Econometric Theory, **0**, 2025, 1–34.
doi:10.1017/S026646662510008X

THIS SHOCK IS DIFFERENT: ESTIMATION AND INFERENCE IN MISSPECIFIED TWO-WAY FIXED EFFECTS PANEL REGRESSIONS

ARTŪRAS JUODIS
University of Amsterdam and Tinbergen Institute

We investigate the properties of the linear two-way fixed effects (FE) estimator for panel data when the underlying data generating process (DGP) does not have a linear parametric structure. The FE estimator is consistent for some pseudo-true value and we characterize the corresponding asymptotic distribution. We show that the rate of convergence is determined by the degree of model misspecification, and that the asymptotic distribution can be non-normal. We propose a novel autoregressive double adaptive wild (AdaWild) bootstrap procedure applicable for a large class of DGPs. Monte Carlo simulations show that it performs well for panels of small and moderate dimensions. We use data from U.S. manufacturing industries to illustrate the benefits of our procedure.

1. INTRODUCTION

In the last decades, the two-way fixed effects estimator (or simply the FE estimator in this article) has become a default method for estimating linear panel data models. Among other things, it became a standard option for causal policy evaluation analysis, as many researchers use this estimator to jointly adjust for unobserved unit-specific and time-specific effects. Thus, it comes as a surprise that no general theoretical results are available that characterize the large sample properties of the FE estimator when the validity of the postulated additive linear model is questionable. This article closes this (unfortunate) knowledge gap.

I would like to thank the Co-Editor Guido Kuersteiner and three anonymous referees for numerous suggestions that greatly improved this article. I would also like to thank Stephane Bonhomme, Tom Boot, Jorg Breitung, Denis Chetverikov, Ivan Fernandez-Val, Yi He, Aureo de Paula, Anders Bredahl Kock, Frank Kleibergen, Ryo Okui, Andres Santos, Vasilis Sarafidis, Gabriela Szini, Gautam Tripathi, Martin Weidner, Joakim Westerlund, as well as the participants of the 2021 International Panel Data Conference, 2023 Annual IAAE Conference (Oslo), and seminar participants at U Amsterdam, Oxford, UCLA, U Duisburg-Essen, U Lund, U Luxembourg, U Cologne, and Tinbergen Econometrics Workshop for helpful comments. Preliminary version of this article was completed while I enjoyed the hospitality of the Department of Economics at the University College London in May 2019. Financial support from the Netherlands Organization for Scientific Research (NWO) under research grant number 451-17-002 is gratefully acknowledged. Artūras Juodis, Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands, e-mail: a.juodis@uva.nl.

In this article, we focus on the properties of the FE estimator within the general framework of misspecified models. The setup we consider is sufficiently general to accommodate both continuous, discrete, or mixed covariates. The price we pay for this generality is that we need to assume that both dimensions of the panel dataset are large, a more typical setting in the industry-level or macro-level panels. In particular, there has been a great amount of interest in the macro panel data literature questioning the validity of the (model-based) additive error components structure. The typical alternative is the multiplicative (interactive/factor) error components structure (e.g., Pesaran, 2006; Bai, 2009; Juodis and Sarafidis, 2022b, among others). However, estimation of such nonlinear models introduces some non-trivial theoretical and empirical trade-offs. For example, the recent results in Juodis, Karabiyik, and Westerlund (2021) suggest that the popular estimator of Pesaran (2006) can be highly sensitive to nuisance parameters that control the strength of the factor estimates.

Given some of the shortcomings of the multiplicative models, it is now evident in the (macro-) panel data literature that it can be beneficial to pre-test for the applicability of the simple additive error-components model (see, e.g., Petrova and Westerlund, 2020; Juodis and Reese, 2022; Kapetanios, Serlenga, and Shin, 2024) instead of completely ignoring these type of models. However, for this purpose one should better understand the statistical properties of the simple FE estimator when the validity of the postulated model is uncertain. For example, when the additive model structure is completely misspecified (as in Galvao and Kato, 2014). This is the goal of this article.

1.1. Contributions

The contributions of this article are two-fold.

i) We characterize the asymptotic properties of the FE estimator where all observed variables are subject to common shocks under a set of minimal assumptions on the underlying DGP. We show that generally the convergence rate of the FE estimator can be as slow as $\sqrt{\min(N,T)}$, and as fast as $\sqrt{NT}$. This article extends the (cross-sectional) independent setup of Galvao and Kato (2014) toward a more realistic setup where independence only holds conditionally on some unobserved common shocks. We show that such an extension greatly complicates analysis, thus justifying that *"this (common) shock is different"*. For this purpose, we use a novel proof strategy (inspired by that of Hahn, Kuersteiner, and Mazzocco, 2022) to derive the joint asymptotic limit theorem as $N,T \to \infty$ (jointly) for dependent averages of the data with the corresponding convergence rates of order $\sqrt{N}$, $\sqrt{T}$, and $\sqrt{NT}$. To the best of our knowledge, this is the first such result in the panel data literature.

ii) We contribute to the literature on multi-way clustering (see, e.g., (Cameron, Gelbach, and Miller, 2011; Thompson, 2011; Menzel, 2021); Chiang, Hansen, and Sasaki, 2024) by proposing a novel inference procedure for double index models where the setup is conditionally independent over cross-sectional dimension, and weakly stationary in time-series dimension. Our bootstrap procedure is of great interest on its own, as it extends the method of Menzel (2021) for panel datasets with temporal dependence.

1.2. Organization and Notation

The organization of this article is as follows. In Section 2, we introduce the model. In Section 3, we provide a formal asymptotic expansion of the FE estimator, while in Section 4, we discuss inference. In Section 5, we conduct a Monte Carlo study to assess the finite sample performance of the proposed procedure while in Section 6, we present an empirical application. Section 7 concludes. All proofs, additional Monte Carlo results, and further discussions are placed in the Supplementary Material.

We denote by N the size of the cross-sectional dimension, and by T the size of the time-series dimension of the data. For a generic vector $z_{i,t}$ we denote its time-series, cross-sectional and overall sample averages by $\bar{z}_i = T^{-1}\sum_{t=1}^{T} z_{i,t}$, $\bar{z}_t = N^{-1}\sum_{i=1}^{N} z_{i,t}$, and $\bar{z} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{i,t}$, respectively. We also use the shorthand notation $\sum_i = \sum_{i=1}^{N}$ and $\sum_t = \sum_{t=1}^{T}$. For any pair of zero-mean cross-sectional variables $(\boldsymbol{v}_i', \boldsymbol{h}_i')'$ denote their corresponding unconditional covariance as $\boldsymbol{\Sigma}_{\boldsymbol{vh}'} = \mathrm{E}[\boldsymbol{v}_i \boldsymbol{h}_i']$. For any pair of zero-mean time-series variables $(\boldsymbol{v}_t', \boldsymbol{h}_t')'$, denote by $\boldsymbol{\Sigma}_{\boldsymbol{vh}',LR}$, their corresponding long-run covariance matrix. We will also make use of the latter definition for any temporally dependent panel variables $(\boldsymbol{v}_{i,t}', \boldsymbol{h}_{i,t}')'$. For any square $[m \times m]$ matrix $\boldsymbol{A}$, denote by $\mathrm{diag}(\boldsymbol{A})$ an $[m \times 1]$ vector with the diagonal elements of $\boldsymbol{A}$. Denote by $\mathbb{I}(\cdot)$ a vector valued indicator function with all statements evaluated element-wise. Let $(\Omega, \mathcal{A}, P)$ be the common probability space. We use the wp1 notation to indicate that the corresponding statement holds with probability 1 (or almost-surely). Finally, let $\Delta < \infty$ denote an arbitrarily large real number independent of N and T.

2. THE TWO-WAY FIXED EFFECTS ESTIMATOR

2.1. The Estimator

Suppose that we have a panel data set of observable variables $\{z_{i,t}\}_{i=1,t=1}^{N,T}$, or simply $\{z_{i,t}\}$. This vector can be further decomposed as $z_{i,t} = (y_{i,t}, \boldsymbol{x}_{i,t}')'$, such that $\{y_{i,t}\}$ is the scalar variable of interest, while $\{\boldsymbol{x}_{i,t}\}$ is the K dimensional vector of policy relevant and additional auxiliary explanatory variables (e.g., covariates and controls). It is standard to estimate the partial (marginal) effect of all elements in $\boldsymbol{x}_{i,t}$ on $y_{i,t}$, while at the same time conditioning on unit-specific and time-specific unobserved characteristics. As a result, it is a common empirical practice,[1] to

[1] See, e.g., Berger et al. (2013) and Voigtländer (2014).

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

assume that $\{y_{i,t}\}$ can be represented by means of the following fully additive linear regression model:

$$y_{i,t} = \boldsymbol{\beta}'\boldsymbol{x}_{i,t} + \eta_i + f_t + v_{i,t}, \tag{1}$$

where scalar variable $\{\eta_i\}$ captures all time-invariant unit-specific unobserved characteristics relevant for $\{y_{i,t}\}$, while $\{f_t\}$ is supposed to account for all common economy-wide shocks, for example, business cycle, productivity, or health-care system shocks that impact $\{y_{i,t}\}$ after conditioning on $\{\boldsymbol{x}_{i,t}\}$. Among other types of regressors, $\boldsymbol{x}_{i,t}$ might contain binary treatment variables $D_{i,t} \in \{0;1\}$ (as in the classical DiD framework).

In the absence of endogeneity between $v_{i,t}$ and $\boldsymbol{x}_{i,t}$, $\boldsymbol{\beta}$ can be estimated using the two-way fixed effects estimator (which we simply label as the FE estimator):

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{x}_{i,t}^{++}}(\widehat{\boldsymbol{x}_{i,t}^{++}})'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\boldsymbol{x}_{i,t}^{++}}\widehat{y_{i,t}^{++}}\right), \tag{2}$$

where $\widehat{\boldsymbol{x}_{i,t}^{++}}$ and $\widehat{y_{i,t}^{++}}$ are defined in terms of the deviations from the corresponding cross-sectional and time-series averages, that is,

$$\widehat{\boldsymbol{x}_{i,t}^{++}} \equiv \boldsymbol{x}_{i,t} - \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{x}_{j,t} - \frac{1}{T}\sum_{s=1}^{T}\boldsymbol{x}_{i,s} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\boldsymbol{x}_{j,s}, \tag{3}$$

and similarly for $\widehat{y_{i,t}^{++}}$.

The FE estimator is consistent for the "true value" $\boldsymbol{\beta}_0$, and has normal asymptotic limit for any fixed T provided that all regressors are strictly exogenous and the additive structure in (1) is correctly specified. If the regressors are only weakly exogenous (pre-determined), then the estimator is not fixed T consistent due to the "Nickell bias" (e.g., Nickell, 1981). However, it remains consistent provided that $T \to \infty$. In that case, the asymptotic distribution is not correctly centered at zero with a non-negligible asymptotic bias present when $N/T \to \kappa \in (0;\infty)$ (see Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003). The non-negligible asymptotic bias corresponds to the leading term of the "Nickell bias".[2]

Unlike time-series demeaning that removed unit-specific effects $\{\eta_i\}$, estimation of the time-effects $\{f_t\}$ is generally non-consequential even as $T \to \infty$. In particular, it is not a source of an additional incidental parameters (IPP) bias, see Hahn and Moon (2006). The latter observation, on the other hand, is solely driven by the linearity of the model and cross-sectional independence of the error terms, and does not generally hold for nonlinear models (see, e.g., Hahn and Newey, 2004; Hahn and Kuersteiner, 2011; Fernández-Val and Lee, 2013; Fernández-Val and Weidner, 2016; Chen, Fernández-Val, and Weidner, 2021b).

[2]For simplicity, from this point onwards, we refer to the leading term of the "Nickell bias" as the "Nickell bias".

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

2.2. Model Misspecification

Overall, except for some recent results in the DiD literature with heterogeneous treatment effects,[3] there is a limited amount of results on the properties of the FE estimator under general types of model misspecification. For example, one can wonder what happens when (1) is just a regression model (linear projection) and does not necessarily have any structural/causal interpretation? More generally, available results have very little to say on the following aspects.

(i) Does the probability limit of the FE estimator have a clear population-level interpretation?
(ii) Is it possible to give (1) interpretation in terms of some linear (population) projection? How should one interpret unobserved components $\{\eta_i\}$, $\{f_t\}$, and $\{v_{i,t}\}$ in such a case?
(iii) How should one proceed with inference in order to account for potential violations of the underlying linear additive model?
(iv) Should one report the same type of standard errors and confidence intervals (CIs) for main empirical models, as well as for the corresponding "robustness checks"?
(v) Is simple clustering, as in the single-index setup of White (1982), a definite solution to these problems?

This article fills these knowledge gaps, under a set of simplifying restrictions. In particular, we limit attention to the sampling scheme where both panel dimensions are large, that is, $N,T \to \infty$ using the joint limit theory of Phillips and Moon (1999). This asymptotic approximation scheme is customary in the panel data literature on misspecified models (e.g., Galvao and Kato, 2014; Okui, 2017).

The contribution of Galvao and Kato (2014) is the closest in its focus to this article. There the same set of questions for the standard FE estimator are addressed in the setting where all cross-sectional units are independent. As such, as we argue later in the article, their results should be seen as a special case of our results discussed in Section 3 (for a specific choice of trivial $\sigma-$algebra of common shocks).

2.3. Main Illustrative Example

In what follows we present the main illustrative example that will be studied throughout this article—a linear model with misspecified error-components structure.

Example 1 (Factor model). Consider the following (factor-augmented) DGP with one regressor

$$y_{i,t} = \beta^* x_{i,t} + \lambda_i f_t + v_{i,t}, \quad x_{i,t} = \gamma_i \pi_t + u_{i,t}, \tag{4}$$

[3]For example, de Chaisemartin and D'Haultfœuille (2020), Callaway and Sant'Anna (2021), Imai and Kim (2021), Sun and Abraham (2021), and Goodman-Bacon (2021).

where the vectors $(v_{i,t}, u_{i,t})'$, $(\gamma_i, \lambda_i)'$, and $(f_t, \pi_t)'$ are mutually independent for all (i,t). Without any loss of generality we further assume that all random variables are mean zero (unconditionally). If $u_{i,t}$ and $v_{i,t}$ are uncorrelated for all (i,t), then the results of this article can be used to establish that the FE estimator is consistent for the "pseudo-true" value β_0, where

$$\beta_0 = \beta^* + \frac{\mathrm{E}[\gamma_i\lambda_i]\,\mathrm{E}[f_t\pi_t]}{\mathrm{E}[\gamma_i^2]\,\mathrm{E}[\pi_t^2] + \mathrm{E}[u_{i,t}^2]}. \tag{5}$$

In this example, if either loadings are uncorrelated, that is, $\mathrm{E}[\gamma_i\lambda_i] = 0$, or factors are uncorrelated, that is, $\mathrm{E}[f_t\pi_t] = 0$, then the FE estimator is consistent for the "true" value β^* that corresponds to the DGP in (4). Intuitively, as $\lambda_i f_t$ is an omitted variable, it does not induce cross-sectional omitted variable bias when $N \to \infty$ as long as $\mathrm{E}[\gamma_i\lambda_i] = 0$, while the time-series omitted variable bias is avoided when $T \to \infty$ and $\mathrm{E}[f_t\pi_t] = 0$. As we consider the asymptotic scheme where both $N, T \to \infty$, consistency with respect to β^* is achieved if either of the two covariances is zero. However, as we will show later in the article, zero-correlation conditions alone are not sufficient to further characterize the asymptotic distribution of the FE estimator. □

This stylized example provides a small snapshot on the type of models covered in this article. Later in the article another example is studied where model misspecification is induced by endogeneity of the regressor. Additional examples, for example, with nonlinearity or measurement error induced model misspecification, together with the corresponding pseudo-true values and all stochastic components relevant for the results in this article are discussed in the Supplementary Matterial. Several examples of DGPs that fall outside of the scope of this article are also discussed there.

3. LARGE SAMPLE RESULTS

In this section, we assume that all common shocks are measurable with respect to the (sub-) σ-algebra $\mathcal{F}$, while all unit-specific time-invariant random variables are measurable with respect to the (sub-) σ-algebra $\mathcal{C}_i$. We denote the (sub-) σ-algebra generated by unit-specific and common shocks by $\mathcal{D}_i = \mathcal{F} \vee \mathcal{C}_i$.

3.1. Primitive Assumptions

Denote by $z_{i,t} = (y_{i,t}, \boldsymbol{x}_{i,t}')'$ a full vector of observables with a typical element $z_{i,t}^{(h)}$ for $h = 1, \ldots, K+1$. The stacked versions of this vector are denoted by $\boldsymbol{Z}_i = (z_{i,1}, \ldots, z_{i,T})'$, for all $i = 1, \ldots, N$.

Assumption 1. The DGP for all (i,t) is such that for some $r \geq 8$ and $\delta > 0$.

(a) Conditionally on $\mathcal{F}$, $\boldsymbol{Z}_i$ are identically distributed and independent (i.i.d.) across i.
(b) σ-algebras $\mathcal{C}_i$ are independent across i, and independent of $\mathcal{F}$.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

(c) Each element $z_{i,t}^{(h)}$ satisfies $\mathrm{E}\left[\left|z_{i,t}^{(h)}\right|^{r+\delta}|\mathcal{D}_i\right] < \Delta$. Moreover, $\{z_{i,t}\}_{t=1}^{T}$ is a ($\mathcal{D}_i$-conditional) α-mixing sequence with mixing coefficient $\alpha_{i,t}(m)$ measurable w.r.t. $\mathcal{D}_i$ such that $\sup_i \sup_t \alpha_{i,t}(m) = \mathcal{O}(m^{-\mu})$ with $\mu = 3(r+\delta)/\delta$.

(d) $\{(z'_{i,t}, (z_{i,t} - \mathrm{E}[z_{i,t}|\mathcal{F}])')'\}_{t=1}^{T}$ is a covariance stationary ($\mathcal{C}_i$-conditional) α-mixing sequence, with mixing coefficients $\widetilde{\alpha}_{i,t}(m)$ measurable w.r.t. $\mathcal{C}_i$ such that $\sup_i \sup_t \widetilde{\alpha}_{i,t}(m) = \mathcal{O}(m^{-\mu})$ with $\mu = 3(r+\delta)/\delta$.

As usual in the large T panel data literature (e.g., Fernández-Val and Weidner, 2016) we need to restrict the temporal (α-mixing) dependence of $\{z_{i,t}\}$ after conditioning on $\mathcal{D}_i$. In the context of correctly specified panel data models, the notion of conditional mixing has been introduced by Hahn and Kuersteiner (2011), and it was later used by Su, Jin, and Zhang (2015), Fernández-Val and Weidner (2016), and Juodis and Sarafidis (2022a) (among others) to study settings with common shocks. Furthermore, the dependence of mixing coefficients on r reflects the natural trade-off between the degree of dependence and the moment bounds of the process. Moreover, similarly to Galvao and Kato (2014) in part (d) we impose strict (conditional on $\mathcal{C}_i$) time-series stationarity of $\{z_{i,t}\}$. As the corresponding statement is not formulated conditional on $\mathcal{F}$, we rule out the settings with non-ergodic $\mathcal{F}$-measurable shocks, for example, as explicitly allowed in the setting of Andrews (2005).

The stochastic restrictions presented above are in general more restrictive than necessary to characterize the asymptotic limit of the FE estimator (both for T fixed and $T \to \infty$), and, in some situations, can be relaxed. For example, the i.i.d. assumption conditional on $\mathcal{F}$ can be relaxed towards exchangeability following Andrews (2005).

3.2. Consistency

In this section, we discuss consistency and convergence rate properties of the FE estimator $\widehat{\boldsymbol{\beta}}$ defined in (2). As an intermediate step, we first derive the asymptotic expansion of $\widehat{\boldsymbol{\Sigma}}_z$, the sample variance-covariance matrix of $\{z_{i,t}\}$ after the TWFE transformation, that is,

$$\widehat{\boldsymbol{\Sigma}}_z = \frac{1}{NT}\sum_i \sum_t (\widehat{z_{i,t}^{++}})(\widehat{z_{i,t}^{++}})'. \tag{6}$$

Before doing so, we introduce some additional notation. In particular, given ($\mathcal{C}_i$-conditional) covariance stationarity of $\{z_{i,t}\}$, we use Hoeffding (1948)-type projection arguments to expand $z_{i,t}$, that is,

$$z_{i,t} = \underbrace{z_{i,t}^{\perp} + z_{i,t}^{+}}_{z_{i,t}^{++}} + \boldsymbol{v}_i^z + \boldsymbol{g}_t^z + \boldsymbol{\mu}^z, \tag{7}$$

where $\boldsymbol{\mu}^z = \mathrm{E}[z_{i,t}]$, $z_{i,t}^{\perp} = z_{i,t} - \mathrm{E}[z_{i,t}|\mathcal{D}_i]$, $z_{i,t}^{+} = \mathrm{E}[z_{i,t}|\mathcal{D}_i] - \boldsymbol{v}_i^z - \boldsymbol{g}_t^z - \boldsymbol{\mu}^z$, and $\boldsymbol{v}_i^z = \mathrm{E}[z_{i,t}|\mathcal{C}_i] - \boldsymbol{\mu}^z$, $\boldsymbol{g}_t^z = \mathrm{E}[z_{i,t}|\mathcal{F}] - \boldsymbol{\mu}^z$. Intuitively, $\{z_{i,t}^{+}\}$ measures the degree

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

of the misspecification of the additive components model, that is, when $\{z_{i,t}\}$ has an additive error components structure then $z_{i,t}^{+} = \mathbf{0}$ wp1. Similarly, let $\boldsymbol{Q}_{i,t} \equiv z_{i,t}^{++}(z_{i,t}^{++})'$, then given covariance stationarity of $\{z_{i,t} - \mathrm{E}[z_{i,t}|\mathcal{F}]\}$, similar decomposition arguments as in (7) can be used to expand $\boldsymbol{Q}_{i,t}$ as follows:

$$\boldsymbol{Q}_{i,t} = \boldsymbol{Q}_{i,t}^{\perp} + \boldsymbol{Q}_{i,t}^{+} + \boldsymbol{U}_i^{\boldsymbol{Q}} + \boldsymbol{G}_t^{\boldsymbol{Q}} + \mathrm{E}[z_{i,t}^{++}(z_{i,t}^{++})']. \tag{8}$$

With this notation at hand, we formulate the first main result of this article.

THEOREM 1. *Let $N,T \to \infty$ (jointly) and $N/T \to \kappa \in (0;\infty)$, then under Assumption 1*

$$\widehat{\boldsymbol{\Sigma}}_z = \mathrm{E}[z_{i,t}^{++}(z_{i,t}^{++})'] + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{U}_i^{\boldsymbol{Q}} + \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{G}_t^{\boldsymbol{Q}} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{Q}_{i,t}^{+} + \mathcal{O}_P((NT)^{-1/2}), \tag{9}$$

where

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{U}_i^{\boldsymbol{Q}} = \mathcal{O}_P(N^{-1/2}), \tag{10}$$

$$\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{G}_t^{\boldsymbol{Q}} + \frac{1}{N}\sum_{i}\boldsymbol{Q}_{i,t}^{+}\right) = \mathcal{O}_P(T^{-1/2}). \tag{11}$$

The following two immediate conclusions can be drawn from this theorem: i) the variance-covariance matrix $\widehat{\boldsymbol{\Sigma}}_z$ is consistent for its population analogue and ii) associated estimation uncertainty upon replacing the infeasible quantities $\{z_{i,t}^{++}\}$ with $\{\widehat{z_{i,t}^{++}}\}$ is of (at most) order $\mathcal{O}_P((NT)^{-1/2})$. Theorem 1 provides a non-trivial extension of the results in Okui (2014), who analyzed the asymptotic properties of (9) in the univariate setting with correctly specified additive error components structure.

Generally, the asymptotic rate in (11) cannot be further improved upon based on Assumption 1 alone. As it is evident from the proof of this theorem, Assumption 1 can only be used to establish that the sum of the two corresponding components is an α-mixing sequence, not that the two components individually form α-mixing sequences.

As the FE estimator is just a nonlinear transformation of $\widehat{\boldsymbol{\Sigma}}_z$, the asymptotic properties of the FE estimator follow directly subject to the usual population *no multicollinearity* condition.

Assumption 2. $\boldsymbol{\Sigma}_x = \mathrm{E}[\boldsymbol{x}_{i,t}^{++}(\boldsymbol{x}_{i,t}^{++})']$ is positive definite.

This assumption excludes various types of trivial "low-rank" regressors from the model. The next corollary confirms that the FE estimator $\widehat{\boldsymbol{\beta}}$ is consistent with respect to appropriately defined pseudo-true value $\boldsymbol{\beta}_0$.

COROLLARY 1. *Let $N,T \to \infty$ (jointly) and $N/T \to \kappa \in (0;\infty)$, then under Assumptions 1 and 2*

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \mathcal{O}_P(\min(N,T)^{-1/2}), \tag{12}$$

where

$$\boldsymbol{\beta}_0 \equiv \mathrm{E}[\boldsymbol{x}_{i,t}^{++}(\boldsymbol{x}_{i,t}^{++})']^{-1}\,\mathrm{E}[\boldsymbol{x}_{i,t}^{++}y_{i,t}^{++}]. \tag{13}$$

Hence, under fairly primitive conditions on the DGP[4] the FE estimator is consistent for $\boldsymbol{\beta}_0$. Moreover, Corollary 1 indicates that the estimator is at least $\sqrt{\min(N,T)}$-consistent. This is in sharp contrast with the setup of Galvao and Kato (2014), where the minimal convergence rate is always $\sqrt{N}$. The potential reduction in the convergence rate is solely explained by the presence of $\{\boldsymbol{G}_t^{Q}\}$-type components in (8), that are by construction/assumption excluded from the model of Galvao and Kato (2014). In Supplementary Material, we show that $\boldsymbol{\beta}_0$ can be usefully interpreted as the solution to a population partial linear projection problem for cross-sectionally demeaned variables.

3.3. High-Level Conditions

In what follows we introduce, a set of additional regularity conditions that are sufficient to rigorously describe the asymptotic distribution of the FE estimator. These additional assumptions are needed to accommodate serial (temporal) dependence in $\{z_{i,t}\}$ generated either by common shocks and/or conditionally independent unit-specific idiosyncratic shocks.[5]

In order to describe the asymptotic properties of the FE estimator, we need to first characterize asymptotic properties for the sample average of the corresponding influence function

$$\frac{1}{NT}\sum_i\sum_t \widehat{\boldsymbol{x}_{i,t}^{++}}(\widehat{y_{i,t}^{++}} - \boldsymbol{\beta}_0'\widehat{\boldsymbol{x}_{i,t}^{++}}). \tag{14}$$

For this purpose, we define the (infeasible) influence function that corresponds to the FE estimator as

$$\boldsymbol{w}_{i,t} \equiv \boldsymbol{x}_{i,t}^{++}(y_{i,t}^{++} - \boldsymbol{\beta}_0'\boldsymbol{x}_{i,t}^{++}), \tag{15}$$

[4] As it was discussed in Section 3.1, these conditions are only marginally stronger than those considered by Galvao and Kato (2014) in the setting without common shocks.

[5] In particular, these assumptions are redundant for the setup with exchangeable double arrays as in Menzel (2021) and Fernández-Val, Freeman, and Weidner (2021), as they are implied by the Aldous–Hoover–Kallenberg representation.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

for all (i,t). Here, the term in the brackets can be defined as the reduced form projection "error term"

$$\varepsilon_{i,t} \equiv y_{i,t}^{++} - \boldsymbol{\beta}_0' \boldsymbol{x}_{i,t}^{++}, \tag{16}$$

such that $\mathrm{E}[\boldsymbol{x}_{i,t}^{++}\varepsilon_{i,t}] = \mathbf{0}_K$, $\mathrm{E}[\varepsilon_{i,t}] = 0$, and $\mathrm{E}[\boldsymbol{w}_{i,t}] = \mathbf{0}_K$.[6]

Using (8), the infeasible influence function $\boldsymbol{w}_{i,t}$ can be expanded as

$$\boldsymbol{w}_{i,t} = \boldsymbol{w}_{i,t}^{\perp} + \boldsymbol{w}_{i,t}^{+} + \boldsymbol{v}_i + \boldsymbol{g}_t, \tag{17}$$

with all terms defined accordingly as transformations of the corresponding elements of $\boldsymbol{Q}_{i,t}$, for example, $\boldsymbol{g}_t = (\mathbf{0}_K, \boldsymbol{I}_K)\boldsymbol{G}_t^{\boldsymbol{Q}}(1, -\boldsymbol{\beta}_0')'$. As $\boldsymbol{w}_{i,t}$ has the "multi-way clustering" structure, the asymptotic properties of

$$\frac{1}{NT}\sum_i \sum_t \boldsymbol{w}_{i,t}, \tag{18}$$

depend on the corresponding averages of the four individual components in (17). Unfortunately, Assumption 1 is not sufficiently informative to establish distributional results for all four components. The need for additional restrictions is best illustrated by the fact that, while Assumption 1 is sufficient to establish rates of convergence in Theorem 1, it is silent about the asymptotic distribution of

$$\overline{\boldsymbol{w}^{+}} = \frac{1}{NT}\sum_i \sum_t \boldsymbol{w}_{i,t}^{+}. \tag{19}$$

In particular, α- mixing restrictions in Assumption 1 (c) are not informative about the asymptotic properties of this term, as $\overline{\boldsymbol{w}^{+}}$ is non-stochastic upon conditioning on $\mathcal{D}_i$.

In this article, we address this problem by imposing additional (high-level) restrictions on the stochastic behavior of the individual components of the influence function $\{\boldsymbol{w}_{i,t}\}$.

Remark 1. We would like to note that the approach pursued in this article is by no means the only approach that can be used to study asymptotic properties of the FE estimator. In particular, there are at least two alternative approaches that can be used instead: i) imposing high-level conditions on the full matrix $\{\boldsymbol{Q}_{i,t}\}$ and ii) imposing low-level parametric restrictions on $\{z_{i,t}\}$ directly. The latter approach would bring our analysis closer to the setup of Chiang et al. (2024), and postulate that $z_{i,t} = \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_t, \boldsymbol{\varepsilon}_{i,t})$ for some unobserved function $\boldsymbol{f}(\cdot,\cdot,\cdot)$ and corresponding vectors of unknown finite dimensions. The approach we take in this article is less restrictive as we only need to specify restrictions on the Hoeffding-type decomposition of $\{\boldsymbol{w}_{i,t}\}$ without the need to specify directly restrictions on $\{\boldsymbol{\alpha}_i\}$, $\{\boldsymbol{\lambda}_t\}$, $\{\boldsymbol{\varepsilon}_{i,t}\}$. Evidently, for our high-level sufficient conditions

[6]Using similar decomposition arguments to those in (7), it can be seen that $\varepsilon_{i,t} = \varepsilon_{i,t}^{++}$ wp1.

specified below should be case-by-case verified for any specific DGP of the form $z_{i,t} = \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_t, \boldsymbol{\varepsilon}_{i,t})$.[7]

In what follows we revisit, Example 1 to illustrate the type of structures/DGPs that restrictions on $\{\boldsymbol{w}_{i,t}^{+}\}$ should accommodate.

Example 1 (Continued). In this example,

$$y_{i,t}^{++} = \beta^* x_{i,t}^{++} + \lambda_i f_t + v_{i,t}, \tag{20}$$

$$x_{i,t}^{++} = \gamma_i \pi_t + u_{i,t}. \tag{21}$$

Hence,

$$\varepsilon_{i,t} = \lambda_i f_t + v_{i,t} + (\beta^* - \beta_0)(\gamma_i \pi_t + u_{i,t}), \tag{22}$$

$$w_{i,t} = \varepsilon_{i,t}(\gamma_i \pi_t + u_{i,t}). \tag{23}$$

The corresponding components of (17) decomposition are provided by[8]

$$w_{i,t}^{\perp} = u_{i,t}(v_{i,t} + \lambda_i f_t) + (2(\beta^* - \beta_0)u_{i,t} + v_{i,t})\gamma_i \pi_t + (\beta^* - \beta_0)(u_{i,t}^2 - 1), \tag{24}$$

$$w_{i,t}^{+} = (\lambda_i \gamma_i - \mathrm{E}[\lambda_i \gamma_i])(f_t \pi_t - \mathrm{E}[f_t \pi_t]) - (\gamma_i^2 - \mathrm{E}[\gamma_i^2])(\pi_t^2 - \mathrm{E}[\pi_t^2])(\beta_0 - \beta^*), \tag{25}$$

$$g_t = \mathrm{E}[\lambda_i \gamma_i](f_t \pi_t - \mathrm{E}[f_t \pi_t]) - (\beta_0 - \beta^*)\,\mathrm{E}[\gamma_i^2](\pi_t^2 - \mathrm{E}[\pi_t^2]), \tag{26}$$

$$\nu_i = \mathrm{E}[f_t \pi_t](\lambda_i \gamma_i - \mathrm{E}[\lambda_i \gamma_i]) - (\beta_0 - \beta^*)\,\mathrm{E}[\pi_t^2](\gamma_i^2 - \mathrm{E}[\gamma_i^2]). \tag{27}$$

In this example, Assumption 1 will be satisfied if we further assume that e.g., $\{(v_{i,t}, u_{i,t}, \pi_t, f_t)'\}$ is a stationary (jointly) α- mixing sequence of corresponding size.[9] Moreover, we need to assume that $(v_{i,t}, u_{i,t})'$ has a finite $8+\delta$ moment, while $(\pi_t, f_t)'$ and $(\gamma_i, \lambda_i)'$ are uniformly bounded for all (i,t).

Note how components $\{\nu_i\}$,$\{g_t\}$, $\{w_{i,t}^{+}\}$ are solely determined from the misspecification of the two-way error components structure in $\{z_{i,t}\}$. However, this not true in general. Our next example illustrates that even if $\boldsymbol{z}_{i,t}^{+} = \boldsymbol{0}$ wp1 (i.e., the additive error components structure is correctly specified), the FE estimator can have expansion of the influence function similar to that in Example 1.

Example 2 (Model with endogenous regressor). Consider the following DGP with one endogenous regressor

$$y_{i,t} = \beta^* x_{i,t} + \eta_i^{(y)} + f_t^{(y)} + \gamma_i v_{i,t}, \quad x_{i,t} = \eta_i^{(x)} + f_t^{(x)} + \pi_t u_{i,t}, \tag{28}$$

[7] As we argue in one of the counterexamples in the Supplementary Material, direct restrictions on $\{\boldsymbol{\alpha}_i\}$, $\{\boldsymbol{\lambda}_t\}$, $\{\boldsymbol{\varepsilon}_{i,t}\}$ alone are not sufficient to rule out some pathological cases, where asymptotic distribution of the estimator is non-standard. This is also illustrated by the fact that Assumption 3(iv) in Chiang et al. (2024) includes additional high-level restrictions that explicitly exclude pathological cases with singular long-run variances.

[8] Here, we use explicitly the fact that for this example $\mathrm{E}[u_{i,t}v_{i,t}] = 0$ holds.

[9] Note, that cross-sectional independence of $\{(v_{i,t}, u_{i,t})'\}$ is imposed in Section 2.3.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

where the vectors $(v_{i,t}, u_{i,t})'$, $(\gamma_i, \eta_i^{(y)}, \eta_i^{(x)})'$, and $(\pi_t, f_t^{(y)}, f_t^{(x)})'$ are mutually independent for all (i,t). The results of this article can be used to establish that the FE estimator is consistent for the "pseudo-true" value:

$$\beta_0 = \beta^* + \frac{\mathrm{E}[\gamma_i]\,\mathrm{E}[\pi_t]\,\mathrm{E}[v_{i,t}u_{i,t}]}{\mathrm{E}[\pi_t^2]\,\mathrm{E}[u_{i,t}^2]}. \tag{29}$$

For example, when $\mathrm{E}[\gamma_i] = 0$ and/or $\mathrm{E}[\pi_t] = 0$, the FE estimator is consistent for the "true" value β^*. More generally, in this example

$$y_{i,t}^{++} = \beta^* x_{i,t}^{++} + \gamma_i v_{i,t}, \tag{30}$$

$$x_{i,t}^{++} = \pi_t u_{i,t}, \tag{31}$$

$$w_{i,t} = (\beta^* - \beta_0)\pi_t^2 u_{i,t}^2 + \gamma_i \pi_t v_{i,t} u_{i,t}. \tag{32}$$

Using this construction, it is easy to see that in terms of the decomposition (17)

$$w_{i,t}^{\perp} = (\beta^* - \beta_0)\pi_t^2 (u_{i,t}^2 - \mathrm{E}[u_{i,t}^2]) + \gamma_i \pi_t (v_{i,t}u_{i,t} - \mathrm{E}[v_{i,t}u_{i,t}]), \tag{33}$$

$$w_{i,t}^{+} = (\gamma_i - \mathrm{E}[\gamma_i])(\pi_t - \mathrm{E}[\pi_t])\,\mathrm{E}[v_{i,t}u_{i,t}]), \tag{34}$$

$$g_t = (\beta^* - \beta_0)(\pi_t^2 - \mathrm{E}[\pi_t^2])\,\mathrm{E}[u_{i,t}^2] + \mathrm{E}[\gamma_i](\pi_t - \mathrm{E}[\pi_t])\,\mathrm{E}[v_{i,t}u_{i,t}], \tag{35}$$

$$\nu_i = (\gamma_i - \mathrm{E}[\gamma_i])\,\mathrm{E}[\pi_t]\,\mathrm{E}[v_{i,t}u_{i,t}]. \tag{36}$$

For this example, Assumption 1 will be satisfied if we further assume that $\{(v_{i,t}, u_{i,t}, \pi_t, f_t^{(y)}, f_t^{(x)})'\}$ is a stationary (jointly) α- mixing sequence of corresponding size. Moreover, we need to assume that $(v_{i,t}, u_{i,t})'$ has a finite $8+\delta$ moment, while $(\pi_t, f_t^{(y)}, f_t^{(x)})'$ and $(\gamma_i, \eta_i^{(y)}, \eta_i^{(y)})'$ are uniformly bounded for all (i,t).[10] Evidently, Assumption 2 is satisfied as long as

$$\mathrm{E}[(x_{i,t}^{++})^2] = \mathrm{E}[(\pi_t u_{i,t})^2] = \mathrm{E}[\pi_t^2]\,\mathrm{E}[u_{i,t}^2] > 0. \tag{37}$$

In these two examples, $\{\boldsymbol{w}_{i,t}^{+}\}$ exhibits two key features: i) it is a bivariate function (in terms of $\mathcal{C}_i$ and $\mathcal{D}$ measurable random variables) of some of the unit-specific and time-specific random variables that impact $\{\boldsymbol{z}_{i,t}\}$ (but not necessarily of all of them); ii) it has a factor (low-rank) structure for some fixed rank R.

Motivated by these two observations, we assume that these restrictions also hold for more general models. In particular, we assume that $\boldsymbol{w}_{i,t}^{+}$ can be (smoothly) approximated in terms of some low-dimensional $\mathcal{C}_i$ and $\mathcal{F}$ measurable unobserved variables.[11]

[10] It is easy to see that for this example (and some other examples studied in the Supplementary Material), Assumption 1 is unnecessarily restrictive due to the invariance of the FE estimator to $(f_t^{(y)}, f_t^{(x)}, \eta_i^{(y)}, \eta_i^{(x)})$.

[11] For example, this assumption is naturally satisfied if the data follows a finite dimensional non-linear quantile factor model as in Chen, Dolado, and Gonzalo (2021a).

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

Assumption 3. The second degree projection term of $\boldsymbol{w}_{i,t}$ ($\boldsymbol{w}_{i,t}^{+}$) satisfies

$$\boldsymbol{w}_{i,t}^{+} = \boldsymbol{w}(\boldsymbol{c}_i, \boldsymbol{d}_t), \tag{38}$$

where: i) $\boldsymbol{c}_i \in \mathbb{R}^{d_c}$ is $\mathcal{C}_i$-measurable, uniformly bounded random variable; ii) $\boldsymbol{d}_t \in \mathbb{R}^{d_d}$; iii) $\{\boldsymbol{d}_t\}$ is a $\mathcal{F}$-measurable, uniformly bounded sequence of random variables. Furthermore, for any $\boldsymbol{c}$ and $\boldsymbol{d}$

$$\boldsymbol{w}(\boldsymbol{c}, \boldsymbol{d}) = \sum_{\ell=1}^{\infty} \boldsymbol{\gamma}_\ell \phi^{(\ell)}(\boldsymbol{c}) \psi^{(\ell)}(\boldsymbol{d}), \tag{39}$$

where $\{\phi^{(\ell)}(\cdot)\}_{\ell \geq 1}$ and $\{\psi^{(\ell)}(\cdot)\}_{\ell \geq 1}$ are orthonormal functions in $L^2(\mathbb{R}^{d_c})$ and $L^2(\mathbb{R}^{d_d})$, respectively.

The models in Examples 1 and 2 trivially satisfy this assumption by appropriately defining $\{\boldsymbol{c}_i\}$ and $\{\boldsymbol{d}_t\}$ in terms of linear combinations of random variables in (25) and (34). For more details, we refer to the corresponding discussion in the Supplementary Material.

Assumption 3 is generally only required if the convergence rate of the FE estimator is $\sqrt{NT}$.[12] In other cases it is redundant, as the asymptotic distribution of the FE estimator will be primarily determined by the averages of $\{\boldsymbol{g}_t\}$ and $\{\boldsymbol{v}_i\}$.

Finally, we impose an additional set of regularity conditions on the individual components of (17). We conform with the notation introduced in Section 1 and set $\overline{\boldsymbol{g}} = T^{-1}\sum_t \boldsymbol{g}_t$ and $\overline{\boldsymbol{w}_i^{\perp}} = T^{-1}\sum_t \boldsymbol{w}_{i,t}^{\perp}$.

Assumption 4. For all (N,T), including $N,T \to \infty$, the following conditions are satisfied.

(a) $\{\boldsymbol{v}_i\}$ is such that: $\boldsymbol{v}_i = \boldsymbol{0}_K$ wp1 or the variance matrix $\boldsymbol{\Sigma}_{\boldsymbol{v}\boldsymbol{v}'}$ is non-singular.
(b) $\sum_{\ell=1}^{\infty} \|\boldsymbol{\gamma}_\ell\|^2 < \Delta$.
(c) $\{\boldsymbol{g}_t\}$ is such that: $\boldsymbol{g}_t = \boldsymbol{0}_K$ wp1 or the limiting long-run variance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{g}\boldsymbol{g}',LR} \equiv \lim_{T\to\infty} \mathrm{E}[T\overline{\boldsymbol{g}}\,\overline{\boldsymbol{g}}']$, is non-singular and non-stochastic.
(d) $\{(\boldsymbol{g}_t', \boldsymbol{d}_t')'\}$ is a covariance-stationary α-mixing sequence, with mixing coefficients $\tilde{\alpha}_t(m)$ such that $\sup_t \widetilde{\alpha}_t(m) = \mathcal{O}(m^{-\mu})$ with $\mu = 3(r+\delta)/\delta$ and r as in Assumption 1.
(e) The limiting long-run variance of $\{\boldsymbol{w}_{i,t}^{\perp}\}$, $\boldsymbol{\Sigma}_{\boldsymbol{w}^{\perp}(\boldsymbol{w}^{\perp})',LR} \equiv \lim_{T\to\infty} \mathrm{E}[T\overline{\boldsymbol{w}_i^{\perp}}(\overline{\boldsymbol{w}_i^{\perp}})']$, is non-singular and non-stochastic and is such that $\mathrm{E}[T\overline{\boldsymbol{w}_i^{\perp}}(\overline{\boldsymbol{w}_i^{\perp}})'|\mathcal{F}] - \boldsymbol{\Sigma}_{\boldsymbol{w}^{\perp}(\boldsymbol{w}^{\perp})',LR} = o_P(1)$ when $T \to \infty$.

These conditions control the degeneracy level of the conditional expectations of $\{\boldsymbol{w}_{i,t}\}$, such that the resulting convergence rate of FE estimator is restricted to: $\sqrt{N}$, $\sqrt{T}$, or $\sqrt{NT}$. In particular, following Davezies, D'Haultfœuille, and Guyonvarch (2021) and Menzel (2021) we label the setting where $\boldsymbol{v}_i = \boldsymbol{g}_t = \boldsymbol{0}_K$ wp1 as "degenerate", and otherwise as "non-degenerate". In part (d) we assume

[12]Subject to a very mild additional regularity condition that $\{\boldsymbol{g}_t\}$ and $\{\boldsymbol{w}_{i,t}^{+}\}$ are jointly $\mathcal{C}_i$-conditional α- mixing sequences.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

that all $\mathcal{F}$-measurable sequences are jointly stationary α-mixing sequences. This permits a rather straightforward asymptotic limit theory for time averages of the corresponding quantities.

In this assumption, we rule out the possibility of any "small variances" in the DGP, that is, we assume that the corresponding (long-run) variances of $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{g}_t\}$ are non-degenerate in the limit. As an alternative to this restrictive DGP, we could follow Menzel (2021) and consider a more general setup where the population covariance matrices of $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{g}_t\}$ are drifting sequences in N,T. For example, Assumption 4 (a) can be easily relaxed without major effects on the asymptotic distribution of the FE estimator. On the other hand, relaxing non-singularity of $\boldsymbol{\Sigma}_{\boldsymbol{gg}',LR}$ in part (c) is less straightforward, see the corresponding discussion in the Supplementary Material.

Finally, we illustrate the meaning of these restrictions for Examples 1 and 2.

Example 1 (Continued). For the sake of illustration, we further assume that $\mathrm{E}[\gamma_i\lambda_i]=0$, such that $\beta_0=\beta^*$. Assumption 4 (a) is satisfied immediately provided that we assume that the variance covariance matrices of $\{(\gamma_i,\lambda_i)'\}$ and $\{(f_t,\pi_t)'\}$ are non-singular. Part (c) is satisfied immediately as $g_t=0$ wp1, while part (d) is satisfied given (previously) imposed mixing assumptions on $\{(f_t,\pi_t)'\}$. Finally, part (e) needs to be verified/checked for a given DGP of $\{(f_t,\pi_t)'\}$ and $\{(u_{i,t},v_{i,t})'\}$. For example, if $\{(f_t,\pi_t)'\}$ is an uncorrelated sequence in t, this condition is immediately satisfied if $\{u_{i,t}v_{i,t}\}$ has a non-zero long-run variance.

Example 2 (Continued). For the sake of illustration, we further assume that $\mathrm{E}[\pi_t]=0$, such that $\beta_0=\beta^*$. Assumption 4 (a) is satisfied trivially as $v_i=0$ wp1. Part (b) is satisfied directly given the structure of $\{w_{i,t}^{+}\}$, while parts (c) and (d) are satisfied as long as $\{(\pi_t-\mathrm{E}[\pi_t])\}$ has a non-zero long-run variance and $\mathrm{E}[\gamma_i]\neq 0$ is a fixed constant. Finally, part (e) is satisfied as long as long-run variances of both $\{\pi_t\}$ and $\{(v_{i,t}u_{i,t}-\mathrm{E}[v_{i,t}u_{i,t}])\}$ are non-zero.

3.4. Asymptotic Distribution

Our next theorem characterizes the asymptotic distribution of the FE estimator. Before doing so, we define the following two quantities that will characterize the asymptotic bias of the FE estimator

$$\boldsymbol{b}_{N,T}^{(1)}=\frac{1}{NT}\sum_{i}\sum_{t}\sum_{s\neq t}\boldsymbol{x}_{i,t}^{++}\varepsilon_{i,s}, \tag{40}$$

$$\boldsymbol{b}_{N,T}^{(2)}=\frac{1}{NT}\sum_{t}\sum_{i}\sum_{j\neq i}\boldsymbol{x}_{i,t}^{+}\varepsilon_{j,t}^{+}. \tag{41}$$

THEOREM 2. *Under Assumptions 1-4 with proportional asymptotics $N,T\to\infty$ (jointly) and $N/T\to\kappa\in(0;\infty)$ the asymptotic distribution of the FE estimator can take the following two forms:*

i) *(Non-degenerate limit) when either* $\boldsymbol{\Sigma}_{\boldsymbol{gg}',LR}$ *and/or* $\boldsymbol{\Sigma}_{\boldsymbol{vv}'}$ *are non-singular matrices then*

$$\sqrt{\min(N,T)}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0+\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\left(\frac{1}{T}\boldsymbol{b}_{N,T}^{(1)}+\frac{1}{N}\boldsymbol{b}_{N,T}^{(2)}\right)\right)\xrightarrow{d}\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\xi}, \tag{42}$$

where $\boldsymbol{\xi}\sim N(\mathbf{0}_K,\min(1,\kappa^{-1})\boldsymbol{\Sigma}_{\boldsymbol{vv}'}+\min(1,\kappa)\boldsymbol{\Sigma}_{\boldsymbol{gg}',LR})$.

ii) *(Degenerate limit) otherwise*

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0+\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\left(\frac{1}{T}\boldsymbol{b}_{N,T}^{(1)}+\frac{1}{N}\boldsymbol{b}_{N,T}^{(2)}\right)\right)\xrightarrow{d}\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\zeta}, \tag{43}$$

with

$$\boldsymbol{\zeta}=^d\boldsymbol{\zeta}^{\perp}+\sum_{\ell=1}^{\infty}\boldsymbol{\gamma}_{\ell}\psi_{\infty}^{(\ell)}\phi_{\infty}^{(\ell)}. \tag{44}$$

Here, all random variables in $\boldsymbol{\zeta}$ *are jointly Gaussian, group-wise independent, and satisfy: a)* $\boldsymbol{\zeta}^{\perp}\sim N\left(\mathbf{0}_K,\boldsymbol{\Sigma}_{\boldsymbol{w}^{\perp}(\boldsymbol{w}^{\perp})',LR}\right)$*; b)* $\phi_{\infty}^{(\ell)}\sim N(0,1)$*; c)* $\mathrm{cov}\left(\phi_{\infty}^{(\ell)},\phi_{\infty}^{(\ell')}\right)=0$ *for all* $\ell\neq\ell$*; d)* $\psi_{\infty}^{(\ell)}\sim N\left(0,\sigma^2_{\psi^{(\ell)},LR}\right)$*; e)* $\mathrm{cov}\left(\psi_{\infty}^{(\ell)},\psi_{\infty}^{(\ell')}\right)=\sum_{s=-\infty}^{\infty}\mathrm{E}[\psi^{(\ell)}(\boldsymbol{d}_s)\psi^{(\ell')}(\boldsymbol{d}_0)]$.

For both i) and ii): $\boldsymbol{b}_{N,T}^{(1)}=T^{-1}\sum_t\sum_{s\neq t}\mathrm{E}[\boldsymbol{x}_{i,t}^{++}\varepsilon_{i,s}|\mathcal{F}]+o_P(1)=\mathcal{O}_P(1)+o_P(1)$, *and* $\boldsymbol{b}_{N,T}^{(2)}=\mathcal{O}_P(1)$.

Theorem 2 fully characterizes the leading variance and bias components of the FE estimator. In the non-degenerate (or "strong misspecification") case we find that the asymptotic distribution of the FE estimator is asymptotically normal, with a potentially slow $\sqrt{\min(N,T)}$ convergence rate. The asymptotic variance-covariance matrix, has a standard double-clustering structure as in Thompson (2011) and Chiang et al. (2024). On the other hand, for setting ii), the leading term of the asymptotic distribution can be non-normal (a product of independent multivariate normal random variables) if both $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{g}_t\}$ are zero vectors wp1, but $\boldsymbol{w}_{i,t}^{+}$ is non-negligible.

Part ii) of this theorem also covers all correctly specified linear models with (potentially) weakly-exogenous $\{\boldsymbol{x}_{i,t}\}$. In this respect, this result extends the results in Hahn and Moon (2006) and Chudik, Pesaran, and Yang (2018) to many types of correctly specified models with $\mathcal{D}_i$ measurable heteroscedasticity and weak-dependence structures in the idiosyncratic errors.

Our general setting is responsible for the presence of two additional components/bias terms that are not asymptotically negligible when the convergence rate is $\sqrt{NT}$. The first one is the serial-correlation, or the "Nickel bias" common to all dynamic models (irrespective of the degree of misspecification) estimated after the FE transformation. As it is evident from Theorem 2, this component can be present even if the additive error components structure of the model is correctly specified.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

However, there is clear difference between our formulation of the "Nickel bias", and the one usually used in the literature. Most papers formulate this bias unconditionally, that is by assuming that:[13]

$$\boldsymbol{b}_{N,T}^{(1)} = T^{-1}\sum_{t}\sum_{s\neq t}\mathrm{E}[\boldsymbol{x}_{i,t}^{++}\varepsilon_{i,s}] + o_P(1). \tag{45}$$

As a result, one can further show (see, e.g., Galvao and Kato, 2014):

$$\boldsymbol{b}_{N,T}^{(1)} = \sum_{k=-\infty}^{\infty}\mathrm{E}[\boldsymbol{x}_{i,1}^{++}\varepsilon_{i,1+k}] + o_P(1). \tag{46}$$

If the approximation in (45) is correct, then the overall "Nickell bias" can be consistently estimated, or it can be effectively removed using the Half Panel Jackknife (HPJ) bias-correction put forward by Dhaene and Jochmans (2015) (as later done in this article).

Unfortunately, we are not aware of any primitive conditions that ensure that the approximation in (45) is appropriate (at least when the limiting distribution is degenerate) for the setting with common (stochastic) shocks.[14] However, as illustrated in the Supplementary Material, this restriction can be checked as long as additional restrictions on the DGP (especially $\mathcal{F}$) are imposed. We leave this unsatisfactory result as it is, and in what follows implicitly assume that approximation in (46) holds.

The second bias term - $\boldsymbol{b}_{N,T}^{(2)}$ - is solely determined by misspecification of the additive error components structure in the model. In particular, by definition, $\varepsilon_{i,t}$ can be cross-sectionally correlated, causing this component to be non-negligible asymptotically. For example, from the definition of $\boldsymbol{b}_{N,T}^{(2)}$, it can be seen that $\boldsymbol{b}_{N,T}^{(2)} = \boldsymbol{0}_K$ wp1 when either: i) the additive error components model in $\boldsymbol{x}_{i,t}$ or $y_{i,t}$ is correctly specified; ii) $\mathcal{F}$ is a trivial σ-algebra.

Fortunately, it can be shown that $\boldsymbol{b}_{N,T}^{(2)}$ is asymptotically negligible under fairly mild additional regularity conditions.

PROPOSITION 1. *Let Assumptions 1-4 be satisfied, and let further assume that: i) Assumption 1(d) also holds for the stacked vector* $\{(\boldsymbol{z}_{1,t}', \ldots, \boldsymbol{z}_{N,t}')'\}_{t=1}^{T}$ *sequence conditionally* $\mathcal{C}^N = \vee_{i=1}^{N}\mathcal{C}_i$*; ii)* $\mathrm{E}[\boldsymbol{x}_{i,t}^{++}\varepsilon_{i,t}|\mathcal{C}_i] = \boldsymbol{0}_K$ *wp1 also implies that* $\mathrm{E}[\boldsymbol{x}_{i,t}^{++}\varepsilon_{j,t}|\mathcal{C}^N] = \boldsymbol{0}_K$ *for all (i,j), then:*

$$\boldsymbol{b}_{N,T}^{(2)} = \mathcal{O}_P(T^{-1/2}). \tag{47}$$

[13]For example, Chudik et al. (2018) consider a deterministic form of heteroscedasticity that generates a bias-term similar to that in $\boldsymbol{b}_{N,T}^{(1)}$. However, their provided high-level condition (19) eventually assumes away the possibility that heteroscedasticity might play a role in the limit.

[14]For example, the β- mixing condition on $\{z_{i,t}\}$, as e.g., used by Chiang et al. (2024), together with the Yoshihara's inequality is not sufficient to make the desired claim.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

In other words, misspecification-induced cross-sectional dependence is asymptotically negligible, once we appropriately extend the definition of "degeneracy" towards more general functions of the data.

Remark 2 (Comparison with Galvao and Kato, 2014). The results in Theorem 2 naturally extend the results in Galvao and Kato (2014), as for their setting $\boldsymbol{w}_{i,t}^{+} = \mathbf{0}_K$ and $\boldsymbol{g}_t = \mathbf{0}_K$ wp1, for all $i = 1, \ldots, N$, and $t = 1, \ldots, T$ by construction. Thus, the asymptotic distribution is always normal and the corresponding convergence rate is always in between $\sqrt{N}$ and $\sqrt{NT}$. Moreover, as in their setting $\mathcal{F}$ is a trivial σ-algebra it follows that (45) is satisfied wp1 and $\boldsymbol{b}_{N,T}^{(2)} = \mathbf{0}_K$ wp1 (by construction).

3.5. Bias-Correction

In many applications, T is typically smaller (or of similar magnitude) than N, and in such situations, inference procedures that neglect the "Nickel bias" $\boldsymbol{b}_{N,T}^{(1)}$ may be inaccurate. Thus, it is generally preferred to account for this bias term. For this purpose, we use the HPJ bias-correction approach put forward by Dhaene and Jochmans (2015).The HPJ estimator $\widetilde{\boldsymbol{\beta}}$ is defined as

$$\widetilde{\boldsymbol{\beta}} = 2\widehat{\boldsymbol{\beta}} - (\widehat{\boldsymbol{\beta}}_{S_1} + \widehat{\boldsymbol{\beta}}_{S_2})/2, \tag{48}$$

where $\widehat{\boldsymbol{\beta}}_{S_l}$ for $l = 1, 2$ are the corresponding FE estimators based on the half samples of $\{z_{i,t} : i = 1, \ldots, N; t \in S_l\}$. In particular, assuming that T is even, $S_1 = \{1, \ldots, T/2\}$ and $S_2 = \{T/2 + 1, \ldots, T\}$.

The HPJ approach was used (or suggested as a viable alternative to bias-correction techniques) in Fernández-Val and Weidner (2016); Galvao and Kato (2014); Okui and Yanagi (2019), and Juodis et al. (2021), among others. The setups and applications in the aforementioned papers differ substantially, highlighting the broad range of econometric problems where the Panel Jackknife approach is applicable.

Theorem 2 is readily available to derive the asymptotic expansion for the half-panel estimators. Building upon the available results in Chambers (2013) (for pure time-series components) and Galvao and Kato (2014)/Dhaene and Jochmans (2015) (for panel components), it can be shown that the half panel jackknife bias-correction has no variance effects under the maintained stationarity assumptions in Assumptions 1 and 4. This conclusion holds irrespective of the degree of model misspecification.

Despite the asymptotic equivalence between the FE and the HPJ-FE estimators, evidence from multiple known Monte Carlo studies indicate that variances of the FE and the HPJ-FE can differ substantially in finite samples (see, e.g., Galvao and Kato, 2014; Dhaene and Jochmans, 2015). This variance increase can be related to the higher-order properties of the HPJ procedure, see Hahn et al. (2024) for a comparative study of higher-order properties of different bias-correction procedures. In Section 4.4, we show how in practice this additional finite-sample variation can be accounted for.

4. AUTOREGRESSIVE DOUBLE ADAPTIVE WILD (ADAWILD) BOOTSTRAP

In what follows we introduce the AdaWild bootstrap inference procedure. As discussed previously in this paper, the asymptotic distribution of the FE estimator (around the pseudo-true value $\boldsymbol{\beta}_0$) is determined by the properties of the corresponding (infeasible) influence function

$$\boldsymbol{w}_{i,t} \equiv \boldsymbol{x}_{i,t}^{++}\varepsilon_{i,t}, \tag{49}$$

as previously defined in (15). For the sake of simplicity (and tractability of the corresponding derivations) in this section we assume that $\{\boldsymbol{w}_{i,t}\}$ is directly observed. While we believe that the procedure developed below is also applicable for feasible (estimated) influence function, a formal proof of this conjecture is well beyond the scope of this article and would require a separate paper on its own.[15]

Assumption 5. $\{\boldsymbol{w}_{i,t}\}_{i=1,t=1}^{N,T}$ is observed.

As a result, the proposed inference procedure is applicable to any (subject to regularity conditions similar to those in Assumptions 1–4) vector $\{\boldsymbol{w}_{i,t}\}$ of the form:

$$\boldsymbol{w}_{i,t} = \boldsymbol{w}_{i,t}^{\perp} + \boldsymbol{w}_{i,t}^{+} + \boldsymbol{\nu}_i + \boldsymbol{g}_t + \boldsymbol{\mu}, \tag{50}$$

where, as in Menzel (2021), $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{w}_{i,t}]$ is the main parameter of interest.

In a nutshell, our suggested procedure combines several types of Wild (or multiplier) bootstrap methods, and is inspired by the procedure (with model selection) of Menzel (2021) for the framework with double exchangeable arrays.[16] Unlike Menzel (2021), our proposed inference procedure is suitable for temporarily dependent (over t) and conditionally independent (over i) data as motivated by Assumption 1.

4.1. The Bootstrap DGP

Let $\widehat{\boldsymbol{\nu}}_i = \overline{\boldsymbol{w}}_i - \overline{\boldsymbol{w}}$, $\widehat{\boldsymbol{g}}_t = \overline{\boldsymbol{w}}_t - \overline{\boldsymbol{w}}$, and $\widehat{\boldsymbol{w}_{i,t}^{++}} = \boldsymbol{w}_{i,t} - \widehat{\boldsymbol{\nu}}_i - \widehat{\boldsymbol{g}}_t - \overline{\boldsymbol{w}}$ are the "plug-in" estimates for the components in (17). Let $\{\omega_{i,b}\}_{i=1}^{N}$ and $\{\omega_{t,b}\}_{t=1}^{T}$ be two mutually independent sequences of bootstrap weights (whose characteristics are discussed formally in Section 4.2) for any $b = 1,\ldots,B$ bootstrap iteration.

Our suggested bootstrap procedure is based on the bootstrap draws of the form

$$\boldsymbol{w}_{i,t,b}^{*} = \overline{\boldsymbol{w}} + \omega_{i,b}(\boldsymbol{d}_{\boldsymbol{\nu}} \odot \widehat{\boldsymbol{\nu}}_i) + \omega_{t,b}(\boldsymbol{d}_{\boldsymbol{g}} \odot \widehat{\boldsymbol{g}}_t) + \omega_{i,b}\omega_{t,b}\widehat{\boldsymbol{w}_{i,t}^{++}}, \tag{51}$$

for all (i,t) and $b = 1,\ldots,B$. Here, $\odot$ is the Hadamard (element-wise) product, while $\boldsymbol{d}_{\boldsymbol{\nu}}$ and $\boldsymbol{d}_{\boldsymbol{g}}$ are binary (model selection) indicators (formally defined in Section 4.2).

[15] The complexity of derivations can be appreciated by noting that even extensive derivations in Chiang et al. (2024) are not complete, as they do not account for the estimation uncertainty stemming from cross-sectional demeaning. Their derivations only account for time-series demeaning of the data.

[16] In Supplementary Material, we compare the procedure put forward in this section with the one proposed by Menzel (2021).

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

The proposed bootstrap DGP directly mimics the true decomposition in (17), except for the presence of binary model selector indicators, $\boldsymbol{d_v}$ and $\boldsymbol{d_g}$ in (51). The presence of model selection indicators makes the procedure "adaptive" to the degree of model misspecification (clustering), i.e., it is used to determine if the asymptotic distribution is better approximated by case *i*) or case *ii*) of Theorem 2.

Without adapting to the degree of model misspecification, the resulting procedure can be shown to be conservative when convergence rate is $\sqrt{NT}$. For further details we refer to the corresponding discussion in Menzel (2021).[17]

The resulting bootstrap statistic is just a sample average of the form

$$\overline{\boldsymbol{w}}_b^* = \frac{1}{NT}\sum_i \sum_t \boldsymbol{w}_{i,t,b}^*. \tag{52}$$

Given the bootstrap sample $\{\overline{\boldsymbol{w}}_b^*\}_{b=1}^B$, bootstrap based CIs for $\overline{\boldsymbol{w}}$ can be easily constructed, see Section 4.2.

4.2. Regularity Conditions

In what follows we first specify the DGPs for bootstrap weights $\{\omega_{i,b}\}_{i=1}^N$ and $\{\omega_{t,b}\}_{t=1}^T$. Given that in Assumption 1 cross-sectional and time-series dimensions are not treated symmetrically, the two sets of corresponding bootstrap weights are also generated differently. In particular, given the conditional independence assumption over the cross-sectional dimension (Assumption 1 (*a*)), the cross-sectional weights $\{\omega_{i,b}\}_{i=1}^N$ are drawn independently over i from $N(0,1)$.

Using the Autoregressive Wild Bootstrap (AWB) of Friedrich, Smeekes, and Urbain (2020) the time-series weights $\{\omega_{t,b}\}_{t=1}^T$ are generated recursively[18]

$$\omega_{t,b} = \gamma\omega_{t-1,b} + \sqrt{1-\gamma^2}\xi_{t,b}. \tag{53}$$

For some random initial condition $\omega_{0,b} = \xi_{0,b}$ and a sequence of independent innovations $\{\xi_{t,b}\}_{t=0}^T$. For primarily technical reasons, we suggest to draw all $\xi_{t,b}$ from a standard normal distribution with truncation at $[-M;M]$ (for some arbitrarily large value $0 < M < \Delta$). As the sequence $\{\omega_{t,b}\}_{t=0}^T$ is uniformly bounded and integrable, the corresponding proof of the bootstrap CLT is simplified considerably.[19]

Our next assumption (equivalent to that of Friedrich et al., 2020) imposes high-level conditions on the bootstrap autoregressive parameter γ.

[17] Intuitively the model selection step serves the same purpose as the negative variance adjustment term in the double-clustered estimators of Cameron et al. (2011); Thompson (2011), and more recently by Chiang et al. (2024).

[18] Overall, the AWB belongs to the general class of dependent wild bootstrap procedures covered in Doukhan et al. (2015).

[19] In practice this restriction plays no major role, as M can be set arbitrarily large (e.g., $M = 30$). As a result, in terms of actual bootstrap computations no real adjustments are generally needed. In contrast, Friedrich et al. (2020) suggested non-truncated weights $\xi_{t,b} \sim N(0,1)$, as their setting is purely time-series.

Assumption 6. The bootstrap parameter γ satisfies

$$\gamma = \theta^{1/\upsilon(T)} \tag{54}$$

such that for some $\theta \in (0;1)$

$$\upsilon(T) \to \infty, \quad T^{-1/2}\upsilon(T) \to 0. \tag{55}$$

The parameter γ can be seen as the corresponding "bandwidth" of AWB (when comparing with the Dependent Wild Bootstrap of Shao, 2010). In particular, Assumption 6 implies that $\gamma \to 1$, as $T \to \infty$. In applications we suggest setting $\gamma = 0.4$, as our preliminary (unreported) simulation results indicate that this value is more suitable for the typical dataset dimensions considered in applied work (as opposed to e.g., $\gamma = 0.2$ and $\gamma = 0.6$ also considered by Friedrich et al., 2020).

Finally, we discuss how the model selection indicators $\boldsymbol{d}_{\boldsymbol{g}}$ and $\boldsymbol{d}_{\boldsymbol{\nu}}$ can be constructed. Let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ be the sample variance-covariance matrix of $\widehat{\boldsymbol{w}_{i,t}^{++}}$, that is,

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \frac{1}{NT}\sum_{i}\sum_{t}\widehat{\boldsymbol{w}_{i,t}^{++}}\left(\widehat{\boldsymbol{w}_{i,t}^{++}}\right)'. \tag{56}$$

Then, the procedure of Menzel (2021) can be extended to the multivariate setting as follows:

$$\boldsymbol{d}_{\boldsymbol{g}} = \mathbb{I}\left(\operatorname{diag}\left(\frac{N}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{g}}_t\widehat{\boldsymbol{g}}_t' - \kappa_{\boldsymbol{g},N}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right) \geq \boldsymbol{0}\right), \quad \boldsymbol{d}_{\boldsymbol{\nu}} = \mathbb{I}\left(\operatorname{diag}\left(\frac{T}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\nu}}_i\widehat{\boldsymbol{\nu}}_i' - \kappa_{\boldsymbol{\nu},T}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\right) \geq \boldsymbol{0}\right). \tag{57}$$

Here, $\kappa_{\boldsymbol{g},N} \geq 0, \kappa_{\boldsymbol{\nu},T} \geq 0$ are some appropriately chosen non-decreasing sequences. Scaling by $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ ensures *scale* invariance of the corresponding model selection procedure.

Assumption 7. As $N, T \to \infty$ (jointly):

$$\kappa_{\boldsymbol{g},N} = o_P(N), \quad \kappa_{\boldsymbol{\nu},T} = o_P(T). \tag{58}$$

In general, any sequences satisfying Assumption 7 will consistently select the model that satisfies Assumption 4. However, for the procedure to be suitable for the settings with "small" covariance matrices, the choices of $\kappa_{\boldsymbol{g},N}$ and $\kappa_{\boldsymbol{\nu},T}$ have to be further refined. For example, small $\kappa_{\boldsymbol{g},N}$ ($\kappa_{\boldsymbol{\nu},T}$) favors a "more misspecified" model with slower convergence rates (and most likely wider CIs). On the other hand, large $\kappa_{\boldsymbol{g},N}$ ($\kappa_{\boldsymbol{\nu},T}$) favors a "less misspecified" model with a faster convergence rate and narrower CIs. Based on some preliminary simulation results we suggest the following values in the model selection step

$$\kappa_{\boldsymbol{g},N} = 0.5\ln(N), \quad \kappa_{\boldsymbol{\nu},T} = 0.5\ln(T). \tag{59}$$

The specific choices of $\kappa_{\boldsymbol{g},N}$ and $\kappa_{\boldsymbol{\nu},T}$, ensure that the additive effects $\{\boldsymbol{\nu}_i\}$ and $\{\boldsymbol{g}_t\}$ are allowed to have local-to-zero drifting variances, hence permitting certain deviations from Assumption 4.

4.3. Bootstrap Consistency

In this section, we prove consistency (in a sense clarified below) of the AdaWild procedure. We base our proof strategy upon Lemma 1 in Bücher and Kojadinovic (2019), which implies that the usual bootstrap convergence (conditional on the data) is implied by the unconditional joint convergence of the given statistic (in this case $\overline{\boldsymbol{w}}$) and two realizations of the bootstrap statistics, for example, $((\overline{\boldsymbol{w}}_1^*)', (\overline{\boldsymbol{w}}_2^*)')'$. This way, the characteristic function based proof strategy of Theorem 2, can be also used to prove bootstrap consistency.

For the proof strategy in Bücher and Kojadinovic (2019) to be applicable, we need to impose the following (mild) regularity condition on the sample paths of any two arbitrary copies of the time-series bootstrap weights, $\{\omega_{t,1}\}_{t=0}^T$ and $\{\omega_{t,2}\}_{t=0}^T$.

Assumption 8. For any two copies $\{\omega_{t,1}\}_{t=0}^T$ and $\{\omega_{t,2}\}_{t=0}^T$ of the bootstrap process (53) the joint process $\boldsymbol{w}_{i,t}^{\perp **} = (1, \omega_{t,1}, \omega_{t,2})' \otimes (\boldsymbol{w}_{i,t}^{\perp})$ satisfies

$$\mathrm{E}[T\overline{\boldsymbol{w}_i^{\perp **}}(\overline{\boldsymbol{w}_i^{\perp **}})'|\mathcal{F}^*] - \boldsymbol{\Sigma}_{\boldsymbol{w}^{\perp **}(\boldsymbol{w}^{\perp **})', LR} = o_P(1), \quad T \to \infty, \tag{60}$$

where $\overline{\boldsymbol{w}_i^{\perp **}} = T^{-1}\sum_t \boldsymbol{w}_{i,t}^{\perp **}$, $\mathcal{F}^* = \mathcal{F} \vee \sigma(\{\omega_{t,1}\}_{t=0}^T) \vee \sigma(\{\omega_{t,2}\}_{t=0}^T)$, and $\boldsymbol{\Sigma}_{\boldsymbol{w}^{\perp **}(\boldsymbol{w}^{\perp **})', LR} = \lim_{T\to\infty} \mathrm{E}[T\overline{\boldsymbol{w}_i^{\perp **}}(\overline{\boldsymbol{w}_i^{\perp **}})']$.

Assumption 8 generalizes Assumption 4 (e) to the extended vector $\{\boldsymbol{w}_{i,t}^{\perp **}\}$.

Before stating the main result of this section, lets us denote by $\mathbb{P}^*(\cdot)$ the bootstrap distribution function given realization of $\{\boldsymbol{w}_{i,t}\}$.

THEOREM 3. *Under Assumptions 1–8 with $(N,T) \to \infty$ (jointly)*

$$\sup_{\boldsymbol{x}} \left| \mathbb{P}^*\left(r_{N,T}(\overline{\boldsymbol{w}}_b^* - \overline{\boldsymbol{w}}) \le \boldsymbol{x}\right) - \mathbb{P}\left(r_{N,T}(\overline{\boldsymbol{w}} - \mathrm{E}[\boldsymbol{w}_{i,t}]) \le \boldsymbol{x}\right) \right| \xrightarrow{p} 0, \tag{61}$$

where $r_{N,T}$ is either:

i) $\sqrt{\min(N,T)}$ *in the case of a non-degenerate limit when either $\boldsymbol{\Sigma}_{\boldsymbol{gg}',LR}$ and/or $\boldsymbol{\Sigma}_{\boldsymbol{vv}'}$ are non-singular;*
ii) $\sqrt{NT}$ *in the case of a degenerate limit.*

This theorem confirms that the proposed bootstrap procedure replicates the asymptotic distribution of the sample average $\overline{\boldsymbol{w}}$ in the two cases described in Theorem 2. Hence, asymptotically valid inference can be conducted using the suggested bootstrap procedure.

4.4. Feasible Implementation of AdaWild

In this section, we discuss how bootstrap CIs can be constructed in practice for any linear combination $\boldsymbol{r} \in \mathbb{R}^K$ of the FE and HPJ-FE estimators. In particular, Algorithm 1 summarizes the procedure of the former estimator, while Algorithm 2 of the latter.[20]

For the HPJ-FE estimator, we suggest that for $\widetilde{\boldsymbol{\beta}}$ a modified version of the AdaWild bootstrap statistic is used

$$\widetilde{\boldsymbol{w}}_b^* = 2\overline{\boldsymbol{w}}_b^* - (\overline{\boldsymbol{w}}_{b,S_1}^* + \overline{\boldsymbol{w}}_{b,S_2}^*)/2. \tag{62}$$

Here, the half-sample averages $\overline{\boldsymbol{w}}_{b,S_1}^*$ and $\overline{\boldsymbol{w}}_{b,S_2}^*$ are generated as in (51) while keeping $\boldsymbol{d}_{\boldsymbol{g}}$ and $\boldsymbol{d}_{\boldsymbol{v}}$ (as well as $\{\omega_{i,b}\}_{i=1}^N$ and $\{\omega_{t,b}\}_{t=1}^T$) fixed in both sub-samples. This implementation maximizes the finite sample variation associated with time-series demeaning of $\{z_{i,t}\}$ in each sub-sample, while other sources of estimation uncertainty (i.e., $\boldsymbol{d}_{\boldsymbol{g}}$, $\boldsymbol{d}_{\boldsymbol{v}}$, and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$) are kept fixed.

All bootstrap CIs are constructed using the feasible counterpart of $\boldsymbol{w}_{i,t}$, that is,

$$\boldsymbol{w}_{i,t} \equiv \widehat{\boldsymbol{x}_{i,t}^{++}}(\widehat{y_{i,t}^{++}} - \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{x}_{i,t}^{++}}). \tag{63}$$

In Section 5, we investigate the finite sample properties of the corresponding feasible AdaWild bootstrap procedures.

Algorithm 1 Bootstrap CI algorithm for the FE estimator.

Given the FE estimator $\widehat{\boldsymbol{\beta}}$, construct observed/feasible influence functions $\{\boldsymbol{w}_{i,t}\}$ as $\boldsymbol{w}_{i,t} = \widehat{\boldsymbol{x}_{i,t}^{++}}\widehat{\varepsilon}_{i,t}$, where $\widehat{\varepsilon}_{i,t} = \widehat{y_{i,t}^{++}} - \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{x}_{i,t}^{++}}$.
Calculate $\boldsymbol{d}_{\boldsymbol{g}}$, and $\boldsymbol{d}_{\boldsymbol{v}}$ for given choices of $\kappa_{\boldsymbol{g},N}$ and $\kappa_{\boldsymbol{v},T}$.
for $b = 1,\ldots,B$ **do**
  Generate bootstrap weights $\{\omega_{i,b}\}_{i=1}^N$, $\{\omega_{t,b}\}_{t=1}^T$ for a given choice of γ;
  Construct bootstrap samples as in Eq.(51);
  Construct the corresponding bootstrap sample statistic $\overline{\boldsymbol{w}}_b^*$ as in Eq.(52);
  Construct the perturbed (bootstrap) FE estimator:

$$\widehat{\boldsymbol{\beta}}_b^* = \widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\overline{\boldsymbol{w}}_b^*.$$

end for
For any $\boldsymbol{r} \in \mathbb{R}^K$ construct $(1-\alpha)\%$ nominal coverage bootstrap intervals as:

$$I_{\alpha,\boldsymbol{r}'\boldsymbol{\beta}_0} = [\boldsymbol{r}'\widehat{\boldsymbol{\beta}} + \widehat{q}_{\alpha/2,\boldsymbol{r}}, \boldsymbol{r}'\widehat{\boldsymbol{\beta}} + \widehat{q}_{1-\alpha/2,\boldsymbol{r}}],$$

where

$$\widehat{q}_{\alpha,\boldsymbol{r}} = \inf\{u \in \mathbb{R} : \mathbb{P}^*[\boldsymbol{r}'(\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}}) \leq u] \geq \alpha\},$$

is the α- quantile of B bootstrap statistics $\boldsymbol{r}'(\widehat{\boldsymbol{\beta}}_b^* - \widehat{\boldsymbol{\beta}})$.

---

[20] For practical considerations and simplicity we consider non-pivotal percentile bootstrap CIs for this purpose. While pivotal quantities are generally preferred to their non-pivotal counterparts, the former involve explicit estimation of the long-run variances of $\{\boldsymbol{g}_t\}$ and $\{\boldsymbol{w}_{i,t}^{\perp}\}$. Moreover, we are not aware of any construction of pivotal quantities that are nuisance parameters free when the asymptotic distribution is degenerate as in Theorem 2.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

Algorithm 2 Bootstrap CI algorithm for the HPJ-FE estimator.

Given the three FE estimators ($\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\beta}}_{S_1}$, $\widehat{\boldsymbol{\beta}}_{S_2}$) construct three sets of feasible influence functions $\{\boldsymbol{w}_{i,t}\}$, $\{\boldsymbol{w}_{i,t}^{S_1}\}$, and $\{\boldsymbol{w}_{i,t}^{S_2}\}$ as in Step 1 of Algorithm 1.

Calculate $\boldsymbol{d}_{\boldsymbol{g}}$, and $\boldsymbol{d}_{\boldsymbol{v}}$ for a given choice of $\kappa_{\boldsymbol{g},N}$ and $\kappa_{\boldsymbol{v},T}$.

for $b = 1,\ldots,B$ **do**

Generate bootstrap weights $\{\omega_{i,b}\}_{i=1}^{N}$, $\{\omega_{t,b}\}_{t=1}^{T}$ for a given choice of γ;

Construct bootstrap samples as in Eq.(51) for every set of influence functions $\{\boldsymbol{w}_{i,t}\}$, $\{\boldsymbol{w}_{i,t}^{S_1}\}$, and $\{\boldsymbol{w}_{i,t}^{S_1}\}$ for a given set of bootstrap weights and $\boldsymbol{d}_{\boldsymbol{g}}$, and $\boldsymbol{d}_{\boldsymbol{v}}$;

Construct sample statistics $\overline{\boldsymbol{w}}_b^*$, $\overline{\boldsymbol{w}}_{b,S_1}^*$, and $\overline{\boldsymbol{w}}_{b,S_2}^*$ as in Eq.(52);

Using Eq.(62) define the perturbed bootstrap FE-HPJ estimator:

$$\widetilde{\boldsymbol{\beta}}_b^* = \widetilde{\boldsymbol{\beta}} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\widetilde{\boldsymbol{w}}_b^*.$$

end for

For any $\boldsymbol{r} \in \mathbb{R}^K$ construct $(1-\alpha)\%$ nominal coverage bootstrap intervals as:

$$I_{\alpha,\boldsymbol{r}'\boldsymbol{\beta}_0} = [\boldsymbol{r}'\widetilde{\boldsymbol{\beta}} + \widehat{q}_{\alpha/2,\boldsymbol{r}}, \boldsymbol{r}'\widetilde{\boldsymbol{\beta}} + \widehat{q}_{1-\alpha/2,\boldsymbol{r}}],$$

where

$$\widehat{q}_{\alpha,\boldsymbol{r}} = \inf\{u \in \mathbb{R} : \mathbb{P}^*[\boldsymbol{r}'(\widetilde{\boldsymbol{\beta}}_b^* - \widetilde{\boldsymbol{\beta}}) \leq u] \geq \alpha\},$$

is the α- quantile of B bootstrap statistics $\boldsymbol{r}'(\widetilde{\boldsymbol{\beta}}_b^* - \widetilde{\boldsymbol{\beta}})$.

5. SIMULATION STUDY

5.1. The Setup

In this section, we numerically illustrate some of the properties of the proposed AdaWild procedure. We base our Monte Carlo experiments on the DGP of Example 1:

$$y_{i,t} = \lambda_i f_t + v_{i,t}, \quad x_{i,t} = \gamma_i \pi_t + u_{i,t}. \tag{64}$$

Without loss of generality, we assume that all stochastic variables have zero mean. In order to account for all combinations of potential convergence rates of the FE estimator, the factors and the factor loadings are generated in the following way:

$$\gamma_i = \rho_\lambda \lambda_i + \sqrt{1-\rho_\lambda^2}\gamma_i^{\perp}, \quad \pi_t = \rho_f f_t + \sqrt{1-\rho_f^2}\pi_t^{\perp}. \tag{65}$$

Here, $\lambda_i \sim N(0,1)$, $\gamma_i^{\perp} \sim N(0,1)$, while $\{f_t\}$ and $\{\pi_t^{\perp}\}$ are two independent, zero-mean Gaussian AR(1) processes with a common AR parameter $\alpha_q = 0.6$ and a unit variance. This way, $\{(f_t,\pi_t)'\}$ are allowed to be serially and contemporaneously correlated.

The idiosyncratic component $\{u_{i,t}\}$ is a Gaussian AR(1) process with an autoregressive parameter α_u. In what follows we set $\alpha_u = \alpha_q/2 = 0.3$, to ensure that

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

Table 1. Monte Carlo design parameters

ρ_f	ρ_λ	Convergence rate	Asymptotic distribution	β_0	FE bias
0.0	0.0	$\sqrt{NT}$	Non-normal	0	0
0.5	0.0	$\sqrt{N}$	Normal	0	0
0.0	0.5	$\sqrt{T}$	Normal	0	0
0.5	0.5	$\sqrt{\min(N,T)}$	Normal	0.125	≈ -0.13

Note: Here, $\beta_0 = \rho_f \rho_\lambda / 2$. See example 1 and the Supplementary Material for further details on β_0 and b. Here, "FE Bias" is given by $-\Sigma_x^{-1} b$.

in at least one setup the bias term b is non-zero, see Supplementary Material for the corresponding derivations. Finally, for simplicity we assume that $\{v_{i,t}\}$ is i.i.d. Gaussian random variable. All initial conditions for dynamic processes are drawn from their corresponding stationary distributions.

The adequacy of our proposed inference procedure (with respect to the pseudo-true value β_0) is investigated by setting $\rho_f = \{0.0; 0.5\}$ and $\rho_\lambda = \{0.0; 0.5\}$. Table 1 provides a summary of the design parameters and their implications for the asymptotic distribution of the FE estimator.

For each specification, $M = 10000$ Monte Carlo replications are performed. We report empirical rejection frequencies for nominal level of 5% for sample sizes of $N = \{50, 100, 200, 500\}$ and $T = \{20, 50, 100\}$ using $B = 999$ bootstrap replications. We report the mean bias and the RMSE of the HPJ-FE estimator, as well as the average length of the bootstrap CIs. All aforementioned statistics are reported after scaling by the parametric rate $\sqrt{NT}$. We also report the mean values of model selection variables d_g and d_ν as defined in (57).

The Supplementary Material contains further Monte Carlo results for the multiple extensions of the current setup.

Below we separately discuss the simulation results for each of the designs from Table 1.

5.2. Results: Panel Convergence Rate $\sqrt{NT}$

The results for the setup with $\rho_f = 0.0$ and $\rho_\lambda = 0.0$ are summarized in Table 2. For this setup the asymptotic distribution of the FE estimator is non-normal, while the rate of convergence is $\sqrt{NT}$. Below we summarize the main findings.

Estimation. As can be expected, the estimator is almost completely free from any finite sample bias, even after $\sqrt{NT}$ scaling. The RMSE is mostly dominated by the variance component, which stabilizes for larger values of N and T at around ≈ 1.25.

Inference. The proposed AdaWild bootstrap procedure performs well, with empirical rejection frequencies close to the nominal level of 5% even for very small sample sizes. However, in most cases inference is mildly under-sized.

Table 2. Estimation and inference results for $\rho_f = 0.0$ and $\rho_\lambda = 0.0$

Design		Estimation		Inference		Model selection	
N	T	Bias	RMSE	Size	Length	$\#d_g = 1$	$\#d_\nu = 1$
50	20	0.0032	1.4403	0.0231	6.9462	0.0447	0.3105
50	50	−0.0048	1.3516	0.0345	5.7509	0.0336	0.1463
50	100	−0.0176	1.2486	0.0411	5.1949	0.0294	0.0938
100	20	−0.0049	1.4541	0.0207	6.7938	0.0179	0.2965
100	50	0.0014	1.2994	0.0299	5.6460	0.0129	0.1410
100	100	−0.0049	1.2504	0.0458	5.1264	0.0118	0.0937
200	20	−0.0228	1.4240	0.0190	6.6797	0.0079	0.2801
200	50	−0.0172	1.3162	0.0335	5.6186	0.0070	0.1380
200	100	−0.0065	1.2364	0.0402	5.1132	0.0060	0.0923
500	20	−0.0085	1.4293	0.0218	6.6327	0.0033	0.2748
500	50	−0.0138	1.2789	0.0297	5.6182	0.0021	0.1426
500	100	0.0277	1.2446	0.0397	5.1326	0.0029	0.0958

Note: Here, "Bias" is the mean bias of the scaled bias-corrected FE estimator. "RMSE" is the corresponding Root Mean Squared Error of the scaled bias-corrected FE estimator. "Size" is empirical rejection frequency for the null hypothesis $\beta_0 = \rho_f\rho_\lambda/2$. "Length" is the average length of scaled AdaWild confidence interval. $\#d_g = 1$ is the fraction of replications with $d_g = 1$. $\#d_\nu = 1$ is the fraction of replications with $d_\nu = 1$.

Comparing these results with the corresponding outcomes without bias-correction (in the Supplementary Material), we conclude that this is primarily driven by the adjustment of the CIs to account for the HPJ bias-correction. The length of the corresponding CIs narrows as N and T increase and stabilizes at around ≈ 5.

Model Selection. For all values of (N,T) our model selection procedure correctly selects the model without $\{g_t\}$ component, that is, without time-effects, in $\{w_{i,t}\}$. For $N > 200$ the fraction of simulations with $d_g = 1$ is already below 1%. On the other hand, as we only consider moderate values of T, the procedure over-selects the model with non-zero $\{\nu_i\}$ component. Performance of this procedure improves greatly as T increases.

5.3. Results: Cross-sectional Convergence Rate $\sqrt{N}$

The results for the setup with $\rho_f = 0.5$ and $\rho_\lambda = 0.0$ are summarized in Table 3. For this setup the asymptotic distribution of the FE estimator is normal, while the convergence rate is "cross-sectional" $\sqrt{N}$. Below we summarize the main findings.

Estimation. As in the previous setup with a faster convergence rate, the estimator is almost completely free from any finite sample bias, even after $\sqrt{NT}$ scaling. The RMSE is fully dominated by the variance, which increases in T. This is in line with the fact that the convergence rate is $\sqrt{N}$.

Table 3. Estimation and inference results for $\rho_f = 0.5$ and $\rho_\lambda = 0.0$

Design		Estimation		Inference		Model selection	
N	T	Bias	RMSE	Size	Length	$\#d_g = 1$	$\#d_\nu = 1$
50	20	0.0099	1.7875	0.0175	8.6154	0.0336	0.6725
50	50	0.0041	2.2388	0.0308	9.6167	0.0221	0.8154
50	100	0.0098	2.7754	0.0380	11.6631	0.0173	0.9570
100	20	−0.0047	1.8058	0.0156	8.6848	0.0144	0.6921
100	50	0.0116	2.1545	0.0260	9.6737	0.0080	0.8307
100	100	0.0090	2.7301	0.0298	11.8501	0.0066	0.9667
200	20	−0.0119	1.7710	0.0149	8.6200	0.0061	0.6941
200	50	0.0025	2.1906	0.0274	9.7524	0.0040	0.8411
200	100	0.0222	2.7687	0.0276	12.0014	0.0022	0.9740
500	20	−0.0088	1.7965	0.0138	8.6572	0.0020	0.6941
500	50	0.0245	2.2245	0.0248	9.8738	0.0015	0.8563
500	100	−0.0251	2.7574	0.0268	12.0135	0.0007	0.9742

Note: See Table 2 for an explanation.

Inference. Empirical rejection frequencies based on bootstrap CIs are close to the nominal level of 5% even for very small sample sizes. However, overall the procedure is rather conservative, as all numbers are (again) smaller than 5%. As expected, the length of the corresponding CIs widens at the rate $\sqrt{T}$.

Model Selection. Similarly to previous setup with $\sqrt{NT}$ convergence rate, for all values of (N,T) our model selection procedure selects the model without $\{g_t\}$. On the other hand, already for small values of T the procedure correctly includes $\{\nu_i\}$ in the bootstrap model for the majority ($> 67\%$) of the replications. The selection rate increases to $> 95\%$ as T increases.

5.4. Results: Time-series Convergence Rate $\sqrt{T}$

The results for the setup with $\rho_f = 0.0$ and $\rho_\lambda = 0.5$ are summarized in Table 4. For this setup the asymptotic distribution of the FE estimator is normal, while the convergence rate is a "time-series" $\sqrt{T}$. Below we summarize the main findings.

Estimation. The results almost perfectly mirror those in Table 3, except that the RMSE increases in N in this setup. Furthermore, the estimator exhibits a more visible finite sample bias for larger values of the N/T ratio, that the HPJ bias-correction is not able to fully account for.

Inference. For small values of T (especially with a large N/T ratio) the proposed inference procedure does not approximate well the finite sample distribution of the FE estimator. This is not surprising as the convergence rate is $\sqrt{T}$ only, and the wild autoregressive bootstrap procedure should properly capture the temporal

Table 4. Estimation and inference results for $\rho_f = 0.0$ and $\rho_\lambda = 0.5$

Design		Estimation		Inference		Model selection	
N	T	Bias	RMSE	Size	Length	$\#d_g = 1$	$\#d_\nu = 1$
50	20	−0.0445	3.3229	0.0770	11.5604	0.7116	0.2561
50	50	0.0009	3.1461	0.0690	11.2555	0.8370	0.1058
50	100	−0.0097	3.0048	0.0665	10.8595	0.8880	0.0629
100	20	0.0469	4.4759	0.0987	14.7254	0.8924	0.2411
100	50	−0.0557	4.3421	0.0811	14.8138	0.9730	0.1101
100	100	0.0263	4.1260	0.0786	14.2621	0.9900	0.0607
200	20	0.0785	6.1412	0.1014	19.5184	0.9850	0.2282
200	50	−0.0016	5.9081	0.0927	19.6530	0.9994	0.0984
200	100	0.0889	5.6209	0.0900	18.9698	0.9999	0.0599
500	20	0.0970	9.6957	0.1214	29.3018	0.9994	0.2234
500	50	−0.0759	9.1774	0.1049	29.5511	1.0000	0.0997
500	100	−0.0685	8.8175	0.1049	28.7799	1.0000	0.0588

Note: See Table 2 for an explanation.

dependence in $\{g_t\}$. As a result, as it is common for time-series models, in most cases rejection frequencies are larger than the nominal 5% level (i.e., the CIs undercover the true value). The situation somewhat improves as T increases for fixed value of N, as predicted by the implied $\sqrt{T}$ convergence rate.

Model Selection. Model selection rates for $\{\nu_i\}$ almost completely mirror those in Table 2, and will not be discussed any further. As for the $\{g_t\}$ component, our proposed procedure correctly selects the model with time-effects in $\{w_{i,t}\}$ for all combinations of (N,T) in at least 70% of all replications. This rate increases substantially once N increases (as predicted by the theory). Furthermore, one can observe that this rate also increases in T, suggesting the role of the convergence rate of the FE estimator is also non-negligible in this setup.

5.5. Results: Minimal Convergence Rate $\sqrt{\min(N,T)}$

The results for the setup with $\rho_f = 0.5$ and $\rho_\lambda = 0.5$ are summarized in Table 5. For this setup the asymptotic distribution of the FE estimator is normal, while the convergence rate is $\sqrt{\min(N,T)}$. This design is especially challenging as both $\{\nu_i\}$ and $\{g_t\}$ have a non-negligible effect for the asymptotic distribution of the FE estimator. Below we summarize the main findings.

Estimation. This design inherits some of the results discussed previously for $\sqrt{N}$ and $\sqrt{T}$ convergence rates. However, because of the non-negligible time-series bias b this design is more complex than a simple combination of previous results. In particular, despite the use of the HPH bias-correction, the finite sample bias is still visible and increases with N, and decreases with T.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

Table 5. Estimation and inference results for $\rho_f = 0.5$ and $\rho_\lambda = 0.5$

Design		Estimation		Inference		Model selection	
N	T	Bias	RMSE	Size	Length	$\#d_g = 1$	$\#d_\nu = 1$
50	20	−0.1880	3.2077	0.1117	11.8502	0.6512	0.6276
50	50	−0.0415	3.3836	0.0962	12.8764	0.7974	0.7766
50	100	−0.0494	3.6872	0.0882	14.2196	0.8395	0.9374
100	20	−0.1378	4.4053	0.1247	14.8170	0.8585	0.6408
100	50	−0.0911	4.3135	0.0985	15.7188	0.9618	0.7928
100	100	−0.0602	4.5177	0.0803	16.6853	0.9807	0.9497
200	20	−0.0840	5.9981	0.1260	19.3678	0.9779	0.6579
200	50	−0.0046	5.7628	0.1033	19.9699	0.9992	0.8103
200	100	−0.1202	5.7906	0.0895	20.4272	1.0000	0.9526
500	20	−0.2681	9.1182	0.1473	28.0073	0.9991	0.6604
500	50	−0.0056	8.8576	0.1183	28.9254	1.0000	0.8193
500	100	−0.1020	8.5635	0.1017	28.7945	1.0000	0.9601

Note: See Table 2 for an explanation.

Inference. Inferential results for this design are qualitatively comparable to those in the $\sqrt{T}$ design from Table 4. In particular, the results in Table 5 indicate that the time-series component $\{g_t\}$ plays an important role in the asymptotic distribution of the FE estimator. However, as predicted by a slow $\sqrt{\min(N,T)}$ convergence rate, bootstrap CIs (on average) are at least as wide for all combinations of (N,T) as those in the previous designs. However, the CIs are still relatively narrow or centered at the wrong point, as the corresponding rejection frequencies are in the range of $9\% - 15\%$. However, these numbers mark a major improvement as compared to the similar results without HPJ bias correction as summarized in the Supplementary Material.

Model Selection. Model selection rates for $\{g_t\}$ and $\{\nu_i\}$ mirror those in Tables 3 and 4, respectively. Overall, we observe that it is more challenging to correctly select the model for $\{\nu_i\}$, than for $\{g_t\}$. There are two main explanations for this pattern: first, the values of T we consider are smaller than those for N; second, serial correlation of the unobserved components has a sizeable effect for the construction of the selector variable d_ν (see also a corresponding remark in the Supplementary Material).

5.6. Summary

Overall, we find that Theorem 2 serves as a good approximation for the finite sample distribution of the FE estimator. However, our proposed AdaWild bootstrap procedure works best when the convergence rate is either $\sqrt{N}$ or $\sqrt{NT}$.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

As expected, the performance of the AdaWild procedure is less satisfactory with either $\sqrt{T}$ or $\sqrt{\min(N,T)}$ convergence rates. While in those cases the model selection procedure works effectively in selecting the correct model, the AWB procedure under-estimates sampling uncertainty. As indicated by additional Monte Carlo simulations summarized in the Supplementary Material, these over-rejection patterns are solely due to the presence of serially correlated regressors and factors. Hence, for the settings with mildly correlated factors empirical rejection rates are expected to be closer to the nominal rates.

6. EMPIRICAL ILLUSTRATION

In what follows we illustrate the AdaWild procedure on a real dataset. For this purpose, we employ the novel dataset of Voigtländer (2014), who analyzed the skill bias of technical change in U.S. manufacturing using U.S. input-output data, combined with the NBER Manufacturing Industries Database. The dataset spans $N = 313$ sectors of the US economy over the period of $1958 - 2005$ ($T = 48$).[21]

6.1. The Model

Following Yin, Liu, and Lin (2021), we investigate the potential causes of the historically increasing wage inequality between high-skilled and low-skilled workers in the U.S. manufacturing industries. We consider the simplified version of the empirical model of Yin et al. (2021)

$$\ln\left(\frac{w_{L.i,t}}{w_{H.i,t}}\right) = \alpha_i + \beta_1 \ln(\sigma_{i,t}) + \beta_2 \ln\left(\frac{H_{i,t}}{L_{i,t}}\right) + \eta_i + \lambda_t + \varepsilon_{i,t}, \tag{66}$$

where the variable $\frac{w_{L.i,t}}{w_{H.i,t}}$ is the relative wage of low-skilled workers to high skilled workers. The regressors of interest are: a) $\sigma_{i,t}$, the input skill intensity measure; b) $\frac{H_{i,t}}{L_{i,t}}$, the ratio of high and low skilled workers in the sector i. The input skill intensity measure $\sigma_{i,t}$ was originally constructed by Voigtländer (2014) using the weighted average of white-collar workers in other industries than i, with weights calculated using the Input-Output expenditures.

This dataset has been used by Yin et al. (2021) and Juodis (2022) in the context of models estimated by CCE estimators. Hence,we compare the results of the FE estimators[22] with those reported by Juodis (2022) for the CCE and the regularized CCE (rCCE) estimators.[23]

[21] The dataset is available at the data repository of the *Journal of Business and Economic Statistics*. We consider a balanced sub-sample as in Juodis (2022).

[22] Also used in Ciccone and Peri (2005) and Voigtländer (2014).

[23] Note that the results for the CCE and the rCCE estimators are not in any way adjusted for (potential) misspecification of the corresponding linear model.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

Table 6. Estimation results for the restricted empirical model of $y_{i,t} = \ln\left(w_{L.i,t}/w_{H.i,t}\right)$

	Estimator			
$\boldsymbol{x}_{i,t}$	CCE	rCCE	FE	HPJ-FE
$\ln(\sigma_{i,t})$	-0.63	-0.62	-0.71	-0.41
(Bootstrap)	$(-1.01; -0.22)$	$(-1.03; -0.21)$	$(-1.54; 0.10)$	$(-1.57; 0.75)$
(CCM)			$(-1.17; -0.24)$	$(-0.88; 0.05)$
(HAC)			$(-1.70; 0.28)$	$(-1.40; 0.57)$
(MW)			$(-1.21; -0.20)$	$(-0.92; 0.09)$
$\ln\left(H_{i,t}/L_{i,t}\right)$	0.40	0.36	0.23	0.15
(Bootstrap)	(0.33; 0.47)	(0.29; 0.41)	(0.17; 0.33)	(0.06; 0.28)
(CCM)			(0.18; 0.29)	(0.09; 0.21)
(HAC)			(0.14; 0.33)	(0.06; 0.24)
(MW)			(0.14; 0.32)	(0.06; 0.24)

Note: 95% CIs in the parentheses. HAC and MW are based on the Newey–West estimator with lag length $L = 4$. AdaWild procedure is implemented as described in Section 4.4. $\sigma_{i,t}$, the input skill intensity measure; $H_{i,t}/L_{i,t}$ the ratio of high and low skilled workers in the i-th sector.

6.2. Estimation Results

Estimation results are presented in Table 6. As advocated by Juodis (2022), for the CCE and the rCCE estimators we use the cross-sectional bootstrap based CIs. For the FE and the HPJ-FE estimator we report four types of CIs: i) bootstrap (AdaWild); ii) "CCM" of Arellano (1987); iii) "HAC" as in Driscoll and Kraay (1998); iv) "MW" similar to that in Thompson (2011) and Chiang et al. (2024). For more details regarding the implementation of ii)–iv), we refer to the Supplementary Material.

Point estimates. The FE point estimates differ somewhat from the CCE counterparts, as the estimated coefficient for $\ln(\sigma_{i,t})$ is smaller (but larger in absolute value), while the ones on $\ln\left(w_{L.i,t}/w_{H.i,t}\right)$ are also smaller in absolute values. HPJ correction substantially adjusts point estimates towards 0. Interestingly enough, these point estimates mostly fall outside the corresponding CIs for CCE procedures.

Confidence Intervals. The CIs for FE and HPJ-FE estimates of $\ln\left(w_{L.i,t}/w_{H.i,t}\right)$ do not vary much, irrespective of the method used. It is noticeable that the CCM approach generally results in tightest CIs. These CIs, however, can be grossly misleading for this regressor, as our model selection procedure indicates that $d_g = 1$ (i.e., $\{\widehat{\boldsymbol{g}}_t\}$ is included in the bootstrap sample). Not surprisingly, the remaining three procedures (AdaWild, HAC, MW) provide fairly comparable CIs, as they are all consistent in this setup.

https://doi.org/10.1017/S026646662510008X Published online by Cambridge University Press

The above conclusions do not hold for the CIs of $\ln(\sigma_{i,t})$. For this regressor, the CCM and the MW approaches result in comparable CIs while the CIs based on HAC CIs and AdaWild procedures are noticeably longer.

Overall, these results indicate that, depending on the degree of model misspecification, AdaWild based CIs can either closely resemble CIs based on normal approximation and clustered standard errors (either in one or multiple dimensions), or provide completely different results. This conclusion just confirms theoretical predictions of this article.

7. CONCLUDING REMARKS

In this article, we study the properties of the linear Two-Way FE estimator when the underlying DGP of observables is left almost completely unspecified. In particular, we do not restrict the way regressors, unit-level heterogeneity, and common shocks interact in the DGP. Conditional independence of all variables (once appropriately conditioned on common shocks) as well as time-series stationarity of the data are the only major restrictions imposed throughout the article.

We show that the FE estimator is consistent for a well-defined pseudo-true value under appropriate stationarity assumptions. Using a novel asymptotic approximation theory with $N, T \to \infty$ jointly, we prove that the asymptotic distribution can be normal as well as non-normal. The corresponding convergence rate can be as slow as $\sqrt{\min(N,T)}$ and as fast as $\sqrt{NT}$. We use two stylized examples to illustrate how these properties depend on the way the common shocks enter the model for observables.

We argue that uniform non-conservative inference is impossible in this setup, as no inference procedure can accommodate various degrees of convergence rates and different types of asymptotic distributions. As a solution to this problem we suggest the AdaWild bootstrap procedure that is point-wise consistent. The procedure is intuitive and easy to implement by practitioners. Our Monte Carlo results indicate that this procedure paired with the HPJ bias-correction well approximates the finite sample distribution of the FE estimator, as long as the time-series component in the influence function is not dominant. The performance of the suggested procedure deteriorates when the time-series component dominates the asymptotic distribution.

The main results in this article are derived under a few simplifying assumptions. As we mentioned in the article, some of these regularity conditions are imposed solely for technical reasons and are expected to be inconsequential for the main conclusion of this article. It remains as an open question in the literature whether there exists a sufficiently rich generalization of the Aldous–Hoover–Kallenberg decomposition for the temporarily dependent setup. Finally, it also remains as an open question under which (interpretable) conditions the "Nickell bias" is asymptotically non-stochastic. These are some questions we are currently working on in follow-up projects.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S026646662510008X.

REFERENCES

Alvarez, J., & Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121–1159.

Andrews, D. W. K. (2005). Cross-section regression with common shocks. *Econometrica*, 73, 1551–1585.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77, 1229–1279.

Berger, D., Easterly, W., Nunn, N., & Satyanath, S. (2013). Commercial imperialism? Political influence and trade during the cold war. *American Economic Review*, 103, 863–896.

Bücher, A., & Kojadinovic, I. (2019). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability*, 32, 1145–1165.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225, 200–230.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29, 238–249.

Chambers, M. J. (2013). Jackknife estimation of stationary autoregressive models. *Journal of Econometrics*, 172, 142–157.

Chen, L., Dolado, J. J., & Gonzalo, J. (2021a). Quantile factor models. *Econometrica*, 89, 875–910.

Chen, M., Fernández-Val, I., & Weidner, M. (2021b). Nonlinear factor models for network and panel data," *Journal of Econometrics*, 220, 296–324.

Chiang, H. D., Hansen, B. E., & Sasaki, Y. (2024). Standard Errors for Two-Way Clustering with Serially Correlated Time Effects, *The Review of Economics and Statistics*, 1–40.

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

Ciccone, A., & Peri, G. (2005). Long-run substitutability between more and less educated workers: evidence from U.S. States, 1950–1990. *The Review of Economics and Statistics*, 87, 652–663.

Davezies, L., D'Haultfœuille, X., & Guyonvarch, Y. (2021). Empirical process results for exchangeable arrays. *Annals of Statistics*, 49, 845–862.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110, 2964–2996.

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

Doukhan, P., Lang, G., Leucht, A., & Neumann, M. H. (2015). Dependent wild bootstrap for the empirical process. *Journal of Time Series Analysis*, 36, 290–314.

Driscoll, J. C., & Kraay, A. C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80, 549–560.

Fernández-Val, I., Freeman, H., & Weidner, M. (2021). Low-rank approximations of nonseparable panel models. *Econometrics Journal*, 24(2), C40–C77.

Fernández-Val, I., & Lee, J. (2013). Panel data models with nonadditive unobserved heterogeneity: estimation and inference. *Quantitative Economics*, 4, 453–481.

Fernández-Val, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics*, 192, 291–312.

Friedrich, M., Smeekes, S., & Urbain, J.-P. (2020). Autoregressive wild bootstrap inference for nonparametric trends. *Journal of Econometrics*, 214, 81–109.

Galvao, A. F., & Kato, K. (2014). Estimation and inference for linear panel data models under misspecification when both n and T are large. *Journal of Business & Economic Statistics*, 32, 285–309.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225, 254–277.

Hahn, J., D. W. Hughes, G. Kuersteiner, and W. K. Newey (2024). Efficient bias correction for cross-section and panel data. *Quantitative Economics*, 15, 783–816.

Hahn, J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4), 1639–1657.

Hahn, J., & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel data models with fixed effects. *Econometric Theory*, 27, 1152–1191.

Hahn, J., Kuersteiner, G., & Mazzocco, M. (2022). Joint time series and cross-section limit theory under mixingale assumptions. *Econometric Theory*, 38, 942–958.

Hahn, J., & Moon, H. R. (2006). Reducing bias of MLE in a dynamic panel model," *Econometric Theory*, 22, 499–512.

Hahn, J., & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4), 1295–1319.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19, 293–325.

Imai, K., & I. S. Kim (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29, 405–415.

Juodis, A. (2022). A regularization approach to common correlated effects estimation. *Journal of Applied Econometrics*, 37, 788–810.

Juodis, A., H. Karabiyik, & Westerlund, J. (2021). On the robustness of the pooled CCE estimator. *Journal of Econometrics*, 220, 325–348.

Juodis, A., & Reese, S. (2022). The incidental parameters problem in testing for remaining cross-section correlation. *Journal of Business and Economic Statistics*, 40, 1191–1203.

Juodis, A., & Sarafidis, V. (2022a). An incidental parameters free inference approach for panels with common factors. *Journal of Econometrics*, 229, 19–54.

Juodis, A., & Sarafidis, V. (2022b). A linear estimator for factor-augmented fixed-T panels with endogenous regressors. *Journal of Business and Economic Statistics*, 40, 1–15.

Kapetanios, G., Serlenga, L., & Shin, Y. (2024). An LM test for the conditional independence between regressors and factor loadings in panel data models with interactive effects. *Journal of Business and Economic Statistics*, 42, 743–761.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*, 89, 2143–2188.

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

Okui, R. (2014). Asymptotically unbiased estimation of autocovariances and autocorrelations with panel data in the presence of individual and time effects. *Journal of Time Series Econometrics*, 6, 129–181.

Okui, R. (2017). Misspecification in dynamic panel data models and model-free inferences. *Japanese Economic Review*, 68, Edition 3, 283–304.

Okui, R., & Yanagi, T. (2019). Panel data analysis with heterogeneous dynamics. *Journal of Econometrics*, 212, 451–475.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74, 967–1012.

Petrova, Y., & Westerlund, J. (2020). Fixed effects demeaning in the presence of interactive effects in treatment effects regressions and elsewhere. *Journal of Applied Econometrics*, 35, 960–964.

Phillips, P. C. B., & Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67, 1057–1111.

Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association*, 105, 218–235.

Su, L., Jin, S., & Zhang, Y. (2015). Specification test for panel data models with interactive fixed effects. *Journal of Econometrics*, 186, 222–244.
Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225, 175–199.
Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics*, 99, 1–10.
Voigtländer, N. (2014). Skill bias magnified: Intersectoral linkages and white-collar labor demand in U.S. manufacturing. *The Review of Economics and Statistics*, 96, 495–513.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.
Yin, S.-Y., Liu, C.-A., & Lin, C.-C. (2021). Focused information criterion and model averaging for large panels with a multifactor error structure. *Journal of Business & Economic Statistics*, 39, 54–68.