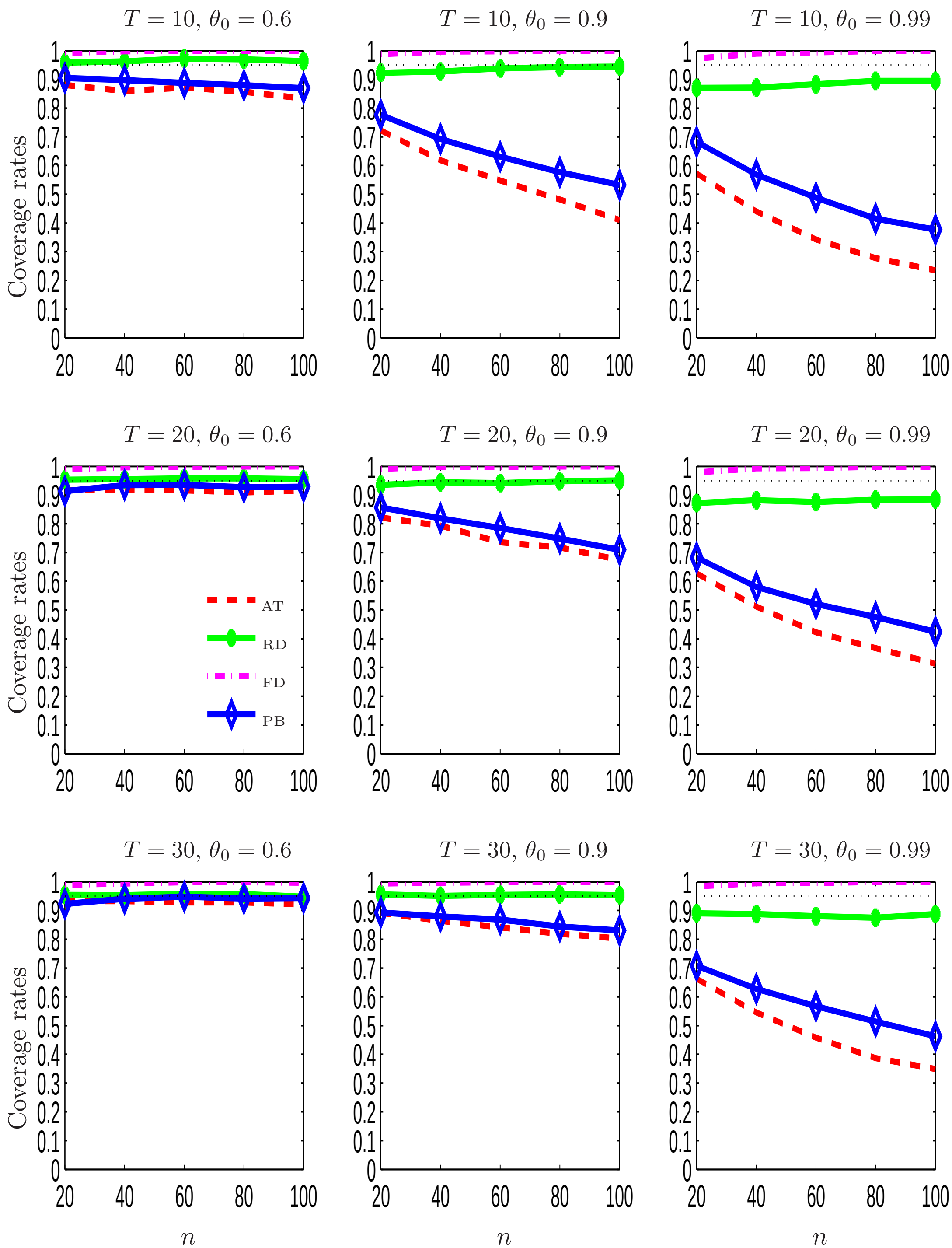

Figure 3: Coverage rates of nominal 95% symmetric intervals based on $\hat{\hat{\theta}}$

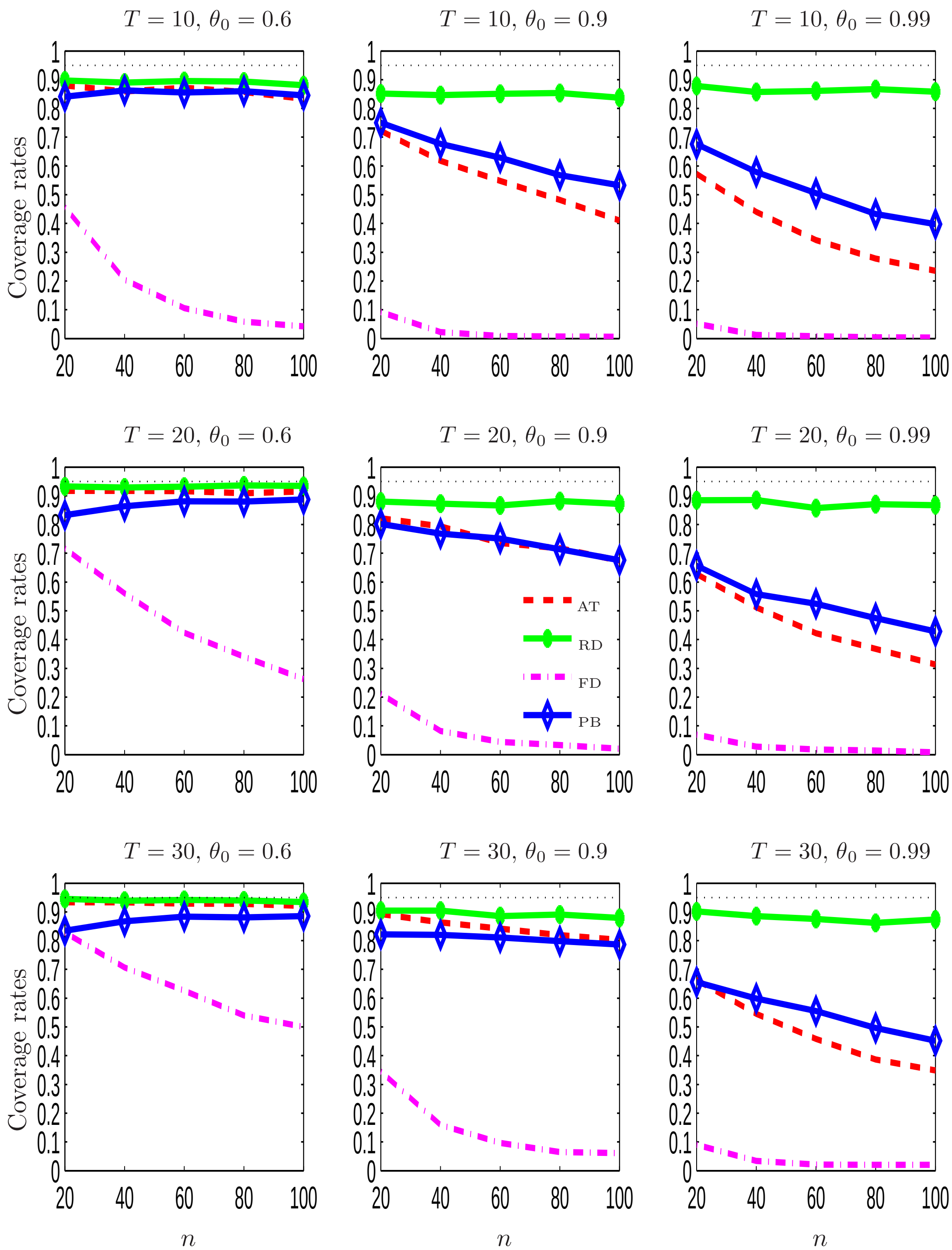

Figure 4: Coverage rates of nominal 95% equal-tailed confidence intervals based on $\hat{\hat{\theta}}$

by construction. Hence, we consider bootstrap symmetric intervals in Figures 1 and 3 and bootstrap equal-tailed intervals in Figures 2 and 4. Each figure contains nine plots, where each plot shows the actual coverage rates across different values of n for a given combination of T and θ_0. Specifically, we vary T across rows ($T \in \{10, 20, 30\}$) and θ_0 across columns ($\theta_0 \in \{0.3, 0.6, 0.9\}$ for Figures 1-2 and, $\theta_0 \in \{0.6, 0.9, 0.99\}$ for Figures 3-4). All intervals are based on t-statistics studentized with an heteroskedasticity-robust standard error.

Figure 1 shows that the asymptotic theory-based intervals that rely on the biased fixed-effects estimator can be severely distorted, especially as n increases. This is entirely expected because these intervals rely on the $N(0,1)$ distribution, which does not take into account the presence of the incidental parameter bias. We only include these intervals here as a benchmark for the PB and the FD bootstrap methods, which also fail to capture this bias. The results for these methods show that they indeed follow closely the intervals based on the asymptotic standard normal distribution. Figure 1 also shows that the RD bootstrap symmetric intervals outperform all the remaining intervals, essentially eliminating the coverage distortions for $\theta_0 = 0.3$ and 0.6. For these values of θ_0, the RD bootstrap shows very little sensitivity to increases of n, which reflects the fact that it contains a built-in incidental parameter bias correction. When $\theta_0 = 0.9$, the RD bootstrap rates deteriorate (with distortions increasing as a function of n), but it still dominates the remaining methods. Distortions decrease as a function of T. As Hahn and Kuersteiner (2002) show, the limiting distribution of $\hat{\theta}$ (and its rate of convergence) changes when $\theta_0 = 1$, which explains the deterioration of all methods in the vicinity of one. The comparison of Figure 1 with Figure 2 shows that equal-tailed intervals based on the RD bootstrap outperform the symmetric intervals, especially when θ_0 is large (and close to one).

Figure 3 shows that asymptotic theory-based intervals that rely on the bias-corrected estimator $\hat{\hat{\theta}}$ can be severely distorted in finite samples, especially if θ_0 is large. In particular, large distortions arise when $\theta_0 \in \{0.9, 0.99\}$. For instance, if $T = 10$ and $\theta_0 = 0.9$, the coverage rate of a 95% asymptotic theory-based interval varies between 70% and 40% for values of n between 20 and 100. These rates increase to around 90% to 80% when $T = 30$. When $\theta_0 = 0.99$, these numbers deteriorate by a lot, varying between 70% and 35% when $T = 30$. When $\theta_0 = 0.6$, the asymptotic theory works much better, but there are still noticeable coverage distortions when $T = 10$ (rates are around 90% in this case). By comparison, the RD bootstrap symmetric intervals are much less distorted for all combinations of n, T and θ_0. For $\theta_0 \in \{0.6, 0.9\}$, this method essentially eliminates all the coverage distortions noted for the asymptotic theory-based intervals. When $\theta_0 = 0.99$, rates deteriorate but not by much, remaining around 90% for all values of n and T. The PB tends to follow the asymptotic theory-based intervals when $\theta_0 = 0.6$, but it outperforms these intervals when θ_0 increases. Symmetric intervals tend to ouperform equal-tailed intervals for these two methods, as the comparison of Figures 3 and 4 shows.

The FD bootstrap symmetric intervals are too conservative for all combinations of n, T and θ_0. The reason for this behavior is that the FD bootstrap distribution is incorrectly centered at $-D =$

$\sqrt{\rho}\,(1+\theta_0) > 0$. Thus, the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ is shifted to the right of that of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$, implying that the bootstrap quantiles of the absolute value of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ will be systematically larger than those of the original finite sample distribution (centered at zero). Instead, the equal-tailed FD intervals tend to undercover, reflecting the fact that the bootstrap distribution is to the right of the true distribution. As n increases, this pushes the center of the bootstrap distribution further to the right, explaining the deterioration of the results for large values of n.

6 Conclusion

The main contribution of this paper is to study the validity of the bootstrap for inference on a stationary linear dynamic panel model with individual specific fixed effects. We consider three bootstrap methods: the recursive-design wild bootstrap, the fixed-design wild bootstrap and the pairs bootstrap. These methods are a natural generalization to the panel context of the bootstrap methods considered by Gonçalves and Kilian (2004) in the pure time series autoregressive model.

A crucial difference between the pure time series context and the panel context considered here is the presence of the incidental parameter bias due to the estimation of the fixed effects. We show that only the recursive-design bootstrap is able to capture this bias whereas the other two methods fail to do so. Thus, in contrast with the recursive-design wild bootstrap, the fixed-design and the pairs bootstrap do not consistently estimate the distribution of the standard biased fixed effects estimator and cannot be used for bias correction.

Although bootstrap intervals based on the biased fixed effects estimates are asymptotically valid if obtained with the recursive-design bootstrap, refinements can be obtained if bootstrap inference is based on the bias-corrected estimates. Our results show that the recursive-design is valid in this context whereas the fixed-design bootstrap is not. An interesting finding is that the invalidity of the pairs bootstrap to estimate the distribution of the biased fixed effects estimator does not prevent this method to be valid when applied to the bias-corrected estimates.

An important limitation of the present setup is the fact that we do not allow for additional regressors x_{it}.When these regressors are strictly exogeneous, a recursive-design bootstrap that fixes x_{it} at their original values should be able to capture the incidental bias. We have confirmed this by simulations (not reported here). Providing a proof of this result is outside the scope of this paper and is left for future research. The validity of the pairs bootstrap when applied to a bias-corrected estimator under the presence of extra regressors has recently been studied by Kaffo (2013) in the more general context of nonlinear dynamic models.

Further extensions of this work include the proposal of bootstrap methods that are robust to nonstationarity, where a form of the grid bootstrap can be useful, and a study of the higher order properties of the recursive-design bootstrap using Edgeworth expansions. These extensions are outside the scope of the present paper and are left for future research.

References

Alvarez, J., and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

Billingsley, P. (1995): *Probability and measure*, Wiley Series in Probability and Mathematical Statistics. 3rd ed., John Wiley & Sons Inc.

Brillinger, D. R. (1981): *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, second edition edn.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*. Springer, second edition edn.

Bun, M. J., and M. A. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business & Economic Statistics*, 23, 200–210.

Chang, Y., and J. Y. Park (2003): "A Sieve Bootstrap For The Test Of A Unit Root," *Journal of Time Series Analysis*, 24(4), 379–400.

Davidson, R., and E. Flachaire (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146(1), 162–169.

Engle, R. F., N. Shephard, and K. Sheppard (2008): "Fitting vast dimensional time-varying covariance models," Discussion paper.

Everaert, G., and L. Pozzi (2007): "Bootstrap-based bias correction for dynamic panels," *Journal of Economic Dynamics and Control*, 31(4), 1160–1184.

Fernandez-Val, I., and M. Weidner (2013): "Individual and time effects in nonlinear panel models with large N,T," CeMMAP working papers CWP60/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Galvão, A. F., and K. Kato (2013): "Estimation and Inference for Linear Panel Data Models under Misspecification when Both N and T are Large," forthcoming in the *Journal of Business and Economic Statistics*.

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

Gonçalves, S., and L. Kilian (2004): "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 123(1), 89–120.

Gonçalves, S., and B. Perron (2014): "Bootstrapping Factor-Augmented Regression Models," forthcoming in the *Journal of Econometrics*.

Gonçalves, S., and H. White (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100, 970–979.

Gouriroux, C., P. C. Phillips, and J. Yu (2010): "Indirect inference for dynamic panel models," *Journal of Econometrics*, 157(1), 68–77.

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(4), 1639–1657.

——— (2011): "Bias Reduction For Dynamic Nonlinear Panel Models With Fixed Effects," *Econometric Theory*, 27(06), 1152–1191.

Hansen, C. B. (2007): "Asymptotic properties of a robust variance matrix estimator for panel data when T is large," *Journal of Econometrics*, 141(2), 597–620.

Hounkannounon, B. (2010): "Bootstrap for panel regression models with random effects," Discussion paper.

Kaffo, M. (2013): "Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects," *Unpublished manuscript*.

Kapetanios, G. (2008): "A Bootstrap Procedure for Panel Datasets with Many Cross-sectional Units," *Econometrics Journal*, 11(2), 377–395.

Kiviet, J. F. (1995): "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models," *Journal of Econometrics*, 68(1), 53–78.

Kreiss, J. (1997): "Asymptotical Properties of Residual Bootstrap for Autoregressions," Manuscript, institute for mathematical stochastics, technical university of braunschweig, germany.

Lee, Y. (2012): "Bias in dynamic panel models under time series misspecifcation," *Journal of Econometrics*.

Liu, R. (1988): "Bootstrap procedure under some non-i.i.d. models.," *Annals of Statistics*, 16, 1696–1708.

Liu, R., and K. Singh (1992): "Efficiency and robustness in resampling," *Annals of Statitics*, 20, 370–384.

Moon, H. R., and P. C. B. Phillips (2004): "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," *Econometrica*, 72(2), 467–522.

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

Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–26.

Pakel, C., N. Shephard, and K. Sheppard (2011): "Nuisance parameters, composite likelihoods and a panel of GARCH models," *Statistica Sinica*, 21(1), 307–329.

Phillips, P. C., and D. Sul (2007): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," *Journal of Econometrics*, 137(1), 162–188.

Wu, C. (1986): "Jackknife, bootstrap and other resampling methods in regression analysis.," *Annals of Statistics*, 14, 1261–1295.

A Appendix A: proofs of results in Section 2

Throughout this Appendix, we let Δ denote a generic constant independent of n and T. Given a matrix A, we let $|A| = \left(tr\left(A'A\right)\right)^{1/2}$. The following results are instrumental in the proofs that follow. They correspond to Lemmas 1 and 2 in Hansen (2007) respectively.

Theorem A.1 *Suppose Z_{iT} are independent across i for all T with $E\left(Z_{iT}\right) = \mu_{iT}$ and $E\left|Z_{iT}\right|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. Then $\frac{1}{n}\sum_{i=1}^{n}\left(Z_{iT} - \mu_{iT}\right) \to^P 0$ as $n, T \to \infty$ jointly.*

Theorem A.2 *For $k \times 1$ vectors Z_{iT}, suppose Z_{iT} are independent across i for all T with $E\left(Z_{iT}\right) = 0$, $E\left(Z_{iT}Z_{iT}^{'}\right) = \Omega_{iT}$, and $E\left|Z_{iT}\right|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$. Assume $\Omega = \lim_{n,T}\frac{1}{n}\sum_{i=1}^{n}\Omega_{iT}$ is positive definite with minimum eigenvalue $\lambda_{min} > 0$. Then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{iT} \to^d N\left(0, \Omega\right)$ as $n, T \to \infty$ jointly.*

We first provide some auxiliary lemmas, followed by the proof of Theorem 2.1. The proof of the auxiliary lemmas follows at the end.

Lemma A.1 *Under Assumption A1, for fixed l, $p \in \mathbb{N}$, **(i)** $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it-l}\varepsilon_{it-p} \to^P \sigma^2 1_{\{l=p\}}$; and **(ii)** $\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p} \to^P 0$.*

Lemma A.2 *Under Assumption A1, for fixed $k \in \mathbb{N}$, $\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\varepsilon_{it}\varepsilon_{it-1}, \ldots, \varepsilon_{it}\varepsilon_{it-k}\right) \to^d N\left(0, \Omega_k\right)$, where $\Omega_k \equiv \left[\tau_{lp}\right]_{l,p=1,\ldots,k}$.*

Lemma A.2 is the analog of Lemma A.1 of Gonçalves and Kilian (2004) (henceforth GK (2004)). To state the following lemma, we need to introduce some notation. In particular, let $u_{it} = \sum_{l=0}^{\infty}\theta_0^l\varepsilon_{it-l}$, which is well defined given that $|\theta_0| < 1$. It follows that

$$y_{it-1} = \frac{\alpha_i}{1-\theta_0} + \sum_{l=1}^{\infty}\theta_0^{l-1}\varepsilon_{it-l} \equiv \frac{\alpha_i}{1-\theta_0} + u_{it-1}, \tag{12}$$

for all (i, t). Therefore,

$$A_{nT} \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i-}\right)^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^2 - \frac{1}{n}\sum_{i=1}^{n}\bar{u}_{i-}^2 \equiv A_{nT1} - A_{nT2},$$

where $\bar{u}_{i-} = \frac{1}{T}\sum_{t=1}^{T}u_{it-1}$. The next lemma establishes the consistency of A_{nT}.

Lemma A.3 *Under Assumption A1, **(i)** $A_{nT1} \to^P A \equiv \sigma^2\left(1-\theta_0^2\right)^{-1}$; **(ii)** $A_{nT2} \to^P 0$; and **(iii)** $A_{nT} \to^P A$.*

Our next lemma establishes the limiting distribution of

$$B_{nT} \equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1} - \bar{y}_{i-}) (\varepsilon_{it} - \bar{\varepsilon}_i) = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}\varepsilon_{it} - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}\bar{\varepsilon}_i \equiv B_{nT1} - B_{nT2}.$$

Lemma A.4 *Under Assumption A1, **(i)** $B_{nT1} \to^d N(0, B)$, where $B = \sum_{l,p=1}^{\infty} \theta_0^{l+p-2} \tau_{lp}$; **(ii)** $B_{nT2} \to^P -A \cdot D$, where $A = \sigma^2 \left(1 - \theta_0^2\right)^{-1}$ and $D = -\sqrt{\rho}\,(1 + \theta_0)$; and **(iii)** $B_{nT} \to^d N(A \cdot D, B)$.*

Proof of Theorem 2.1. The proof follows from Lemmas A.3 and A.4 by Slutsky's theorem.

Proof of Lemma A.1 (i) For fixed $l, p \in \mathbb{N}$, let $Z_{iT}^{lp} = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it-l}\varepsilon_{it-p}$, $i = 1, \ldots, n$. We check that $\{Z_{iT}^{lp}\}$ satisfies the conditions of Theorem A.1. First, $\{Z_{iT}^{lp}\}$ are independent across i for all T with $E\left(Z_{iT}^{lp}\right) = \sigma_i^2 1_{\{l=p\}}$. Second, we show that $E\left|Z_{iT}^{lp}\right|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i and T. Taking $\delta = 1$, by repeated application of the Cauchy-Schwartz inequality, we can show that $E\left(Z_{iT}^{lp}\right)^2 \leq E(\varepsilon_{it})^4 \leq \Delta < \infty$. Thus, $\frac{1}{n} \sum_{i=1}^{n} \left(Z_{iT}^{lp} - \sigma_i^2 1_{\{l=p\}}\right) \to_P 0$ as $n, T \to \infty$ jointly. The result follows by noting that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 = \sigma^2$ by A1(iv). To prove part (ii), define for fixed $l, p \in \mathbb{N}$, $Z_{iT}^{lp} = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \varepsilon_{it-l}\varepsilon_{is-p}$. Then, $\{Z_{iT}^{lp}\}$ are independent across i for all T with $E\left(Z_{iT}^{lp}\right) = \mu_{iT}^{lp}$, where $\mu_{iT}^{lp} = 0$ for $l \leq p$ and for $l - p \geq T$, and $\mu_{iT}^{lp} = \frac{T-l-p}{T^2}\sigma_i^2$ for $l - p \in \{1, \ldots, T-1\}$. By repeated application of the Cauchy-Schwartz inequality, we can show that $E\left(Z_{iT}^{lp}\right)^2 \leq E(\varepsilon_{it})^4 \leq \Delta < \infty$, which proves that Z_{iT}^{lp} verifies the conditions of Theorem A.1. To end the proof of (ii), note that by definition of μ_{iT}^{lp},

$$\frac{1}{n} \sum_{i=1}^{n} \mu_{iT}^{lp} = \frac{T-l-p}{T^2} \left(\frac{1}{n} \sum_{i=1}^{n} \sigma_i^2\right) 1_{\{l-p\in\{1,\ldots,T-1\}\}} \to 0$$

as $n, T \to \infty$ jointly, for all $l, p \in \mathbb{N}$, given that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 = \sigma^2$.

Proof of Lemma A.2 For fixed $k \in \mathbb{N}$, let $Z_{iT}^{k} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_{it}\varepsilon_{it-1}, \ldots, \varepsilon_{it}\varepsilon_{it-k})'$, $i = 1, \ldots, n$. We check that Z_{iT}^{k} satisfies the conditions of Theorem A.2. First, Z_{iT}^{k} are independent across i for all T with $E\left(Z_{iT}^{k}\right) = 0$. Second, $E\left(Z_{iT}^{k} Z_{iT}^{k\prime}\right) = [\tau_{ilp}]_{l,p=1,\ldots,k} \equiv \Omega_{ik}$ for all i since by assumption $E\left(\varepsilon_{it}^2 \varepsilon_{it-l}\varepsilon_{it-p}\right) = \tau_{ilp}$ for all t and all l, p. Third, we show that for fixed $k \in \mathbb{N}$, $E\left|Z_{iT}^{k}\right|^{2\delta} \leq \Delta < \infty$, uniformly in i for some $\delta > 1$ (we take $\delta = 2$). By the $c - r$ inequality,

$$E\left|Z_{iT}^{k}\right|^4 = E\left(\sum_{l=1}^{k}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{it-l}\right)^2\right)^2 \leq k \sum_{l=1}^{k} E\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{it-l}\right)^4 = k \sum_{l=1}^{k} \frac{1}{T^2} \sum_{t_1,\ldots,t_4=1}^{T} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right),$$

where we let $z_{it}^l = \varepsilon_{it}\varepsilon_{it-l}$ for all $1 \le l \le k$. Noting that $E\left(z_{it}^l\right) = 0$ and given the definition of the fourth order joint cumulant (see Brillinger (1981), p. 19), we have that

$$\begin{aligned} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right) &= E\left(z_{t_1}^l z_{t_2}^l\right) E\left(z_{t_3}^l z_{t_4}^l\right) + E\left(z_{t_1}^l z_{t_3}^l\right) E\left(z_{t_2}^l z_{t_4}^l\right) + E\left(z_{t_1}^l z_{t_4}^l\right) E\left(z_{t_2}^l z_{t_3}^l\right) \\ &\quad + cum\left(z_{it_1}^l, z_{it_2}^l, z_{it_3}^l, z_{it_4}^l\right). \end{aligned}$$

By the m.d.s assumption, $E\left(z_t^l z_s^l\right) = E\left(\varepsilon_{it}\varepsilon_{it-l}\varepsilon_{is}\varepsilon_{is-l}\right) = \tau_{ill} 1_{\{t=s\}}$ for any (t, s), which implies that

$$\frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right) = 3\tau_{ill}^2 + \frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} cum\left(z_{it_1}^l, z_{it_2}^l, z_{it_3}^l, z_{it_4}^l\right).$$

Given the strict stationarity assumption, $cum\left(z_{it_1}^l, z_{it_2}^l, z_{it_3}^l, z_{it_4}^l\right) = cum\left(z_{it_1-t_4}^l, z_{it_2-t_4}^l, z_{it_3-t_4}^l, z_{i0}^l\right)$, which implies that

$$\frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right) = 3\tau_{ill}^2 + \frac{1}{T^2}\sum_{t_4=1}^{T}\left\{\sum_{t_1,t_2,t_3=1}^{T} cum\left(z_{it_1-t_4}^l, z_{it_2-t_4}^l, z_{it_3-t_4}^l, z_{i0}^l\right)\right\},$$

where the expression in curly brackets is $O(1)$ uniformly in i, l and t_4, given A1(vii) (applied with $l_1 = l_2 = l_3 = l_4 = l$). This shows that $\frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right)$ is uniformly bounded in i, l and T and hence, for a fixed $k \in \mathbb{N}$, $E\left|Z_{iT}^k\right|^4 \le \Delta < \infty$ uniformly in i and T. Also,

$$\lim_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ik} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\tau_{ilp}\right]_{l,p=1,\ldots,k} = \left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tau_{ilp}\right]_{l,p=1,\ldots,k} = \left[\tau_{lp}\right]_{l,p=1,\ldots,k} \equiv \Omega_k,$$

where Ω_k is positive definite with minimum eigenvalue $\lambda_{min} > 0$ since by assumption, $\tau_{ll} > 0$ for all l. Thus, the conditions of Theorem A.2 are verified, ending the proof.

Proof of Lemma A.3. The proof of part (i) follows from Lemma A.1(i) using the same steps as the proof that $A_{1n} \to^P 0$ in Theorem 3.1 in GK (p. 108). To prove (ii), which is new in our panel context, we use the definition of $\bar{u}_{i-}$ to decompose A_{nT2} as follows:

$$\begin{aligned} A_{nT2} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)^2 \\ &= \frac{1}{T}\left\{\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it-l}\varepsilon_{it-p}\right)\right\} + 2\left\{\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right\} \\ &\equiv a_{1,nT} + 2a_{2,nT}. \end{aligned}$$

Given part (i), we have $a_{1,nT} = (1/T) \times A_{nT1} = o_P(1)$. Next we show that $a_{2,nT} = o_P(1)$. For fixed $m \in \mathbb{N}$, define $a_{2,nT}^m = \sum_{l=1}^{m}\sum_{p=1}^{m}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)$. By Lemma A.1(ii), it follows that that $a_{2,nT}^m \to 0$ for all $m \in \mathbb{N}$. Thus, it suffices to show that $\lim_{m\to\infty}\limsup_{n,T\to\infty} P\left(\left|a_{2,nT} - a_{2,nT}^m\right| > \delta\right) = 0$, for all $\delta > 0$ (see Brockwell and Davis (1991)'s Proposition 6.3.9). By Markov's inequality,

$$P\left(\left|a_{2,nT} - a_{2,nT}^m\right| > \delta\right) \leq \frac{1}{\delta} E\left|a_{2,nT} - a_{2,nT}^m\right|$$
$$\leq \frac{1}{\delta} E\left|\sum_{l=m+1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right| + \frac{1}{\delta} E\left|\sum_{l=1}^{m}\sum_{p=m+1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right|$$
$$\leq \frac{2}{\delta}\sum_{l=m+1}^{\infty}\sum_{p=1}^{\infty}|\theta_0|^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left|\varepsilon_{it-l}\varepsilon_{is-p}\right|\right) \leq \left(\sum_{l=m+1}^{\infty}|\theta_0|^{l-1}\right)K \to 0 \text{ as } m \to \infty,$$

given the absolute summability of θ_0^{l-1} and the fact that $E\left|\varepsilon_{it-l}\varepsilon_{is-p}\right| \leq \Delta < \infty$ uniformly. This completes the proof of (ii). (iii) follows from (i) and (ii).

Proof of Lemma A.4 Part (i) follows from Lemma A.1 and the cross sectional independence assumption, using arguments similar to those used in the proof of Theorem 3.1 of GK (2004) (see part (ii) of their proof). To prove (ii) (which is specific to the fixed effects OLS estimator), note that we can show that the following decomposition holds:

$$\begin{aligned}
B_{nT2} &\equiv \frac{1}{T\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{l=1}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right)\\
&= \frac{1}{T\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{l=1}^{t-1}\theta_0^{l-1}\varepsilon_{it-l} + \sum_{l=t}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right)\\
&= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right) + \sqrt{\frac{n}{T}}\frac{1-\theta_0^T}{1-\theta_0}\left\{\sum_{l=1}^{\infty}\theta_0^{l-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{i1-l}\right)\right\}\\
&\equiv \mathcal{B}_{nT2.1} + \mathcal{B}_{nT2.2}.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathcal{B}_{nT2.1} &= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=1}^{T-l}\varepsilon_{is} + \sum_{s=T-l+1}^{T}\varepsilon_{is}\right)\\
&= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2 + \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T}\varepsilon_{is}\right)\\
&\equiv b_1 + b_2.
\end{aligned}$$

For fixed $m \in \mathbb{N}$, define

$$b_{1,m} = \sqrt{\frac{n}{T}}\sum_{l=1}^{m-1}\theta_0^{l-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2\right) = \sqrt{\frac{n}{T}}\sum_{l=1}^{m-1}\theta_0^{l-1}\left(\frac{1}{n}\sum_{i=1}^{n}Z_{iT,l}\right),$$

where $Z_{iT,l} \equiv T^{-1}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2$. For fixed l, we can show that $\frac{1}{n}\sum_{i=1}^{n}Z_{iT,l} \to^P \sigma^2$ by an application of Lemma A.1. In particular, we can use the same arguments as in Lemma A.2 to show that $E\left|Z_{iT,l}\right|^2$ is uniformly bounded by relying on Assumption A1 (vi). Thus, $b_{1,m} \to^P \sqrt{\rho}\sum_{l=1}^{m-1}\theta_0^{l-1}\sigma^2 =$

$\sqrt{\rho}\sigma^2 \dfrac{1-\theta_0^{m-1}}{1-\theta_0} \equiv D_m$ and $D_m \to \sqrt{\rho}\sigma^2 \dfrac{1}{1-\theta_0} \equiv -A \cdot D$ as $m \to \infty$, where $A \equiv \sigma^2 / \left(1-\theta_0^2\right)$ and $D \equiv -\sqrt{\rho}\,(1+\theta_0)$. In addition, by Markov's inequality, we have

$$P\left(|b_1 - b_{1,m}| > \delta\right) \leq \frac{1}{\delta}\sqrt{\frac{n}{T}} \sum_{l=m}^{T-1} |\theta_0|^{l-1} \left(\frac{1}{n}\sum_{i=1}^{n} E\left(Z_{iT,l}\right)\right) = \frac{1}{\delta}\sqrt{\frac{n}{T}} \sum_{l=m}^{T-1} |\theta_0|^{l-1} \left(\frac{T-l}{T}\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2\right).$$

It follows that $\lim_{m\to\infty} \lim \sup_{n,T\to\infty} P\left(|b_1 - b_{1,m}| > \delta\right) = 0$ since $n/T \to \rho$, $|\theta_0|^{l-1}$ is absolutely summable and $\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \to \sigma^2$. Let us turn to b_2. For fixed m, define

$$b_{2,m} = \sqrt{\frac{n}{T}} \sum_{l=1}^{m-1} \theta_0^{l-1} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T}\left(\sum_{t=1}^{T-l} \varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T} \varepsilon_{is}\right) \equiv \sqrt{\frac{n}{T}} \sum_{l=1}^{m-1} \theta_0^{l-1} \frac{1}{n}\sum_{i=1}^{n} Y_{iT,l},$$

where $Y_{iT,l}$ are independent across i, $E\left(Y_{iT,l}\right) = 0$ and $E\left|Y_{iT,l}\right|^2 \leq E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_{it}\right)^4 \leq \Delta$ by Assumption A1 (v) and (vi). Thus, by Theorem A.1, $\frac{1}{n}\sum_{i=1}^{n} Y_{iT,l} = o_P(1)$ and therefore, $b_{2,m} = o_P(1)$. Finally, by Markov's inequality, we have

$$\begin{aligned} P\left(|b_2 - b_{2,m}| > \delta\right) &\leq \frac{1}{\delta}\sqrt{\frac{n}{T}} E\left|\sum_{l=m}^{T-1} \theta_0^{l-1} \frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l} \varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T} \varepsilon_{is}\right)\right| \\ &\leq \frac{1}{\delta}\sqrt{\frac{n}{T}} \sum_{l=m}^{T-1} |\theta_0|^{l-1} \frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_{it}\right)^2 = \frac{1}{\delta}\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2\right)\sum_{l=m}^{T-1} |\theta_0|^{l-1}, \end{aligned}$$

which implies that $\lim_{m\to\infty} \lim\sup_{n,T\to\infty} P\left(|b_2 - b_{2,m}| > \delta\right) = 0$ for any $\delta > 0$. To complete the proof of Lemma A.4 (ii), we note that $E\left(\mathcal{B}_{nT2.2}\right) = 0$ and we can show that $Var\left(\mathcal{B}_{nT2.2}\right) = O\left(1/nT\right) = o\left(1\right)$. Part (iii) follows from (i) and (ii) by Slutsky's theorem.

B Appendix B: Proofs of results in Section 3

B.1 Proofs of results in Section 3.1

Throughout this section, $y_{it}^* = \hat{\alpha}_i + \hat{\theta} y_{it-1}^* + \varepsilon_{it}^*$, where $\varepsilon_{it}^* = \hat{\varepsilon}_{it} \cdot \eta_{it}$, with η_{it} are i.i.d.(0, 1) and $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta} y_{it-1}$.

B.1.1 Auxiliary lemmas

Lemma B.1 *Under Assumption A1, for fixed $k, l \in \mathbb{N}$,* ***(i)*** $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T} \varepsilon_{it-k}^{*2} \to^{P^*} \sigma^2$*;* ***(ii)*** $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T} \varepsilon_{it-k}^{*}\varepsilon_{it}^{*} \to^{P^*} 0$*; and* ***(iii)*** $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=max(k,l)+1}^{T} \varepsilon_{it}^{*2}\varepsilon_{it-k}^{*}\varepsilon_{it-l}^{*} \to^{P^*} \tau_{kl} 1_{\{k=l\}}$*, in probability, where* $\tau_{kl} = E\left(\varepsilon_{it}^2 \varepsilon_{it-k}\varepsilon_{it-l}\right)$.

Lemma B.2 *Under Assumption A1, for all* $k \in \mathbb{N}$, $\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \left(\varepsilon_{it}^* \varepsilon_{it-1}^*, \ldots, \varepsilon_{it}^* \varepsilon_{it-k}^*\right)' \to^{d^*} N\left(0, \tilde{\Omega}_k\right)$, *in probability, where* $\tilde{\Omega}_k \equiv diag\left(\tau_{11}, \ldots, \tau_{kk}\right)$.

For the next lemma, let $y_{i0}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}}$. It follows that for fixed $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$y_{it}^* = \hat{\theta}^t \frac{\hat{\alpha}_i}{1-\hat{\theta}} + \frac{1-\hat{\theta}^t}{1-\hat{\theta}} \hat{\alpha}_i + \sum_{s=0}^{t-1} \hat{\theta}^s \varepsilon_{it-s}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}} + \sum_{s=0}^{t-1} \hat{\theta}^s \varepsilon_{it-s}^* \equiv \frac{\hat{\alpha}_i}{1-\hat{\theta}} + u_{it}^*.$$

Therefore,

$$A_{nT}^* = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^* - y_{i-}^*\right)^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^{*2} - \frac{1}{n} \sum_{i=1}^{n} \bar{u}_{i-}^{*2} \equiv A_{nT1}^* - A_{nT2}^*,$$

where $\bar{u}_{i-}^* = \frac{1}{T} \sum_{t=1}^{T} u_{it-1}^*$ and $u_{it-1}^* = \sum_{s=0}^{t-1-1} \hat{\theta}^s \varepsilon_{it-1-s}^* = \sum_{s=1}^{t-1} \hat{\theta}^{s-1} \varepsilon_{it-s}^*$.

Lemma B.3 *Under Assumption A1,* ***(i)*** $A_{nT1}^* \to^{P^*} A \equiv \frac{\sigma^2}{1-\theta_0^2}$; ***(ii)*** $A_{nT2}^* \to^{P^*} 0$; *and* ***(iii)*** $A_{nT}^* \to^{P^*} A$, *in probability.*

Similarly, if we define $B_{nT}^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right)$, given the definition of y_{it-1}^*, we have

$$B_{nT}^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^* \varepsilon_{it}^* - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^* \bar{\varepsilon}_i^* \equiv B_{nT1}^* - B_{nT2}^*. \tag{13}$$

Lemma B.4 *Under Assumption A1,* ***(i)*** $B_{nT1}^* \to^{d^*} N\left(0, \tilde{B}\right)$; ***(ii)*** $B_{nT2}^* \to^{P^*} -A \cdot D$; *and* ***(iii)*** $B_{nT}^* \to^{d^*} N\left(A \cdot D, \tilde{B}\right)$, *in probability, where* $\tilde{B} = \sum_{l=1}^{\infty} \theta_0^{2l-2} \tau_{ll}$, *and* A *and* D *are defined as in Lemma A.4.*

B.1.2 Proofs

Proof of Theorem 3.1. The result follows from Lemmas B.3 and B.4, Theorem A.1 and Polya's Theorem, given that the normal distribution is everywhere continuous. Note that Assumption A1 needs to be strengthened by A1(v′) in order for $\tilde{B} = B$.

Proof of Theorem 3.2. We show that (1) $\sqrt{nT}(\tilde{\theta}^* - \hat{\theta}) \to^{d^*} N(D, C)$ in probability; and (2) for some $\delta > 0$, $E^*\left(\left|\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right|^{1+\delta}\right) = O_P(1)$. Starting with (1), we can write $\sqrt{nT}(\tilde{\theta}^* - \hat{\theta}) = \sqrt{nT}(\hat{\theta}_{rd}^* - \hat{\theta}) + R_{nT}^*$, with $R_{nT}^* = -\sqrt{nT}(\hat{\theta}_{rd}^* - \hat{\theta}) 1_{\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}^* - \bar{y}_{i-}^*)^2 < \frac{\eta}{2}\right\}}$, given the definition of $\tilde{\theta}^*$ (with $\delta = \frac{\eta}{2}$ and $\eta \in (0, \frac{\sigma^2}{1-\theta_0^2})$). By Theorem 3.1, $\sqrt{nT}(\hat{\theta}_{rd}^* - \hat{\theta}) = O_{P^*}(1)$, in probability, and

$$E^*\left(1_{\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it-1}^* - \bar{y}_{i-}^*)^2 < \frac{\eta}{2}\right\}}\right) = P^*\left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1}^* - \bar{y}_{i-}^*)^2 < \frac{\eta}{2}\right) \to^P 0,$$

given (3). By Markov's inequality, we conclude that $R_{nT}^* = o_{P^*}(1)$ in probability. To prove (2), we let $\delta = 1$ and define $S = \left\{ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1}^* - \bar{y}_{i-}^*)^2 \geq \frac{\eta}{2} \right\}$. Then, given the definition of $\tilde{\theta}^*$, we have

$$
\begin{aligned}
E^* \left(\left| \sqrt{nT} (\tilde{\theta}^* - \hat{\theta}) \right|^2 \right) &= E^* \left(\left| \sqrt{nT} (\hat{\theta}_{rd}^* - \hat{\theta}) 1_S \right|^2 \right) \\
&= E^* \left(\left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1}^* - \bar{y}_{i-}^*)^2 \right)^{-2} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1}^* - \bar{y}_{i-}^*)(\varepsilon_{it}^* - \bar{\varepsilon}_i^*) \right)^2 1_S \right) \\
&\leq \frac{4}{\eta^2} E^* \left(\left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1}^* - \bar{y}_{i-}^*)(\varepsilon_{it}^* - \bar{\varepsilon}_i^*) \right)^2 \right) \equiv \frac{4}{\eta^2} E^*(B_{nT}^{*2}),
\end{aligned}
$$

where B_{nT}^* can be decomposed as $B_{nT}^* = B_{1nT}^* - B_{2nT}^*$, with B_{1nT}^* and B_{2nT}^* given in equation (13). We now show that $E^*(B_{nT}^{*2}) = O_P(1)$. We have that $E^* \left(B_{nT}^{*2} \right) \leq 2 \left(E^* \left(B_{1nT}^{*2} \right) + E^* \left(B_{2nT}^{*2} \right) \right)$, where $E^* \left(B_{1nT}^{*2} \right) = Var^* \left(B_{1nT}^* \right) \to^P \tilde{B}$, so $E^* \left(B_{1nT}^{*2} \right) = O_P(1)$. For the second term, note that

$$
B_{2nT}^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^* \bar{\varepsilon}_i^* = \mathcal{B}_{nT2.1}^* + \mathcal{B}_{nT2.2}^*,
$$

where $\mathcal{B}_{nT2.1}^*$ and $\mathcal{B}_{nT2.2}^*$ are defined in the proof of Lemma B.4. As we argue in that proof, $E^* \left(\mathcal{B}_{nT2.2}^{*2} \right) \to^P 0$, so we are left to prove that $E^* \left(\mathcal{B}_{nT2.1}^{*2} \right) = O_P(1)$. Given the definition of $\mathcal{B}_{nT2.1}^*$,

$$
\begin{aligned}
E^* \left(\mathcal{B}_{nT2.1}^{*2} \right) &= \frac{1}{nT^3} \sum_{i,j=1}^{n} \sum_{l,p=1}^{T-1} \hat{\theta}^{l+p-2} E^* \left[\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 \left(\sum_{s=1}^{T-p} \varepsilon_{js}^* \right)^2 \right] \\
&= \frac{1}{nT^3} \sum_{i=1}^{n} \sum_{l,p=1}^{T-1} \hat{\theta}^{l+p-2} E^* \left[\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 \left(\sum_{s=1}^{T-p} \varepsilon_{is}^* \right)^2 \right] \\
&\quad + \frac{1}{nT^3} \sum_{i \neq j} \sum_{l,p=1}^{T-1} \hat{\theta}^{l+p-2} E^* \left[\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 \right] E^* \left[\left(\sum_{s=1}^{T-p} \varepsilon_{js}^* \right)^2 \right] \equiv b_1^* + b_2^*.
\end{aligned}
$$

Now,

$$
\begin{aligned}
b_1^* &= \frac{1}{nT^3} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{2l-2} E^* \left[\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^4 \right] + 2 \frac{1}{nT^3} \sum_{i=1}^{n} \sum_{l>p}^{T-1} \hat{\theta}^{l+p-2} E^* \left[\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 \left(\sum_{s=1}^{T-p} \varepsilon_{is}^* \right)^2 \right] \\
&= b_{11}^* + b_{12}^*
\end{aligned}
$$

For b_{11}^*, using the fact that $E^* |\eta_{it}|^4 \leq \Delta < \infty$,

$$
b_{11}^* \leq \frac{(1+\Delta)}{nT^3} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{2l-2} \left\{ \sum_{t=1}^{T-l} \hat{\varepsilon}_{it}^4 + 3 \sum_{t \neq s}^{T-l} \hat{\varepsilon}_{it}^2 \hat{\varepsilon}_{is}^2 \right\} \leq \frac{3(1+\Delta)}{T} \left\{ \frac{1}{nT^2} \sum_{i=1}^{n} \sum_{t,s=1}^{T} \hat{\varepsilon}_{it}^2 \hat{\varepsilon}_{is}^2 \right\} \left(\sum_{l=1}^{T} \hat{\theta}^{2l-2} \right) = O_P(\frac{1}{T}),
$$

given that the terms in brackets are $O_P(1)$. Similarly,

$$\begin{aligned}
b_{12}^* &= 2\frac{1}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-p}\varepsilon_{it}^* + \sum_{t=T-p+1}^{T-l}\varepsilon_{it}^*\right)^2\left(\sum_{s=1}^{T-p}\varepsilon_{is}^*\right)^2\right] \\
&\leq \frac{4}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-p}\varepsilon_{it}^*\right)^4 + \left(\sum_{t=T-p+1}^{T-l}\varepsilon_{it}^*\right)^2\left(\sum_{s=1}^{T-p}\varepsilon_{is}^*\right)^2\right] \\
&\leq \frac{4(1+\Delta)}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}\left\{3\sum_{t,s=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2 + \left(\sum_{t=T-p+1}^{T-l}\hat{\varepsilon}_{it}^2\right)\left(\sum_{s=1}^{T-p}\hat{\varepsilon}_{is}^2\right)\right\} \\
&\leq \frac{16(1+\Delta)}{T}\left\{\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2\right\}\left(\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}\right) = O_P(\frac{1}{T}) = O_P(1).
\end{aligned}$$

Finally, for b_2^* we have

$$b_2^* = \frac{1}{nT^3}\sum_{i\neq j}^{n}\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}\left(\sum_{t=1}^{T-l}\hat{\varepsilon}_{it}^2\right)\left(\sum_{s=1}^{T-p}\hat{\varepsilon}_{js}^2\right) \leq \frac{n}{T}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\right)^2\left(\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}\right) = O_P(1).$$

This complete the proof of Theorem 3.2.

Proof of Lemma 3.1. From Lemma B.3, $\hat{A}_{rd}^* \to^{P^*} A$. Hence, it suffices to show that $\hat{B}_{rd}^* \to^{P^*} \tilde{B}$, in probability. We can write $\tilde{\varepsilon}_{it}^* - \bar{\tilde{\varepsilon}}_i^* = \varepsilon_{it}^* - \bar{\varepsilon}_i^* - \left(\hat{\theta}_{rd}^* - \hat{\theta}\right)\left(y_{it-1}^* - \bar{y}_{i-}^*\right)$, where $\tilde{\varepsilon}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\theta}_{rd}^* y_{it-1}^*$ and $\varepsilon_{it}^* = y_{it}^* - \hat{\alpha}_i - \hat{\theta} y_{it-1}^*$. Thus,

$$\hat{B}_{rd}^* = \hat{B}_1^* + \hat{B}_2^* + \hat{B}_3^*, \text{ with}$$

$\hat{B}_1^* = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2(\varepsilon_{it}^* - \bar{\varepsilon}_i^*)^2$, $\hat{B}_2^* = -2\left(\hat{\theta}_{rd}^* - \hat{\theta}\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^3(\varepsilon_{it}^* - \bar{\varepsilon}_i^*)$ and $\hat{B}_3^* = \left(\hat{\theta}_{rd}^* - \hat{\theta}\right)^2\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^4$.

We show: (a) $\hat{B}_1^* \to^{P^*} B$, (b) $\hat{B}_1^* \to^{P^*} 0$ and (c) $\hat{B}_1^* \to^{P^*} 0$. Starting with (a), note that

$$\begin{aligned}
\hat{B}_1^* &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2(\varepsilon_{it}^* - \bar{\varepsilon}_i^*)^2 \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(u_{it-1}^* - \bar{u}_{i-}^*\right)^2(\varepsilon_{it}^* - \bar{\varepsilon}_i^*)^2 \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(u_{it-1}^{*2} - 2u_{it-1}^*\bar{u}_{i-}^* + \bar{u}_{i-}^{*2}\right)\left(\hat{\varepsilon}_{it}^{*2} - 2\varepsilon_{it}^*\bar{\varepsilon}_i^* + \bar{\varepsilon}_i^{*2}\right) \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*2}\varepsilon_{it}^{*2} + R_{nT}^*,
\end{aligned}$$

where

$$R_{nT}^{*} = -\frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*2}\varepsilon_{it}^{*}\bar{\varepsilon}_{i}^{*}+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*2}\bar{\varepsilon}_{i}^{*2}-\frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*}\bar{u}_{i-}^{*}\varepsilon_{it}^{*2}+\frac{4}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*}\bar{u}_{i-}^{*}\varepsilon_{it}^{*}\bar{\varepsilon}_{i}^{*}$$
$$+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\bar{u}_{i-}^{*2}\varepsilon_{it}^{*2}-\frac{3}{n}\sum_{i=1}^{n}\bar{u}_{i-}^{*2}\bar{\varepsilon}_{i}^{*2}\equiv -R_{nT1}^{*}+R_{nT2}^{*}-R_{nT3}^{*}+R_{nT4}^{*}+R_{nT5}^{*}-R_{nT6}^{*}.$$

By arguments similar to those of the proof of Corollary 3.1. of Gonçalves and Kilian (2004), one can show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*2}\varepsilon_{it}^{*2}\to^{P^*}\tilde{B}$. To show that $R_{nT}^{*}\to^{P^*}0$ in probability, it suffices that $E^{*}\left(\left|R_{nTj}^{*}\right|\right)\to^{P}0$ for $j=1,2,3,5,4,6$. For $j=1$,

$$|R_{nT1}^{*}|\leq 2\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}^{*4}\right]^{1/2}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^{*2}\bar{\varepsilon}_{i}^{*2}\right]^{1/2}\equiv A_{1}^{*}\times A_{2}^{*}.$$

Let us start with A_1^*. Since $u_{it-1}^{*}=\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}$,

$$E^{*}\left|A_{1}^{*}\right|^{2} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}\right)^{*4}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p,q,r=1}^{t-1}\hat{\theta}^{s+p+q+r-4}E^{*}\left(\varepsilon_{it-s}^{*}\varepsilon_{it-p}^{*}\varepsilon_{it-q}^{*}\varepsilon_{it-r}^{*}\right)$$
$$\leq \frac{\Delta}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p=1}^{t-1}\hat{\theta}^{2s+2p-4}\hat{\varepsilon}_{it-s}^{2}\hat{\varepsilon}_{it-p}^{2}\leq\frac{\Delta}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\hat{\varepsilon}_{it-s}^{2}\hat{\varepsilon}_{it-p}^{2},$$

where $\hat{\varepsilon}_{it}=0\ \forall t\leq 0$. Therefore,

$$E^{*}\left|A_{1}^{*}\right|^{2} \leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it-s}^{2}\hat{\varepsilon}_{it-p}^{2}\right)$$
$$\leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it-s}^{4}\right)^{1/2}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it-p}^{4}\right)^{1/2}$$
$$\leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{4}\right)=O_{P}\left(1\right),$$

given that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{4}=O_{P}\left(1\right)$ under Assumption A1 and the fact that $\hat{\theta}-\theta_{0}=o_{P}(1)$ with $|\theta_0|<1$. To conclude that $R_{nT1}^{*}\to^{P^*}0$, it suffices to show that $E^{*}\left(|A_{2}^{*}|\right)\to^{P}0$. This can be done as follows:

$$E^{*}\left|A_{2}^{*}\right|^{2} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}E^{*}\left(\varepsilon_{it}^{*2}\bar{\varepsilon}_{i}^{*2}\right)=\frac{1}{nT^{3}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{p,q=1}^{T}E^{*}\left(\varepsilon_{it}^{*2}\varepsilon_{ip}^{*}\varepsilon_{iq}^{*}\right)$$
$$= \frac{1}{nT^{3}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{p=1}^{T}\hat{\varepsilon}_{it}^{2}\hat{\varepsilon}_{ip}^{2}=\frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}\right)^{2}\right\}=O_{P}\left(\frac{1}{T}\right).$$

Similar arguments can be applied to R^*_{nT}, $j = 2, 3, 5, 4, 6$. For $\hat{B}^*_2$ and $\hat{B}^*_3$, one can easily show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3\left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)$ and $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$ are $O_{P^*}(1)$.

Proof of Lemma B.1. The proof follows closely that of Lemma A.2 in GK (2004) and therefore we skip the details, only mentioning the changes introduced in the panel context. As in GK (2004), for part (i), we can write

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon^{*2}_{it} - \sigma^2 = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^2_{it}\left(\eta^2_{it} - 1\right)\right] + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^2_{it} - \sigma^2\right] \equiv F^*_1 + F_2,$$

where now $\hat{\varepsilon}_{it} = \varepsilon_{it} + (\alpha_i - \hat{\alpha}_i) + \left(\theta_0 - \hat{\theta}\right)y_{it-1}$ depends also on $(\alpha_i - \hat{\alpha}_i)$, new to the fixed effects estimator. Thus, to show that $F_2 = o_P(1)$, we need to use the fact that $\sup_{1\le i\le n}|\hat{\alpha}_i - \alpha_i| = o_P(1)$ under our assumptions. Since $E\left[\left(\sum_{t=1}^{T}\varepsilon_{it}\right)^2\right] = \sum_{t=1}^{T}E\left(\varepsilon^2_{it}\right) = O(T)$, it follows that $\sum_{t=1}^{T}\varepsilon_{it} = O_P\left(\sqrt{T}\right)$ uniformly in i, and therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} = O_P(1)$ uniformly in i. Also, given that $\frac{1}{T}\sum_{t=1}^{T}y_{it-1} = O_P(1)$ uniformly in i and $\hat{\theta} - \theta_0 = o_P(1)$, we have

$$\begin{aligned}
\sup_{1\le i\le n}|\hat{\alpha}_i - \alpha_i| &= \sup_{1\le i\le n}\left|\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right) - \left(\hat{\theta} - \theta_0\right)\frac{1}{T}\sum_{t=1}^{T}y_{it-1}\right| \\
&\le \frac{1}{\sqrt{T}}\sup_{1\le i\le n}\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\right| + \left|\hat{\theta} - \theta_0\right|\sup_{1\le i\le n}\left|\frac{1}{T}\sum_{t=1}^{T}y_{it-1}\right| \\
&= \frac{1}{\sqrt{T}}O_P(1) + o_P(1)\,O_P(1) = o_P(1).
\end{aligned}$$

The proof that $E^*\left(F^{*2}_1\right) = o_P(1)$ follows exactly the same steps as the proof in GK (2004), with the only difference that we again rely on the uniform convergence (over i) of $\hat{\alpha}_i$ towards α_i (in addition to the convergence of $\hat{\theta}$ towards θ_0) to show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it} = O_P(1)$. The proof of (ii) and (iii) follow similarly. In particular, to prove (iii) we show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=max(k,l)+1}^{T}\varepsilon^2_{it}\varepsilon_{it-k}\varepsilon_{it-l} \to^P \tau_{kl}$ by verifing the conditions of Theorem A.1.

Proof of Lemma B.2. For fixed $k \in \mathbb{N}$, we check that $Z^{*k}_{iT} = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\left(\varepsilon^*_{it}\varepsilon^*_{it-1}, \ldots, \varepsilon^*_{it}\varepsilon^*_{it-k}\right)'$ satisfies the conditions of Theorem A.2, conditionally on the original sample with probability converging to one. First, $\{Z^{*k}_{iT}\}$ are (conditionally) independent across i for all T with $E^*\left(Z^{*k}_{iT}\right) = 0$. Second,

$$E^*\left(Z^{*k}_{iT}Z^{*k\prime}_{iT}\right) = diag\left(\frac{1}{T}\sum_{t=k+1}^{T}\hat{\varepsilon}^2_{it}\hat{\varepsilon}^2_{it-1}, \ldots, \frac{1}{T}\sum_{t=k+1}^{T}\hat{\varepsilon}^2_{it}\hat{\varepsilon}^2_{it-k}\right) \equiv \hat{\Omega}_{iT}.$$

Under our assumptions, $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^{2}\hat{\varepsilon}_{it-p}^{2}\to^{P}\tau_{pp}$, $p=1,\ldots,k$, which implies that $plim_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_{iT}=\tilde{\Omega}_k$, where $\tilde{\Omega}_k$ is positive definite with minimum eigenvalue $\lambda_{min}>0$ since $\tau_{rr}>0$ for all $r\geq 1$. Lastly, we can show that $E^*\left\|Z_{iT}^{*k}\right\|^{2\delta}=O_P(1)$, uniformly in i for $\delta=2$. In particular, by the $c-r$ inequality (with $r=2$),

$$
\begin{aligned}
E^*\left\|Z_{iT}^{*k}\right\|^{4} &= E^*\left(\sum_{l=1}^{k}\left(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\varepsilon_{it}^{*}\varepsilon_{it-l}^{*}\right)^{2}\right)^{2}\leq k^{2-1}\sum_{l=1}^{k}E^*\left(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\varepsilon_{it}^{*}\varepsilon_{it-l}^{*}\right)^{4}\\
&= k\sum_{l=1}^{k}\frac{1}{T^2}\sum_{t_1,\ldots,t_4=k+1}^{T}E^*\left(\varepsilon_{it_1}^{*}\varepsilon_{it_1-l}^{*}\varepsilon_{it_2}^{*}\varepsilon_{it_2-l}^{*}\varepsilon_{it_3}^{*}\varepsilon_{it_3-l}^{*}\varepsilon_{it_4}^{*}\varepsilon_{it_4-l}^{*}\right)\\
&\leq \Delta k\sum_{l=1}^{k}\frac{1}{T^2}\left(\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^{4}\hat{\varepsilon}_{it-l}^{4}+3\sum_{t\neq s}^{T}\hat{\varepsilon}_{it}^{2}\hat{\varepsilon}_{it-l}^{2}\hat{\varepsilon}_{is}^{2}\hat{\varepsilon}_{is-l}^{2}\right)=O_P(1),
\end{aligned}
$$

given that $1/T\sum_{t=1+k}^{T}\hat{\varepsilon}_{it}^{4}\hat{\varepsilon}_{it-l}^{4}=O_P(1)$ under Assumption A1. Note also the use of the definition of $\varepsilon_{it}^{*}=\hat{\varepsilon}_{it}\eta_{it}$ and the i.i.d. properties of η_{it} to justify the fact that the only non-zero contributions to the sum in the second equality are when (1) $t_1=t_2=t_3=t_4$; (2) $t_1=t_2\neq t_3=t_4$; (3) $t_1=t_3\neq t_2=t_4$; (4) $t_1=t_4\neq t_2=t_3$.

Proof of Lemma B.3 The proof of (i) follows the same arguments of the proof of Lemma A.4 of GK (2004), by replacing their Lemma A.2 with our Lemma B.1 to justify the convergence in probability of $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T}\varepsilon_{it-k}^{*2}$ towards σ^2 and of $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T-l}\varepsilon_{it-k}^{*}\varepsilon_{it}^{*}$ towards zero. Part (iii) follows from (i) and (ii). Part (ii) is new to the panel context considered here, so we provide more details. First, recall that $u_{it-1}^{*}=\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}$, which implies that

$$
\bar{u}_{i-}^{*}\equiv\frac{1}{T}\sum_{t=1}^{T}u_{it-1}^{*}=\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}\right)=\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\underbrace{\left(\frac{1}{T}\sum_{t=1}^{T-l}\varepsilon_{it}^{*}\right)}_{\equiv\chi_{il}^{*}}=\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\chi_{il}^{*}.
$$

Hence,

$$
\begin{aligned}
A_{nT2}^{*} &= \frac{1}{n}\sum_{i=1}^{n}\bar{u}_{i-}^{*2}=\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\chi_{il}^{*}\right)^{2}=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\chi_{il}^{*2}+\frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l-1)}\hat{\theta}^{(k-1)}\chi_{il}^{*}\chi_{il+k}^{*}\\
&\equiv \mathcal{A}_1^{*}+\mathcal{A}_2^{*}.
\end{aligned}
$$

Given the definition of χ_{il}^{*}, we have that

$$
\mathcal{A}_1^{*}=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T^2}\sum_{t=1}^{T-l}\varepsilon_{it}^{*2}+2\frac{1}{T^2}\sum_{k=1}^{T-l-1}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)\equiv a_{11}^{*}+a_{12}^{*}.
$$

Using Lemma B.1.(i), and following the proof of Lemma A.4 of GK(2004), we can show that

$$a_{11}^{*}=\frac{1}{T}\left\{\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l}\varepsilon_{it}^{*2}\right)\right\}=O_{P^{*}}\left(\frac{1}{T}\right)=o_{P^{*}}\left(1\right).$$

For the second term, we have that

$$a_{12}^{*}=\frac{2}{T}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-l-1}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)=\frac{2}{T}\sum_{l=1}^{T-1}\sum_{k=1}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right).$$

For fixed m, let

$$a_{12,m}^{*}=\frac{2}{T}\sum_{l=1}^{m-1}\sum_{k=1}^{m-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right).$$

By Lemma B.1.(ii), we have that $\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\rightarrow^{P^{*}}0$, in probability. Since $\hat{\theta}\rightarrow^{P}\theta_{0}$, it follows that $a_{12,m}^{*}\rightarrow^{P^{*}}0$, in probability. To conclude that $a_{12}^{*}\rightarrow^{P^{*}}0$, in probability, it suffices to show that $\lim_{m\rightarrow\infty}\lim\sup_{n,T\rightarrow\infty}P^{*}\left(\left|a_{12}^{*}-a_{12,m}^{*}\right|>\delta\right)=o_{P}\left(1\right)$. We have that

$$\begin{aligned}a_{12}^{*}-a_{12,m}^{*}&=\frac{2}{T}\sum_{l=m}^{T-1}\sum_{k=1}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)+\frac{2}{T}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)\\&=R_{12.1,m}^{*}+R_{12.2,m}^{*}.\end{aligned}$$

By the triangle inequality,

$$E^{*}\left|R_{12.1,m}^{*}\right|\leq\frac{2}{T}\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\sum_{k=1}^{T-l-1}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}E^{*}\left|\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right|\right)$$

$$\leq\frac{2}{T}\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\sum_{k=1}^{T-l-1}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\left|\hat{\varepsilon}_{it}\hat{\varepsilon}_{it+k}\right|E^{*}\left|\eta_{it}\eta_{it+k}\right|\right)\leq2\Delta\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}\right)\left(\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\right),$$

where we have used the fact that $E^{*}\left|\eta_{it}\eta_{it+k}\right|\leq\Delta$ and Cauchy-Schwartz's inequality to justify the third inequality. Under Assumption A1, we have that $\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}=O_{P}\left(1\right)$ whereas $\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\rightarrow^{P}$ $\theta_{0}^{2(m-1)}/\left(1-\theta_{0}^{2}\right)$, which converges to 0 as $m\rightarrow\infty$ since $\left|\theta_{0}\right|<1$. This shows that $\lim_{m\rightarrow\infty}\lim\sup_{n,T\rightarrow\infty}E^{*}\left|R_{12.1,m}^{*}\right|=$

$o_P(1)$. For $R^*_{12.2,m}$,

$$\begin{aligned}
E^*\left|R^*_{12.2,m}\right|^2 &\leq \frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\sum_{p=1}^{m-1}\sum_{q=m-p}^{T-p-1}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\sum_{j=1}^{n}\sum_{s=1}^{T-p-q}E^*\left(\varepsilon^*_{it}\varepsilon^*_{it+k}\varepsilon^*_{js}\varepsilon^*_{js+q}\right)\right)\\
&= \frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\sum_{p=1}^{m-1}\sum_{q=m-p}^{T-p-1}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-q)}E^*\left(\varepsilon^{*2}_{it}\varepsilon^*_{it+k}\varepsilon^*_{it+q}\right)\right)\\
&= \frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\sum_{k=\max(m-l,m-p)}^{\min(T-l-1,T-p-1)}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-q)}E^*\left(\varepsilon^{*2}_{it}\varepsilon^{*2}_{it+k}\right)\right)\\
&\leq 4\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it}\right)\frac{1}{nT^2}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\hat{\theta}^{2(l+p-2)},
\end{aligned}$$

which converges to 0 as $n,T\to\infty$ since under Assumption 1, we have that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it}=O_P(1)$ and

$$p\lim_{n,T\to\infty}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\hat{\theta}^{2(l+p-2)}=\left(\frac{1-\theta_0^{2(m-1)}}{1-\theta_0^2}\right)^2\to\frac{1}{1-\theta_0^2}\text{ as }m\to\infty,$$

showing that $\lim_{m\to\infty}\lim\sup_{n,T\to\infty}E^*\left|R^*_{12.2,m}\right|^2=o_P(1)$. This ends the proof of $\mathcal{A}^*_1=o_{P^*}(1)$. For $\mathcal{A}^*_2$, we have that

$$\begin{aligned}
\mathcal{A}^*_2 &= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l}\varepsilon^*_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon^*_{is}\right)\\
&= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l-k}\varepsilon^*_{it}+\frac{1}{T}\sum_{t=T-l-k+1}^{T-l}\varepsilon^*_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon^*_{is}\right)\\
&= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l-k}\varepsilon^*_{it}\right)^2+\frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=T-l-k+1}^{T-l}\varepsilon^*_{it}\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon^*_{is}\right)\\
&\equiv a^*_{21}+a^*_{22}.
\end{aligned}$$

Now,

$$\begin{aligned}
a^*_{21} &= \frac{2}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon^{*2}_{it}\right)+\frac{4}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\sum_{p=1}^{T-l-k-1}\hat{\theta}^{(l+k-2)}\left(\frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k-p}\varepsilon^*_{it}\varepsilon^*_{it+p}\right)\\
&\equiv a^*_{21.1}+a^*_{21.2}.
\end{aligned}$$

By Lemma B.1 (i), and following the proof of Lemma A.4 of GK(2004), we can show that

$$a^*_{21.1}=\frac{2}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon^{*2}_{it}\right)=O_{P^*}\left(\frac{1}{T}\right)=o_{P^*}(1).$$

The proof that $a^*_{21.2}=o_{P^*}(1)$ follows by showing that $E^*\left|a^*_{21.2}\right|^2=o_P(1)$. For a^*_{22}, we use Markov's inequality and apply the same reasoning as that used to show that $b_2=o_P(1)$ in the proof of Lemma

A.4.

Proof of Lemma B.4. Part (i) follows by the same arguments used by GK (2004) to prove their Lemma A.5, given our Assumption A1 and the fact that $\sup_i |\hat{\alpha}_i - \alpha_i| = o_P(1)$ and $\hat{\theta} \to^P \theta_0$. Part (iii) follows trivially from parts (i) and (ii). Part (ii) is the new bias term, which we consider in more detail here. First, recall that $\sum_{t=1}^{T} u_{it-1}^* = \left(\sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon_{it}^* \right)$, which implies that

$$B_{nT2}^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^* \bar{\varepsilon}_i^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \left(\sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon_{it}^* \right) \left(T^{-1} \sum_{s=1}^{T} \varepsilon_{is}^* \right),$$

given the definition of $\bar{\varepsilon}_i^*$. It follows that

$$\begin{aligned} B_{nT2}^* &= \frac{1}{T\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right) \left(\sum_{s=1}^{T-l} \varepsilon_{is}^* + \sum_{s=T-l+1}^{T-l} \varepsilon_{is}^* \right) \\ &= \frac{1}{T\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 + \frac{1}{T\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right) \left(\sum_{s=T-l+1}^{T-l} \varepsilon_{is}^* \right) \equiv \mathcal{B}_{nT2.1}^* + \mathcal{B}_{nT2.2}^*. \end{aligned}$$

For fixed l, we can write

$$\left(\sum_{t=1}^{T-l} \varepsilon_{it}^* \right)^2 = \sum_{t=1}^{T-l} \varepsilon_{it}^{*2} + 2 \sum_{k=1}^{T-l-1} \sum_{t=1}^{T-l-k} \varepsilon_{it}^* \varepsilon_{it+k}^*,$$

which implies that

$$\mathcal{B}_{nT2.1}^* = \frac{1}{T\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon_{it}^{*2} + \frac{2}{T\sqrt{nT}} \sum_{i=1}^{n} \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{k=1}^{T-l-1} \sum_{t=1}^{T-l-k} \varepsilon_{it}^* \varepsilon_{it+k}^* \equiv b_1^* + b_2^*.$$

Using arguments similar to those applied in the proof of Lemma B.3, we can show that $b_1^* \to^{P^*} \sqrt{\rho} \frac{\sigma^2}{1-\theta_0}$. For b_2^*, we have that

$$\begin{aligned} E^* |b_2^*|^2 &= 4 \left(\frac{n}{T} \right) \sum_{l=1}^{T-1} \sum_{k=1}^{T-l-1} \sum_{p=1}^{T-1} \sum_{q=1}^{T-p-1} \hat{\theta}^{l+p-2} \frac{1}{n^2 T^2} \sum_{i,j=1}^{n} \sum_{t=1}^{T-l-k} \sum_{s=1}^{T-p-q} E^* \left(\varepsilon_{it}^* \varepsilon_{it+k}^* \varepsilon_{js}^* \varepsilon_{js+q}^* \right) \\ &= 4 \left(\frac{n}{T} \right) \sum_{l=1}^{T-1} \sum_{k=1}^{T-l-1} \sum_{p=1}^{T-1} \sum_{q=1}^{T-p-1} \hat{\theta}^{l+p-2} \frac{1}{n^2 T^2} \sum_{i=1}^{n} \sum_{t=1}^{\min(T-l-k, T-p-q)} E^* \left(\varepsilon_{it}^{*2} \varepsilon_{it+k}^* \varepsilon_{it+q}^* \right) \\ &= 4 \left(\frac{n}{T} \right) \sum_{l=1}^{T-1} \sum_{k=1}^{\min(T-l-1, T-p-1)} \sum_{p=1}^{T-1} \hat{\theta}^{l+p-2} \frac{1}{n^2 T^2} \sum_{i=1}^{n} \sum_{t=1}^{\min(T-l-k, T-p-k)} \hat{\varepsilon}_{it}^2 \hat{\varepsilon}_{it+k}^2 \\ &\leq 4 \frac{1}{nT} \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^4 \right) \left(\frac{n}{T} \right) T \sum_{l=1}^{T-1} \sum_{p=1}^{T-1} \hat{\theta}^{l+p-2} = O_P \left(\frac{1}{n} \right). \end{aligned}$$

Using similar arguments, we can show that $E^* |\mathcal{B}_{nT2.2}^*|^2 \to^P 0$, which completes the proof of Lemma B.4.

B.2 Proof of Theorem 3.3

We show that

$$B_{nT}^{*} \equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-}\right) \left(\varepsilon_{it}^{*} - \bar{\varepsilon}_{i}^{*}\right) \to^{d^*} N\left(0, B\right)$$

in probability, where $\varepsilon_{it}^{*} = \hat{\varepsilon}_{it} \cdot \eta_{it}$, with η_{it} are i.i.d.$(0,1)$. We can write

$$B_{nT}^{*} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}\varepsilon_{it}^{*} - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}\bar{\varepsilon}_{i}^{*} \equiv B_{nT1}^{*} - B_{nT2}^{*}.$$

Writing $B_{nT1}^{*} = n^{-1/2} \sum_{i=1}^{n} Z_{iT}^{*}$, with $Z_{iT}^{*} \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\varepsilon_{it}^{*}$, we verify that the conditions of Theorem A.2 hold with probability converging to one. First, $\{Z_{iT}^{*}\}$ are independent across i for all T with $E^{*}\left(Z_{iT}^{*}\right) = 0$ and $E^{*}\left(Z_{iT}^{*2}\right) = \frac{1}{T} \sum_{t=1}^{T} u_{it-1}^{2}\hat{\varepsilon}_{it}^{2} \equiv \Omega_{iT}$. Moreover, for $\delta = 2$, and using the independence of ε_{it}^{*} across i and t, we have that

$$\begin{aligned} E^{*}\left(Z_{iT}^{*2+\delta}\right) &= E^{*}\left(\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\varepsilon_{it}^{*}\right)^{4}\right) = \frac{1}{T^{2}} \sum_{t,s,p,q=1}^{T} u_{it-1}u_{is-1}u_{ip-1}u_{iq-1}E^{*}\left(\varepsilon_{it}^{*}\varepsilon_{is}^{*}\varepsilon_{ip}^{*}\varepsilon_{iq}^{*}\right) \\ &\leq \frac{3}{T^{2}} \sum_{t,s=1}^{T} u_{it-1}^{2}u_{is-1}^{2}E^{*}\left(\varepsilon_{it}^{*2}\varepsilon_{is}^{*2}\right) \leq \frac{3\Delta}{T^{2}} \sum_{t=1}^{T} u_{it-1}^{4}\hat{\varepsilon}_{it}^{4} + \frac{6\Delta}{T^{2}} \sum_{t>s=1}^{T} u_{it-1}^{2}u_{is-1}^{2}\hat{\varepsilon}_{it}^{2}\hat{\varepsilon}_{is}^{2} = O_{P}\left(1\right) \end{aligned}$$

given that $E^{*}\left|\eta_{it}\right|^{4} \leq \Delta < \infty$. Finally, we can show that $\frac{1}{n} \sum_{i=1}^{n} \Omega_{iT} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^{2}\hat{\varepsilon}_{it}^{2} \to^{P} B$. To complete the proof, we show that $B_{nT2}^{*} = \sqrt{\frac{n}{T}} \frac{1}{n} \sum_{i=1}^{n} Z_{iT}^{*} \to^{P^*} 0$ by verifying that the conditions of Theorem A.1 apply to

$$Z_{iT}^{*} \equiv \frac{1}{T} \sum_{t=1}^{T} u_{it-1}\bar{\varepsilon}_{i}^{*} = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{is}^{*}\right).$$

Given that ε_{is}^{*} are independent across i, so are Z_{iT}^{*}. Moreover, $E^{*}\left(Z_{iT}^{*}\right) = 0$ and for $\delta = 1$,

$$E^{*}\left(Z_{iT}^{*1+\delta}\right) = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\right)^{2} E^{*}\left(\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \varepsilon_{is}^{*}\right)^{2}\right) = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\right)^{2} \left(\frac{1}{T} \sum_{s=1}^{T} \hat{\varepsilon}_{is}^{2}\right) = O_{P}\left(1\right).$$

B.3 Proof of Theorem 3.4

Let $I_1, \ldots, I_n$ be i.i.d. random variables uniformly distributed on $\{1, \ldots, n\}$, and let

$$\left(y_{it}^{*}, y_{it-1}^{*}\right) = \left(y_{I_i t}, y_{I_i t-1}\right),\ t = 1, \ldots, T,\ i = 1, \ldots, n.$$

Define $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta} y_{it-1}$, $\hat{\varepsilon}_{it}^{*} = y_{it}^{*} - \hat{\alpha}_{i}^{*} - \hat{\theta} y_{it-1}^{*}$ and $\varepsilon_{it}^{*} = y_{it}^{*} - \alpha_{i}^{*} - \theta_0 y_{it-1}^{*}$, where $\alpha_{i}^{*} = \alpha_{I_i}$. We show that (a) $A_{nT}^{*} \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^{*} - \bar{y}_{i-}^{*}\right)^{2} \to^{P^*} A$ and (b) $B_{nT}^{*} \equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^{*} - \bar{y}_{i-}^{*}\right) \left(\hat{\varepsilon}_{it}^{*} - \bar{\hat{\varepsilon}}_{i}^{*}\right) \to^{d^*}$

$N(0,B)$, in probability. Recall that $y_{it-1} = \frac{\alpha_i}{1-\theta_0} + u_{it-1}$. Similarly, define $\mu_i \equiv E(y_{it-1}) = \frac{\alpha_i}{1-\theta_0}$ and $\mu_i^* = \mu_{I_i}$. Then, for (a),

$$\begin{aligned} A_{nT}^* &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^* - \bar{y}_{i-}^* + \mu_i^*\right)^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\left(y_{it-1}^* - \mu_i^*\right) - \left(\frac{1}{T}\sum_{s=1}^{T}\left(y_{is-1}^* - \mu_i^*\right)\right)\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^*\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{s=1}^{T}\left(y_{is-1}^* - \mu_i^*\right)\right)^2 \equiv A_{nT1}^* - A_{nT2}^*. \end{aligned}$$

We show that (a1) $A_{nT1}^* \to^{P^*} A$ and (a2) $A_{nT2}^* \to^{P^*} 0$. For (a1), we let $Z_{iT}^* = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^*\right)^2$, which implies that $A_{nT1}^* = \frac{1}{n}\sum_{i=1}^{n} Z_{iT}^*$, and we use Theorem A.1. Notice that $\{Z_{iT}^*\}$ are independent across i for all T with

$$E^*\left(Z_{iT}^*\right) = \frac{1}{T}\sum_{t=1}^{T} E^*\left(\left(y_{I_it-1} - \mu_{I_i}\right)^2\right) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n}\sum_{j=1}^{n}\left(y_{jt-1} - \mu_j\right)^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it-1}^2 \to^P A.$$

Also, for $\delta = 1$,

$$E^*\left(Z_{iT}^{*1+\delta}\right) = \frac{1}{T^2}\sum_{t,s=1}^{T} E^*\left(\left(y_{it-1}^* - \mu_i^*\right)^2\left(y_{is-1}^* - \mu_i^*\right)^2\right) = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s=1}^{T} u_{it-1}^2 u_{is-1}^2 = O_P(1).$$

For (a2), define $\tilde{Z}_{iT}^* = \left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^*\right)\right)^2$ and let $A_{nT2}^* = \frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{iT}^*$, where the $\{\tilde{Z}_{iT}^*\}$ are independent across i for all T with

$$E^*\left(\tilde{Z}_{iT}^*\right) = \frac{1}{T^2}\sum_{t,s=1}^{T} E^*\left(\left(y_{it-1}^* - \mu_i^*\right)\left(y_{is-1}^* - \mu_i^*\right)\right) = \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s=1}^{T} u_{it-1}u_{is-1} = \frac{1}{n}\sum_{i=1}^{n}\bar{u}_{i-}^2 \to^P 0,$$

by Lemma A.3. The result follows by showing that $E^*\left(\tilde{Z}_{iT}^{*1+\delta}\right) = \frac{1}{nT^4}\sum_{i=1}^{n}\sum_{t,s,p,q=1}^{T} u_{it-1}u_{is-1}u_{ip-1}u_{iq-1} = O_P(1)$ for $\delta = 1$. Next we show (b). With our notations, B_{nT}^* can be rewritten as

$$\begin{aligned} B_{nT}^* &= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right) - \frac{1}{nT}\sum_{j=1}^{n}\sum_{s=1}^{T}\left(y_{js-1} - \bar{y}_{j-}\right)\left(\varepsilon_{js} - \bar{\varepsilon}_j\right)\right\} \\ &+ \left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i-}\right)^2 - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2\right\}\sqrt{nT}\left(\hat{\theta} - \theta_0\right) \equiv B_{nT}^{*\prime} + R_{nT}^*. \end{aligned}$$

Using (a) and Theorem 2.1, we have $R^*_{nT} = o_{P^*}(1) O_P(1) = o_{P^*}(1)$. Therefore, (b) follows if we prove that $B^{*\prime}_{nT} \to_{d_{P^*}} N(0, B)$ in probability. Noting that $\sum_{t=1}^{T} \left(y^*_{it-1} - \bar{y}^*_{i-}\right)\left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right) = \sum_{t=1}^{T} \left(y^*_{it-1} - \mu^*_i\right)\left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)$,

$$
\begin{aligned}
B^{*\prime}_{nT} &= \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ \left(y^*_{it-1} - \mu^*_i\right)\left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right) - \frac{1}{nT} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\left(\varepsilon_{js} - \bar{\varepsilon}_j\right) \right\} \\
&= \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ \left(y^*_{it-1} - \mu^*_i\right)\varepsilon^*_{it} - \frac{1}{nT} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\varepsilon_{js} \right\} \\
&\quad - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left\{ \left(y^*_{it-1} - \mu^*_i\right)\bar{\varepsilon}^*_i - \frac{1}{nT} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\bar{\varepsilon}_j \right\} \equiv B^*_{nT1} - B^*_{nT2},
\end{aligned}
$$

Therefore, it suffices to show that (b1) $B^*_{nT1} \to^{d^*} N(0, B)$ and (b2) $B^*_{nT2} \to^{P^*} 0$ in probability. For (b1), we verify the conditions of Theorem A.2 with $B^*_{nT1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z^*_{iT}$ and

$$
Z^*_{iT} \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} z^*_{it} \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(y^*_{it-1} - \mu^*_i\right)\varepsilon^*_{it} - \frac{1}{n\sqrt{T}} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\varepsilon_{js} \equiv q^*_{iT} - E^*\left(q^*_{iT}\right).
$$

Notice that $\{Z^*_{iT}\}$ are independent across i for all T with $E^*\left(Z^*_{iT}\right) = 0$ and $\Omega^*_{iT} \equiv E^*\left(Z^{*2}_{iT}\right) = E^*\left(q^*_{iT}\right)^2 - \left(E^*\left(q^*_{iT}\right)\right)^2$, where

$$
\Omega^* \equiv \frac{1}{n} \sum_{i=1}^{n} \Omega^*_{iT} = \frac{1}{n} \sum_{i=1}^{n} E^*\left(q^*_{iT}\right)^2 - \frac{1}{n} \sum_{i=1}^{n} \left(E^*\left(q^*_{iT}\right)\right)^2 \equiv \Omega^*_1 + \Omega^*_2.
$$

By Lemma A.4 (i),

$$
\Omega^*_2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{n\sqrt{T}} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\varepsilon_{js} \right)^2 = \frac{1}{n} \left(\frac{1}{\sqrt{nT}} \sum_{j=1}^{n} \sum_{s=1}^{T} u_{js-1}\varepsilon_{js} \right)^2 = O_P\left(\frac{1}{n}\right).
$$

Moreover,

$$
\Omega^*_1 = \frac{1}{n} \sum_{i=1}^{n} E^*\left(q^*_{iT}\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_{it-1}\varepsilon_{it} \right)^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u^2_{it-1}\varepsilon^2_{it} + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t \neq s} u_{it-1}\varepsilon_{it}u_{is-1}\varepsilon_{is},
$$

where the first term converges to B in probability and the second term is an $o_P(1)$ given Assumption A1(vii) in particular. Thus, $\Omega^* \to^{P} B$. The result follows by showing that $E^*\left(q^{*2+\delta}_{iT}\right) = O(1)$ for $\delta = 2$. To prove (b2), we proceed similarly but verify the conditions of Theorem A.1 instead. We omit the details to conserve space.

C Appendix C: Proofs of results in Section 4

Proof of Theorem 4.1. The proof follows from Theorem 3.1 and the fact that $\hat{\theta}^*_{rd} \to^{P^*} \theta_0$ in probability.

Proof of Theorem 4.2. The proof follows from Theorem 3.4 and the fact that $\hat{\theta}^*_{pb} \to^{P^*} \theta_0$ in probability.

Proof of Lemma 4.1. From the proof of Theorem 3.4, $\hat{A}^*_{pb} \to^{P^*} A$. Hence, it suffices to show that $\hat{B}^*_{pb} \to^{P^*} B$, in probability. We can write $\tilde{\varepsilon}^*_{it} - \bar{\tilde{\varepsilon}}^*_i = \hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i - \left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right)\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$, where $\tilde{\varepsilon}^*_{it} = y^*_{it} - \hat{\alpha}^*_i - \hat{\hat{\theta}}^*_{pb} y^*_{it-1}$ with $\hat{\alpha}^*_i = \bar{y}^*_i - \hat{\hat{\theta}}^*_{pb}\bar{y}^*_{i-}$ and $\hat{\varepsilon}^*_{it} = y^*_{it} - \breve{\alpha}_i - \hat{\hat{\theta}} y^*_{it-1}$ with $\breve{\alpha}_i = \hat{\alpha}_{I_i}$. As in the proof of Lemma 3.1, we can write

$$\hat{B}^*_{pb} = \hat{B}^*_1 + \hat{B}^*_2 + \hat{B}^*_3, \text{ with}$$

$$\hat{B}^*_1 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\tilde{\varepsilon}^*_{it} - \bar{\tilde{\varepsilon}}^*_i\right)^2, \hat{B}^*_2 = -2\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right)\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i\right) \text{ and}$$

$$\hat{B}^*_3 = \left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right)^2 \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4.$$

Given the pairs bootstrap DGP, one can show that $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i\right)$ and $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$ are $O_{P^*}(1)$ terms, and therefore, $\hat{B}^*_2$ and $\hat{B}^*_3$ are $o_{P^*}(1)$ terms. For $\hat{B}^*_1$, we also define $\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i = \varepsilon^*_{it} - \bar{\varepsilon}^*_i - \left(\hat{\hat{\theta}} - \theta_0\right)\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$, where $\varepsilon^*_{it} = y^*_{it} - \alpha^*_i - \theta_0 y^*_{it-1}$ with $\tilde{\alpha}^*_i = \alpha_{I_i}$. This implies that

$$\hat{B}^*_1 = \chi^*_1 + \chi^*_2 + \chi^*_3, \text{ with}$$

$$\chi^*_1 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2, \ \chi^*_2 = -2\left(\hat{\hat{\theta}} - \theta_0\right)\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right) \text{ and}$$

$\chi^*_3 = \left(\hat{\hat{\theta}} - \theta_0\right)^2 \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$. As before, one can show that given the bootstrap DGP of the pairwise bootstrap, $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)$ and $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$ are $O_{P^*}(1)$ and therefore, $\hat{\chi}^*_2$ and $\hat{\chi}^*_3$ are $o_{P^*}(1)$. Let us turn to $\hat{\chi}^*_1$. Since we have resampled only in the cross section,

$$\begin{aligned} E^*\left|\hat{\chi}^*_1\right| &= \frac{1}{n}\sum_{i=1}^n E^*\left\{\frac{1}{T}\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2\right\} = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y_{it-1} - \bar{y}_{i-}\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y_{it-1} - \mu_i\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 + \frac{2}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y_{it-1} - \mu_i\right)\left(\mu_i - \bar{y}_{i-}\right)\left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \\ &\quad + \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(\mu_i - \bar{y}_{i-}\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \equiv B_1 + B_2 + B_3, \end{aligned}$$

where $\mu_i = E\left(y_{it}\right) = \frac{\alpha_i}{1-\theta_0}$. By Cauchy-Schwartz inequality,

$$|B_2| \leq \left(\frac{2}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(\mu_i - \bar{y}_{i-}\right) 2\right)^{1/2} \left(\frac{2}{nT}\sum_{i=1}^n\sum_{t=1}^T \left(y_{it-1} - \mu_i\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^4\right)^{1/2} \to^P 0$$

since $\frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)^2\left(\varepsilon_{it}-\bar{\varepsilon}_i\right)^4 = O_P(1)$ and $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mu_i-\bar{y}_{i-}\right)2 = \frac{1}{n}\sum_{i=1}^{n}\bar{u}_{i-}^2 \to 0$ by Lemma A.3. One can also show that $B_3 = o_P(1)$. Finally,

$$B_1 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)^2\varepsilon_{it}^2 - \frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)^2\varepsilon_{it}\bar{\varepsilon}_i + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)^2\bar{\varepsilon}_i^2$$

where the first term obviously converges in probability to B while the remaining terms converge to 0 by making use of the Cauchy-Schwartz inequality.