Federal Reserve Bank of Dallas
Globalization and Monetary Policy Institute
Working Paper No. 281
http://www.dallasfed.org/assets/documents/institute/wpapers/2016/0281.pdf

Half-Panel Jackknife Fixed Effects Estimation of Panels with Weakly Exogenous Regressors*

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
Department of Economice & USC Dornsife INET, University of Southern California, USA and Trinity College, Cambridge, UK

Jui-Chung Yang
USC Dornsife INET, University of Southern California

August 2016

Abstract

This paper considers estimation and inference in fixed effects (FE) panel regression models with lagged dependent variables and/or other weakly exogenous (or predetermined) regressors when N (the cross section dimension) is large relative to T (the time series dimension). The paper first derives a general formula for the bias of the FE estimator which is a generalization of the Nickell type bias derived in the literature for the pure dynamic panel data models. It shows that in the presence of weakly exogenous regressors, inference based on the FE estimator will result in size distortions unless N/T is sufficiently small. To deal with the bias and size distortion of FE estimator when N is large relative to T, the use of half-panel Jackknife FE estimator is proposed and its asymptotic distribution is derived. It is shown that the bias of the proposed estimator is of order T^{-2}, and for valid inference it is only required that $N/T^3 \to 0$, as $N,T \to \infty$ jointly. Extensions to panel data models with time effects (TE), for balanced as well as unbalanced panels, are also provided. The theoretical results are illustrated with Monte Carlo evidence. It is shown that the FE estimator can suffer from large size distortions when $N > T$, with the proposed estimator showing little size distortions. The use of half-panel jackknife FE-TE estimator is illustrated with two empirical applications from the literature.

JEL codes: C12, C13, C23

* Alexander Chudik, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75006. 214-922-5769. alexander.chudik@dal.frb.org. M. Hashem Pesaran, Department of Economics, University of Southern California, 3620 South Vermont Avenue, Kaprielian Hall 300, Los Angeles, CA 90089-0253. pesaran@usc.edu. Jui-Chung Yang, USC Dornslife Institute for New Economic Thinking, 3620 South Vermont Avenue, Kaprielian (KAP) Hall, 364C, Los Angeles, CA 90089-0253. 979-739-1484. juichuny@usc.edu. We are grateful to Geert Dhaene, Jinyong Hahn, Cheng Hsiao, Koen Jochmans, Ron Smith, Donggyu Sul, and Martin Weidner for helpful comments and suggestions.

1 Introduction

This paper considers the application of the split-panel jackknife method recently proposed by Dhaene and Jochmans (2015b) for panel data models with fixed effects (FE). It focusses on linear panel data models with lagged dependent variables and/or weakly exogenous regressors where N (the cross section dimension) is large relative to T (the time series dimension).[1] It is well known that standard FE estimators in such models suffer from small T bias and their use in inference can lead to large size distortions. The analysis of Dhaene and Jochmans (2015b) requires that $N/T \to \kappa$ for some $0 < \kappa < \infty$, as $N, T \to \infty$ jointly, which excludes the case of N large relative to T. In this paper we propose a bias-corrected jackknife FE estimator which only requires that $N/T^3 \to 0$ as $N, T \to \infty$ jointly, and is therefore appropriate in the case of many cross-country empirical applications in the literature where N is typically much larger than T. We derive exact expressions for small-T bias of the FE and half-panel jackknife bias-corrected FE estimators (as $N \to \infty$). After the split-panel jackknife procedure, the bias is reduced from $O\left(T^{-1}\right)$ to $O\left(T^{-2}\right)$, and a valid inference can be done by using a consistent estimator of the asymptotic variance which is also proposed. Our analysis is also sufficiently general and applies equally to dynamic panels as well as panels with weakly exogenous regressors with or without dynamics. Also, unlike analytical bias correction or bootstrap procedures developed only for the case of lagged dependent variable models, in which the model of the lagged dependent variable is already specified by default, the proposed estimator does not require the researcher to fully specify the nature of the weak exogeneity of the regressors, and as a result it is applicable to a wider class of models that generate correlation between errors and future values of the regressors. We allow for the weakly exogenous regressors to follow general linear stationary processes with possibly heterogeneous coefficients, and only require the correlation of the regressors and the future errors to decay exponentially. This is important since the alternative approach of modelling the dependent variable and the regressors jointly as vector autoregressions involves much stronger assumptions when T is small relative to N, such as homogeneity of coefficients and dynamics on the processes generating the weakly exogenous regressors. Our approach also allows for inclusion of strictly exogenous regressors with nonstationary or non-linear processes.

Following the seminal work of Nickell (1981), it is well known that the standard FE estimator suffers from small T bias in the case of dynamic panels. What is less recognized in the literature is that this small-T bias exists regardless of whether the lags of the dependent variables are included or not, so long as one or more of the regressors are weakly exogenous. Moreover, in such cases the inference based on standard FE estimators will be invalid and can result in large size distortions unless $N/T \to 0$, as $N, T \to \infty$ jointly.

There are a number of methods in the literature that can handle weakly exogenous regressors, the most prominent of which is the GMM (Generalized Method of Moments) procedure, developed

[1] Some authors prefer to refer to weakly exogenous regressors as predetermined. See, for example, Arellano and Bond (1991). We shall provide a formal definition in the context of panel data models in Section 2. In the econometric literature the concept of weak exogeneity was introduced by Engle et al. (1983) within a likelihood framework.

by Arellano and Bond (1991), Ahn and Schmidt (1995), Arellano and Bover (1995), and Blundell and Bond (1998), among others. The GMM approach is applicable when T is fixed, as $N \to \infty$. The condition of fixed T precludes cases where N and T tend to infinity jointly. It is also well known that the GMM method can suffer from weak/many instruments problem, particularly in cases where T takes moderate values. The jackknife bias correction considered in this paper requires $N, T \to \infty$, but it allows T to rise at a much slower rate than N, which makes the method attractive also for panels where the time dimension is small relative to N. A related class of estimators proposed in the literature is the one based on analytical and bootstrapped bias corrections. Analytical bias-corrections exploit on an asymptotic bias formula or its approximation and are considered by Bruno (2005), Bun (2003), Bun and Carree (2005 and 2006), Bun and Kiviet (2003), Hahn and Kuersteiner (2002), Hahn and Moon (2006), and Kiviet (1995 and 1999).[2] Bootstrap and simulation based bias corrections are considered, for example, by Everaert and Ponzi (2007) and Phillips and Sul (2003 and 2007). All these bias-correction methods have been developed for the case of models with lagged dependent variables only, in which the model of the lagged dependent variable is already specified by default. As a result, these methods do not readily extend to more general settings where one or more of the regressors (in addition to the lagged dependent variable) are weakly exogenous, without specifying a model for such regressors. Choi, Mark, and Sul (2010) consider common recursive mean adjustment to overcome the small-T bias. However, this approach requires $\left(\log^2 T\right)(N/T) \to 0$ for consistency which is not satisfied when N is larger than T. Han, Phillips, and Sul (2014) propose X-differencing for estimation of autoregressive panel data models regardless of the N/T ratio, but it is unclear if this approach can be generalized beyond the pure autoregressive panels. Last but not least, jackknife bias-correction can be used to tackle the consequences of weakly exogenous regressors. As noted earlier, Dhaene and Jochmans (2015b) consider a split-panel jackknife method, which is applicable without the need to specify the model for regressors, but the authors require $N/T \to \kappa$ for some $0 < \kappa < \infty$, as $N, T \to \infty$ jointly, which excludes the case of N large relative to T.[3] More recently, the modified profile likelihood method has also been applied to panel data models with lagged dependent variables, but it is assumed that the other included regressors are strictly exogenous. See Bartolucci et al. (2016) and Dhaene and Jochmans (2015a), for example.

In this paper we provide new results for the bias of FE estimators and extend the half-panel jackknife method studied by Dhaene and Jochmans (2015b) in a number of directions. First, we derive exact expressions for the small-T bias of the FE and half-panel jackknife bias-corrected FE estimators (as $N \to \infty$), which are respectively $O\left(T^{-1}\right)$ and $O\left(T^{-2}\right)$. We also provide a rigorous derivation of the asymptotic distribution of the proposed jackknife estimator and give a consistent estimator of its variance. Most importantly, we show that even if T is much smaller than N, so long

[2] Hahn and Newey (2004) propose analytical and jackknife bias correction procedures for nonlinear panel data models, assuming independently distributed data (both cross-sectionally and over time), that do not apply to dynamic panel data models.

[3] Perhaps it should be clarified that the requirement for N/T to converge to a constant in the paper by Dhaene and Jochmans (2015b) is due to the fact that these authors consider a general (possibly nonlinear) panel data models.

as $T = KN^{\epsilon}$, for some $0 < K < \infty$ and $\epsilon > 1/3$, and inference using the half-panel jackknife FE estimator will be valid so long as $N/T^3 \to 0$, as N and $T \to \infty$. As a result the jackknife estimator is applicable to panels with large N and moderate T sample sizes. We also consider panel data models with time effects (TE) and propose a FE-TE half-panel jackknife estimator, and extend the analysis of FE and FE-TE models to unbalanced panels. This latter extension is non-trivial and, at the same time, particularly important in empirical research where available panel data sets are generally unbalanced.

The drawbacks of the FE estimator and the satisfactory properties of the half-panel jackknife bias-corrected FE estimator are illustrated in a series of Monte Carlo experiments in cases where N is large relative to T in a number of different set ups, including models with weakly exogenous regressors, with and without lagged dependent variables, with and without time effects, and for balanced as well as unbalanced panels. Specifically we considered the following sample size combinations, $N = 30, 60, 100, 200, 500, 1000$, and $T = 30, 60, 100, 200$. In contrast to the FE and FE-TE estimators, the proposed half jackknife estimators perform well (in the mean square error sense, size and power) even if N is much larger than T, so long as N/T^3 is sufficiently small. This is to be contrasted with the FE and FE-TE estimators that perform well only if T is larger than N.[4]

The proposed FE-TE jackknife estimator is illustrated and compared to the FE-TE estimator in the case of two different empirical applications from the literature. The first application is a cross-country analysis by Berger et al. (2013) on the extent of US political influence on bilateral trades of US and foreign countries during the Cold War, and the second application considers the influential contribution of Donohue and Levitt (2001) on the relationship between legalized abortion and crime across the US states. In the case of Berger et al. (2013) study, the jackknife FE-TE estimates are in line with the FE-TE estimates and in fact suggest that the effect of CIA (Central Intelligence Agency) intervention on the US exports is even larger, with a higher level of statistical significance, than the estimates based on FE-TE. We also find some important differences in the case of the coefficients of other control variables. For example, after the jackknife bias correction, we find that having democracy, sharing a contiguous border, sharing a common language and participating the GATT (General Agreement on Tariffs and Trade) have positive and statistically significant effects on trades in all directions, while many of the estimates reported by Berger et al. (2013) for these coefficients are statistically insignificant. For the latter study, Donohue and Levitt (2001) found that legalized abortion in 1970s has been one of the main causes of the substantial decline in violent crime, property crime and murders observed in the US during 1990s. For this example, first, as a baseline model, we use our half-panel jackknife estimator to estimate the same model as Donohue and Levitt (2001) and second we allow for dynamics by adding the lagged crime

[4] We also considered using GMM estimators in our comparisons, but it soon became apparent that even for panels with moderate time dimensions we had to deal with many moment conditions, often larger than the number of available observations. For example, in the case of panels with two weakly exogenous regressors and $T = 30$, we have $T(T-1) = 870$ moment conditions, and as a result standard GMM procedures are likely to perform poorly, and must be augmented with some form of selection/shrinkage applied to the moment conditions. This is an area of ongoing research and will not be pursued in this paper, which focusses on the FE approach with T taking moderate values relative to N.

rate variable to the covariates. Our half-panel jackknife baseline estimates do not alter the main conclusion of Donohue and Levitt (2001). However, after allowing for dynamics, we find that the abortion variable continues to be statistically significant for violent and murder crimes but not for property crimes.

The remainder of the paper is organized as follows. The panel data model and its assumptions are set out in Section 2. Exact analytical expressions for small-T bias of the FE and the half-panel jackknife FE estimators are derived and their orders established in Section 3. This section also considers the problem of inference when N and $T \to \infty$ jointly, and provides a consistent estimator of the variance of the half-panel jackknife FE estimator. Section 4 considers extensions of the basic model to panel data models with fixed and time effects, as well as to unbalanced panels. Section 5 describes the Monte Carlo experiments and reports a summary of the main findings. Section 6 provides the empirical applications. Some concluding remarks are offered in Section 7. Proofs are relegated to the Appendix, and an online Supplement provides the full set of results for the Monte Carlo experiments that we have conducted.

Notations

K denotes a generic positive finite constant that does not depend on the sample size (N and T). K can take different values at different instances in the paper. $O\left(.\right)$ and $o\left(.\right)$ denote the Big O and Little o notations, respectively. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$ if there exists a positive finite constant K such that $\left|f_n\right|/g_n \leq K$ for all n. $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. $\overset{p}{\to}$ denotes convergence in probability.

2 Panel data model and its assumptions

We begin by considering the balanced fixed effects panel data model

$$y_{it} = \mu_i + \mathbf{z}_{it}'\boldsymbol{\alpha} + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it},\ i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{1}$$

where y_{it} is the dependent variable for the cross section unit i and time period t, μ_i is the unit-specific fixed effect, and $\mathbf{z}_{it}$ is a vector of strictly exogenous regressors, $\mathbf{x}_{it}$ is a $k \times 1$ vector of weakly exogenous stationary regressors, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown homogeneous slope coefficients, and u_{it} is the unit-specific error term error. We allow the strictly exogenous regressors to follow a general (linear or non-linear) process. They could include deterministic variables, such as seasonal dummies or time trends, and can even include unit root processes. But to simplify the analysis we assume that the weakly exogenous regressors, $\mathbf{x}_{it}$, have the following decomposition:

$$\mathbf{x}_{it} = \boldsymbol{\mu}_{ix} + \boldsymbol{\omega}_{it},\ i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{2}$$

where $\boldsymbol{\mu}_{ix}$ is a $k \times 1$ vector of fixed effects, and $\boldsymbol{\omega}_{it}$ is the $k \times 1$ vector of stochastic components, assumed to follow the general linear processes

$$\boldsymbol{\omega}_{it} = \sum_{s=0}^{\infty} \mathbf{A}_{is} \mathbf{v}_{i,t-s}, \tag{3}$$

in which $\{\mathbf{A}_{is},$ for $i = 1, 2, ..., N; s = 0, 1, ...\}$ are $k \times k$ matrices of coefficients, and $\mathbf{v}_{it}$, for $i = 1, 2, ..., N; t = 1, 2, ..., T$, are $k \times 1$ vectors of regressor-specific innovations.[5]

We adopt the following assumptions:

ASSUMPTION 1 (Idiosyncratic errors) *Errors u_{it}, for $i = 1, 2, ..., N; t = 1, 2, ..., T$ are independently distributed with zero means, $E(u_{it}) = 0$, possibly heteroskedastic variances, $E(u_{it}^2) = \sigma_{ui}^2 < K$, and uniformly bounded fourth moments, $E(u_{it}^4) < K$ for all i and t.*

ASSUMPTION 2 (Regressor innovations) *Innovations $\mathbf{v}_{it}$, for $i = 1, 2, ..., N; t = 1, 2, ..., T$ are $IID(\mathbf{0}_{k\times 1}, \mathbf{I}_k)$ with uniformly bounded fourth moments.*

ASSUMPTION 3 (Fixed effects) *Fixed effects, μ_i and $\boldsymbol{\mu}_{ix}$, are bounded such that $|\mu_i| < K$ and $\|\boldsymbol{\mu}_{ix}\| < K$, for all i if they are non-stochastic, and $E|\mu_i| < K$ and $E\|\boldsymbol{\mu}_{ix}\| < K$, if they are stochastic.*

ASSUMPTION 4 (Weak exogeneity conditions) *For all i and t*

(a)

$$E(\mathbf{v}_{i,t+h} u_{it}) = \begin{cases} \mathbf{0}_{k\times 1}, & \text{for } h \leq 0, \\ \boldsymbol{\gamma}_{iuv}(h), & \text{for } h > 0, \end{cases}, \tag{4}$$

where

$$\|\boldsymbol{\gamma}_{iuv}(h)\| < K\rho^h, \text{ for } h > 0, \tag{5}$$

and for some $0 < \rho < 1$, in which $\|\boldsymbol{\gamma}_{iuv}(h)\| = \sqrt{\boldsymbol{\gamma}'_{iuv}(h)\boldsymbol{\gamma}_{iuv}(h)}$ is the Euclidean norm of $\boldsymbol{\gamma}_{iuv}(h)$.

(b) *Conditions (4) and (5) hold and, in addition, $\mathbf{v}_{i,t+h}$ is independently distributed of u_{it} for all i, t and all $h \leq 0$.*

ASSUMPTION 5 (Regressors) *Consider the processes (2)-(3) for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, and let the $k \times k$ coefficient matrices, $\mathbf{A}_{is}$, satisfy*

$$\|\mathbf{A}_{is}\| < K\rho^s, \tag{6}$$

[5] More general processes for $\mathbf{x}_{it}$ can be entertained, but this will not be pursued in this paper.

for $i = 1, 2, ..., N,\ s = 0, 1, ...$ *where* $0 < \rho < 1$, *and* $\|\mathbf{A}_{is}\| = tr\left(\mathbf{A}_{is}'\mathbf{A}_{is}\right)$ *is the Frobenius norm of* $\mathbf{A}_{is}$. *Further, let* $\mathbf{\Gamma}_i(h)$ *denote the autocovariance matrix function of* $\boldsymbol{\omega}_{it}$, *given by*

$$\mathbf{\Gamma}_i(h) = \sum_{s=0}^{\infty} \mathbf{A}_{i,s+h}\mathbf{A}_{is}', \text{ for, } i = 1, 2, ..., N; \tag{7}$$

and set

$$\bar{\mathbf{\Gamma}}(h) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{\Gamma}_i(h), \text{ for } h = 0, 1, 2..., \tag{8}$$

where $\bar{\mathbf{\Gamma}}(0)$ *and* $\{\mathbf{\Gamma}_i(0), \text{for } i = 1, 2, ..., N\}$ *are* $k \times k$ *nonsingular matrices.*

Remark 1 *The panel data model (1)-(3) allows for a number of specifications, including the lagged dependent variable models, where one or more lags of the dependent variable feature among the* k *regressors in the vector* $\mathbf{x}_{it}$.

Remark 2 *Assumption 1 is standard and allows for heteroskedastic errors. It also rules out cross-sectionally dependent errors. Although a weak form of error cross-sectional dependence (as defined in Chudik et al., 2011) would not affect consistency of the FE estimator when* N *and* T *are large, it will affect the asymptotic variance and inference, see Pesaran and Tosetti (2011) for further discussion.*

Remark 3 *Apart from a minimal boundedness requirement, Assumption 3 imposes no other restrictions on the fixed effects,* μ_i *and* $\boldsymbol{\mu}_{ix}$. *Specifically, they are allowed to be cross-sectionally dependent as well as correlated with each other and the error terms,* u_{it} *and* $\mathbf{v}_{it}$.

Remark 4 *Assumptions 4-5 in addition to Assumptions 1-2 control the degree of serial correlation in regressors as well as the degree of dependence between the errors and the future values of the regressors. Both the serial correlation and the extent to which current errors are correlated with future regressors decay exponentially. The former is a direct consequence of serially uncorrelated* $\mathbf{v}_{it}$ *and the condition (6). The latter is established in the next proposition.*

Proposition 1 *(Weak exogeneity of* $\boldsymbol{\omega}_{it}$*) Suppose Assumptions 4.a and 5 hold. Then,*

$$E\left(\boldsymbol{\omega}_{i,t+h}u_{it}\right) = \begin{cases} \mathbf{0}_{k\times 1}, & \text{for } h \leq 0, \\ \boldsymbol{\gamma}_i(h), & \text{for } h > 0, \end{cases}, \tag{9}$$

and for all t, *where* $\boldsymbol{\omega}_{it}$ *is defined in (2), and* $\boldsymbol{\gamma}_i(h)$ *satisfies*

$$\|\boldsymbol{\gamma}_i(h)\| < K\rho^h, \text{ for } h = 1, 2, ..., \tag{10}$$

for some $0 < \rho < 1$.

The proof of this and other propositions are provided in the Appendix.

3 Small-T bias of FE estimator

In this sub-section we derive the small-T bias of the FE estimator of $\boldsymbol{\beta}$ when $\mathbf{x}_{it}$ is weakly exogenous, and follows the linear stationary processes as defined by the decomposition (2) and Assumption 5. But to simplify the derivations, and without loss of generality, we abstract form $\mathbf{z}_{it}$ in (11), and consider the following model

$$y_{it} = \mu_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it},\ i = 1, 2, ..., N; t = 1, 2, ..., T, \tag{11}$$

It is clear that a sub-set of $\mathbf{x}_{it}$ can be strictly exogenous, but the above assumptions on $\mathbf{x}_{it}$ require such strictly exogenous regressors to follow stationary linear processes, which could be restrictive in practice. However, it is easily seen that our analysis can also accommodate additional strictly exogenous regressors that follow possibly non-linear or non-stationary processes. As noted earlier they could also include deterministic processes. Inclusion of strictly exogenous regressors can affect the rate of convergence of the FE estimator of $\boldsymbol{\alpha}$, but not that of the FE estimator of $\boldsymbol{\beta}$. For example, adding a strictly exogenous regressor which is integrated of order 1(or $I(1)$) to model (11), yields the rate of convergence of $T\sqrt{N}$ for the FE estimator of α (the coefficient of the $I(1)$ regressor), but does not alter the standard $\sqrt{NT}$ convergence rate of the FE estimator of $\boldsymbol{\beta}$.

3.1 Bias of the FE estimator

The FE estimator of $\boldsymbol{\beta}$ in model (11) is given by

$$\hat{\boldsymbol{\beta}}_{FE}= \hat{\mathbf{Q}}_{FE}^{-1}\hat{\mathbf{q}}_{FE}, \tag{12}$$

where

$$\hat{\mathbf{Q}}_{FE} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})'}{NT},\ \hat{\mathbf{q}}_{FE} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})\, y_{it}}{NT}, \tag{13}$$

and $\bar{\mathbf{x}}_{i\cdot} = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$. In what follows, we assume that there exist $N_0, T_0 > 0$ such that the FE estimator $\hat{\boldsymbol{\beta}}_{FE}$ is well defined for all $N \geq N_0$ and all $T \geq T_0$, and the probability limit $p\lim_{N\to\infty}\hat{\boldsymbol{\beta}}_{FE}$ for any given $T \geq T_0$ is also well defined. To this end, we require the following invertibility conditions to hold.

ASSUMPTION 6 (Existence of FE estimators) *There exists $N_0, T_0 > 0$ such that for all $N > N_0$ and all $T \geq T_0$, $\hat{\mathbf{Q}}_{FE}$ defined by (13) is positive definite, and matrix $\bar{\mathbf{\Gamma}}(0) - T^{-1}\bar{\mathbf{\Psi}}_T$ is invertible, where $\bar{\mathbf{\Gamma}}(0)$ is defined by (8), and $\bar{\mathbf{\Psi}}_T$ is given by*

$$\bar{\mathbf{\Psi}}_T = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{\Psi}_{iT}, \tag{14}$$

with

$$\boldsymbol{\Psi}_{iT} = \boldsymbol{\Gamma}_i(0) + \sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\left[\boldsymbol{\Gamma}_i(h) + \boldsymbol{\Gamma}_i'(h)\right]. \tag{15}$$

Remark 5 *The smallest value of T for which Assumption 6 can be satisfied is 2, because it takes at least 2 time periods for $\boldsymbol{\beta}$ to be identified in the panel data model (11) without imposing any further restrictions on the fixed effects.*

Remark 6 *The invertibility of $\hat{\mathbf{Q}}_{FE}$ is required for $\hat{\boldsymbol{\beta}}_{FE}$ to be well defined, while the invertibility of $\left(\bar{\boldsymbol{\Gamma}}(0) - \frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)$ is required for $p\lim_{N\to\infty}\hat{\boldsymbol{\beta}}_{FE}$ to be well defined.*

We derive the small-T bias of the FE estimator next.

Proposition 2 *(Small-T bias of the FE estimator) Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ are generated by the panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), and Assumptions 1-3, 4.a, and 5-6 hold. Then for any fixed $T \geq T_0$, where T_0 is given by Assumption 6, we have*

$$Bias_T\left(\hat{\boldsymbol{\beta}}_{FE}\right) \equiv \lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) = -\frac{1}{T}\left(\bar{\boldsymbol{\Gamma}}(0) - \frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)^{-1}\bar{\boldsymbol{\chi}}_T, \tag{16}$$

where the FE estimator $\hat{\boldsymbol{\beta}}_{FE}$ is defined by (12), $\bar{\boldsymbol{\Gamma}}(0)$ and $\bar{\boldsymbol{\Psi}}_T$ are defined by (8) and (14), respectively,

$$\bar{\boldsymbol{\chi}}_T = \sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\bar{\boldsymbol{\gamma}}(h), \tag{17}$$

and

$$\bar{\boldsymbol{\gamma}}(h) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\gamma}_i(h). \tag{18}$$

In addition, $Bias_T\left(\hat{\boldsymbol{\beta}}_{FE}\right) = O\left(T^{-1}\right)$.

The result in the above proposition extends the well-known Nickell bias (Nickell, 1981) and covers both dynamic panel data models as well as static models with weakly exogenous regressors. In what follows we illustrate these features by means of two simple examples.

Example 1 *Consider the pure first-order autoregressive panel data model*

$$y_{it} = \mu_i + \beta x_{it} + u_{it},$$

where $x_{it} = y_{i,t-1}$. For this specification, (using the notations in (2)) we have

$$x_{it} = (1-\beta)^{-1}\mu_i + \omega_{it},$$

where $\omega_{it} = \beta\omega_{i,t-1} + u_{i,t-1}$*. It is now easily seen that*

$$\begin{aligned}\gamma_i(h) &= E\left(\omega_{i,t+h}u_{it}\right) = \sigma_{ui}^2\beta^{h-1}, \text{ for } h = 1, 2...\\ \gamma_i(h) &= E\left(\omega_{i,t+h}u_{it}\right) = 0, \text{ for } h = 0, -1, ...\end{aligned}$$

Using (7) and noting that $\mathbf{A}_{is} = \sigma_{ui}\beta^s$*, we also have*

$$\mathbf{\Gamma}_i\left(h\right) = \sum_{s=0}^{\infty} \mathbf{A}_{i,s+h}\mathbf{A}'_{is} = \frac{\sigma_{ui}^2\beta^h}{1-\beta^2}.$$

Hence, using the above results in (8), (14), (17) and (18) we obtain

$$\begin{aligned}\bar{\boldsymbol{\gamma}}\left(h\right) &= \bar{\sigma}_u^2\beta^{h-1}, \; \bar{\boldsymbol{\chi}}_T = \bar{\sigma}_u^2\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\beta^{h-1},\\ \bar{\mathbf{\Gamma}}\left(0\right) &= \frac{\bar{\sigma}_u^2}{1-\beta^2}, \text{ and } \bar{\mathbf{\Psi}}_T = \frac{\bar{\sigma}_u^2}{1-\beta^2} + \frac{2\bar{\sigma}_u^2}{1-\beta^2}\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\beta^h.\end{aligned}$$

where $\bar{\sigma}_u^2 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \sigma_{ui}^2$*. Now using the above results in (16) we have*

$$Bias_T\left(\hat{\beta}_{FE}\right) = \frac{-\left(1-\beta^2\right)f_T\left(\beta\right)}{T}\frac{1}{1-\frac{1}{T}-\frac{2}{T}\beta f_T\left(\beta\right)},$$

where

$$f_T\left(\beta\right) = \sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\beta^{h-1} = \frac{1}{1-\beta} - \frac{1}{T}\frac{1-\beta^T}{\left(1-\beta\right)^2}. \tag{19}$$

Also, after some algebra, it is easily seen that the expression for the bias can be written as

$$Bias_T\left(\hat{\beta}_{FE}\right) = -\left(\frac{1+\beta}{T}\right)\left\{\frac{1-\frac{1}{T}\left(\frac{1-\beta^T}{1-\beta}\right)}{1-\frac{1}{T}-\frac{2\beta}{T(1-\beta)}\left[1-\frac{1}{T}\frac{1-\beta^T}{1-\beta}\right]}\right\}, \tag{20}$$

which is the expression first derived by Nickell (1981) for the case of homogeneous error variances. But the above derivations show that the same expression for the bias obtains even if the error variances are heterogeneous.

Example 2 *Consider now the following static panel data model*

$$y_{it} = \mu_i + \beta x_{it} + u_{it}, \tag{21}$$

where

$$x_{it} = c_{ix} + \rho x_{i,t-1} + \kappa y_{i,t-1} + \varepsilon_{it}, \tag{22}$$

for $i = 1, 2, ..., N$ *and* $t = 1, 2, ..., T$*, and suppose* $|\varphi| = |\rho + \kappa\beta| < 1$*,* $u_{it} \sim IID\left(0, \sigma_{ui}^2\right)$*,* $\varepsilon_{it} \sim$

$IID\left(0,\sigma_{\varepsilon i}^2\right)$, and that u_{it} is independently distributed of $\varepsilon_{i't'}$ for any i, i' and any t, t'. This model is also a special case of (2), (3) and (11). Substituting (21) in (22) for $y_{i,t-1}$, we obtain

$$\begin{aligned} x_{it} &= (c_{ix}+\kappa\mu_i)+\varphi x_{i,t-1}+\varepsilon_{it}+\kappa u_{i,t-1}, \\ &= \mu_{ix}+\omega_{it}, \end{aligned}$$

where $\varphi=\rho+\kappa\beta$, $\mu_{ix}=(c_{ix}+\kappa\mu_i)/(1-\varphi)$, and

$$\omega_{it}=(1-\varphi L)^{-1}\varepsilon_{it}+(1-\varphi L)^{-1}\kappa u_{i,t-1}. \tag{23}$$

Hence, we obtain

$$\gamma_i(h)=E(\omega_{i,t+h}u_{it})=\kappa\varphi^{h-1}\sigma_{ui}^2, \text{ for } h=1,2,...,$$

and

$$\bar{\gamma}(h)=\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\gamma_i(h)=\kappa\varphi^{h-1}\bar{\sigma}_u^2, \text{ for } h=1,2,..., \tag{24}$$

where $\bar{\sigma}_u^2=\lim_{N\to\infty}N^{-1}\sum_{i=1}^N\sigma_{ui}^2$. Substituting (24) in (17), yields

$$\bar{\chi}_T=\kappa\bar{\sigma}_u^2 f_T(\varphi), \tag{25}$$

where

$$f_T(\varphi)=\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\varphi^{h-1}=\frac{1}{1-\varphi}-\frac{1}{T}\frac{\left(1-\varphi^T\right)}{(1-\varphi)^2}. \tag{26}$$

The remaining terms of the small-T bias formula (16) are

$$\bar{\Gamma}(h)=\frac{\varphi^h}{1-\varphi^2}\left(\bar{\sigma}_\varepsilon^2+\kappa^2\bar{\sigma}_u^2\right), \tag{27}$$

for $h=0,1,2,...,$ and

$$\bar{\Psi}_T=\frac{\bar{\sigma}_\varepsilon^2+\kappa^2\bar{\sigma}_u^2}{1-\varphi^2}\left[1+2\varphi f_T(\varphi)\right]. \tag{28}$$

Now using (25), (27) for $h=0$, and (28) in (16), the exact small-T bias of $\hat{\beta}_{FE}$ for this example is given by

$$Bias_T\left(\hat{\beta}_{FE}\right)=-\left(\frac{1}{T-1}\right)\left(\frac{\kappa\eta^2}{1+\kappa^2\eta^2}\right)\frac{\left(1-\varphi^2\right)f_T(\varphi)}{1-\frac{2\varphi}{T-1}f_T(\varphi)}, \tag{29}$$

where $\eta=\bar{\sigma}_u/\bar{\sigma}_\varepsilon$. Also recall that $\varphi=\rho+\kappa\beta$, and $f_T(\varphi)$ is given by (26). It is clear from (29) that there is no bias when regressors are strictly exogenous, namely when $\kappa=0$. Moreover, $Bias_T\left(\hat{\beta}_{FE}\right)\to 0$ as $\eta\to 0$. The exact bias function is plotted in Figure 1 for several choices of the parameter values. The sign of the bias depends on κ, since $f_T(\varphi)$ and $1-2\varphi f_T(\varphi)/(T-1)$ are both positive for $\varphi\in(-1,1)$. The shape of the bias function is nontrivial. Noting that $\bar{\Gamma}(0)$,

$\bar{\Psi}_T$, and $\bar{\chi}_T$ are all $O(1)$, the $O\left(T^{-1}\right)$ approximation to $Bias_T\left(\hat{\beta}_{FE}\right)$ is given by

$$Bias_T\left(\hat{\beta}_{FE}\right) = -\frac{1}{T}\left(1+\rho+\kappa\beta\right)\frac{\kappa\eta^2}{1+\kappa^2\eta^2} + O\left(T^{-2}\right).$$

The $O\left(T^{-1}\right)$ approximation of the bias increases with ρ, κ, and β and the variance ratio η^2, and it has a larger approximation error when φ is in the neighborhood of 1.

Remark 7 *In empirical research, such as the one on abortion and crime by Donohue and Levitt (2001) which we re-examine in some detail below (see Section 6.2), the investigators use lagged values of the regressors in the hope of avoiding the endogeneity problem. However, in the context of panel data models with fixed effects such a strategy does not solve the problem, and could even accentuate it. As an illustration suppose (y_{it}, x_{it}) are generated as in Example 2. Substituting (22) in (21), we have*

$$y_{it} = (\mu_i + \beta c_{ix}) + \theta x_{i,t-1} + \beta\kappa y_{i,t-1} + \beta\varepsilon_{it} + u_{it},$$

where $\theta = \beta\rho$, and, using (23) in the lagged (21),

$$y_{i,t-1} = (\mu_i + \beta\mu_{ix}) + \beta\omega_{i,t-1} + u_{i,t-1},$$

which can also be rewritten as,

$$y_{it} = \mu_i^* + \theta x_{i,t-1} + u_{it}^*, \tag{30}$$

where $\mu_i^* = \mu_i + \beta c_{ix} + \beta\kappa\left(\mu_i + \beta\mu_{ix}\right)$, and $u_{it}^* = \kappa\beta^2\omega_{i,t-1} + \beta\kappa u_{i,t-1} + \beta\varepsilon_{it} + u_{it}$. It is clear that u_{it}^* and $x_{i,t-1} = \mu_{ix} + \omega_{i,t-1}$ are uncorrelated only when the regressor is strictly exogenous ($\kappa = 0$) in which case the FE regression of y_{it} on $x_{i,t-1}$ consistently estimates the parameter $\theta = \beta\rho$. However, in the weakly exogenous case ($\kappa \neq 0$), we obtain

$$p\lim_{N\to\infty}\left(\hat{\theta}_{FE} - \theta\right) = \kappa\beta^2 + O\left(T^{-1}\right),$$

where $\kappa\beta^2$ is the bias from the correlation of $x_{i,t-1}$ and u_{it}^*, and the $O\left(T^{-1}\right)$ term is the weak exogeneity bias due to the correlation between the error term u_{it}^* and future regressors.

3.2 Half-panel jackknife FE estimator

Assume that T is even and consider the following half-panel jackknife FE estimator of $\boldsymbol{\beta}$,

$$\tilde{\boldsymbol{\beta}}_{FE} = 2\hat{\boldsymbol{\beta}}_{FE} - \frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{a,FE} + \hat{\boldsymbol{\beta}}_{b,FE}\right), \tag{31}$$

where $\hat{\boldsymbol{\beta}}_{FE}$ is the FE estimator defined in (12) using the full sample of T time periods, and $\hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$ are the FE estimators using the first $T/2$ and the last $T/2$ observations, respectively.

Specifically, the FE estimators $\hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$ are

$$\hat{\boldsymbol{\beta}}_{a,FE} = \hat{\mathbf{Q}}_{a,FE}^{-1}\hat{\mathbf{q}}_{a,FE}, \text{ and } \hat{\boldsymbol{\beta}}_{b,FE} = \hat{\mathbf{Q}}_{b,FE}^{-1}\hat{\mathbf{q}}_{b,FE}, \tag{32}$$

where

$$\hat{\mathbf{Q}}_{a,FE} = \sum_{i=1}^{N}\sum_{t=1}^{T/2}\frac{(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a})(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a})'}{NT/2},\ \hat{\mathbf{q}}_{a,FE} = \sum_{i=1}^{N}\sum_{t=1}^{T/2}\frac{(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a})\,y_{it}}{NT/2}, \tag{33}$$

$$\hat{\mathbf{Q}}_{b,FE} = \sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\frac{(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot b})(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot b})'}{NT/2},\ \hat{\mathbf{q}}_{b,FE} = \sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\frac{(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot b})\,y_{it}}{NT/2}, \tag{34}$$

and $\bar{\mathbf{x}}_{i\cdot a} = 2T^{-1}\sum_{t=1}^{T/2}\mathbf{x}_{it}$, and $\bar{\mathbf{x}}_{i\cdot b} = 2T^{-1}\sum_{t=T/2+1}^{T}\mathbf{x}_{it}$ are the temporal averages of regressors over the first and the second half sub-samples. In the following exposition, we assume that the sample is sufficiently large so that $\hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$ are well defined (similarly to Assumption 6 for the full sample FE estimator).

ASSUMPTION 7 (Existence of half-panel FE estimators) *There exists $N_0, T_0 > 0$ such that for all $N > N_0$ and all even $T \geq T_0$, the matrices $\hat{\mathbf{Q}}_{a,FE}$,$\hat{\mathbf{Q}}_{b,FE}$ and $\left(\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right)$ are invertible, where $\hat{\mathbf{Q}}_{a,FE}$ and $\hat{\mathbf{Q}}_{b,FE}$ are defined by (33) and (34), respectively, $\bar{\mathbf{\Gamma}}(0)$ is defined by (8), and $\bar{\mathbf{\Psi}}_{T/2}$ is defined by (14) with T replaced by $T/2$.*

Using the same arguments as in the derivation of the small-T bias of the full-sample FE estimator $\hat{\boldsymbol{\beta}}_{FE}$ in Proposition 2, we obtain (under Assumptions 1-3, 4.a, 5, and 7) the following small-T biases of the half-panel FE estimates $\hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$:

$$\lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{a,FE} - \boldsymbol{\beta}\right) = \lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{b,FE} - \boldsymbol{\beta}\right) = -\frac{2}{T}\left(\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right)^{-1}\bar{\boldsymbol{\chi}}_{T/2},$$

and hence

$$\lim_{N\to\infty}\frac{1}{2}\left[E\left(\hat{\boldsymbol{\beta}}_{a,FE} - \boldsymbol{\beta}\right) + E\left(\hat{\boldsymbol{\beta}}_{b,FE} - \boldsymbol{\beta}\right)\right] = -\frac{2}{T}\left(\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right)^{-1}\bar{\boldsymbol{\chi}}_{T/2}. \tag{35}$$

Using (16) and (35) we now obtain the following small-T bias of the jackknife FE estimator $\tilde{\boldsymbol{\beta}}_{FE}$,

$$\lim_{N\to\infty} E\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) = -\frac{2}{T}\left(\bar{\mathbf{\Gamma}}(0) - \frac{1}{T}\bar{\mathbf{\Psi}}_{T}\right)^{-1}\bar{\boldsymbol{\chi}}_{T} + \frac{2}{T}\left(\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right)^{-1}\bar{\boldsymbol{\chi}}_{T/2}. \tag{36}$$

The above expression depends, in a complicated manner, on the 'average degree of serial correlation' in $\mathbf{x}_{it}$, represented by $\bar{\mathbf{\Gamma}}(0)$, $\bar{\mathbf{\Psi}}_{T}$ and $\bar{\mathbf{\Psi}}_{T/2}$, and the 'average degree of correlation' between u_{it} and $\mathbf{x}_{it'}$, represented by $\bar{\boldsymbol{\chi}}_{T}$ and $\bar{\boldsymbol{\chi}}_{T/2}$. Consider the special case where $\mathbf{x}_{it}$ are (on average) serially uncorrelated so that $\bar{\boldsymbol{\Psi}}_T = \bar{\boldsymbol{\Psi}}_{T/2} = \mathbf{0}_{k\times k}$. Then

$$Bias_T\left(\tilde{\boldsymbol{\beta}}_{FE}\right) = -\frac{2}{T}\bar{\boldsymbol{\Gamma}}\left(0\right)^{-1}\left[\bar{\boldsymbol{\chi}}_T - \bar{\boldsymbol{\chi}}_{T/2}\right]. \tag{37}$$

But using (17),

$$\begin{aligned}
\bar{\boldsymbol{\chi}}_T - \bar{\boldsymbol{\chi}}_{T/2} &= \sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\bar{\boldsymbol{\gamma}}\left(h\right) - \sum_{h=1}^{T/2-1}\left(1-\frac{2h}{T}\right)\bar{\boldsymbol{\gamma}}\left(h\right)\\
&= \left[\left(1-\frac{1}{T}\right)\bar{\boldsymbol{\gamma}}\left(1\right) - \left(1-\frac{2}{T}\right)\bar{\boldsymbol{\gamma}}\left(1\right)\right]\\
&\quad + \left[\left(1-\frac{2}{T}\right)\bar{\boldsymbol{\gamma}}\left(2\right) - \left(1-\frac{4}{T}\right)\bar{\boldsymbol{\gamma}}\left(2\right)\right]\\
&\quad \vdots\\
&\quad + \left[\left(1-\frac{T/2-1}{T}\right)\bar{\boldsymbol{\gamma}}\left(T/2-1\right) - \left(1-\frac{2\left(T/2-1\right)}{T}\right)\bar{\boldsymbol{\gamma}}\left(T/2-1\right)\right]\\
&\quad + \sum_{h=T/2}^{T-1}\left(1-\frac{h}{T}\right)\bar{\boldsymbol{\gamma}}\left(h\right),
\end{aligned}$$

and after some simplifications we obtain

$$\begin{aligned}
\bar{\boldsymbol{\chi}}_T - \bar{\boldsymbol{\chi}}_{T/2} &= \left[\frac{1}{T}\bar{\boldsymbol{\gamma}}\left(1\right) + \frac{2}{T}\bar{\boldsymbol{\gamma}}\left(2\right) + ... + \frac{T/2-1}{T}\bar{\boldsymbol{\gamma}}\left(T/2-1\right)\right]\\
&\quad + \sum_{h=T/2}^{T-1}\left(1-\frac{h}{T}\right)\bar{\boldsymbol{\gamma}}\left(h\right).
\end{aligned} \tag{38}$$

Using Proposition 1, it readily follows from (38) that

$$\left\|\bar{\boldsymbol{\chi}}_T - \bar{\boldsymbol{\chi}}_{T/2}\right\| = O\left(\frac{1}{T}\right), \tag{39}$$

and therefore in view of (37) we have $Bias_T\left(\tilde{\boldsymbol{\beta}}_{FE}\right) = O\left(T^{-2}\right)$. In the more general case where the regressors, $\mathbf{x}_{it}$, are serially correlated the bias of the half-panel jackknife FE estimator is of the same order, $O\left(T^{-2}\right)$, as established in the following proposition.

Proposition 3 *(Small-T bias of the half-panel jackknife FE estimator) Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ is generated by the panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), and Assumptions 1-3, 4.a, and 5-7 hold. Then, for any fixed even $T \geq T_0$, where T_0 is chosen so that Assumptions 6 and 7 are satisfied, the small-T bias of the half-panel jackknife FE estimator $\tilde{\boldsymbol{\beta}}_{FE}$ defined in (31) is given by (36) and it is of order $O\left(T^{-2}\right)$.*

Example 3 *Using the set-up of Example 2, the exact small-T bias of* $\tilde{\beta}_{FE}$ *is given by*

$$Bias_T\left(\tilde{\beta}_{FE}\right) = -\frac{2}{T}\left(\bar{\Gamma}(0) - \frac{1}{T}\bar{\Psi}_T\right)^{-1}\bar{\chi}_T + \frac{2}{T}\left(\bar{\Gamma}(0) - \frac{2}{T}\bar{\Psi}_{T/2}\right)^{-1}\bar{\chi}_{T/2},$$

where $\bar{\chi}_{T/2}$, $\bar{\Gamma}(0)$, *and* $\bar{\Psi}_T$ *are given by (25),(27), and (28), respectively. After substituting these terms, the small-T bias of* $\tilde{\beta}_{FE}$ *is*

$$Bias_T\left(\tilde{\beta}_{FE}\right) = 2\frac{\left(1-\varphi^2\right)\kappa\eta^2}{1+\kappa^2\eta^2}\left[\frac{1}{T-2}\frac{f_{T/2}(\varphi)}{1-\frac{4\varphi}{T-2}f_{T/2}(\varphi)} - \frac{1}{T-1}\frac{f_T(\varphi)}{1-\frac{2\varphi}{T-1}f_T(\varphi)}\right],$$

where $f_T(\varphi)$ *is defined in (25), and* $\varphi = \rho + \kappa\beta$*. It is easily verified that this bias is of order* $O\left(T^{-2}\right)$ *by noting that* $f_{T/2}(\varphi) - f_T(\varphi) = O\left(T^{-1}\right)$*, and, therefore, the term in the square bracket is* $O\left(T^{-2}\right)$*. Figure 1 plots* $Bias_T\left(\tilde{\beta}_{FE}\right)$ *as a function of* κ *for* $T = 30$*, and* $\rho = 0.2$ *or* $0,5$ *and* $\eta = 1$ *or* 2*. As can be seen the magnitude of* $Bias_T\left(\tilde{\beta}_{FE}\right)$ *is very small for all admissible values of* κ*, except for those values of* κ*,*β *and* ρ *that result in* φ *close to unity.*

3.3 Asymptotic distribution of half-panel jackknife FE estimator

Suppose now that $N, T \to \infty$ jointly and note that

$$\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) = 2\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) - \frac{1}{2}\left[\left(\hat{\boldsymbol{\beta}}_{a,FE}-\boldsymbol{\beta}\right) + \left(\hat{\boldsymbol{\beta}}_{b,FE}-\boldsymbol{\beta}\right)\right],$$

where $\hat{\boldsymbol{\beta}}_{FE}, \hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$ are FE estimators for the full sample and the two half sub-samples - all obtainable from the general formula in (12). Let

$$\mathbf{z}_{FE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}\right)u_{it}, \tag{40}$$

$$\mathbf{z}_{a,FE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a}\right)u_{it}, \tag{41}$$

and

$$\mathbf{z}_{b,FE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b}\right)u_{it}, \tag{42}$$

and recall that $\hat{\mathbf{Q}}_{FE}$, $\hat{\mathbf{Q}}_{a,FE}$, and $\hat{\mathbf{Q}}_{b,FE}$ are defined by (13), (33) and (34), respectively. Then we have

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta} &= 2\hat{\mathbf{Q}}_{FE}^{-1}\mathbf{z}_{FE}-\frac{1}{2}\left(\hat{\mathbf{Q}}_{a,FE}^{-1}\mathbf{z}_{a,FE}+\hat{\mathbf{Q}}_{b,FE}^{-1}\mathbf{z}_{b,FE}\right)\\ &= \hat{\mathbf{Q}}_{FE}^{-1}\left(2\mathbf{z}_{FE}-\frac{1}{2}\mathbf{z}_{a,FE}-\frac{1}{2}\mathbf{z}_{b,FE}\right)\\ &\quad+\frac{1}{2}\left(\hat{\mathbf{Q}}_{FE}^{-1}-\hat{\mathbf{Q}}_{a,FE}^{-1}\right)\mathbf{z}_{a,FE}\\ &\quad+\frac{1}{2}\left(\hat{\mathbf{Q}}_{FE}^{-1}-\hat{\mathbf{Q}}_{b,FE}^{-1}\right)\mathbf{z}_{b,FE}. \end{aligned} \tag{43}$$

Consider the properties of $\hat{\mathbf{Q}}_{FE}-\hat{\mathbf{Q}}_{a,FE}$ when $N,T\to\infty$ jointly, and note that $\hat{\mathbf{Q}}_{FE}-\hat{\mathbf{Q}}_{a,FE}$ can be written as

$$\begin{aligned}\hat{\mathbf{Q}}_{FE}-\hat{\mathbf{Q}}_{a,FE} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\right)\mathbf{x}_{it}'-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a}\right)\mathbf{x}_{it}'\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left[\mathbf{x}_{it}\mathbf{x}_{it}'-\bar{\mathbf{x}}_{i\cdot}\mathbf{x}_{it}'-2\left(\mathbf{x}_{it}\mathbf{x}_{it}'-\bar{\mathbf{x}}_{i\cdot a}\mathbf{x}_{it}'\right)\right]\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\left(\mathbf{x}_{it}\mathbf{x}_{it}'-\bar{\mathbf{x}}_{i\cdot}\mathbf{x}_{it}'\right)\\ &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\mathbf{x}_{it}\mathbf{x}_{it}'-\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_{i\cdot}\bar{\mathbf{x}}_{i\cdot a}'/2+\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_{i\cdot a}\bar{\mathbf{x}}_{i\cdot a}'\\ &\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'-\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_{i\cdot}\bar{\mathbf{x}}_{i\cdot a}'/2\\ &= \frac{1}{2N}\sum_{i=1}^{N}\left(\frac{2}{T}\sum_{t=T/2+1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'-\frac{2}{T}\sum_{t=1}^{T/2}\mathbf{x}_{it}\mathbf{x}_{it}'\right)-\frac{1}{N}\sum_{i=1}^{N}\left(\bar{\mathbf{x}}_{i\cdot}\bar{\mathbf{x}}_{i\cdot a}'-\bar{\mathbf{x}}_{i\cdot a}\bar{\mathbf{x}}_{i\cdot a}'\right).\end{aligned}$$

But under Assumptions 2, 3 and 5,

$$\begin{aligned}&\frac{2}{T}\sum_{t=T/2+1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}'-\frac{2}{T}\sum_{t=1}^{T/2}\mathbf{x}_{it}\mathbf{x}_{it}'\xrightarrow{p}\mathbf{0}_{k\times k},\quad\text{and}\\ &\bar{\mathbf{x}}_{i\cdot}\bar{\mathbf{x}}_{i\cdot a}'-\bar{\mathbf{x}}_{i\cdot a}\bar{\mathbf{x}}_{i\cdot a}'\xrightarrow{p}\mathbf{0}_{k\times k},\end{aligned}$$

uniformly in i, as $T\to\infty$. Hence

$$\hat{\mathbf{Q}}_{FE}-\hat{\mathbf{Q}}_{a,FE}\xrightarrow{p}\mathbf{0}_{k\times k},$$

as $N,T\to\infty$ jointly, without any restrictions on the relative rates of N and T. Since also

$$p\lim_{N,T\to\infty}\hat{\mathbf{Q}}_{FE}=\mathbf{Q}=\bar{\mathbf{\Gamma}}\left(0\right), \tag{44}$$

is nonsingular, it follows that

$$\left(\hat{\mathbf{Q}}_{FE}^{-1}-\hat{\mathbf{Q}}_{a,FE}^{-1}\right)\xrightarrow{p}\mathbf{0}_{k\times k},\text{ as }N,T\to\infty\text{, jointly.}$$

Similarly, $\left(\hat{\mathbf{Q}}_{FE}^{-1}-\hat{\mathbf{Q}}_{b,FE}^{-1}\right)\xrightarrow{p}\mathbf{0}_{k\times k}$, as $N,T\to\infty$ jointly. Using these results in (43) and noting that $\sqrt{NT}\mathbf{z}_{a,FE}=O_p\left(1\right)$, and $\sqrt{NT}\mathbf{z}_{b,FE}=O_p\left(1\right)$, we obtain

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right)\overset{d}{\sim}\sqrt{NT}\mathbf{Q}^{-1}\left(2\mathbf{z}_{FE}-\frac{1}{2}\mathbf{z}_{a,FE}-\frac{1}{2}\mathbf{z}_{b,FE}\right),$$

as $N,T\to\infty$ jointly. Substituting back the expressions for $\mathbf{z}_{FE}$, $\mathbf{z}_{a,FE}$ and $\mathbf{z}_{b,FE}$ (given by (40), (41) and (42)), we have

$$\begin{aligned}\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right)\overset{d}{\sim}\quad&\frac{1}{\sqrt{NT}}\mathbf{Q}^{-1}\Bigg[2\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\right)u_{it}\\&-\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a}\right)u_{it}-\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot b}\right)u_{it}\Bigg].\end{aligned}$$

Let $\mathbf{d}_{ita}=\mathbf{x}_{it}-(2\bar{\mathbf{x}}_{i\cdot}-\bar{\mathbf{x}}_{i\cdot a})$, $\mathbf{d}_{itb}=\mathbf{x}_{it}-(2\bar{\mathbf{x}}_{i\cdot}-\bar{\mathbf{x}}_{i\cdot b})$, and

$$\mathbf{d}_{it}=I\left(t\leq T/2\right)\mathbf{d}_{ita}+I\left(t>T/2\right)\mathbf{d}_{itb}, \tag{45}$$

and note that $\sum_{t=1}^{T}\mathbf{d}_{it}=\mathbf{0}_{k\times 1}$. Hence,

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right)\overset{d}{\sim}\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}u_{it}.$$

Let $T=CN^{\epsilon}$ for some $0<C<\infty$ and $\epsilon>0$, and let $N\to\infty$. Since,

$$E\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}u_{it}\right)=-\sqrt{NT}\frac{2}{T}\left(\bar{\boldsymbol{\chi}}_{NT}-\bar{\boldsymbol{\chi}}_{N,T/2}\right),$$

where

$$\bar{\boldsymbol{\chi}}_{NT}=\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\bar{\boldsymbol{\gamma}}_{N}\left(h\right),\ \bar{\boldsymbol{\gamma}}_{N}\left(h\right)=N^{-1}\sum_{i=1}^{N}\boldsymbol{\gamma}_{i}\left(h\right),$$

and $\bar{\boldsymbol{\chi}}_{NT} - \bar{\boldsymbol{\chi}}_{N,T/2} = O\left(T^{-1}\right)$. Then

$$E\left[\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}u_{it}\right] = O\left(\frac{\sqrt{NT}}{T^2}\right) = O\left(N^{-\frac{3\epsilon-1}{2}}\right),$$

and $E\left[\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right)\right] \rightarrow \mathbf{0}_{k\times 1}$ when $\epsilon > 1/3$. Hence, for $T = CN^{\epsilon}$, with $\epsilon > 1/3$, and $N \rightarrow \infty$, we obtain

$$AsyVar\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE}\right) = \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}, \tag{46}$$

where

$$\mathbf{R} = \lim_{N,T\rightarrow\infty}\frac{1}{NT}E\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{d}_{it}\mathbf{d}'_{js}u_{it}u_{js}\right).$$

Although $\mathbf{d}_{it}$ is defined as a function of the regressors, it does not depend on the regressor fixed effects. Specifically, after using (2) in the definition of $\mathbf{d}_{it}$, we obtain $\mathbf{d}_{it} = \mathbf{b}_{it}$, where

$$\mathbf{b}_{it} = I\left(t \leq T/2\right)\mathbf{b}_{ita} + I\left(t > T/2\right)\mathbf{b}_{itb}, \tag{47}$$

$\mathbf{b}_{ita} = \boldsymbol{\omega}_{it} - \left(2\bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{i\cdot a}\right)$, $\mathbf{b}_{itb} = \boldsymbol{\omega}_{it} - \left(2\bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{i\cdot b}\right)$, and $\bar{\boldsymbol{\omega}}_{i\cdot}, \bar{\boldsymbol{\omega}}_{i\cdot a}$ and $\bar{\boldsymbol{\omega}}_{i\cdot b}$ are the full and the two sub-samples temporal averages defined in the same way as $\bar{\mathbf{x}}_{i\cdot}, \bar{\mathbf{x}}_{i\cdot a}$ and $\bar{\mathbf{x}}_{i\cdot b}$. Since $\{\mathbf{b}_{it}, u_{it}\}$ are cross-sectionally independent, then $E\left(\mathbf{b}_{it}\mathbf{b}'_{js}u_{it}u_{js}\right) = \mathbf{0}_{k\times k}$ for $i \neq j$, and

$$\mathbf{R} = \lim_{N,T\rightarrow\infty}\frac{1}{NT}E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{b}_{it}\mathbf{b}'_{is}u_{it}u_{is}\right), \tag{48}$$

which does not depend on the fixed effects. Hence, $\mathbf{R}$ can be estimated consistently by

$$\hat{\mathbf{R}}_{FE}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}\mathbf{d}'_{it}\hat{u}^2_{it,FE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\mathbf{h}}_{it,FE}\hat{\mathbf{h}}'_{it,FE}, \tag{49}$$

where $\hat{\mathbf{h}}_{it,FE} = \mathbf{d}_{it}\hat{u}_{it,FE}$ and $\hat{u}_{it,FE} = \left(y_{it} - \bar{y}_{i\cdot}\right) - \tilde{\boldsymbol{\beta}}_{FE}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}\right)$, in which $\bar{y}_{i\cdot}$ is the temporal average of y_{it} defined in the same way as $\bar{\mathbf{x}}_{i\cdot}$.

The next proposition establishes consistency of $\hat{\mathbf{R}}_{FE}$ and summarizes the earlier result on asymptotic distribution.

Proposition 4 *(Asymptotic distribution of half-panel jackknife FE estimator) Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ is generated by panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), Assumptions 1-3, 4.b, and 5-7 hold, and $N, T \rightarrow \infty$ jointly such that $T = CN^{\epsilon}$, for some $0 < C < \infty$ and $\epsilon > 1/3$. Then the asymptotic distribution of the half-panel jackknife FE estimator $\tilde{\boldsymbol{\beta}}_{FE}$ defined by (31) is given by*

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}\right) \overset{d}{\sim} N\left(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\right),$$

where $\mathbf{Q}$ *and* $\mathbf{R}$ *are defined by (44) and (48), respectively. A consistent estimator of the asymptotic variance of* $\tilde{\boldsymbol{\beta}}_{FE}$ *is given by*

$$\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE}\right) = \hat{\mathbf{Q}}_{FE}^{-1}\hat{\mathbf{R}}_{FE}\hat{\mathbf{Q}}_{FE}^{-1}, \tag{50}$$

where $\hat{\mathbf{Q}}_{FE}$ *and* $\hat{\mathbf{R}}_{FE}$ *are defined by (13) and (49), respectively.*

4 Extensions

4.1 Models with fixed and time effects

Consider now panel data models with fixed (μ_i) and time effects (δ_t)

$$y_{it} = \mu_i + \delta_t + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}. \tag{51}$$

The regressors are generated as before, but are now generalized to have time effects, $\boldsymbol{\delta}_{tx}$, of their own, namely

$$\mathbf{x}_{it} = \boldsymbol{\mu}_{ix} + \boldsymbol{\delta}_{tx} + \boldsymbol{\omega}_{it},\ \boldsymbol{\omega}_{it} = \sum_{s=0}^{\infty}\mathbf{A}_{is}\mathbf{v}_{i,t-s}, \tag{52}$$

The time effects, δ_t and $\boldsymbol{\delta}_{tx}$, are assumed to satisfy the following assumption.

ASSUMPTION 8 (Time effects) *Time effects,* δ_t *and* $\boldsymbol{\delta}_{tx}$*, for* $t = 1, 2, ..., T$*, can be non-stochastic or stochastic. In either case,* $E\left|\delta_t\right| < K$ *and* $E\left\|\boldsymbol{\delta}_{tx}\right\| < K$*.*

The fixed and time effect (FE-TE) estimators for the full sample and the half-samples are given by

$$\hat{\boldsymbol{\beta}}_{FE-TE} = \hat{\mathbf{Q}}_{FE-TE}^{-1}\hat{\mathbf{q}}_{FE-TE}, \tag{53}$$

$$\hat{\boldsymbol{\beta}}_{a,FE-TE} = \hat{\mathbf{Q}}_{a,FE-TE}^{-1}\hat{\mathbf{q}}_{a,FE-TE},\ \text{and}\ \hat{\boldsymbol{\beta}}_{b,FE-TE} = \hat{\mathbf{Q}}_{b,FE-TE}^{-1}\hat{\mathbf{q}}_{b,FE-TE}, \tag{54}$$

where

$$\hat{\mathbf{Q}}_{FE-TE} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}\right)'}{NT}, \tag{55}$$

$$\hat{\mathbf{Q}}_{a,FE-TE} = \sum_{i=1}^{N}\sum_{t=1}^{T/2}\frac{\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{a}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{a}\right)'}{NT/2}, \tag{56}$$

$$\hat{\mathbf{Q}}_{b,FE-TE} = \sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\frac{\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{b}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{b}\right)'}{NT/2}, \tag{57}$$

and

$$\hat{\mathbf{q}}_{FE-TE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}-\bar{\mathbf{x}}_{\cdot t}+\bar{\mathbf{x}}\right)y_{it},\ \hat{\mathbf{q}}_{a,FE-TE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot a}-\bar{\mathbf{x}}_{\cdot t}+\bar{\mathbf{x}}_{a}\right)y_{it},$$

$$\hat{\mathbf{q}}_{b,FE-TE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot b}-\bar{\mathbf{x}}_{\cdot t}+\bar{\mathbf{x}}_{b}\right)y_{it},$$

in which $\bar{\mathbf{x}}_{\cdot t} = N^{-1}\sum_{i=1}^{N}\mathbf{x}_{it}$ is the cross section average of regressors at a point in time t, $\bar{\mathbf{x}} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}$ is the overall (double) average of regressors for the full sample, and $\bar{\mathbf{x}}_{a} = 2\,(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_{b} = 2\,(NT)^{-1}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\mathbf{x}_{it}$ are the overall averages of regressors over the two sub-samples.

The FE-TE estimators are well defined only when $\hat{\mathbf{Q}}_{FE-TE}$, $\hat{\mathbf{Q}}_{a,FE-TE}$, and $\hat{\mathbf{Q}}_{b,FE-TE}$ are invertible. This is postulated in the following assumption, which corresponds to Assumptions 6 and 7.

ASSUMPTION 9 (Existence of FE-TE estimators) *There exists $N_0, T_0 > 0$ such that for all $N > N_0$ and all even $T \geq T_0$, $\hat{\mathbf{Q}}_{FE-TE}$, $\hat{\mathbf{Q}}_{a,FE-TE}$, and $\hat{\mathbf{Q}}_{b,FE-TE}$, defined by (55), (56), and (57), respectively, are positive definite, and matrices $\left[\bar{\mathbf{\Gamma}}(0) - \frac{1}{T}\bar{\mathbf{\Psi}}_T\right]$ and $\left[\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right]$ are invertible, where $\bar{\mathbf{\Gamma}}(0)$ and $\bar{\mathbf{\Psi}}_T$ are defined in Assumptions 5 and 6.*

The next proposition establishes that the small-T bias of the FE-TE estimator and its half-panel jackknife bias-corrected version is identical to the small-T biases obtained in the context of the panel data model without time effects considered earlier.

Proposition 5 *(Small-T bias of the FE-TE estimator and its half-panel jackknife version) Suppose y_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, is generated by the panel data model (51)-(52), and Assumptions 1-3, 4.a, 5, and 8-9 hold. Then, for any even $T \geq T_0$, where T_0 is given by Assumption 9, we have*

$$\lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{FE-TE} - \boldsymbol{\beta}\right) = -\frac{1}{T}\left[\bar{\mathbf{\Gamma}}(0) - \frac{1}{T}\bar{\mathbf{\Psi}}_T\right]^{-1}\bar{\boldsymbol{\chi}}_T = O\left(T^{-1}\right), \tag{58}$$

and

$$\lim_{N\to\infty} E\left(\tilde{\boldsymbol{\beta}}_{FE-TE} - \boldsymbol{\beta}\right) = -\frac{2}{T}\left[\bar{\mathbf{\Gamma}}(0) - \frac{1}{T}\bar{\mathbf{\Psi}}_T\right]^{-1}\bar{\boldsymbol{\chi}}_T + \frac{2}{T}\left[\bar{\mathbf{\Gamma}}(0) - \frac{2}{T}\bar{\mathbf{\Psi}}_{T/2}\right]^{-1}\bar{\boldsymbol{\chi}}_{T/2} = O\left(T^{-2}\right), \tag{59}$$

where $\hat{\boldsymbol{\beta}}_{FE-TE}$ is defined in (53), its half-panel jackknife bias-corrected version $\tilde{\boldsymbol{\beta}}_{FE-TE}$ is given by

$$\tilde{\boldsymbol{\beta}}_{FE-TE} = 2\hat{\boldsymbol{\beta}}_{FE-TE} - \frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{a,FE-TE} + \hat{\boldsymbol{\beta}}_{b,FE-TE}\right), \tag{60}$$

$\hat{\boldsymbol{\beta}}_{a,FE-TE}$ and $\hat{\boldsymbol{\beta}}_{b,FE-TE}$ are defined in (54), $\bar{\mathbf{\Gamma}}(0)$ is defined in Assumption 5, $\bar{\mathbf{\Psi}}_T$ is defined in (14), and $\bar{\boldsymbol{\chi}}_T$ is defined by (17). $\bar{\mathbf{\Psi}}_{T/2}$ and $\bar{\boldsymbol{\chi}}_{T/2}$ are obtained from the expressions for $\bar{\mathbf{\Psi}}_T$ and $\bar{\boldsymbol{\chi}}_T$ by replacing T with $T/2$.

Let

$$\mathbf{z}_{FE-TE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}\right)u_{it}, \tag{61}$$

$$\mathbf{z}_{a,FE-TE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{a}\right)u_{it},\ \mathbf{z}_{b,FE-TE} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}_{b}\right)u_{it}.$$

Then similar to (43), the half-panel jackknife bias-corrected FE-TE estimator $\tilde{\boldsymbol{\beta}}_{FE-TE}$ can be written as

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_{FE-TE} - \boldsymbol{\beta} &= \hat{\mathbf{Q}}_{FE-TE}^{-1}\left(2\mathbf{z}_{FE-TE} - \frac{1}{2}\mathbf{z}_{a,FE-TE} - \frac{1}{2}\mathbf{z}_{b,FE-TE}\right) \\ &\quad + \frac{1}{2}\left(\hat{\mathbf{Q}}_{FE-TE}^{-1} - \hat{\mathbf{Q}}_{a,FE-TE}^{-1}\right)\mathbf{z}_{a,FE-TE} + \frac{1}{2}\left(\hat{\mathbf{Q}}_{FE-TE}^{-1} - \hat{\mathbf{Q}}_{b,FE-TE}^{-1}\right)\mathbf{z}_{b,FE-TE}.\end{aligned} \tag{62}$$

Using results (A.1) and (A.2) of Lemma 1 in the Appendix, and noting that

$$\hat{\mathbf{Q}}_{FE-TE} \xrightarrow{p} \mathbf{Q} = \bar{\boldsymbol{\Gamma}}(0), \text{ as } N, T \to \infty \text{ jointly,}$$

where $\bar{\boldsymbol{\Gamma}}(0)$ is invertible under Assumption 5, we obtain

$$\left(\hat{\mathbf{Q}}_{FE-TE}^{-1} - \hat{\mathbf{Q}}_{a,FE-TE}^{-1}\right) \xrightarrow{p} \mathbf{0}_{k\times k} \text{ and } \left(\hat{\mathbf{Q}}_{FE-TE}^{-1} - \hat{\mathbf{Q}}_{b,FE-TE}^{-1}\right) \xrightarrow{p} \mathbf{0}_{k\times k},$$

as $N, T \to \infty$ jointly. Using these results in (62) and noting that (as in the FE case) $\sqrt{NT}\mathbf{z}_{a,FE-TE} = O_p(1)$ and $\sqrt{NT}\mathbf{z}_{b,FE-TE} = O_p(1)$, then we have

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE-TE} - \boldsymbol{\beta}\right) \overset{d}{\sim} \sqrt{NT}\mathbf{Q}^{-1}\left(2\mathbf{z}_{FE-TE} - \frac{1}{2}\mathbf{z}_{a,FE-TE} - \frac{1}{2}\mathbf{z}_{b,FE-TE}\right),$$

as $N, T \to \infty$ jointly. Substituting the expressions for $\mathbf{z}_{FE-TE}$, $\mathbf{z}_{a,FE-TE}$, and $\mathbf{z}_{b,FE-TE}$, we obtain

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE-TE} - \boldsymbol{\beta}\right) \overset{d}{\sim} \mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}^{*}u_{it}, \tag{63}$$

where $\mathbf{d}_{it}^{*} = I(t \leq T/2)\,\mathbf{d}_{ita}^{*} + I(t > T/2)\,\mathbf{d}_{itb}^{*}$, and

$$\mathbf{d}_{ita}^{*} = \mathbf{x}_{it} - (2\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{i\cdot a}) - \bar{\mathbf{x}}_{\cdot t} + (2\bar{\mathbf{x}} - \bar{\mathbf{x}}_{a}),\ \mathbf{d}_{itb}^{*} = \mathbf{x}_{it} - (2\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{i\cdot b}) - \bar{\mathbf{x}}_{\cdot t} + (2\bar{\mathbf{x}} - \bar{\mathbf{x}}_{b}).$$

Consequently, we propose the following estimator of the asymptotic variance of $\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE-TE}$:

$$\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE-TE}\right) = \hat{\mathbf{Q}}_{FE-TE}^{-1}\hat{\mathbf{R}}_{FE-TE}\hat{\mathbf{Q}}_{FE-TE}^{-1}, \tag{64}$$

where $\hat{\mathbf{Q}}_{FE-TE}$ is given by (55),

$$\hat{\mathbf{R}}_{FE-TE}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}^{*}\mathbf{d}_{it}^{*\prime}\hat{u}_{it,FE-TE}^{2}, \tag{65}$$

and

$$\hat{u}_{it,FE-TE}=\left(y_{it}-\bar{y}_{i\cdot}-\bar{y}_{\cdot t}-\bar{y}\right)-\tilde{\boldsymbol{\beta}}_{FE-TE}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}-\bar{\mathbf{x}}_{\cdot t}+\bar{\mathbf{x}}\right),$$

in which $\bar{y}_{\cdot t}$ and $\bar{y}$ are defined in the same way as $\bar{\mathbf{x}}_{\cdot t}$ and $\bar{\mathbf{x}}$.

The following proposition establishes sufficient conditions for the consistency of $\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE-TE}\right)$ and for the asymptotic unbiasedness of $\tilde{\boldsymbol{\beta}}_{FE-TE}$.

Proposition 6 *(Asymptotic distribution of jackknife two-way FE estimator) Suppose y_{it}, for $i=1,2,...,N$ and $t=1,2,...,T$, is generated by panel data model (51)-(52), Assumptions 1-3, 4.b, 5, and 8 hold, and $N,T\to\infty$ jointly such that $T=CN^{\epsilon}$, for some $0<C<\infty$ and $\epsilon>1/3$. Then the asymptotic distribution of the half-panel jackknife two-way FE estimator, $\tilde{\boldsymbol{\beta}}_{FE-TE}$, defined by (60), is given by*

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE-TE}-\boldsymbol{\beta}\right)\overset{d}{\sim}N\left(\mathbf{0},\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\right), \tag{66}$$

where $\mathbf{Q}=\bar{\boldsymbol{\Gamma}}(0)$, $\bar{\boldsymbol{\Gamma}}(0)$ is defined in Assumption (5) and $\mathbf{R}$ is defined by (48). A consistent estimator of the asymptotic variance of $\tilde{\boldsymbol{\beta}}_{FE-TE}$ is given by

$$\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE-TE}\right)=\hat{\mathbf{Q}}_{FE-TE}^{-1}\hat{\mathbf{R}}_{FE-TE}\hat{\mathbf{Q}}_{FE-TE}^{-1}, \tag{67}$$

where $\hat{\mathbf{Q}}_{FE-TE}$ and $\hat{\mathbf{R}}_{FE-TE}$ are defined by (55) and (65), respectively.

4.2 Unbalanced panels

In this sub-section we consider the extension of the jackknife procedure to unbalanced panels. This is an important extension for empirical analysis since most data sets are unbalanced. Suppose that for cross section unit i we have observations on y_{it} and $\mathbf{x}_{it}$ over the period $t=T_{fi},T_{fi}+1,...,T_{li}$, where T_{fi} and T_{li} are the first and last time periods for which data is available for this cross section unit. Let $T_i=T_{fi}-T_{li}+1$, $T_{\max}=\max T_i$, $T_{\min}=\min T_i$, and denote the average number of time series observations available by $\bar{T}_N=N^{-1}\sum_{i=1}^{N}T_i$, and its limit by $\bar{T}=\lim_{N\to\infty}\bar{T}_N$. Moreover, without any loss of generality, let the first time period of the panel be $1=\min_i T_{fi}$ and the last $T=\max_i T_{li}$. In this setting, gaps in the data are not allowed, but the panel could be unbalanced at both ends of the time period $(1,T)$.

4.2.1 Unbalanced FE panels

Consider the FE panel data model (11) first. The FE estimator in the unbalanced panel data setting is given by

$$\hat{\boldsymbol{\beta}}_{FEu}=\hat{\mathbf{Q}}_{FEu}^{-1}\hat{\mathbf{q}}_{FEu}, \tag{68}$$

where

$$\hat{\mathbf{Q}}_{FEu}=\frac{1}{\sum_{i=1}^{N}T_i}\sum_{i=1}^{N}\sum_{t=T_{fi}}^{T_{li}}\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]',\ \hat{\mathbf{q}}_{FEu}=\frac{1}{\sum_{i=1}^{N}T_i}\sum_{i=1}^{N}\sum_{t=T_{fi}}^{T_{li}}\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]y_{it} \tag{69}$$

and $\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)=\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\mathbf{x}_{it}$. Let

$$\vartheta_i=\frac{T_i}{\bar{T}_N}, \tag{70}$$

and initially assume that $|\vartheta_i|<K$ for all i. Also, as before, we require $\hat{\boldsymbol{\beta}}_{FE}$ and its large-N probability limit to exist. This is ensured by the following assumption, which replaces the earlier Assumption 6.

ASSUMPTION 10 (Existence of FE estimators for unbalanced panels) *There exists $N_0, T_0>0$ such that for all $N>N_0$ and all $T_i\geq T_0$, $\hat{\mathbf{Q}}_{FEu}$ defined in (69) is a positive definite matrix, $\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i\}}$ and $\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)$ are invertible, where*

$$\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\boldsymbol{\Gamma}_i\left(0\right), \tag{71}$$

$$\bar{\boldsymbol{\Psi}}_{\{T_i\}}=\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\boldsymbol{\Psi}_{iT_i}, \tag{72}$$

ϑ_i is defined in (70), $\boldsymbol{\Gamma}_i\left(0\right)$ is defined in Assumption 5, and

$$\boldsymbol{\Psi}_{iT_i}=\boldsymbol{\Gamma}_i\left(0\right)+\sum_{i=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\left[\boldsymbol{\Gamma}_i\left(h\right)+\boldsymbol{\Gamma}_i'\left(h\right)\right]. \tag{73}$$

The following proposition derives the bias of $\hat{\boldsymbol{\beta}}_{FE}$ when the time dimensions of the panel $\{T_i\}$ are all fixed and $N\to\infty$.

Proposition 7 *(Small-T bias of the FE estimator for unbalanced panels) Suppose y_{it}, for $t=T_{fi},T_{fi}+1,...,T_{li}$, $i=1,2,...,N$, is generated by the panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), and Assumptions 1-3, 4.a, 5, and 10 hold. Then for any given $\{T_i\}$ such that $T_i\geq T_0$ for all i, where T_0 is given by Assumption 10, we have*

$$\lim_{N\to\infty}\left(\hat{\boldsymbol{\beta}}_{FEu}-\boldsymbol{\beta}\right)=\frac{1}{\bar{T}}\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i\}}\right]^{-1}\bar{\boldsymbol{\chi}}_{\{T_i\}}=O\left(\bar{T}^{-1}\right),$$

where the unbalanced FE estimator, $\hat{\boldsymbol{\beta}}_{FEu}$, is defined by (68), $\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)$ and $\bar{\boldsymbol{\Psi}}_{\{T_i\}}$ are defined by (71) and (72), respectively,

$$\bar{\boldsymbol{\chi}}_{\{T_i\}}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{i=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\boldsymbol{\gamma}_i\left(h\right)=O\left(\bar{T}^{-1}\right), \tag{74}$$

and $\boldsymbol{\gamma}_i(h)$ is defined by (9).

Proposition 7 establishes that the bias of the FE estimator in the case of unbalanced panels is of order $O\left(\bar{T}^{-1}\right)$, where $\bar{T}$ is the limit of the average time dimension as $N \to \infty$. Moreover, the bias also depends, among other factors, on the degree to which the panel is unbalanced as characterized by the distribution of ϑ_i over i.

A simple way of implementing the half-panel jackknife bias correction is to assume that T_i are all even and to divide the unbalanced sample into two unbalanced sub-samples; the first sub-sample (denoted by subscript a) consisting of the first $T_i/2$ observations for cross section unit i, $i = 1, 2, ..., N$, and the second sub-sample (denoted by subscript b) consisting of the last $T_i/2$ observations of the same i^{th} unit. Specifically, let (using, for simplicity, the same notations as in the case of balanced panels)

$$\hat{\boldsymbol{\beta}}_{a,FEu}=\hat{\mathbf{Q}}_{a,FEu}^{-1}\hat{\mathbf{q}}_{a,FEu} \text{ and } \hat{\boldsymbol{\beta}}_{b,FEu}= \hat{\mathbf{Q}}_{b,FEu}^{-1}\hat{\mathbf{q}}_{b,FEu}, \tag{75}$$

where

$$\hat{\mathbf{Q}}_{a,FEu} = \frac{2}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=T_{fi}}^{T_{fi}+T_i/2-1} \left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a}(T_i)\right]\left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a}(T_i)\right]', \tag{76}$$

$$\hat{\mathbf{q}}_{a,FEu} = \frac{2}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=T_{fi}}^{T_{fi}+T_i/2-1} \left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot a}(T_i)\right] y_{it},$$

and

$$\hat{\mathbf{Q}}_{b,FEu} = \frac{2}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=T_{fi}+T_i/2}^{T_{li}} \left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b}(T_i)\right]\left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b}(T_i)\right]', \tag{77}$$

$$\hat{\mathbf{q}}_{b,FEu} = \frac{2}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=T_{fi}+T_i/2}^{T_{li}} \left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot b}(T_i)\right] y_{it},$$

$$\bar{\mathbf{x}}_{i\cdot a}(T_i) = \frac{2}{T_i} \sum_{t=T_{fi}}^{T_{fi}+T_i/2-1} \mathbf{x}_{it}, \quad \bar{\mathbf{x}}_{i\cdot b}(T_i) = \frac{2}{T_i} \sum_{t=T_{fi}+T_i/2}^{T_{li}} \mathbf{x}_{it}.$$

We continue to assume that the sample is sufficiently large so that the half-panel FE estimates $\hat{\boldsymbol{\beta}}_{a,FE}$ and $\hat{\boldsymbol{\beta}}_{b,FE}$ are well defined (similar to Assumption 7 for the balanced panel).

ASSUMPTION 11 (Existence of half-panel FE estimators for unbalanced panels) *There exists $N_0, T_0 > 0$ such that for all $N > N_0$ and all even $T_i \geq T_0$, the matrices $\hat{\mathbf{Q}}_{a,FEu}$, $\hat{\mathbf{Q}}_{b,FEu}$, and $\bar{\boldsymbol{\Gamma}}_{\vartheta}(0) - 2\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i/2\}}$ are invertible, where $\hat{\mathbf{Q}}_{a,FEu}$ and $\hat{\mathbf{Q}}_{b,FEu}$ are defined in (76) and (77), respectively, and $\bar{\boldsymbol{\Gamma}}_{\vartheta}(0)$ and $\bar{\boldsymbol{\Psi}}_{\{T_i/2\}}$ are defined by (71) and (72), respectively.*

Using the same arguments as in the proof of Proposition 7, we obtain

$$\lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{FE,a}-\boldsymbol{\beta}\right)=\lim_{N\to\infty} E\left(\hat{\boldsymbol{\beta}}_{FE,b}-\boldsymbol{\beta}\right)=\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-2\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i/2\}}\right]^{-1}\bar{\boldsymbol{\chi}}_{\{T_i/2\}},$$

where $\bar{\boldsymbol{\chi}}_{\{T_i/2\}}$ is defined in (74), specifically

$$\bar{\boldsymbol{\chi}}_{\{T_i/2\}}=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{2\vartheta_i}{T_i}\sum_{i=1}^{T_i/2-1}\left(1-\frac{2h}{T_i}\right)\boldsymbol{\gamma}_i\left(h\right)=O\left(\bar{T}^{-1}\right). \tag{78}$$

The large-N small-T bias of the half-panel jackknife estimator defined as

$$\tilde{\boldsymbol{\beta}}_{FEu}=2\hat{\boldsymbol{\beta}}_{FEu}-\frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{a,FEu}+\hat{\boldsymbol{\beta}}_{b,FEu}\right), \tag{79}$$

where the unbalanced FE estimator $\hat{\boldsymbol{\beta}}_{FEu}$ is given by (68), and its half-panel counterparts $\hat{\boldsymbol{\beta}}_{a,FEu}$ and $\hat{\boldsymbol{\beta}}_{b,FEu}$ are given by (75), is established in the next proposition.

Proposition 8 *(Small-T bias of half-panel jackknife FE estimator for unbalanced panels) Suppose y_{it}, for $t=T_{fi},T_{fi}+1,...,T_{li}$, $i=1,2,...,N$, is generated by the panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), and Assumptions 1-3, 4.a, 5, and 10-11 hold. Then, for any given $\{T_i\}$ such that $T_i\geq T_0$ and is an even integer for all i, where T_0 is chosen so that Assumptions 6 and 7 are satisfied, we have*

$$\begin{aligned} Bias_T\left(\tilde{\boldsymbol{\beta}}_{FEu}\right) &= \lim_{N\to\infty} E\left(\tilde{\boldsymbol{\beta}}_{FEu}-\boldsymbol{\beta}\right) \\ &= \frac{2}{\bar{T}}\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i\}}\right]^{-1}\bar{\boldsymbol{\chi}}_{\{T_i\}}-\frac{2}{\bar{T}}\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-2\bar{T}^{-1}\bar{\boldsymbol{\Psi}}_{\{T_i/2\}}\right]^{-1}\bar{\boldsymbol{\chi}}_{\{T_i/2\}} \\ &= O\left(\bar{T}^{-1}\bar{T}_h^{-1}\right)=O\left(\bar{T}_h^{-2}\right), \end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_{FEu}$ is defined in (79),

$$\bar{T}_h=\lim_{N\to\infty}\bar{T}_{h,N}, \tag{80}$$

$\bar{T}_{h,N}$ is the sample harmonic mean of T_i,

$$\bar{T}_{h,N}=\left(N^{-1}\sum_{i=1}^{N}\frac{1}{T_i}\right)^{-1}, \tag{81}$$

$\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)$ and $\bar{\boldsymbol{\Psi}}_{\{T_i\}}$ are defined by (71) and (72), respectively, and $\bar{\boldsymbol{\chi}}_{\{T_i\}}$ is defined by (74).

Thus, $Bias_T\left(\tilde{\boldsymbol{\beta}}_{FEu}\right)$ is of order $O\left(\bar{T}_h^{-2}\right)$ in general, and when ϑ_i are bounded below and above (with a possible exception of a finite number of units), then $\bar{T}$ and $\bar{T}_h$ are of the same order of magnitude and $Bias_T\left(\tilde{\boldsymbol{\beta}}_{FEu}\right)=O\left(\bar{T}^{-2}\right)$.

Consider now the case when $N,\{T_i\}\to\infty$ jointly such that $0<K_1<\vartheta_i<K_2<\infty$ for all i, and $T_{\min}=KN^{\epsilon}$, for some $\epsilon>1/3$ and $0<K<\infty$. Under these conditions, $\tilde{\boldsymbol{\beta}}_{FEu}$ is asymptotically unbiased and its variance can be consistently estimated, as

$$\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FEu}\right) = \hat{\mathbf{Q}}_{FEu}^{-1}\hat{\mathbf{R}}_{FEu}\hat{\mathbf{Q}}_{FEu}^{-1},$$

where $\hat{\mathbf{Q}}_{FE}$ is defined in (69), and

$$\hat{\mathbf{R}}_{FEu}=\frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=T_{fi}}^{T_{li}} \mathbf{d}_{it}\mathbf{d}_{it}'\hat{u}_{it,FEu}^2, \tag{82}$$

in which $\mathbf{d}_{it}$ is given by (45) with the averages $\bar{\mathbf{x}}_{i\cdot}$, $\bar{\mathbf{x}}_{i\cdot a}$, and $\bar{\mathbf{x}}_{i\cdot b}$ replaced by their unbalanced sample counterparts $\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)$, $\bar{\mathbf{x}}_{i\cdot a}\left(T_i\right)$, and $\bar{\mathbf{x}}_{i\cdot b}\left(T_i\right)$, respectively, and

$$\hat{u}_{it,FEu} = \left(y_{it} - \bar{y}_i\right) - \tilde{\boldsymbol{\beta}}_{FEu}\left[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right],$$

for $t = T_{fi}, T_{fi}+1, ..., T_{li}$, $i = 1, 2, ..., N$. These results are formally stated in the next proposition.

Proposition 9 *(Asymptotic distribution of the jackknife FE estimator for unbalanced panels) Suppose y_{it}, for $t = T_{fi}, T_{fi}+1, ..., T_{li}$, $i = 1, 2, ..., N$, is generated by panel data model (11) with $\mathbf{x}_{it}$ given by (2)-(3), Assumptions 1-3, 4.a, 5, and 10-11 hold, and $N, \{T_i\} \rightarrow \infty$ jointly such that $T_{\min} = CN^{\epsilon}$, for some $0 < C < \infty$ and $\epsilon > 1/3$, and there exist constants $0 < K_1, K_2 < \infty$ such that $0 < K_1 < \vartheta_i < K_2 < \infty$ for all i except a finite set of cross section units. Then the asymptotic distribution of the half-panel jackknife FE estimator for unbalanced panels, $\tilde{\boldsymbol{\beta}}_{FEu}$, defined by (79) is given by*

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FEu} - \boldsymbol{\beta}\right) \overset{d}{\sim} N\left(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\right),$$

where $\mathbf{Q}$ and $\mathbf{R}$ are defined by (44) and (48), respectively, and asymptotic variance of $\tilde{\boldsymbol{\beta}}_{FEu}$ can be consistently estimated by

$$\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FEu}\right) = \hat{\mathbf{Q}}_{FEu}^{-1}\hat{\mathbf{R}}_{FEu}\hat{\mathbf{Q}}_{FEu}^{-1},$$

where $\hat{\mathbf{Q}}_{FEu}$ and $\hat{\mathbf{R}}_{FEu}$ are defined by (69) and (82), respectively.

4.2.2 Unbalanced FE-TE panels

Consider the FE-TE panel data model (51)-(52) in the case of an unbalanced sample next. The FE-TE estimator in unbalanced panels can be obtained by using two sets of dummy variables to take account of fixed and time effects. Let

$$\bar{y}_{i\cdot}\left(T_i\right) = \frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}} y_{it},\ \bar{y}_{\cdot t}\left(N_t\right) = N_t^{-1}\sum_{i\in S_t} y_{it}, \text{ and } \bar{y}\left(T_1, T_2, ..., T_N\right) = \frac{1}{\sum_{i=1}^{N} T_i}\sum_{i=1}^{N}\sum_{t=T_{fi}}^{T_{li}} y_{it},$$

where S_t is the index set of units with available data for time period t. It is important to highlight that the simple transformation used for the balanced FE-TE panels

$$y_{it}^{*} = y_{it} - \bar{y}_{i\cdot}\left(T_i\right) - \bar{y}_{\cdot t}\left(N_t\right) + \bar{y}\left(T_1, T_2, ..., T_N\right),$$

is no longer valid and does not remove the fixed and time effects when the panel is unbalanced. This simple transformation (applied to all variables) and the dummy variable approach are identical only when the panel is balanced.

It is also important to highlight that de-meaning the variables by subtracting the time averages first and then running a panel regression of $[y_{it} - \bar{y}_{i\cdot}(T_i)]$ on $[\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}(T_i)]$ and time dummies does not filter out the fixed and time effects either. Although empirical papers in the literature often do not report the details of how the FE-TE estimates are computed in the case of unbalanced panels, it is our impression that the latter approach of adding time dummies to a panel regression using demeaned data is often used in practice.

Adding both time and fixed effect dummies in a regression using non-transformed data can be computationally cumbersome when $N + T$ is large. To address such complications, Wansbeek and Kapteyn (1989) proposed a computationally convenient transformation of variables that eliminates the fixed and time effects simultaneously, and is identical to the fixed and time dummy approach.[6] In the Monte Carlo experiments and the empirical applications, we apply Wansbeek and Kapteyn (WK) transformation to compute the FE-TE estimates for unbalanced panels. In the case of jack-knife FE-TE estimators for unbalanced panels, we first construct the two sub-samples as described in sub-section 4.2.1, and then apply WK transformation to eliminate the fixed and time effects from each of the two sub-samples, separately.

However, for theoretical derivations and proofs, the use of WK transformation is rather complicated and in what follows we establish theoretical results for a simplified FE-TE estimation for unbalanced panels, which is based on the insight that in the case of panels with weakly exogenous regressors, it is only the de-meaning across the time dimension that gives rise to the small-T bias. In FE-TE panels with strictly exogenous regressors, indices i and t are interchangeable. This is no longer the case in panels with weakly exogenous regressors, where the FE estimator is subject to the small-T bias, but the TE estimator (in a model with time effects only) is not subject to any bias. Let n denote the number of cross section units with observations on all T time periods, and assume that n/N is bounded away from 0 as $N \to \infty$, namely n and N expand at the same rate. Let S_n be the index set of such n cross section units, and define the simple cross section averages

$$\bar{y}_{\cdot t}\left(n\right) = n^{-1} \sum_{i \in S_n} y_{it} \text{ and } \bar{\mathbf{x}}_{\cdot t}\left(n\right) = n^{-1} \sum_{i \in S_n} \mathbf{x}_{it}. \tag{83}$$

Averaging (51) across the n cross section units in S_n, and then subtracting this average from (51)

[6] See Section 9.4 of Baltagi (2008) for a textbook exposition of Wansbeek and Kapteyn transformation.

yields the following transformed unbalanced FE specification:

$$\tilde{y}_{it} = \tilde{\mu}_i + \boldsymbol{\beta}'\tilde{\mathbf{x}}_{it} + \tilde{u}_{it}, \tag{84}$$

for $t = T_{fi}, T_{fi}+1, ..., T_{li}$, $i = 1, 2, ..., N$, where $\tilde{y}_{it} = y_{it} - \bar{y}_{\cdot t}(n)$, $\tilde{\mu}_i = \mu_i - \bar{\mu}(n)$, $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{\cdot t}(n)$, $\tilde{u}_{it} = u_{it} - \bar{u}_{\cdot t}(n)$, and, similar to (83), $\bar{\mu}(n) = n^{-1}\sum_{i\in S_n}\mu_i$ and $\bar{u}_{\cdot t}(n) = n^{-1}\sum_{i\in S_n} u_{it}$. The transformed model (84) does not exactly correspond to the unbalanced FE model analyzed above, but

$$\bar{u}_{\cdot t}(n) = O_p\left(n^{-1/2}\right), \text{ and } \bar{\boldsymbol{\omega}}_{\cdot t}(n) = n^{-1}\sum_{i\in S_n}\boldsymbol{\omega}_i(n) = O_p\left(n^{-1/2}\right),$$

uniformly in t, $\tilde{\mathbf{x}}_{it} = \tilde{\boldsymbol{\mu}}_{ix} + \boldsymbol{\omega}_{it} + O_p\left(n^{-1/2}\right)$, and $\tilde{y}_{it} = \tilde{\mu}_i + \boldsymbol{\beta}'\left(\tilde{\boldsymbol{\mu}}_{ix} + \boldsymbol{\omega}_{it}\right) + u_{it} + O_p\left(n^{-1/2}\right)$. Hence, (84) corresponds to the unbalanced FE case analyzed above with the exception of the $O_p\left(n^{-1/2}\right)$ terms. It can be established that these terms do not matter for any of the findings above for the unbalanced FE case. Hence, we propose a half-panel jackknife FE-TE estimator by applying the jackknife bias correction procedure to the FE estimator using the transformed variables $\{\tilde{y}_{it}, \tilde{\mathbf{x}}_{it}\}$. The main findings for the half-panel jackknife FE estimator also extend to the half-panel jackknife FE-TE estimator.

5 Small sample properties

Using Monte Carlo techniques, we now investigate the small sample properties of the FE and FE-TE estimators and their half-panel jackknife bias-corrected versions under different set-ups, allowing for different degrees of weak exogeneity of the regressors, with or without lags of the dependent variable, and for both balanced and unbalanced samples. We selectively report some key results. The full set of results are summarized in an online supplement.

5.1 Data generating process (DGP)

Observations on y_{it} and x_{it} are generated *jointly* by

$$y_{it} = \mu_i + \delta_t + \lambda_y y_{i,t-1} + (1-\lambda_y)\beta x_{it} + u_{it}, \tag{85}$$

and

$$x_{it} = (1-\lambda_x)\mu_{ix} + (1-\lambda_x)\kappa_x y_{i,t-1} + \lambda_x x_{i,t-1} + v_{it}, \tag{86}$$

for $i = 1, 2, \ldots, N$ and $t = -99, -98, \ldots, 0, 1, 2, \ldots, T$, using $y_{i,-100} = x_{i,-100} = 0$ as the starting values. The first 100 time observations ($t = -99, -98, \ldots, 0$) are discarded. The fixed effects and the idiosyncratic errors are generated as:

$$\begin{aligned}
\mu_{ix} &\sim IID\mathcal{N}(1,1), \qquad \mu_i = \mu_{ix} + \eta_{yi}, \qquad \eta_{yi} \sim IID\mathcal{N}(1,1), && (87)\\
v_{it} &\sim IID\mathcal{N}\left(0,\sigma_{vi}^2\right), \quad \sigma_{vi}^2 = 0.5 + 0.25\eta_{vi}^2, \quad \eta_{vi}^2 \sim IID\chi^2(2), && (88)\\
u_{it} &\sim IID\mathcal{N}\left(0,\sigma_{ui}^2\right), \quad \sigma_{ui}^2 = 0.5 + 0.25\eta_{ui}^2, \quad \eta_{ui}^2 \sim IID\chi^2(2). && (89)
\end{aligned}$$

This set up allows the fixed effects in the y_{it} and x_{it} equations to be correlated, which in turn induces correlation between μ_i and x_{it}. For the time effects, δ_t, we consider three possibilities: no time effects, linear time effects, and quadratic time effects, namely

$$\delta_t = 0,\ 0.025t, \text{ or } 0.025t - 0.001t^2. \tag{90}$$

We consider three values for λ_y, representing a "static" panel regression with $\lambda_y = 0$, and two dynamic panel regressions with a moderate and high values for $\lambda_y \neq 0$:

$$\lambda_y = 0,\ 0.4, \quad \text{or} \quad 0.8. \tag{91}$$

We also consider three values for the feedback coefficient, κ_x (no feedbacks, a low degree of feedbacks, and a medium degree of feedbacks):

$$\kappa_x = 0,\ 0.2, \quad \text{or} \quad 0.4. \tag{92}$$

Throughout we set $\beta = 0.5$ and $\lambda_x = 0.25$.

In total, we conducted 27 experiments covering all combinations of δ_t, λ_y and κ_x, summarized in Table 1. All experiments were carried out for all $N \in \{30, 60, 100, 200, 500, 1000\}$ and $T \in \{30, 60, 100, 200\}$ combinations, with the number of replications set to $R = 2,000$.

5.2 Experiments without lags of the dependent variable

In the case where $\lambda_y = 0$, the parameter of interest, β $(= 0.5)$, is estimated using the following four estimators:

1. **FE estimator** $\hat{\beta}_{FE}$. In experiments without time effects ($\delta_t = 0$), the FE estimator is defined by (12) and based on the panel regression

   $$y_{it} = \mu_i + \beta x_{it} + e_{it}. \tag{93}$$

   When $\delta_t = 0.025t$ or $0.025t - 0.001t^2$, the FE estimator is based on

   $$y_{it} = \mu_i + gt + \beta x_{it} + e_{it}. \tag{94}$$

   Note that when $\delta_t = 0.025t - 0.001t^2$, the panel regression model (94) is mis-specified.

2. **Half-panel jackknife FE estimator** $\tilde{\beta}_{FE}$ defined by (31).

3. **FE-TE estimator** $\hat{\beta}_{FE-TE}$ is defined by (53), and based on

$$y_{it} = \mu_i + \delta_t + \beta x_{it} + e_{it}. \tag{95}$$

4. **Half-panel jackknife FE-TE estimator** $\tilde{\beta}_{FE-TE}$ defined by (60).

As a benchmark, initially we report the results of Experiment 1, where there are no time effects ($\delta_t = 0$) and the regressor, x_{it}, is strictly exogenous (since $\kappa_x = \lambda_y = 0$). We report bias ($\times 100$), root mean square error (RMSE, $\times 100$), and size (in %) at the 5 percent nominal level, for the estimation of $\beta = 0.5$. In this case, FE and half-panel jackknife FE estimators are both valid for a fixed T and have the same distribution asymptotically as $N \to \infty$. But one would expect the FE estimator to be more efficient in small samples, since in this case the bias correction is not required. This is confirmed by the small sample results reported in Table 2. The FE estimator performs slightly better than its jackknife version (around 2% in terms of RMSEs) when $T = 30$, but for larger values of T both estimators perform very similarly. The inclusion of time dummies to allow for possible time effects does not alter this conclusion. The plot of power functions in Figure 2 for selected values of N and T also show that the two estimators have very similar power performances. These benchmark results are important, since they show that half-panel jackknife FE estimators perform well even if they are applied when bias corrections are not required.

Consider now Experiment 3 where x_{it} is weakly exogenous with $\kappa_x = 0.4$, and $\lambda_y = \delta_t = 0$. The results are summarized in Table 3.[7] It is clear that as compared to $\tilde{\beta}_{FE}$ (the jackknife bias-corrected estimator), the standard FE estimator, $\hat{\beta}_{FE}$ is subject to the small-T bias. For example, for $T = 30$ and $N = 200$, the bias of FE estimator is -0.0013 as compared to 0.0004 for the bias-corrected version. Although, in general, there is a trade off between bias reduction and the variance, the RMSEs in Table 3 clearly show that overall the bias correction has been beneficial, with substantial gains for small T and large N. It is also interesting that the gain in terms of RMSE rises as N is increased relative to T. For example, when $T = 30$ and $N = 200$, the RMSE of the FE estimator is around 43% larger than the RMSE of the bias-corrected estimator, whilst this figure rises to 250% for $T = 30$ and $N = 1,000$. Also, as to be expected, the bias of the FE estimator declines with T and its RMSE falls towards its bias-corrected counterpart. Finally, as predicted by the theory, the FE estimator shows substantial size distortions when N is large relative to T, with its size approaching 70% for $T = 30$ and $T = 1,000$. The results for FE-TE and jackknife FE-TE estimators are given at the lower part of Table 3. The performance of $\hat{\beta}_{FE-TE}$ is similar to $\hat{\beta}_{FE}$, while the jackknife estimators continue to perform well, regardless of the feedbacks. The half-panel jackknife estimator, $\tilde{\beta}_{FE-TE}$, has negligible bias and the correct size for all values of N and T considered. The power functions in Figure 3 also show that the half-panel jackknife FE

[7]Results for $\kappa_x = 0.2$ lie somewhere between the ones reported for $\kappa_x = 0$ and $\kappa_x = 0.4$, and are provided in an online supplement.

and FE-TE estimators perform well, while standard FE and FE-TE estimators are not correctly centered at the true value of β (= 0.5).

Experiments 7 and 9 feature quadratic time effects ($\delta_t = 0.025t - 0.001t^2$), with two choices for the feedback parameter, κ_x. The results for $\kappa_x = 0$ and 0.4 are summarized in Tables 4 and 5, respectively. Note that in these experiments $\hat{\beta}_{FE}$ and $\tilde{\beta}_{FE}$ are mis-specified since they don't allow for the time effects. Interestingly, comparing the size distortion of $\hat{\beta}_{FE}$ in Table 4 (when $\kappa_x = 0$) with the size distortion of $\hat{\beta}_{FE}$ in Table 5 (when $\kappa_x = 0.4$), the latter size distortions are huge, with rejection rates close to 100% for $T > 60$, whereas the former size distortions are very small. In the strictly exogenous case, the size becomes a problem only when T is large ($T > 60$), with the reported over-rejections for the largest values of T considered being less than 13%. The miss-specification of time effects compounded with the small T bias has resulted in very poor performance for $\hat{\beta}_{FE}$ in the weakly exogenous case. Since the presence of time dummies allows for arbitrary time effects, the performance of FE-TE only suffers in the weakly exogenous case (Table 5), and overall the performance of $\hat{\beta}_{FE-TE}$ in Table 5 is similar to $\hat{\beta}_{FE}$ and $\hat{\beta}_{FE-TE}$ in experiments without time dummies in DGP (see Table 3). The half-panel jackknife FE-TE estimator $\tilde{\beta}_{FE-TE}$ performs well regardless of the time effects in the DGP, as expected. The power functions in Figures 4 and 5 also show that when x_{it} is strictly exogenous (Figure 4, with $\kappa_x = 0$), even though bias corrections are not required, the half-panel jackknife FE-TE $\tilde{\beta}_{FE-TE}$ performs almost as well as the non-jackknife FE-TE $\hat{\beta}_{FE-TE}$. But when x_{it} is weakly exogenous (Figure 5, with $\kappa_x = 0.4$), the power functions of FE-TE estimators, $\hat{\beta}_{FE-TE}$, are not centered at the true value of $\beta = 0.5$, while the ones for the half-panel jackknife FE-TE estimator, $\tilde{\beta}_{FE-TE}$, are correctly centered.

5.3 Experiments with lagged dependent variables

In these experiments $\lambda_y = 0.4$ or 0.8, and the parameter of interest is given by the long-run coefficient, $\beta = -b/\phi$, where b and ϕ are estimated using the following dynamic panel regressions:

1. **FE estimator** $\hat{\beta}_{FE}$. When $\delta_t = 0$, $\hat{\beta}_{FE}$ is based on

$$\Delta y_{it} = \mu_i + \phi y_{i,t-1} + b x_{it} + e_{it}, \tag{96}$$

   and when $\delta_t = 0.025t$ or $0.025t - 0.001t^2$, $\hat{\beta}_{FE}$ is based on:

$$\Delta y_{it} = \mu_i + gt + \phi y_{i,t-1} + b x_{it} + e_{it}. \tag{97}$$

   As before, we note that when $\delta_t = 0.025t - 0.001t^2$, the model (97) is mis-specified. β is estimated by

$$\hat{\beta}_{FE} = -\frac{\widehat{b}_{FE}}{\widehat{\phi}_{FE}}. \tag{98}$$

The estimator for the asymptotic variance of $\hat{\beta}_{FE}$ is obtained by the delta method:

$$\widehat{AsyVar}\left(\hat{\beta}_{FE}\right) = \left(\frac{\widehat{b}_{FE}}{\widehat{\phi}_{FE}^2}, -\frac{1}{\widehat{\phi}_{FE}}\right) \widehat{AsyVar}\left(\begin{array}{c} \widehat{\phi}_{FE} \\ \widehat{b}_{FE} \end{array}\right) \left(\frac{\widehat{b}_{FE}}{\widehat{\phi}_{FE}^2}, -\frac{1}{\widehat{\phi}_{FE}}\right)'. \tag{99}$$

2. **Half-panel jackknife FE estimator** $\tilde{\beta}_{FE}$. We first compute the half-panel bias-corrected FE estimators $\tilde{\phi}_{FE}$ and $\tilde{b}_{FE}$ based on $\widehat{\phi}_{FE}$ and $\widehat{b}_{FE}$. $\tilde{\beta}_{FE}$ is obtained as

$$\tilde{\beta}_{FE} = -\frac{\tilde{b}_{FE}}{\tilde{\phi}_{FE}},$$

   and the estimator for the asymptotic variance of $\tilde{\beta}_{FE}$ is obtained by the delta method similar to (99).

3. **FE-TE estimator** $\hat{\beta}_{FE-TE}$ is based on

$$\Delta y_{it} = \mu_i + \delta_t + \phi y_{i,t-1} + bx_{it} + e_{it}. \tag{100}$$

   As in the case of $\hat{\beta}_{FE}$, $\hat{\beta}_{FE-TE} = -\widehat{b}_{FE-TE}/\widehat{\phi}_{FE-TE}$ and its asymptotic variance is obtained by the delta method, as in (99).

4. **Half-panel jackknife FE-TE estimator** $\tilde{\beta}_{FE-TE}$ is computed in the same way as $\tilde{\beta}_{FE}$, but FE-TE estimators are used instead of the FE estimators.

The results for Experiment 12 ($\lambda_y = 0.4$, $\delta_t = 0$ and $\kappa_x = 0.4$) are reported in Table 6. The FE and FE-TE estimators are biased downwards, due to the well-documented downward small-T bias due to the presence of the lagged dependent variable and the weak exogeneity of x_{it}. The small sample bias of FE and FE-TE estimators is duly manifested in large size distortions. As can be seen from Table 6, both FE and FE-TE estimators show size distortions that rise in N and fall in T. The size rises very rapidly in N for any given choice of T. For $N = 1000$ and $T = 30$, the size reaches 99%. The half-panel jackknife estimators, in contrast, are subject to a small positive bias and achieve correct size for all values of N and T considered, with the exception of $N = 1000$ and $T = 30$, where the size is 9%. The plot of the power functions in Figure 6 also show that the half-panel jackknife FE estimator, $\tilde{\beta}_{FE}$, performs well for all selected values of N and T except when $N = 1000$ and $T = 30$, but for the FE estimator the power functions for all N and T combinations are shifted to the left of the true value of β.

Consider now Experiment 18, which features quadratic time effects (with $\delta_t = 0.025t - 0.001t^2$), and $\lambda_y = \kappa_x = 0.4$ as in Experiment 12. The results are reported in Table 7 and Figure 7. $\hat{\beta}_{FE}$ and $\tilde{\beta}_{FE}$ are miss-specified due to the quadratic trends in DGP, and the size distortions of $\hat{\beta}_{FE}$ and $\tilde{\beta}_{FE}$ are huge, with rejection rates 100% in many cases when both N and T are large. Once time dummies are included (considering FE-TE and half-panel jackknife FE-TE estimators), the relative performance of bias uncorrected and bias corrected estimators ($\hat{\beta}_{FE-TE}$ and $\tilde{\beta}_{FE-TE}$) are very similar to the relative performance of $\hat{\beta}_{FE}$ and $\tilde{\beta}_{FE}$ reported in Table 6.

Similar results are also obtained for the other experiments, but to save space those results are provided in an online supplement.

5.4 Experiments with unbalanced panels

We also consider unbalanced panels by dropping $[T/5]$ observations from the beginning and from the end of the sample period for units $i = 1, 2, ..., [N/4]$, and $[T/3]$ observations from the beginning and from the end of the sample periods for units $i = [N/4] + 1, [N/4] + 2, ..., [N/2]$, where $[a]$ denotes the integer part of a. For unbalanced panels we only report the results for Experiment 18 ($\lambda_y = 0.4$, $\delta_t = 0.025t - 0.001t^2$ and $\kappa_x = 0.4$). Again, the results for other experiments are similar to the ones reported for balanced panels. The results are summarized in Table 8 and Figure 8. In the case of models with both fixed and time effects we employ the Wansbeek and Kapteyn (1989) transformation as discussed in sub-section 4.2.2.

As a whole, the findings for the unbalanced panels are similar to the results for the balanced panels reported in Table 7. That is, $\hat{\beta}_{FE}$ and $\tilde{\beta}_{FE}$ are mis-specified and the size distortions are huge, with rejection rates 100% in many cases, $\hat{\beta}_{FE-TE}$ is biased and size-distorted due to weak exogeneity, and $\tilde{\beta}_{FE-TE}$ performs well. However, the RMSE of $\tilde{\beta}_{FE-TE}$ in Table 8 are about 20% larger than the RMSE in Table 7, due to a smaller average value of T in the case of unbalanced panels. Interestingly, the performance of $\tilde{\beta}_{FE-TE}$ is seemingly better in term of size when the samples are unbalanced. For example, when $N = 1000$ and $T = 30$, the size of $\tilde{\beta}_{FE-TE}$ in Table 8 is only 6.70% while in Table 7 it is 9.35%. However, Figure 8 shows that this could be due to the fact that the power functions of $\tilde{\beta}_{FE-TE}$ for the unbalanced panels are flatter than ones for the balanced panels (see Figure 7), due to fewer observations.

6 Empirical illustrations

It is reasonable to ask if the jackknife bias correction makes that much of a difference in practice. In this section we provide two empirical illustrations, one by Berger et al. (2013) on the effect of the US political influence on bilateral trades of US and foreign countries during the Cold War, and a second application by Donohue and Levitt (2001) on the determinants of crimes in the US. The former involves an unbalanced panel of countries in the world economy, and the latter a balanced panel of 48 States in the US.

6.1 Empirical illustration I: Commercial Imperialism

Berger et al. (2013) studied the effect of the US political influence on the bilateral trades of US and foreign countries during the Cold War. Using an annual unbalanced panel of country-level data of 131 countries over the period 1947 – 1989 (43 years), Berger et al. (2013) used the FE-TE estimator to estimate the following panel data regression (equation (7) in their paper):

$$\ln \frac{m_{it}^{US}}{Y_{it}} = \mu_i + \delta_t + \beta USinfluence_{it} + \phi \ln \tau_{it}^{US} - \phi \left(\ln P_t^{US} + \ln P_t^i \right) + \mathbf{X}_{it} \mathbf{\Gamma} + u_{it}, \tag{101}$$

where the dependent variable, $\ln \left(m_{it}^{US} / Y_{it} \right)$, is the natural log of imports into country i in year t from the US normalized by country i's total GDP. $USinfluence_{it}$ is an indicator variable that equals one, in country i in year t, if the CIA (Central Intelligence Agency) either successfully installed a foreign leader or provided covert support for the regime once in power. This valuable dataset was constructed by Berger et al. (2013) according to various studies of the history of the Cold War, typically based on declassified historical documents. $\ln \tau_{it}^{US}$ and $\ln P_t^{US} + \ln P_t^i$ respectively denote the trade costs and the multilateral resistance terms, which are given by the distance between US and country i, and four indicator variables for US and country i sharing a common language (English), sharing a border, both being GATT (General Agreement on Tariffs and Trade) participants, and belonging to a regional trade agreement. $\mathbf{X}_{it}$ is a vector of control variables including the per capita income of country i, an indicator variable for Soviet interventions (constructed in the same manner as CIA interventions), an indicator variable for the change in leadership, a measure of the tenure of the current leader, and an indicator variable for democracy. Berger et al. (2013) also estimated the effects of CIA interventions on log normalized imports from the rest of the world, log normalized exports to the US, and log normalized exports to the rest of the world, with estimating equations derived in an analogous manner as equation (101). Berger et al. (2013) found that the US influence raised the imports from the US to the intervened country but had no effects on imports from the rest of the world, exports to the US, or to the rest of the world.

We apply the half-panel jackknife bias-correction estimator for unbalanced panels (developed in sub-section 4.2.2) to the same dataset of Berger et al. (2013). For countries with odd numbers of observations, we drop the first observations before applying the half-panel jackknife. The results are summarized in Table 9. Column (1.a) shows the estimates reported by Berger et al. (2013) for equation (101). Before jackknife bias-correction, the coefficient of $USinfluence_{it}$ is estimated to be 0.293 and is statistically significant. As it happens the bias corrected estimate of this coefficient at 0.450 is even larger with a higher level of statistical significance. This is in line with the theoretical downward bias of the FE-TE estimators in the presence of weakly exogenous regressors. For other control variables, interestingly, we also find different results from Berger et al. (2013). For example, in column (1.a) both the estimates of the coefficients of the common language indicator and the GATT participants indicator are not statistically significant. However, the bias-corrected estimates given under column (1.b) are both positive and statistically significant. These estimates suggest that sharing a common language and joining GATT would have positive effects on the imports from the US, which makes more sense than the statistically insignificant effects obtained when using FE-TE estimators.

Columns (2.a), (3.a) and (4.a) show the estimates of Berger et al. (2013) with dependent variables log normalized imports from the rest of the world, log normalized exports to the US, and log normalized exports to the rest of the world. Before the jackknife bias correction, the estimates of the coefficients of $USinfluence_{it}$ are statistically insignificant. Berger et al. (2013) argued that the results provided some evidence that the CIA interventions did not create trades in all directions, but only created markets for US exports. Our jackknife estimates (in columns (2.b), (3.b) and (4.b)) support the findings of Berger et al. (2013). After the jackknife bias-correction, the estimates of the coefficients of $USinfluence_{it}$ remain statistically insignificant, but there are some important differences in the case of other coefficients. For example, under columns (3.a) and (3.b), the bias-corrected estimates of the effects of democracy, contiguous border, and GATT participation variables are statistically significant at the 5% level, but none of the estimates reported by Berger et al. (2013) for these coefficients are statistically significant. There are also large differences in the magnitudes of these estimates, which could reflect the extent to which FE-TE estimates could be biased if uncorrected.

6.2 Empirical Illustration II: Abortion and Crime

Donohue and Levitt (2001) studied the effect of legalized abortion on crimes in the US, using a balanced panel of data on 50 US States and the District of Columbia over the period 1985 – 1997 (13 years). These authors estimate the following FE-TE panel data regression (equation (2) in their paper):

$$y_{it} = \ln(crime_{it}) = \mu_i + \delta_t + \beta_1 ABORT_{it} + \boldsymbol{\psi}'\mathbf{x}_{it} + u_{it}, \tag{102}$$

where $\ln(crime_{it})$ is the logarithm of the crime rate per capita in state i and year t. Donohue and Levitt (2001) considered three types of crimes: violent crime, property crime and murders. $ABORT_{it}$, the "effective" legalized abortion rate, and is computed as a weighted average of the abortion rates in which the weights are determined by the fraction of arrests from different age groups. $\mathbf{x}_{it}$ is a vector of control variables, including lagged prisoners and police per capita, a number of variables for state economic conditions, the lagged state welfare generosity, the concealed handgun laws, and per capita beer consumption. Notably, Donohue and Levitt (2001) use the one-year lags of prisoners and police per capita as controls to deal with the endogeneity of these covariates. But as our theoretical analysis shows, lagging the covariates does not eliminate the bias due to possible feedbacks from changes in crimes to policing and imprisonments. Donohue and Levitt (DL) conclude that legalized abortion in 1970s has been one of the main causes of the substantial decline in crime observed in the US during 1990s.

DL study has attracted a great deal of attention with a large number of studies considering different aspects of their analysis ranging from measurement problems, the choice of the control variables, the choice of the abortion variable (whether to focus on aggregate measures of abortion or teenage abortion), data extensions, and possible missing common factors which have led to crime decline not only in the US but across most of the industrialized economies.[8] In this sub-section

[8] See, for example, Joyce (2004, 2009), Foote and Goetz (2008), Moody and Marvell (2010), Belloni et al. (2014),

we focus on the rather narrow estimation and inference issue and adopt DL's original 2001 data set and the associated measurements. To simplify the analysis we follow Belloni et al. (2014) and estimate model (102) only on the 48 contiguous states, and drop the District of Columbia, Alaska and Hawaii from the analysis.[9] However, unlike Belloni et al. (2014) who were concerned with the robustness of DL results to the choice of the covariates, in our analysis we include all the covariates as in Donohue and Levitt (2001).[10]

We estimated model (102) by FE-TE and jackknife FE-TE methods, for all the three crime categories, based on the sub-sample of 48 contiguous states over the period 1985-1997. Given the odd number of available time periods ($T = 13$) we experimented with deleting the first, the last or a random mixture of the first and the last observations to obtain an even number of time periods, needed for implementation of the half-panel jackknife estimator.[11] The results were qualitatively very similar and in Table 10 we only report the jackknife estimators with the first observations (for 1985) deleted. The estimates reported in Donohue and Levitt (2001) are reproduced under columns (1.a), (2.a) and (3.a), the FE-TE estimates are given in columns (1.b), (2.b) and (3.b), and the half-panel jackknife FE-TE estimates are under columns (1.c), (2.c) and (3.c). The FE-TE estimates of the coefficient of the abortion variable, β_1, for violent crime, property crime and murders, are very close to the estimates reported by Donohue and Levitt (2001). But there are some important differences between FE-TE and the jackknife FE-TE estimates, although these differences are quantitative in nature and do not alter DL's main conclusion, with all statistically significant coefficients estimated to be larger in magnitude by the jackknife procedure. For example, for violent crimes, according to DL's own estimates, only the abortion variable has a statistically significant coefficient estimated to be -0.129, while the jackknife FE-TE estimates are statistically significant for prisoners per capital and log state income per capita variables with coefficient much larger than those estimated by FE-TE procedure. A similar picture also emerges for property crime. But for the murder per capita, the estimates based on the jackknife FE-TE procedure are statistically less significant.

Another source of possible bias in the analysis of the relationship between abortion and crime is dynamic mis-specification. DL recognize the importance of the cumulative and persistent effects from abortion to crimes but do not check the robustness of their results to dynamics of crimes as they respond to changing values of the covariates. As a first step towards allowing for such

and Shoesmith (2015). See also responses by Donohue and Levitt (2004, 2008), and Levitt (2004).

[9] The FE-TE estimates of (102) turn out to be highly sensitive to whether District of Columbia, Alaska and Hawaii are included in the analysis. Belloni et al. (2014) estimate a first-differenced version of (102) with time-dummies, and obtain estimates of β_1 which are similar to the original estimates of Donohue and Levitt (2001), when all the control variables are included in the panel regression.

[10] As pointed out by Moody and Marvell (2010), there are also potentially important missing controls such as per cent black, per cent urban, or age distribution by states, often used in crime studies in the US.

[11] We also carried out additional Monte Carlo experiments to see if the performance of the half-panel jackknife FE-TE estimator is adversely affected given the rather small sample ($N = 48$ and $T = 12$) under consideration. We found that our main findings hold and the jackknife FE-TE estimator works well even in such a case. The results are summarized in the online supplement. Recall that for this application $N/T^3 \approx 0.028$ which is sufficiently small as required for the validity of our bias correction.

dynamics we also considered the following simple dynamic panel model:

$$\ln(crime_{it}) = \mu_i + \delta_t + \lambda \ln(crime_{i,t-1}) + \beta_1(1-\lambda)ABORT_{it} + (1-\lambda)\boldsymbol{\psi}'\mathbf{x}_{it} + u_{it}, \qquad (103)$$

where $\ln(crime_{i,t-1})$ is the lagged logarithm of crime rate per capita. It is clear that the FE-TE estimates of β_1 are likely to be biased given the rather small value of T which is now reduced to $T = 12$ due to the presence of lagged values of $\ln(crime_{it})$ amongst the regressors. But the jackknife estimators are likely to be valid even in this application since $N/T^3 = 48/12^3 \approx 0.028$ is sufficiently small. The results are summarized in Table 11, with the estimates of λ given at the bottom of the table. The jackknife FE-TE estimates of λ are statistically significant at the 1% level for all three crime categories. They are also noticeably larger than the FE-TE estimates that are known to be biased downward. The jackknife estimates of λ are quite a bit larger for violent and property crimes as compared to murder crimes. Turing to the short term estimates given by $b_1 = \beta_1(1-\lambda)$ and $\mathbf{b}_2 = (1-\lambda)\boldsymbol{\psi}$, we notice a number of differences as compared to the corresponding estimates in Table 10. Most importantly, using jackknife estimates we find that the abortion variable is no longer statistically significant in the property crimes regressions. Overall, allowing for dynamics shed some doubt on the robustness of DL findings, with mixed results. The abortion variable continues to be statistically significant for violent and murder crimes but not for property crimes. Allowing for dynamics has strengthen the explanatory power of other covariates such as prisoners and police per capita, the unemployment rate (for property crimes), and log state income per capita (for violent and property crimes). The beer consumption per capita is no longer statistically significant. The long run estimates, namely β_1 and $\boldsymbol{\psi}$ in (103), are summarized in Table 12 and give a similar picture.

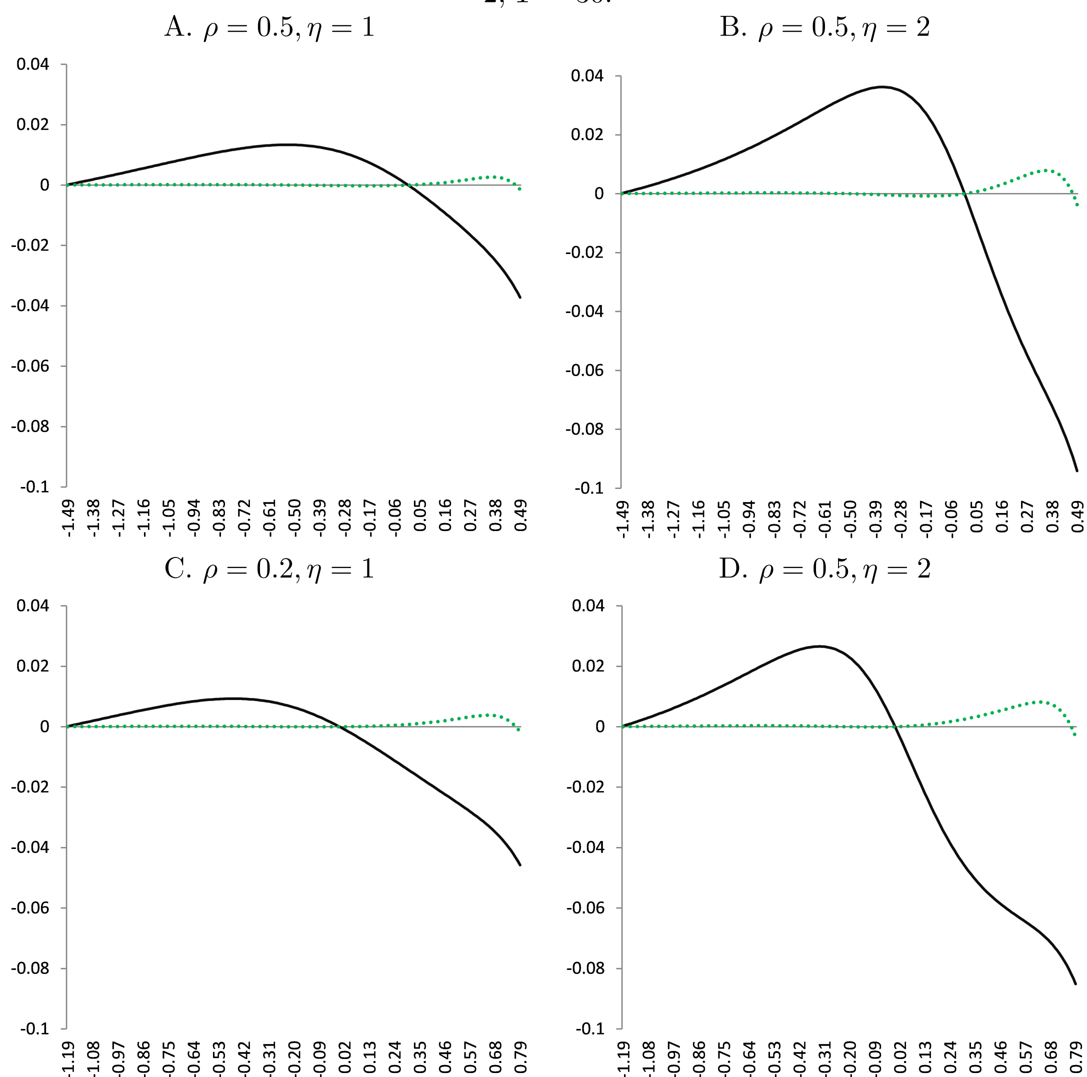
7 Conclusion

In this paper we consider the problem of estimation and inference in panel data models with weakly exogenous regressors when N is relatively large and T is moderate. The existing estimation techniques in the literature have not sufficiently covered this problem since no estimator proposed in the literature is established to deliver a valid inference in this set-up. We have derived exact expressions for the bias of FE and FE-TE estimators when N is large for a given T, and considered the half-panel jackknife method to remedy the bias. We have derived the exact expression for the large-N small-T bias of this particular jackknife method, and we have established that it is asymptotically unbiased as $N, T \to \infty$ jointly such that $T = KN^{\epsilon}$, for some $0 < K < \infty$ and $\epsilon > 1/3$, which makes this method suitable for N large and moderate T. The inference based on the proposed variance estimator in this paper is very good, even for N as large as 1000 and T as small as 30 in the considered set of Monte Carlo experiments. FE and FE-TE estimators, on the other hand, can be grossly oversized in the presence of weakly exogenous regressors (regardless whether the panel includes lagged dependent variable or not), unless N/T is sufficiently small.

The two empirical applications included in the paper illustrate the potential use of the half-panel jackknife method for panel data analysis with dynamics and weakly exogenous regressors.

Overall, our theoretical results backed up with extensive Monte Carlo evidence suggest that jackknife bias correction is a useful remedy to the small-T bias problem of the FE and FE-TE estimators in panels with weakly exogenous regressors when N is large and T is moderate. Moreover, the cost of the half-panel jackknife bias correction seem small when regressors are strictly exogenous. Hence, the jackknife corrected FE and FE-TE estimators are useful additions to the toolkit of applied researchers, particularly since these estimators are also quite easy to implement.

Figure 1: $Bias_T\left(\hat{\beta}_{FE}\right)$ and $Bias_T\left(\tilde{\beta}_{FE}\right)$ as a function of the feedback coefficient κ in Example 2, $T = 30$.



Notes: Solid black line is $Bias_T\left(\hat{\beta}_{FE}\right)$ and dotted green line is $Bias_T\left(\tilde{\beta}_{FE}\right)$. κ is on horizontal axis and the range for κ is chosen so that $\varphi = \rho + \kappa\beta = -0.99, -0.98, ..., 0, 0.01..., 0.98, 0.99$. The remaining parameters are $\beta = 1$, $\rho = 0.5$ (top panel) or 0.2 (bottom panel), and $\eta = 1$ (left panel) or 2 (right panel).

Table 1: List of Monte Carlo Experiments

	DGP:	without lagged dep. variable	with lagged dep. variable	
		$\lambda_y = 0$	$\lambda_y = 0.4$	$\lambda_y = 0.8$
δ_t	κ_x	Exp.	Exp.	Exp.
0	0	1	10	19
0	0.2	2	11	20
0	0.4	3	12	21
$0.025t$	0	4	13	22
$0.025t$	0.2	5	14	23
$0.025t$	0.4	6	15	24
$0.025t - 0.001t^2$	0	7	16	25
$0.025t - 0.001t^2$	0.2	8	17	26
$0.025t - 0.001t^2$	0.4	9	18	27

Notes: DGP with lagged dependent variable is described in Subsection 5.2 and DGP with lagged dependent variable is described in subsection 5.3. δ_t is the time effect in the y_{it} equation, κ_x is the feedback coefficient of $y_{i,t-1}$ in the x_{it} equation, and λ_y is the autoregressive coefficient for the lagged dependent variable in the y_{it} equation.

Table 2: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0$, and $\kappa_x = 0$ (Experiment 1)

	Bias (×100)				RMSE (×100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	0.05	0.06	0.07	-0.01	3.39	2.37	1.76	1.27	5.45	5.80	4.65	5.10
60	0.02	0.01	0.00	-0.04	2.35	1.65	1.24	0.91	5.95	5.70	4.20	5.50
100	0.03	0.01	0.01	-0.02	1.81	1.27	0.96	0.69	5.80	5.10	4.60	5.40
200	0.02	0.00	0.00	-0.01	1.25	0.90	0.68	0.49	5.10	5.20	4.80	5.00
500	0.02	0.01	0.01	0.00	0.80	0.57	0.44	0.30	5.10	5.30	4.80	5.25
1000	0.00	0.00	0.00	-0.01	0.59	0.41	0.31	0.22	5.75	5.45	5.20	5.40
	Half-panel jackknife FE estimator $\widetilde{\beta}_{FE}$											
30	0.01	0.05	0.07	-0.02	3.46	2.39	1.78	1.28	3.90	4.75	4.30	5.35
60	-0.02	0.01	0.00	-0.04	2.41	1.67	1.25	0.92	4.45	4.65	3.90	5.50
100	0.01	0.01	0.00	-0.02	1.85	1.28	0.97	0.70	4.70	3.90	4.70	5.35
200	0.00	0.00	0.00	-0.01	1.30	0.91	0.69	0.49	4.30	4.60	4.45	5.00
500	0.01	0.01	0.01	0.00	0.83	0.58	0.44	0.31	3.75	4.45	4.25	4.75
1000	0.00	0.00	0.00	-0.01	0.61	0.41	0.31	0.22	5.45	5.15	4.85	5.00
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	0.02	0.04	0.07	-0.02	3.43	2.40	1.79	1.29	6.30	6.15	5.70	5.40
60	0.00	0.00	-0.01	-0.04	2.37	1.66	1.25	0.91	6.15	5.60	4.40	5.75
100	0.03	0.01	0.00	-0.02	1.82	1.28	0.96	0.70	5.50	5.40	4.75	5.60
200	0.02	0.00	0.00	-0.01	1.26	0.90	0.68	0.49	5.15	5.25	4.80	5.00
500	0.02	0.01	0.01	0.00	0.80	0.58	0.44	0.30	5.00	5.45	4.90	5.10
1000	0.00	0.00	0.00	-0.01	0.59	0.41	0.31	0.22	5.95	5.40	5.40	5.40
	Half-panel jackknife FE-TE estimator $\widetilde{\beta}_{FE-TE}$											
30	-0.02	0.04	0.06	-0.02	3.52	2.42	1.81	1.29	4.50	5.35	4.95	5.60
60	-0.03	0.00	-0.01	-0.04	2.44	1.68	1.26	0.92	4.60	4.80	3.90	5.90
100	0.01	0.01	0.00	-0.02	1.87	1.28	0.97	0.70	4.80	4.10	4.95	5.30
200	0.00	0.00	0.00	-0.01	1.30	0.91	0.69	0.49	4.35	4.55	4.50	5.10
500	0.01	0.01	0.01	0.00	0.83	0.58	0.44	0.31	3.95	4.65	4.25	4.90
1000	0.00	0.00	0.00	-0.01	0.61	0.41	0.31	0.22	5.35	5.00	4.75	5.10

Notes: DGP is given by $\Delta y_{it} = \mu_i + \delta_t - (1-\lambda_y)\, y_{i,t-1} + (1-\lambda_y)\, \beta x_{it} + u_{it}$, where $x_{it} = (1-\lambda_x)\, \mu_{ix} + (1-\lambda_x)\, \kappa_x y_{i,t-1} + \lambda_x x_{i,t-1} + v_{it}$, $\beta = 0.5$, $\lambda_y = 0$, $\delta_t = 0$, $\mu_i = \mu_{ix} + \eta_{yi}$, $\eta_{yi} \sim IID\mathcal{N}(1,1)$, $u_{it} \sim IID\mathcal{N}\left(0, \sigma_{ui}^2\right)$, $\sigma_{ui}^2 = 0.5 + 0.25\eta_{ui}^2$, $\eta_{ui}^2 \sim IID\chi^2(2)$, $\lambda_x = 0.25$, $\kappa_x = 0$, $\mu_{ix} \sim IID\mathcal{N}(1,1)$, $v_{it} \sim IID\mathcal{N}\left(0, \sigma_{vi}^2\right)$, $\sigma_{vi}^2 = 0.5 + 0.25\eta_{vi}^2$, $\eta_{vi}^2 \sim IID\chi^2(2)$. $R = 2000$. FE and half-panel jackknifed FE are based on equation (93): $y_{it} = \mu_i + \beta x_{it} + u_{it}$. FE-TE and half-panel jackknifed FE-TE are based on equation (95): $y_{it} = \mu_i + \delta_t + \beta x_{it} + u_{it}$.

Figure 2: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0$, and $\kappa_x = 0$ (Experiment 1)

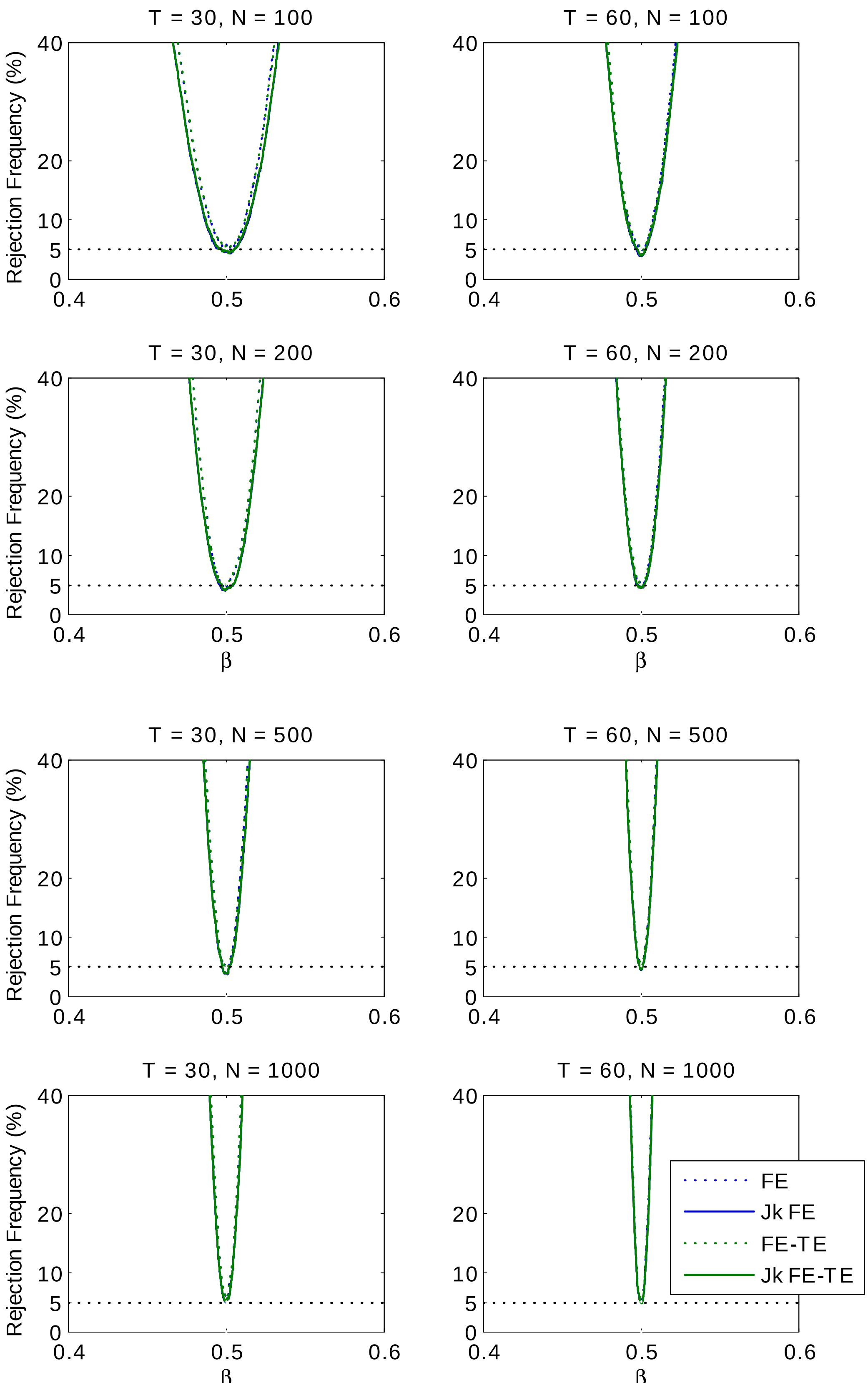


Table 3: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0$, and $\kappa_x = 0.4$ (Experiment 3)

	Bias (×100)				RMSE (×100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	-1.33	-0.63	-0.34	-0.21	3.42	2.30	1.67	1.19	8.05	6.85	5.10	5.80
60	-1.33	-0.66	-0.40	-0.23	2.54	1.66	1.20	0.86	9.85	7.80	5.50	6.95
100	-1.30	-0.65	-0.39	-0.21	2.12	1.35	0.97	0.67	12.00	9.35	7.20	7.10
200	-1.31	-0.66	-0.39	-0.20	1.75	1.05	0.73	0.49	19.70	12.90	8.85	7.85
500	-1.30	-0.64	-0.38	-0.19	1.49	0.83	0.56	0.34	40.35	23.00	16.65	10.40
1000	-1.31	-0.65	-0.38	-0.20	1.42	0.75	0.48	0.28	70.40	42.85	28.35	16.50
	Half-panel jackknife FE estimator $\tilde{\beta}_{FE}$											
30	0.01	0.05	0.06	-0.01	3.27	2.25	1.65	1.18	4.15	5.15	4.50	4.75
60	0.01	0.02	0.00	-0.03	2.28	1.55	1.14	0.84	3.85	4.45	3.85	5.80
100	0.05	0.02	0.01	-0.02	1.75	1.20	0.90	0.65	4.20	4.00	5.35	5.80
200	0.04	0.01	0.01	-0.01	1.22	0.83	0.63	0.45	3.40	4.60	4.40	5.35
500	0.07	0.02	0.02	0.00	0.78	0.54	0.41	0.28	4.15	5.10	4.95	4.65
1000	0.05	0.01	0.01	0.00	0.57	0.39	0.29	0.20	5.00	5.25	5.15	5.40
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	-1.35	-0.64	-0.35	-0.22	3.48	2.33	1.69	1.20	8.50	7.00	5.50	5.60
60	-1.34	-0.67	-0.41	-0.23	2.56	1.67	1.21	0.86	10.85	8.45	6.10	7.15
100	-1.30	-0.65	-0.39	-0.21	2.13	1.35	0.97	0.68	12.15	9.40	7.20	7.15
200	-1.31	-0.66	-0.39	-0.20	1.75	1.05	0.74	0.49	19.90	12.70	9.00	7.80
500	-1.30	-0.64	-0.38	-0.19	1.49	0.83	0.56	0.34	40.85	22.90	16.80	10.50
1000	-1.31	-0.65	-0.39	-0.20	1.42	0.75	0.48	0.28	70.35	42.65	28.60	16.40
	Half-panel jackknife FE-TE estimator $\tilde{\beta}_{FE-TE}$											
30	-0.01	0.04	0.05	-0.02	3.33	2.28	1.67	1.19	4.30	5.30	4.90	5.35
60	0.00	0.01	-0.01	-0.04	2.30	1.56	1.15	0.84	4.20	4.85	3.75	5.25
100	0.05	0.02	0.01	-0.02	1.75	1.20	0.91	0.65	4.85	4.15	5.15	5.60
200	0.04	0.01	0.01	-0.01	1.23	0.83	0.63	0.45	3.85	4.50	4.50	5.20
500	0.07	0.02	0.01	0.00	0.79	0.55	0.41	0.28	4.15	5.20	4.90	4.80
1000	0.05	0.01	0.01	0.00	0.57	0.39	0.29	0.20	5.15	5.20	4.90	5.50

Notes: $\beta = 0.5$, $\lambda_y = 0$, $\delta_t = 0$, and $\kappa_x = 0.4$. For the rest of the settings, see the notes for Table 2.

Figure 3: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0$, and $\kappa_x = 0.4$ (Experiment 3)

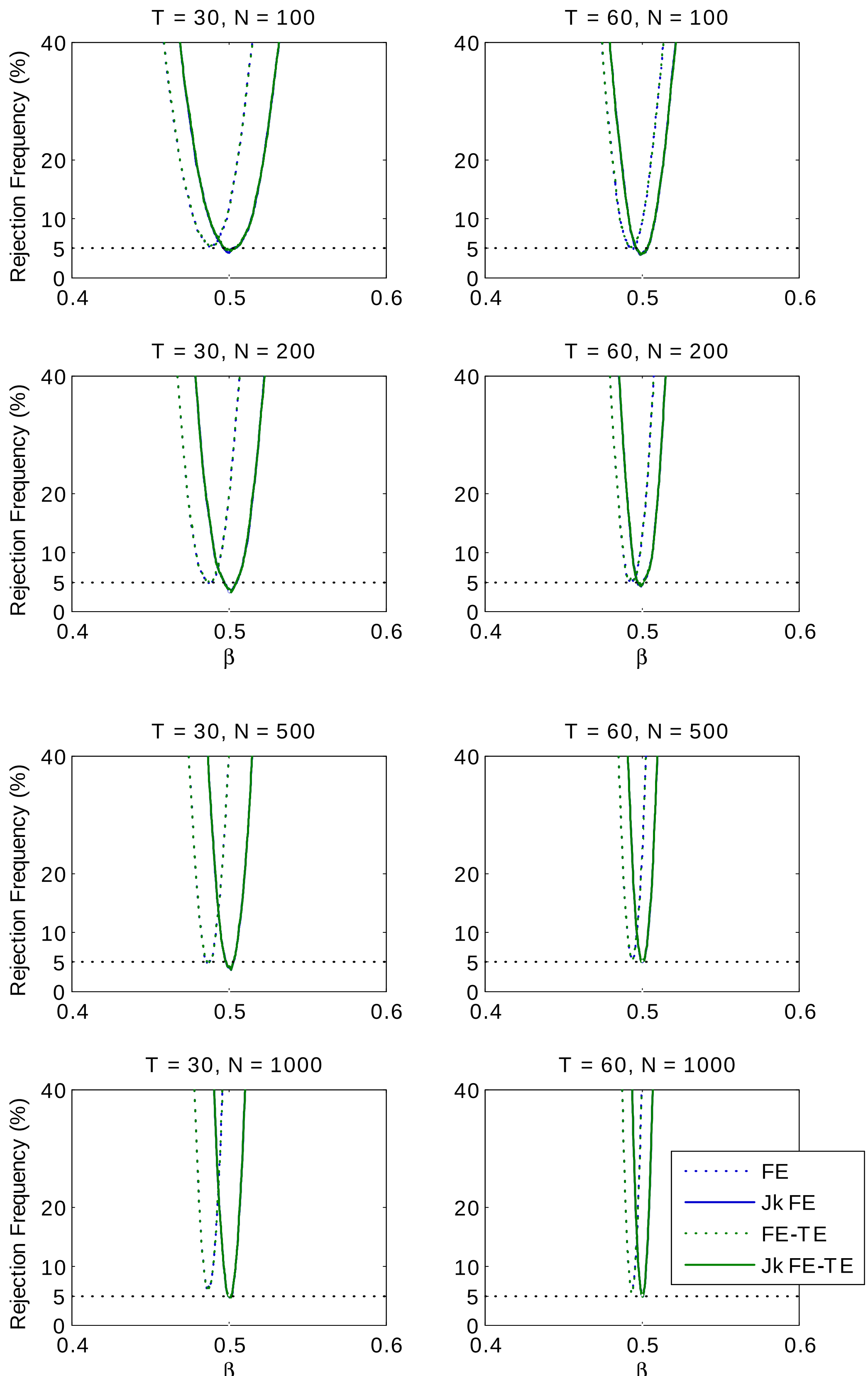


Table 4: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0$ (Experiment 7)

	Bias (×100)				RMSE (×100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	0.04	0.05	0.04	-0.16	3.40	2.50	2.45	4.98	6.00	6.70	7.30	11.75
60	0.01	0.01	0.00	-0.13	2.36	1.75	1.75	3.53	5.85	6.05	7.70	12.40
100	0.03	0.01	-0.01	-0.09	1.81	1.34	1.33	2.70	5.70	5.50	6.40	12.15
200	0.01	0.00	0.02	0.00	1.26	0.95	0.96	1.91	5.35	5.40	7.60	12.10
500	0.02	0.01	0.01	0.01	0.81	0.60	0.61	1.22	5.10	5.55	7.65	11.45
1000	0.00	0.00	0.00	0.02	0.59	0.43	0.43	0.86	6.35	6.20	8.30	11.60
	Half-panel jackknife FE estimator $\widetilde{\beta}_{FE}$											
30	-0.01	0.04	-0.01	-0.28	3.51	2.85	3.77	9.49	4.60	9.35	22.90	39.95
60	-0.02	0.00	-0.02	-0.21	2.45	2.01	2.73	6.70	4.45	9.25	25.45	39.70
100	0.01	0.01	-0.03	-0.14	1.88	1.52	2.05	5.15	4.80	8.30	23.50	41.05
200	0.00	0.00	0.03	0.00	1.32	1.09	1.48	3.65	4.25	8.75	25.00	39.55
500	0.01	0.01	0.01	0.02	0.84	0.69	0.95	2.34	4.00	8.80	24.45	41.50
1000	0.00	0.00	-0.01	0.06	0.62	0.49	0.65	1.64	5.65	9.80	23.80	41.60
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	0.02	0.04	0.07	-0.02	3.43	2.40	1.79	1.29	6.30	6.15	5.70	5.40
60	0.00	0.00	-0.01	-0.04	2.37	1.66	1.25	0.91	6.15	5.60	4.40	5.75
100	0.03	0.01	0.00	-0.02	1.82	1.28	0.96	0.70	5.50	5.40	4.75	5.60
200	0.02	0.00	0.00	-0.01	1.26	0.90	0.68	0.49	5.15	5.25	4.80	5.00
500	0.02	0.01	0.01	0.00	0.80	0.58	0.44	0.30	5.00	5.45	4.90	5.10
1000	0.00	0.00	0.00	-0.01	0.59	0.41	0.31	0.22	5.95	5.40	5.40	5.40
	Half-panel jackknife FE-TE estimator $\widetilde{\beta}_{FE-TE}$											
30	-0.02	0.04	0.06	-0.02	3.52	2.42	1.81	1.29	4.50	5.35	4.95	5.60
60	-0.03	0.00	-0.01	-0.04	2.44	1.68	1.26	0.92	4.60	4.80	3.90	5.90
100	0.01	0.01	0.00	-0.02	1.87	1.28	0.97	0.70	4.80	4.10	4.95	5.30
200	0.00	0.00	0.00	-0.01	1.30	0.91	0.69	0.49	4.35	4.55	4.50	5.10
500	0.01	0.01	0.01	0.00	0.83	0.58	0.44	0.31	3.95	4.65	4.25	4.90
1000	0.00	0.00	0.00	-0.01	0.61	0.41	0.31	0.22	5.35	5.00	4.75	5.10

Notes: $\beta = 0.5$, $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0$. FE and half-panel jackknife FE are based on equation (94): $y_{it} = \mu_i + gt + \beta x_{it} + u_{it}$. FE-TE and half-panel jackknife FE-TE are based on equation (95): $y_{it} = \mu_i + \delta_t + \beta x_{it} + u_{it}$. For the rest of the settings, see the notes for Table 2.

Figure 4: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0$ (Experiment 7)

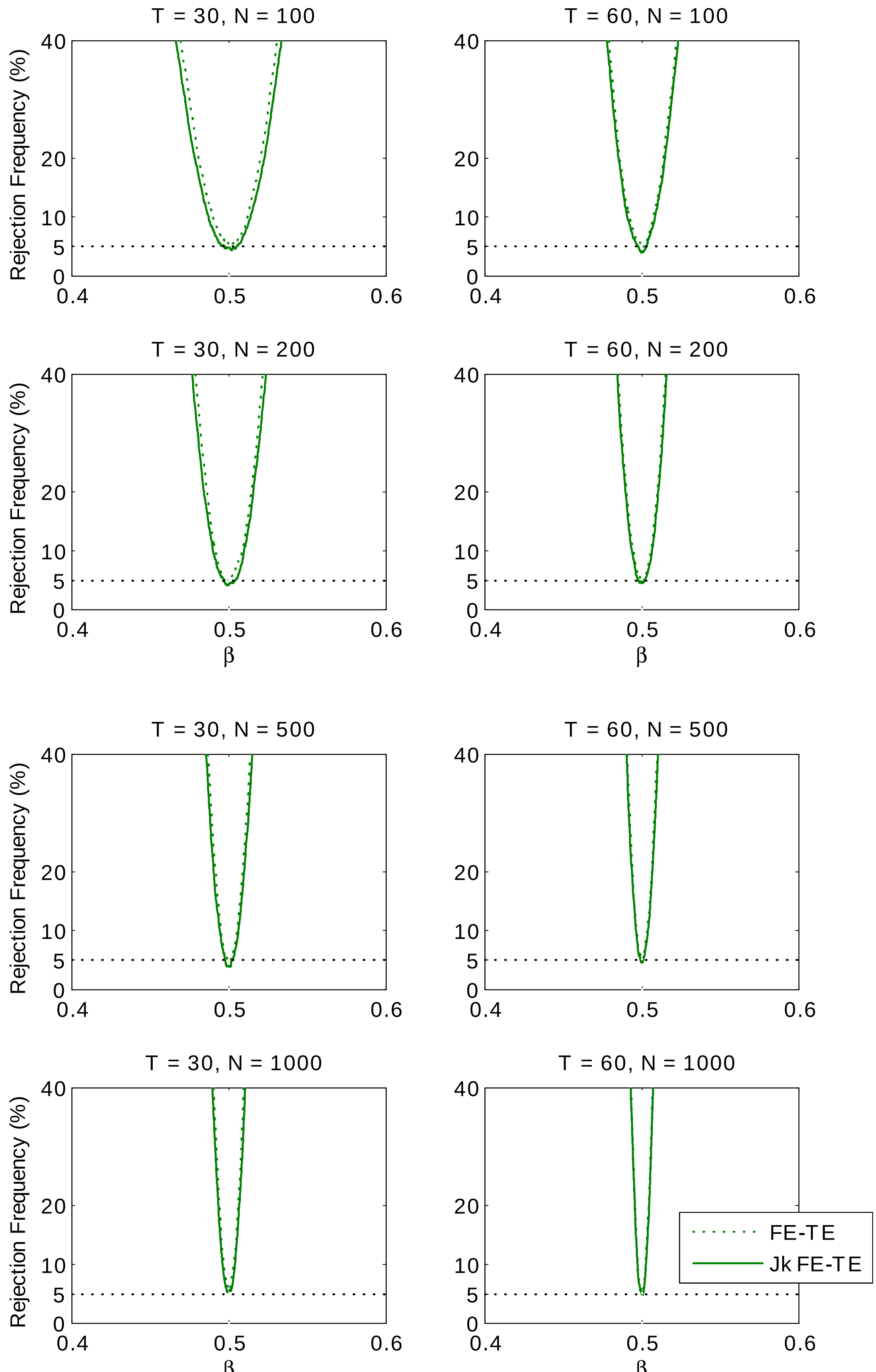


Table 5: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 9)

	Bias (×100)				RMSE (×100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	-1.19	2.22	19.52	126.98	3.39	3.24	19.69	127.04	7.70	19.85	100.00	100.00
60	-1.16	2.18	19.44	126.86	2.47	2.72	19.52	126.89	9.15	30.45	100.00	100.00
100	-1.13	2.20	19.41	126.79	2.02	2.53	19.47	126.81	10.05	46.85	100.00	100.00
200	-1.13	2.19	19.41	126.73	1.62	2.36	19.44	126.74	15.60	74.15	100.00	100.00
500	-1.11	2.21	19.42	126.75	1.34	2.28	19.43	126.76	31.85	97.70	100.00	100.00
1000	-1.13	2.20	19.40	126.74	1.25	2.24	19.40	126.74	57.35	100.00	100.00	100.00
	Half-panel jackknife FE estimator $\tilde{\beta}_{FE}$											
30	0.36	5.60	38.39	234.48	3.38	6.28	38.63	234.58	5.35	66.35	100.00	100.00
60	0.37	5.52	38.28	234.32	2.35	5.87	38.41	234.36	4.75	88.10	100.00	100.00
100	0.42	5.53	38.22	234.20	1.82	5.74	38.29	234.23	5.15	97.65	100.00	100.00
200	0.40	5.53	38.22	234.08	1.30	5.63	38.26	234.09	5.20	100.00	100.00	100.00
500	0.43	5.54	38.21	234.11	0.90	5.58	38.23	234.11	6.95	100.00	100.00	100.00
1000	0.41	5.53	38.18	234.10	0.71	5.55	38.19	234.10	11.30	100.00	100.00	100.00
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	-1.35	-0.64	-0.35	-0.22	3.48	2.33	1.69	1.20	8.50	7.00	5.50	5.60
60	-1.34	-0.67	-0.41	-0.23	2.56	1.67	1.21	0.86	10.85	8.45	6.10	7.15
100	-1.30	-0.65	-0.39	-0.21	2.13	1.35	0.97	0.68	12.15	9.40	7.20	7.15
200	-1.31	-0.66	-0.39	-0.20	1.75	1.05	0.74	0.49	19.90	12.70	9.00	7.80
500	-1.30	-0.64	-0.38	-0.19	1.49	0.83	0.56	0.34	40.85	22.90	16.80	10.50
1000	-1.31	-0.65	-0.39	-0.20	1.42	0.75	0.48	0.28	70.35	42.65	28.60	16.40
	Half-panel jackknife FE-TE estimator $\tilde{\beta}_{FE-TE}$											
30	-0.01	0.04	0.05	-0.02	3.33	2.28	1.67	1.19	4.30	5.30	4.90	5.35
60	0.00	0.01	-0.01	-0.04	2.30	1.56	1.15	0.84	4.20	4.85	3.75	5.25
100	0.05	0.02	0.01	-0.02	1.75	1.20	0.91	0.65	4.85	4.15	5.15	5.60
200	0.04	0.01	0.01	-0.01	1.23	0.83	0.63	0.45	3.85	4.50	4.50	5.20
500	0.07	0.02	0.01	0.00	0.79	0.55	0.41	0.28	4.15	5.20	4.90	4.80
1000	0.05	0.01	0.01	0.00	0.57	0.39	0.29	0.20	5.15	5.20	4.90	5.50

Notes: $\beta = 0.5$, $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$ and $\kappa_x = 0.4$. For the regression equations, see the notes for Table 4. For the rest of the settings, see the notes for Table 2.

Figure 5: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 9)

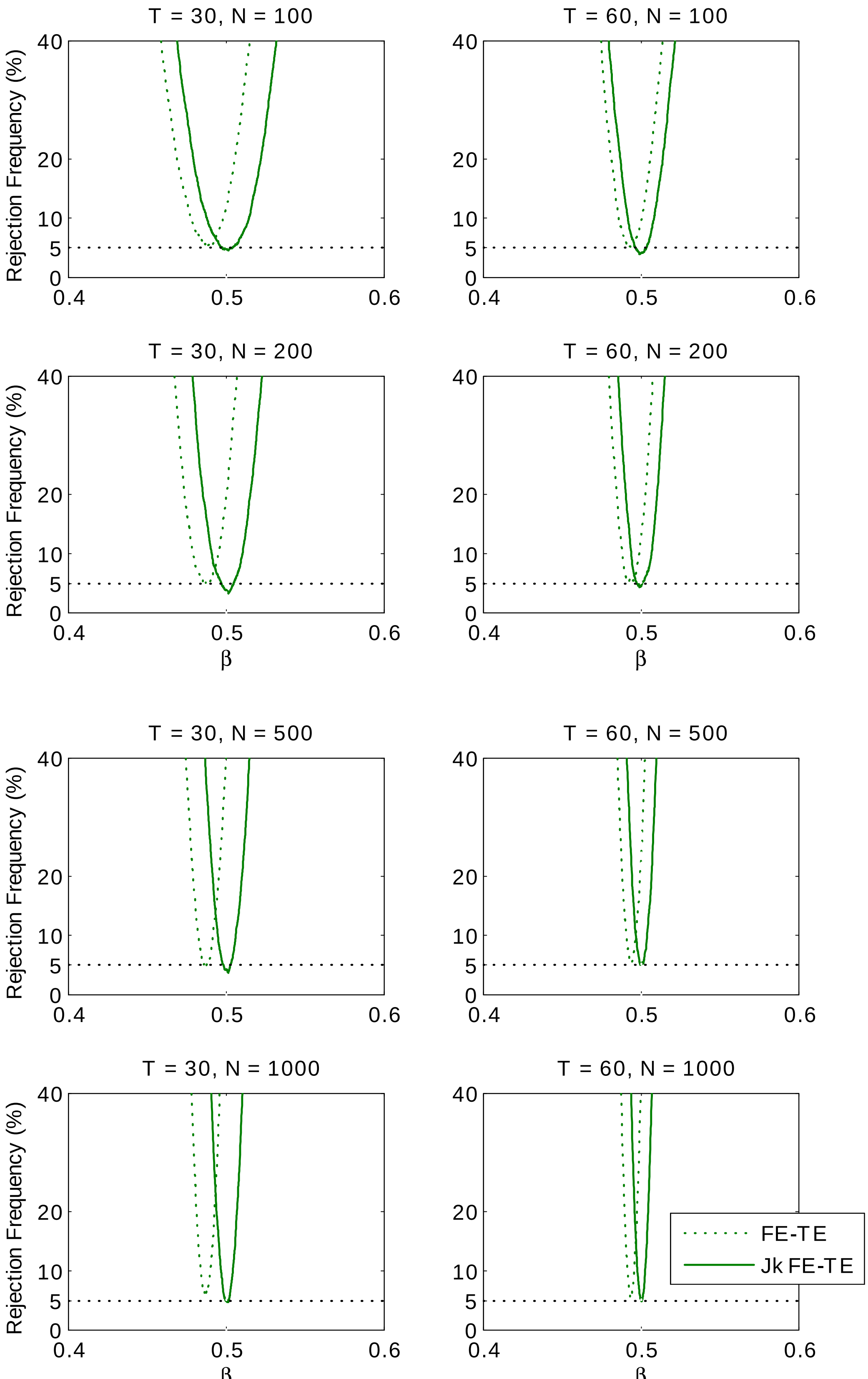


Table 6: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0$, and $\kappa_x = 0.4$ (Experiment 12)

	Bias (×100)				RMSE (×100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	-3.87	-1.89	-1.07	-0.60	6.33	4.07	2.90	2.02	15.35	10.40	7.20	6.35
60	-3.86	-1.93	-1.16	-0.63	5.18	3.14	2.18	1.50	23.00	13.65	8.95	8.10
100	-3.79	-1.91	-1.14	-0.60	4.62	2.70	1.85	1.22	32.40	18.85	11.90	9.50
200	-3.80	-1.91	-1.14	-0.58	4.22	2.33	1.53	0.94	55.00	30.45	18.80	12.15
500	-3.78	-1.88	-1.13	-0.57	3.96	2.07	1.31	0.73	89.75	60.45	39.80	22.65
1000	-3.81	-1.90	-1.14	-0.58	3.90	2.00	1.23	0.66	99.55	87.75	69.20	42.25
	Half-panel jackknife FE estimator $\widetilde{\beta}_{FE}$											
30	0.57	0.24	0.15	0.01	5.85	3.85	2.80	1.97	4.40	4.95	4.40	4.50
60	0.51	0.17	0.05	-0.04	4.09	2.66	1.93	1.40	4.85	4.10	3.75	5.40
100	0.61	0.17	0.07	-0.01	3.14	2.04	1.53	1.08	4.60	3.90	5.10	5.60
200	0.58	0.15	0.07	0.01	2.25	1.42	1.07	0.76	5.10	4.25	4.45	5.20
500	0.62	0.17	0.07	0.01	1.52	0.94	0.70	0.47	6.45	5.60	5.00	4.55
1000	0.58	0.16	0.06	0.01	1.17	0.68	0.49	0.34	9.30	6.15	5.55	4.95
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	-3.90	-1.91	-1.09	-0.61	6.42	4.12	2.94	2.05	16.75	10.90	7.55	6.40
60	-3.87	-1.95	-1.17	-0.64	5.21	3.16	2.20	1.51	23.20	13.80	9.15	8.75
100	-3.79	-1.90	-1.14	-0.60	4.63	2.70	1.85	1.22	32.40	18.95	12.30	9.60
200	-3.79	-1.91	-1.13	-0.58	4.22	2.33	1.53	0.94	54.65	30.85	19.15	11.95
500	-3.78	-1.88	-1.13	-0.57	3.96	2.07	1.31	0.73	89.80	60.70	39.70	22.75
1000	-3.81	-1.90	-1.14	-0.58	3.90	2.00	1.23	0.66	99.55	87.70	69.20	41.95
	Half-panel jackknife FE-TE estimator $\widetilde{\beta}_{FE-TE}$											
30	0.53	0.22	0.14	0.00	5.96	3.91	2.84	2.00	4.95	5.50	4.65	4.80
60	0.49	0.16	0.04	-0.04	4.12	2.67	1.94	1.41	5.40	4.40	4.20	5.35
100	0.60	0.17	0.07	-0.01	3.16	2.05	1.53	1.08	4.95	4.05	5.30	5.65
200	0.58	0.16	0.07	0.01	2.26	1.42	1.07	0.76	5.15	4.55	4.40	5.10
500	0.62	0.17	0.07	0.01	1.52	0.95	0.70	0.47	6.35	5.70	5.20	4.95
1000	0.57	0.16	0.06	0.01	1.17	0.68	0.49	0.34	9.35	6.20	5.50	5.00

Notes: $\beta = 0.5$, $\lambda_y = 0.4$, $\delta_t = 0$, and $\kappa_x = 0.4$. FE and half-panel jackknife FE are based on equation (96): $\Delta y_{it} = \mu_i + \phi y_{i,t-1} + bx_{it} + e_{it}$. FE-TE and half-panel jackknife FE-TE are based on equation (100): $\Delta y_{it} = \mu_i + \delta_t + \phi y_{i,t-1} + bx_{it} + e_{it}$. For the rest of the settings, see the notes for Table 2.

Figure 6: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0$, and $\kappa_x = 0.4$ (Experiment 12)

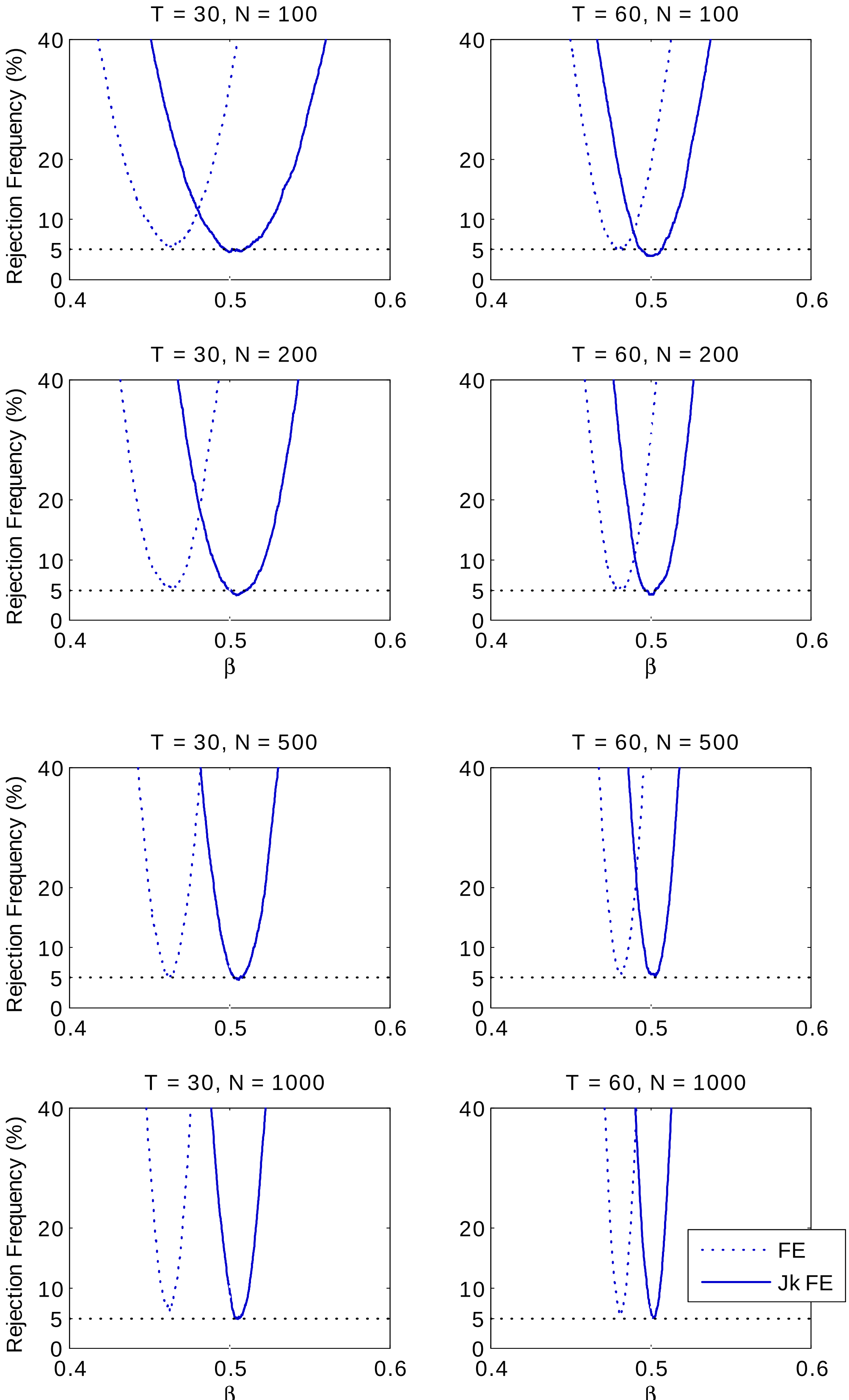


Table 7: Bias ($\times$100), RMSE ($\times$100), and Size (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 18)

	Bias ($\times$100)				RMSE ($\times$100)				Size (%)			
(N,T)	30	60	100	200	30	60	100	200	30	60	100	200
	FE estimator $\widehat{\beta}_{FE}$											
30	-3.42	6.38	48.33	167.31	6.13	7.64	48.58	167.33	13.75	36.75	100.00	100.00
60	-3.36	6.31	48.16	167.27	4.84	6.95	48.29	167.28	18.00	60.90	100.00	100.00
100	-3.26	6.36	48.15	167.26	4.22	6.74	48.23	167.27	25.25	81.65	100.00	100.00
200	-3.25	6.37	48.16	167.24	3.74	6.56	48.20	167.25	42.65	98.15	100.00	100.00
500	-3.22	6.40	48.15	167.25	3.43	6.48	48.17	167.25	78.55	100.00	100.00	100.00
1000	-3.24	6.38	48.13	167.24	3.36	6.42	48.14	167.24	96.40	100.00	100.00	100.00
	Half-panel jackknife FE estimator $\tilde{\beta}_{FE}$											
30	1.88	21.37	2069.45	-1124.19	6.43	22.46	68163.06	2334.83	5.95	94.60	50.50	90.05
60	1.81	20.99	606.02	-1049.33	4.61	21.53	11466.34	1093.71	6.95	99.70	70.05	99.55
100	1.91	20.91	1092.30	-1024.70	3.74	21.24	5040.90	1045.91	8.65	100.00	85.95	100.00
200	1.88	20.88	974.57	-1006.88	2.94	21.04	1361.11	1016.38	12.05	100.00	97.70	100.00
500	1.93	20.86	898.05	-998.06	2.41	20.93	920.89	1001.69	26.25	100.00	100.00	100.00
1000	1.88	20.82	875.31	-994.51	2.16	20.85	884.36	996.31	44.25	100.00	100.00	100.00
	FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30	-3.90	-1.91	-1.09	-0.61	6.42	4.12	2.94	2.05	16.75	10.90	7.55	6.40
60	-3.87	-1.95	-1.17	-0.64	5.21	3.16	2.20	1.51	23.20	13.80	9.15	8.75
100	-3.79	-1.90	-1.14	-0.60	4.63	2.70	1.85	1.22	32.40	18.95	12.30	9.60
200	-3.79	-1.91	-1.13	-0.58	4.22	2.33	1.53	0.94	54.65	30.85	19.15	11.95
500	-3.78	-1.88	-1.13	-0.57	3.96	2.07	1.31	0.73	89.80	60.70	39.70	22.75
1000	-3.81	-1.90	-1.14	-0.58	3.90	2.00	1.23	0.66	99.55	87.70	69.20	41.95
	Half-panel jackknife FE-TE estimator $\tilde{\beta}_{FE-TE}$											
30	0.53	0.22	0.14	0.00	5.96	3.91	2.84	2.00	4.95	5.50	4.65	4.80
60	0.49	0.16	0.04	-0.04	4.12	2.67	1.94	1.41	5.40	4.40	4.20	5.35
100	0.60	0.17	0.07	-0.01	3.16	2.05	1.53	1.08	4.95	4.05	5.30	5.65
200	0.58	0.16	0.07	0.01	2.26	1.42	1.07	0.76	5.15	4.55	4.40	5.10
500	0.62	0.17	0.07	0.01	1.52	0.95	0.70	0.47	6.35	5.70	5.20	4.95
1000	0.57	0.16	0.06	0.01	1.17	0.68	0.49	0.34	9.35	6.20	5.50	5.00

Notes: $\beta = 0.5$, $\lambda_y = 0.4$, $\delta_t = \delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$. FE and half-panel jackknife FE are based on equation (97): $\Delta y_{it} = \mu_i + gt + \phi y_{i,t-1} + bx_{it} + e_{it}$. FE-TE and half-panel jackknife FE-TE are based on equation (100): $\Delta y_{it} = \mu_i + \delta_t + \phi y_{i,t-1} + bx_{it} + e_{it}$. For the rest of the settings, see the notes for Table 2.

Figure 7: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 18)

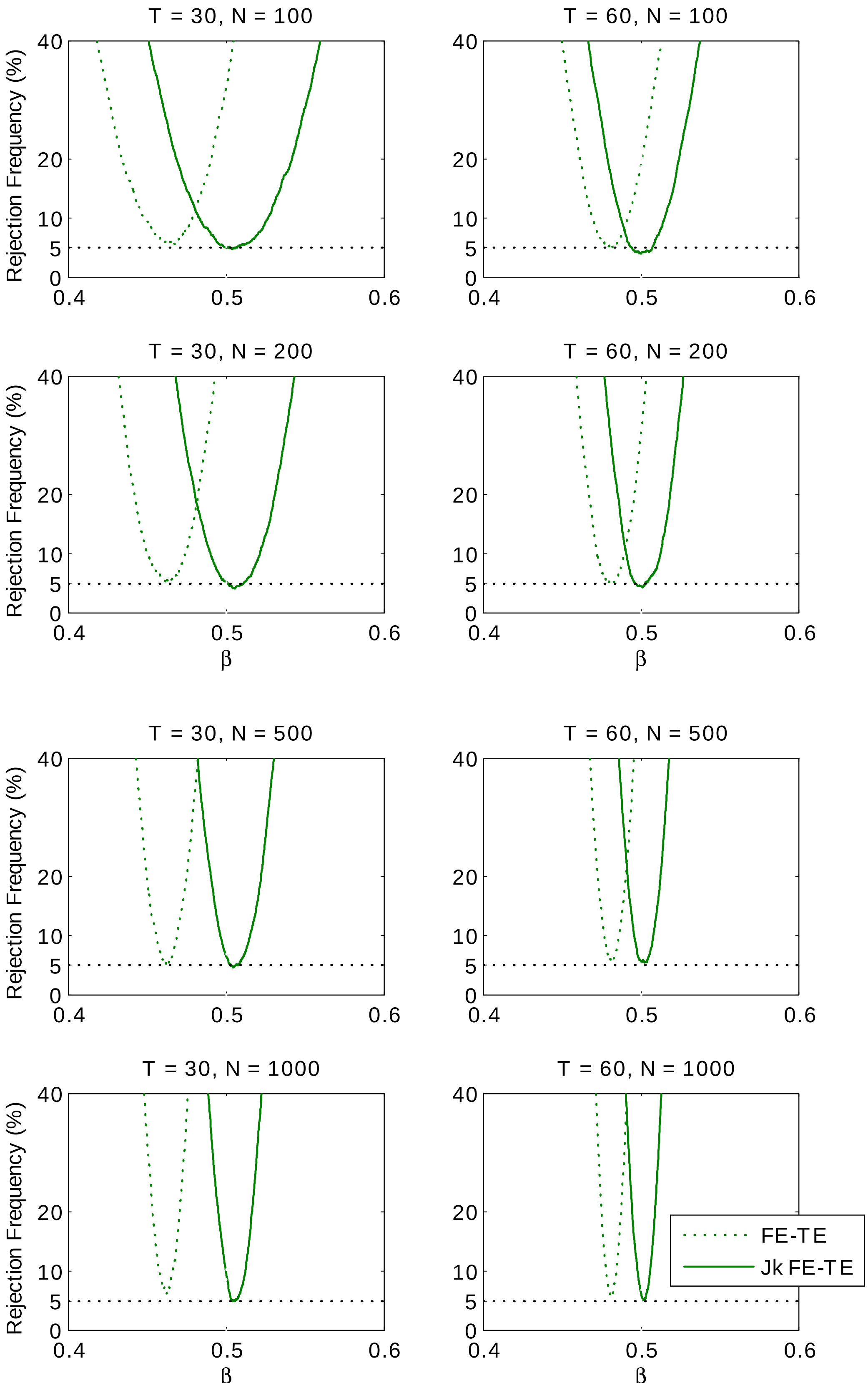


Table 8: Bias (×100), RMSE (×100), and Size (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 18, unbalanced samples)

		Bias (×100)				RMSE (×100)				Size (%)			
	$\max T$	30	60	100	200	30	60	100	200	30	60	100	200
	avg$T$	22	44	73.5	147	22	44	73.5	147	22	44	73.5	147
N	$\min T$	10	20	34	68	10	20	34	68	10	20	34	68
		FE estimator $\widehat{\beta}_{FE}$											
30		-4.95	3.41	36.57	156.99	7.68	5.86	36.96	157.03	17.20	13.15	100.00	100.00
60		-4.83	3.42	36.40	156.83	6.28	4.71	36.61	156.85	24.75	20.70	100.00	100.00
100		-4.68	3.51	36.43	156.80	5.62	4.34	36.56	156.81	36.05	30.85	100.00	100.00
200		-4.67	3.53	36.45	156.77	5.15	3.94	36.51	156.78	58.80	54.00	100.00	100.00
500		-4.61	3.54	36.44	156.80	4.82	3.72	36.47	156.80	92.15	88.25	100.00	100.00
1000		-4.64	3.48	36.40	156.77	4.75	3.58	36.41	156.77	99.90	99.25	100.00	100.00
		Half-panel jackknife FE estimator $\widetilde{\beta}_{FE}$											
30		1.43	15.12	136.85	-1283.45	7.60	16.70	5844.08	8538.55	5.20	68.60	98.65	59.05
60		1.41	14.83	244.64	-1511.99	5.31	15.64	251.41	2133.35	5.70	91.50	100.00	89.45
100		1.50	14.83	238.33	-1374.06	4.20	15.32	242.18	1476.16	6.40	98.95	100.00	98.30
200		1.41	14.81	235.28	-1302.76	3.11	15.04	237.16	1329.32	7.25	100.00	100.00	100.00
500		1.47	14.76	232.56	-1279.45	2.31	14.86	233.24	1289.08	13.25	100.00	100.00	100.00
1000		1.41	14.67	231.48	-1267.79	1.89	14.72	231.83	1272.54	21.50	100.00	100.00	100.00
		FE-TE estimator $\widehat{\beta}_{FE-TE}$											
30		-5.28	-2.64	-1.61	-0.86	7.96	5.05	3.63	2.47	20.40	13.45	9.20	7.65
60		-5.22	-2.60	-1.62	-0.84	6.60	3.88	2.73	1.80	28.45	16.35	11.30	9.30
100		-5.06	-2.53	-1.53	-0.79	5.94	3.40	2.32	1.46	41.45	23.55	15.85	10.15
200		-5.07	-2.53	-1.55	-0.79	5.51	2.97	1.95	1.18	67.30	38.70	25.55	15.70
500		-5.02	-2.52	-1.53	-0.78	5.21	2.72	1.71	0.96	96.85	73.90	52.90	29.80
1000		-5.07	-2.56	-1.55	-0.79	5.16	2.66	1.64	0.89	99.95	96.10	82.40	53.10
		Half-panel jackknife FE-TE estimator $\widetilde{\beta}_{FE-TE}$											
30		0.39	0.24	0.08	-0.03	7.39	4.70	3.45	2.38	5.45	5.80	5.15	5.20
60		0.38	0.20	0.02	-0.03	5.04	3.14	2.34	1.64	5.30	4.25	4.45	5.50
100		0.50	0.23	0.10	0.00	3.85	2.50	1.85	1.27	4.75	4.50	5.05	4.85
200		0.43	0.23	0.09	0.01	2.73	1.72	1.26	0.90	5.25	4.70	4.20	5.20
500		0.50	0.23	0.10	0.01	1.80	1.13	0.83	0.56	5.65	4.65	4.20	5.10
1000		0.43	0.18	0.07	0.01	1.30	0.81	0.59	0.41	6.70	4.95	4.50	5.10

Notes: $\beta = 0.5$, $\lambda_y = 0.8$, $\delta_t = 0$, and $\kappa_x = 0.4$. For the regression equations, see the notes for Table 7. For the rest of the settings, see the notes for Table 2.

Figure 8: Rejection frequency (%) at 5% nominal level with $\lambda_y = 0.4$, $\delta_t = 0.025t - 0.001t^2$, and $\kappa_x = 0.4$ (Experiment 18, unbalanced samples)

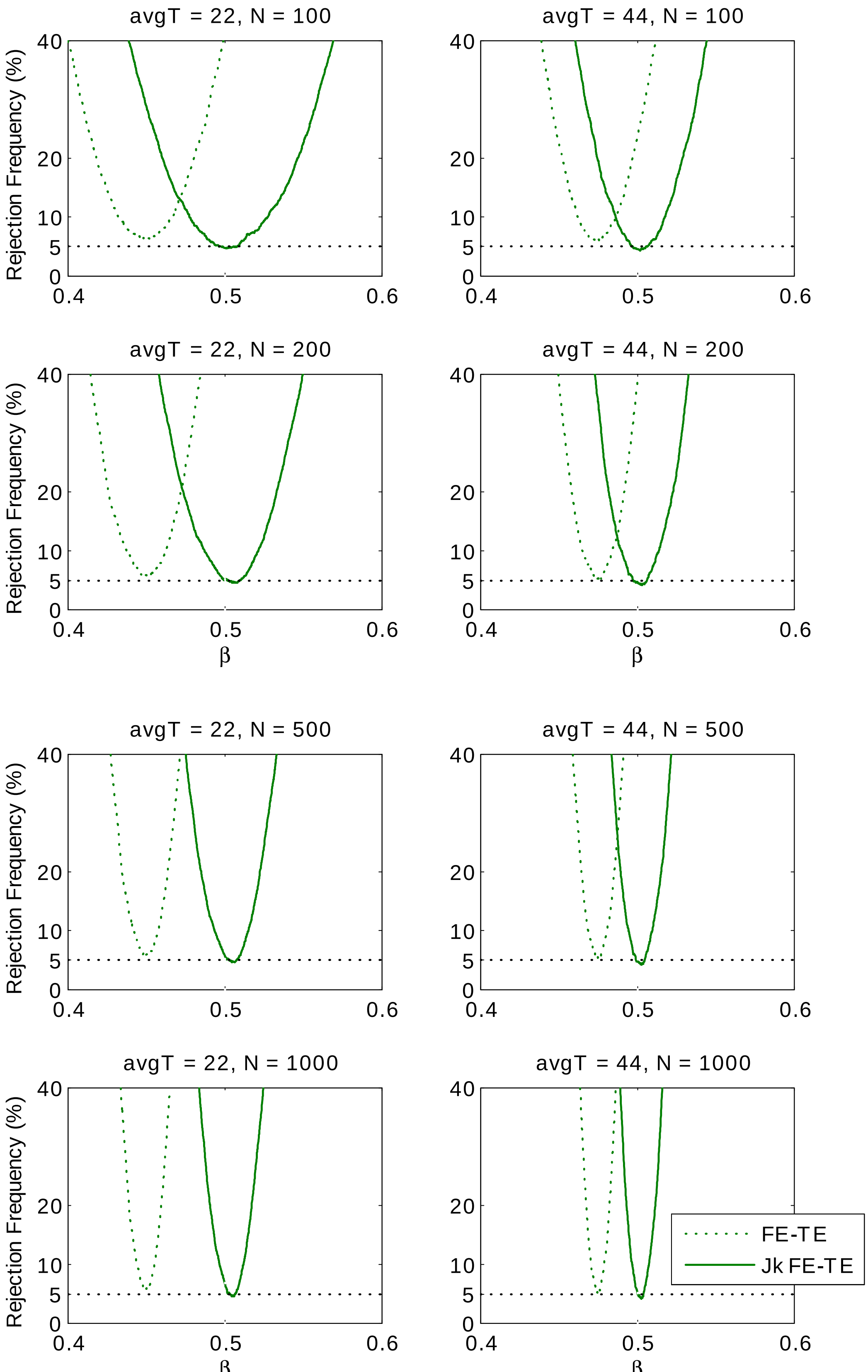


Table 9: Berger et al. (2013) and half-panel jackknife estimates for the effects of US interventions on trade with the US and the rest of the world

	ln normalized imports from the US		ln normalized imports from the world		ln normalized exports to the US		ln normalized exports to the world	
	(1.a)	(1.b)	(2.a)	(2.b)	(3.a)	(3.b)	(4.a)	(4.b)
	BENS[1]	Jackknife FE-TE	BENS[1]	Jackknife FE-TE	BENS[1]	Jackknife FE-TE	BENS[1]	Jackknife FE-TE
US influence	0.293***	0.450***	-0.009	0.041	0.058	0.081	0.000	0.006
	(0.109)	(0.068)	(0.045)	(0.031)	(0.122)	(0.096)	(0.052)	(0.035)
ln per capita income	0.296**	0.469***	0.129	0.290***	1.234***	1.213***	0.647***	0.631***
	(0.148)	(0.106)	(0.111)	(0.072)	(0.239)	(0.163)	(0.134)	(0.075)
Soviet intervention	-1.067**	-1.819***	-0.080	-0.143**	-0.682**	-1.197***	-0.082	-0.154*
	(0.430)	(0.243)	(0.102)	(0.066)	(0.307)	(0.323)	(0.100)	(0.081)
Leader turnover	0.001	0.006	0.026	0.040**	0.028	0.064	0.037*	0.053***
	(0.037)	(0.035)	(0.018)	(0.018)	(0.039)	(0.043)	(0.022)	(0.020)
Leader tenure	0.003	-0.002	0.005**	0.005**	0.013**	0.020***	0.006*	0.008***
	(0.008)	(0.004)	(0.003)	(0.002)	(0.007)	(0.005)	(0.004)	(0.002)
Democracy	0.121*	0.226***	0.069	0.152***	0.065	0.136**	0.082	0.114***
	(0.073)	(0.048)	(0.053)	(0.031)	(0.094)	(0.062)	(0.058)	(0.031)
ln distance	-0.277***	-0.397***	-0.127***	-0.149***	-0.214***	-0.293***	-0.143***	-0.177***
	(0.065)	(0.047)	(0.026)	(0.016)	(0.079)	(0.055)	(0.029)	(0.015)
Contiguous border	2.952*	3.773***	-0.274	-0.267	1.965	3.285**	-0.104	-0.094
	(1.709)	(1.262)	(0.516)	(0.292)	(2.648)	(1.623)	(0.415)	(0.323)
Common language	1.430	5.087***	-0.847**	0.719***	3.676***	5.810***	0.145	0.880***
	(1.204)	(0.793)	(0.343)	(0.227)	(1.280)	(0.960)	(0.355)	(0.242)
GATT participant	0.057	0.840**	-0.075	-0.157***	0.365	0.872**	-0.086	-0.202***
	(0.549)	(0.378)	(0.055)	(0.032)	(0.561)	(0.370)	(0.063)	(0.036)
Regional trade	-1.216**	-2.019***	-1.200***	-1.579***	-1.283	-0.956*	-1.126***	-1.553***
agreement	(0.532)	(0.454)	(0.205)	(0.121)	(0.882)	(0.510)	(0.266)	(0.145)
Observations	4,149	4,110	4,149	4,110	3,922	3,886	3,922	3,886
N	131	131	131	131	131	128	131	128
$\max T$	43	43	43	43	43	42	43	42
avg$T$	31.7	31.4	31.7	31.4	29.9	30.4	29.9	30.4
$\min T$	3	2	3	2	1	2	1	2

Notes: 1. BENS estimates, under columns (1.a), (2.a), (3.a), and (4.a), are taken from columns (3)–(6) in Table 1 of Berger, Easterly, Nunn, and Satyanath (2013). The remaining columns are the half-panel jackknife bias-corrected FE-TE estimates. For the jackknife FE-TE, we drop the first observations for countries with odd numbers of observations. All regressions include country fixed effects and year time effects. The standard errors of the jackknife FE-TE estimates (in parentheses) are computed according to equations (55) and (65).

Table 10: Alternative panel data estimates of the relationship between abortion rates and crime: A static formulation

	ln (violent crime per capita)			ln (property crime per capita)			ln (murder per capita)		
	(1.a) Donohue & Levitt (2001)	(1.b) FE-TE	(1.c) Jackknife FE-TE	(2.a) Donohue & Levitt (2001)	(2.b) FE-TE	(2.c) Jackknife FE-TE	(3.a) Donohue & Levitt (2001)	(3.b) FE-TE	(3.c) Jackknife FE-TE
"Effective" abortion rate	-0.129***	-0.130***	-0.209***	-0.091***	-0.091***	-0.092***	-0.121**	-0.131***	-0.278***
×100	(0.024)	(0.022)	(0.038)	(0.018)	(0.010)	(0.019)	(0.047)	(0.044)	(0.084)
ln (prisoners per capita)	-0.027	-0.124***	-0.272***	-0.159***	-0.198***	-0.272***	-0.231***	-0.264***	-0.330**
$(t-1)$	(0.044)	(0.041)	(0.074)	(0.036)	(0.028)	(0.046)	(0.080)	(0.083)	(0.144)
ln (police per capita)	-0.028	0.132**	-0.108	-0.049	0.071*	-0.134**	-0.300***	0.234*	0.119
$(t-1)$	(0.045)	(0.062)	(0.093)	(0.045)	(0.042)	(0.068)	(0.109)	(0.123)	(0.222)
State unemployment rate	0.069	0.082	0.728	1.310***	1.686***	2.460***	0.968	-0.885	-1.943
(percent unemployed)	(0.505)	(0.551)	(0.741)	(0.389)	(0.284)	(0.402)	(0.794)	(0.987)	(1.355)
ln state income per capita	0.049	0.723***	1.666***	0.084	0.366***	0.600***	-0.098	-0.661	-0.537
	(0.213)	(0.249)	(0.339)	(0.162)	(0.111)	(0.193)	(0.465)	(0.573)	(0.947)
Poverty rate (percent	-0.000	-0.001	0.003	-0.001	-0.001	-0.003	-0.005	0.001	-0.001
below poverty line)	(0.002)	(0.003)	(0.004)	(0.001)	(0.002)	(0.002)	(0.004)	(0.005)	(0.006)
AFDC generosity $(t-15)$	0.008	0.008	0.011	0.002	0.008**	0.013**	-0.000	-0.017	-0.010
	(0.005)	(0.007)	(0.011)	(0.004)	(0.004)	(0.006)	(0.000)	(0.011)	(0.018)
Shall-issue concealed	-0.004	-0.029*	-0.014	0.039***	0.035***	0.058***	-0.015	0.007	0.031
weapons law	(0.012)	(0.018)	(0.025)	(0.011)	(0.011)	(0.015)	(0.032)	(0.031)	(0.049)
Beer consumption per	0.004	0.006	-0.005	0.004	0.006***	0.008***	0.006	-0.003	0.004
capita (gallons)	(0.003)	(0.005)	(0.005)	(0.003)	(0.002)	(0.003)	(0.008)	(0.006)	(0.008)
Observations	663	624	576	663	624	576	663	624	576
N	51	48	48	51	48	48	51	48	48
T	13	13	12	13	13	12	13	13	12

Notes: The regression equation is (102). Columns (1.a), (2.a), and (3.a) are taken from Columns (2), (4), and (6) from Table IV in Donohue and Levitt (2001). Columns (1.b), (2.b), and (3.b) are the FE-TE estimates. Columns (1.c), (2.c), and (3.c) are the half-panel jackknife bias-correction estimates. The FE-TE and jackknife FE-TE estimates exclude the observations for the District of Columbia, Alaska, and Hawaii and use only the data for the 48 contiguous states. Since we need an even number of time series observations to compute the half jackknife FE-TE estimates, we dropped the first observations (1985) and based the estimates on the sample period 1986 – 1997 (which gives $T = 12$). All regressions include country fixed effects and year time effects. The standard errors of the FE-TE estimates (in parentheses) are computed by the Eicker–Huber–White estimator. The standard errors of the jackknife FE-TE estimates are computed according to equations (55) and (65).

Table 11: Alternative panel data estimates of the relationship between abortion rates and crime: A dynamic formulation

	ln (violent crime per capita)		ln (property crime per capita)		ln (murder per capita)	
	(1.a) FE-TE	(1.b) Jackknife FE-TE	(2.a) FE-TE	(2.b) Jackknife FE-TE	(3.a) FE-TE	(3.b) Jackknife FE-TE
"Effective" abortion rate ×100	-0.067***	-0.078***	-0.039***	-0.012	-0.147***	-0.246***
	(0.013)	(0.023)	(0.008)	(0.013)	(0.043)	(0.077)
ln (prisoners per capita) $(t-1)$	-0.076***	-0.160***	-0.073***	-0.099***	-0.249***	-0.251*
	(0.028)	(0.049)	(0.018)	(0.029)	(0.087)	(0.142)
ln (police per capita) $(t-1)$	-0.095***	-0.269***	-0.025	-0.142***	0.228*	0.074
	(0.035)	(0.061)	(0.027)	(0.044)	(0.127)	(0.224)
State unemployment rate	-0.095	-0.012	0.545**	0.562**	-0.836	-2.516*
(percent unemployed)	(0.344)	(0.447)	(0.212)	(0.284)	(1.082)	(1.310)
ln state income per capita	0.541***	0.572***	0.292***	0.388***	-1.006	-0.694
	(0.149)	(0.188)	(0.097)	(0.135)	(0.708)	(0.957)
Poverty rate (percent	0.002	0.003	-0.000	-0.000	0.001	-0.000
below poverty line)	(0.002)	(0.002)	(0.001)	(0.002)	(0.005)	(0.006)
AFDC generosity $(t-15)$	0.003	0.006	0.005**	0.009**	-0.013	-0.009
	(0.004)	(0.006)	(0.003)	(0.004)	(0.011)	(0.017)
Shall-issue concealed	-0.002	0.015	0.018**	0.030***	0.009	0.029
weapons law	(0.001)	(0.017)	(0.008)	(0.011)	(0.032)	(0.048)
Beer consumption per	-0.003	-0.007	0.001	0.001	-0.003	0.005
capita (gallons)	(0.002)	(0.004)	(0.002)	(0.002)	(0.006)	(0.008)
ln (crime per capita) $(t-1)$	0.716***	0.857***	0.683***	0.883***	0.118*	0.306***
	(0.037)	(0.056)	(0.045)	(0.062)	(0.069)	(0.094)
Observations	576	576	576	576	576	576
N	48	48	48	48	48	48
T	12	12	12	12	12	12

Notes: The regression equation is (103). Columns (1.a), (2.a), and (3.a) are the FE-TE estimates. Columns (1.b), (2.b), and (3.b) are the half-panel jackknife bias-correction estimates. See also the notes to Table 10.

Table 12: FE-TE and jackknife FE-TE estimates of the long-run coefficients

	ln (violent crime per capita)		ln (property crime per capita)		ln (murder per capita)	
	(1.a) FE-TE	(1.b) Jackknife FE-TE	(2.a) FE-TE	(2.b) Jackknife FE-TE	(3.a) FE-TE	(3.b) Jackknife FE-TE
"Effective" abortion rate ×100	-0.235***	-0.549**	-0.124***	-0.103	-0.167***	-0.354***
	(0.048)	(0.215)	(0.025)	(0.106)	(0.050)	(0.113)
ln (prisoners per capita) $(t-1)$	-0.267***	-1.124*	-0.230***	-0.848*	-0.283***	-0.362*
	(0.103)	(0.577)	(0.057)	(0.464)	(0.103)	(0.205)
ln (police per capita) $(t-1)$	-0.333***	-1.887**	-0.080	-1.215	0.258*	0.106
	(0.128)	(0.821)	(0.087)	(0.749)	(0.145)	(0.327)
State unemployment rate	-0.336	-0.086	1.722***	4.811*	-0.947	-3.625*
(percent unemployed)	(1.212)	(3.130)	(0.628)	(2.887)	(1.217)	(1.954)
ln state income per capita	1.906***	4.004*	0.922***	3.319	-1.140	-1.001
	(0.545)	(2.068)	(0.323)	(2.189)	(0.803)	(1.402)
Poverty rate (percent	0.006	0.023	-0.001	-0.003	0.002	-0.001
below poverty line)	(0.007)	(0.019)	(0.004)	(0.014)	(0.006)	(0.009)
AFDC generosity $(t-15)$	0.009	0.041	0.016*	0.076	-0.015	-0.013
	(0.013)	(0.045)	(0.008)	(0.054)	(0.012)	(0.025)
Shall-issue concealed	-0.008	0.102	0.057**	0.260*	0.010	0.041
weapons law	(0.040)	(0.121)	(0.024)	(0.137)	(0.036)	(0.070)
Beer consumption per	-0.010	-0.045	0.004	0.008	-0.003	0.007
capita (gallons)	(0.008)	(0.036)	(0.006)	(0.018)	(0.006)	(0.012)
Observations	576	576	576	576	576	576
N	48	48	48	48	48	48
T	12	12	12	12	12	12

Notes: The estimates reported in this table are computed using the estimates of the underlying dynamic panel data regressions summarized in Table 11. See also the notes to that table.

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27.

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

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data* (4 ed.). Wiley.

Bartolucci, F., R. Bellio, A. Salvan, and N. Sartori (2016). Modified profile likelihood for fixed-effects panel data models. *Econometric Reviews 35*, 1271–1289.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*, 608–650.

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

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*, 115–143.

Bruno, G. S. (2005). Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models. *Economics Letters 87*, 361–366.

Bun, M. J. G. (2003). Bias correction in the dynamic panel data model with a nonscalar disturbance covariance matrix. *Econometric Reviews 22*, 29–58.

Bun, M. J. G. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics 23*, 200–210.

Bun, M. J. G. and M. A. Carree (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. *Economics Letters 92*, 220–227.

Bun, M. J. G. and J. Kiviet (2003). On the diminishing returns of higher order terms in asymptotic expansions of bias. *Economic Letters 19*, 145–152.

Choi, C., N. C. Mark, and D. Sul (2010). Bias reduction in dynamic panel data models by common recursive mean adjustment. *Oxford Bulletin of Economics and Statistics 72*(7), 567–599.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Jounal 14*, C45–C90.

Dhaene, G. and K. Jochmans (2015a). Likelihood inference in an autoregression with fixed effects. *Econometric Theory*, forthcoming.

Dhaene, G. and K. Jochmans (2015b). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*(3), 991–1030.

Donohue, III, J. J. and S. D. Levitt (2001). The impact of legalized abortion on crime. *Quarterly Journal of Economics 116*, 379–420.

Donohue, III, J. J. and S. D. Levitt (2004). Further evidence that legalized abortion lowered crime: A reply to Joyce. *Journal of Human Resources 39*, 29–49.

Donohue, III, J. J. and S. D. Levitt (2008). Measurement error, legalized abortion, and the decline in crime: A response to Foote and Goetz. *Quarterly Journal of Economics 123*, 425–440.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica 51*(2).

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

Foote, C. L. and C. F. Goetz (2008). The impact of legalized abortion on crime: Comment. *Quarterly Journal of Economics 123*, 407–423.

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

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

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

Han, C., P. C. B. Phillips, and D. Sul (2014). X-Differencing and dynamic panel model estimation. *Econometric Theory 30*(1), 201–251.

Joyce, T. (2004). Did legalized abortion lower crime? *Journal of Human Resources 39*, 1–28.

Joyce, T. (2009). A simple test of abortion and crime. *Review of Economics and Statistics 91*, 112–123.

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

Kiviet, J. F. (1999). Expectations of expansions for estimators in a dynamic panel data model; some results for weakly-exogenous regressors. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panel Data and Limited Dependent Variables*. Cambridge University Press, Cambridge.

Levitt, S. D. (2004). Understanding why crime fell in the 1990s: Four factors that explain the decline and six that do not. *Journal of Economic Perspectives 18*, 163–190.

Moody, C. E. and T. B. Marvell (2010). On the choice of control variables in the crime equation. *Oxford Bulletin of Economics and Statistics 72*, 696–715.

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

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182–202.

Phillips, P. C. B. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal 6*, 217–259.

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

Shoesmith, G. L. (2015). Crime, teenage abortion, and unwantedness. *Crime & Delinquency*, forthcoming.

Wansbeek, T. and A. Kapteyn (1989). Estimation of the error-components model with incomplete panels. *Journal of Econometrics 41*, 341–361.

A Appendix

A.1 Lemmas (Statements and Proofs)

Lemma 1 *Suppose x_{it}, for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, is generated by (52), and Assumptions 2-3, 5, and 8 hold. Then,*

$$\hat{\mathbf{Q}}_{FE-TE} - \hat{\mathbf{Q}}_{a,FE-TE} \xrightarrow{p} \mathbf{0}_{k\times k}, \tag{A.1}$$

and

$$\hat{\mathbf{Q}}_{FE-TE} - \hat{\mathbf{Q}}_{b,FE-TE} \xrightarrow{p} \mathbf{0}_{k\times k}, \tag{A.2}$$

as $N, T \to \infty$ jointly, where $\hat{\mathbf{Q}}_{FE-TE}$, $\hat{\mathbf{Q}}_{a,FE-TE}$, and $\hat{\mathbf{Q}}_{b,FE-TE}$ are defined in (55)-(57).

Proof. Using $\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}} = \boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}$, where the averages (aggregates) $\bar{\boldsymbol{\omega}}_{i\cdot}$, $\bar{\boldsymbol{\omega}}_{\cdot t}$, and $\bar{\boldsymbol{\omega}}$ are defined in a similar way as the averages $\bar{\mathbf{x}}_{i\cdot}, \bar{\mathbf{x}}_{\cdot t}$, and $\bar{\mathbf{x}}$, we can write $\hat{\mathbf{Q}}_{FE-TE} - \hat{\mathbf{Q}}_{a,FE-TE}$ as

$$\begin{aligned}
\hat{\mathbf{Q}}_{FE-TE} - \hat{\mathbf{Q}}_{a,FE-TE} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}\right)\boldsymbol{\omega}_{it}' \\
&\quad - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\left(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot a} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}_{a}\right)\boldsymbol{\omega}_{it}',
\end{aligned}$$

in which $\bar{\boldsymbol{\omega}}_{i\cdot a}$ and $\bar{\boldsymbol{\omega}}_{a}$ are defined in a similar way as $\bar{\mathbf{x}}_{i\cdot a}$ and $\bar{\mathbf{x}}_{a}$. Re-arranging the terms in the expression above gives

$$\begin{aligned}
\hat{\mathbf{Q}}_{FE-TE} - \hat{\mathbf{Q}}_{a,FE-TE} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{i\cdot}\boldsymbol{\omega}_{it}' + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{i\cdot a}\boldsymbol{\omega}_{it}' \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\boldsymbol{\omega}_{it}' + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{\cdot t}\boldsymbol{\omega}_{it}' \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}\boldsymbol{\omega}_{it}' - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{a}\boldsymbol{\omega}_{it}'.
\end{aligned} \tag{A.3}$$

We focus on the individual rows on the right side of (A.3) below. Consider the first row, which reduces to

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' \xrightarrow{p} \mathbf{0}_{k\times k},$$

as $N, T \to \infty$ jointly, since $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' \xrightarrow{p} \bar{\boldsymbol{\Gamma}}(0)/2$ as well as $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=T/2+1}^{T}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}' \xrightarrow{p} \bar{\boldsymbol{\Gamma}}(0)/2$. Consider next the second row on the right side of (A.3),

$$\begin{aligned}
-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{i\cdot}\boldsymbol{\omega}_{it}' + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{i\cdot a}\boldsymbol{\omega}_{it}' &= -\frac{1}{N}\sum_{i=1}^{N}\bar{\boldsymbol{\omega}}_{i\cdot}\bar{\boldsymbol{\omega}}_{i\cdot}' + \frac{1}{N}\sum_{i=1}^{N}\bar{\boldsymbol{\omega}}_{i\cdot a}\bar{\boldsymbol{\omega}}_{i\cdot a}' \\
&\xrightarrow{p} \mathbf{0}_{k\times k},
\end{aligned}$$

as $N, T \to \infty$ jointly, where we have used $\bar{\boldsymbol{\omega}}_{i\cdot} = O_p\left(T^{-1/2}\right)$ and $\bar{\boldsymbol{\omega}}_{i\cdot a} = O_p\left(T^{-1/2}\right)$ by the covariance stationarity of $\boldsymbol{\omega}_{it}$ with absolute summable autocovariances (uniformly in i). Given the independence of $\boldsymbol{\omega}_{it}$ across i, and the upper bound condition, $\left\|E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'\right)\right\| < K$, we obtain the following result for the expression in the third row on the right side of (A.3),

$$
\begin{aligned}
-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\boldsymbol{\omega}_{it}'+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{\cdot t}\boldsymbol{\omega}_{it}' &= -\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}'+\frac{2}{T}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}' \\
&= \frac{1}{T}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}'-\frac{1}{T}\sum_{t=T/2+1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}' \\
&\xrightarrow{p} \mathbf{0}_{k\times k},
\end{aligned}
$$

where $\bar{\boldsymbol{\omega}}_{\cdot t} = O_p\left(N^{-1/2}\right)$. Last but not least, consider the last row on the right side of (A.3). Using again the covariance-stationarity of $\boldsymbol{\omega}_{it}$ with uniformly absolute summable autocovariances in i and the cross-sectional independence of $\boldsymbol{\omega}_{it}$ across i, we have $\bar{\boldsymbol{\omega}} = O\left(N^{-1/2}T^{-1/2}\right)$, $\bar{\boldsymbol{\omega}}_a = O\left(N^{-1/2}T^{-1/2}\right)$, and

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}\boldsymbol{\omega}_{it}'-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_a\boldsymbol{\omega}_{it}' = \bar{\boldsymbol{\omega}}\bar{\boldsymbol{\omega}}'-\bar{\boldsymbol{\omega}}_a\bar{\boldsymbol{\omega}}_a' \xrightarrow{p} \mathbf{0}_{k\times k},
$$

as $N, T \to \infty$ jointly. Hence, overall

$$
\hat{\mathbf{Q}}_{FE-TE}-\hat{\mathbf{Q}}_{a,FE-TE} \xrightarrow{p} \mathbf{0}_{k\times k},
$$

as $N, T \to \infty$ jointly, which completes the proof of result (A.1). Result (A.2) can be obtained in the same way. ■

A.2 Proofs of Propositions

Proof of Proposition 1. Using (2) we have

$$
E\left(\boldsymbol{\omega}_{i,t+h}u_{it}\right)=\sum_{s=0}^{h-1}\mathbf{A}_{is}E\left(\mathbf{v}_{i,t+h-s}u_{it}\right)+\sum_{s=h+1}^{\infty}\mathbf{A}_{is}E\left(\mathbf{v}_{i,t+h-s}u_{it}\right).
$$

But, under Assumption 4.a,

$$
E\left(\mathbf{v}_{i,t+h-s}u_{it}\right)=\begin{cases}\mathbf{0}_{k\times 1}, & \text{for } s\geq h+1\\ \gamma_{iuv}\left(h-s\right), & \text{for } s<h\end{cases},
$$

where

$$
\left\|\gamma_{iuv}\left(h-s\right)\right\| < K\rho^{h-s} \text{ for } s<h. \tag{A.4}
$$

Hence,

$$
E\left(\boldsymbol{\omega}_{i,t+h}u_{it}\right)=\sum_{s=0}^{h-1}\mathbf{A}_{is}\gamma_{iuv}\left(h-s\right)\equiv\gamma_i\left(h\right),
$$

as desired. Taking the norm of $\gamma_i\left(h\right)$ and using the triangle inequality, (A.4), and condition (6) of Assumption 5, we have

$$
\begin{aligned}
\left\|\gamma_i\left(h\right)\right\| &\leq \sum_{s=0}^{h-1}\left\|\mathbf{A}_{is}\gamma_{iuv}\left(h-s\right)\right\|\leq\sum_{s=0}^{h-1}\left\|\mathbf{A}_{is}\right\|\left\|\gamma_{iuv}\left(h-s\right)\right\| \\
&< \sum_{s=0}^{h-1}K\rho^{s}\cdot K\rho^{h-s}=K^2h\rho^h.
\end{aligned}
$$

Noting that $\rho<1$, there exists $\epsilon=\left(1-\rho\right)/2>0$. Set $\rho_1=\rho+\epsilon<1$, $\rho^*=\rho/\rho_1$ and note that

$$
h\rho^h=\left(h\rho_1^h\right)\left(\rho^*\right)^h.
$$

Since $0 < \rho_1 < 1$, there must exist a positive finite constant K^* such that $\left|K^2 h\rho_1^h\right| < K^*$ for all $h = 1, 2, ...$Therefore

$$\|\gamma_i(h)\| \leq K^* \rho^{*h},$$

as desired, since $\rho^* = \rho/\rho_1 < 1$. This completes the proof. ■

Proof of Proposition 2. Using (2), we have

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})' = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_i)\boldsymbol{\omega}_{it}',$$

where $\bar{\boldsymbol{\omega}}_i = T^{-1}\sum_{t=1}^{T}\boldsymbol{\omega}_{it}$. Since $\boldsymbol{\omega}_{it}$ are cross-sectionally independent, we have for a fixed T and as $N \to \infty$,

$$p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_i)\boldsymbol{\omega}_{it}' = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_i)\boldsymbol{\omega}_{it}'.$$

Hence,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})' \xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})'\right).$$

But, since

$$\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})' = \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'\right] - \bar{\boldsymbol{\omega}}_i\bar{\boldsymbol{\omega}}_i',$$

then

$$\begin{aligned} E\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot})'\right) &= \boldsymbol{\Gamma}_i(0) - \frac{1}{T^2}E\left[(\boldsymbol{\omega}_{i1} + \boldsymbol{\omega}_{i2} + ...\boldsymbol{\omega}_{iT}) \times (\boldsymbol{\omega}_{i1} + \boldsymbol{\omega}_{i2} + ...\boldsymbol{\omega}_{iT})'\right] \\ &= \boldsymbol{\Gamma}_i(0) - \frac{1}{T}\boldsymbol{\Psi}_{iT}, \end{aligned}$$

where $\boldsymbol{\Psi}_{iT}$ is given by (15). Consider $\hat{\mathbf{Q}}_{FE}$ defined in (13). Then, for a fixed T, $\hat{\mathbf{Q}}_{FE} \xrightarrow{p} \mathbf{Q}_T$, as $N \to \infty$, where

$$\begin{aligned} \mathbf{Q}_T &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{\Gamma}_i(0) - \frac{1}{T}\boldsymbol{\Psi}_{iT}\right] \\ &= \bar{\boldsymbol{\Gamma}}(0) - \frac{1}{T}\bar{\boldsymbol{\Psi}}_T, \end{aligned} \quad \text{(A.5)}$$

$\bar{\boldsymbol{\Gamma}}(0) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Gamma}_i(0)$, and $\bar{\boldsymbol{\Psi}}_T = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Psi}_{iT}$. Note that $\left\|\bar{\boldsymbol{\Gamma}}(0)\right\| < K$ and $\left\|\bar{\boldsymbol{\Psi}}_T\right\| < K$ under Assumption 5. In particular, using the triangle inequality and (6), we obtain

$$\begin{aligned} \left\|\bar{\boldsymbol{\Gamma}}(0)\right\| &\leq \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\|\boldsymbol{\Gamma}_i(0)\| \\ &\leq \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{s=0}^{\infty}\|\mathbf{A}_{is}\|^2\right) < K, \end{aligned}$$

and, since $\|\boldsymbol{\Gamma}_i(h)\| \leq \sum_{s=0}^{\infty}\|\mathbf{A}_{is}\|\left\|\mathbf{A}_{i,s-h}'\right\| \leq K\rho^h$,

$$\|\boldsymbol{\Psi}_{iT}\| \leq \|\boldsymbol{\Gamma}_i(0)\| + \sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\left[\|\boldsymbol{\Gamma}_i(h)\| + \left\|\boldsymbol{\Gamma}_i'(h)\right\|\right] < K.$$

Consider now the second term on the right side of (12), and let $\mathbf{z}_{FE}$ be given by (40) and note that $\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta} = \hat{\mathbf{Q}}_{FE}^{-1}\mathbf{z}_{FE}$, where $\hat{\mathbf{Q}}_{FE}$ is invertible under Assumption 6. Recall that for a finite T and as $N \to \infty$, $\hat{\mathbf{Q}}_{FE} \xrightarrow{p} \mathbf{Q}_T$, where $\mathbf{Q}_T$ is given by (A.5). Consider $E(\mathbf{z}_{FE})$ next,

$$E(\mathbf{z}_{FE}) = \frac{1}{N}\sum_{i=1}^{N}\left[E\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\omega}_{it}u_{it}\right) - E(\bar{\boldsymbol{\omega}}_i\bar{u}_i)\right].$$

But under (9), $E(\boldsymbol{\omega}_{it}u_{it}) = \mathbf{0}_{k\times 1}$, and $E(\mathbf{z}_{FE}) = -\frac{1}{N}\sum_{i=1}^{N}E(\bar{\boldsymbol{\omega}}_i\bar{u}_i)$. Under (9),

$$\begin{aligned} E(\bar{\boldsymbol{\omega}}_i\bar{u}_i) &= \frac{1}{T^2}E\left[(\boldsymbol{\omega}_{i,1}+\boldsymbol{\omega}_{i,2}+...+\boldsymbol{\omega}_{i,T})\times(u_{i,1}+u_{i,2}+...+u_{i,T})\right] \\ &= \frac{1}{T}\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\gamma_i(h). \end{aligned}$$

Hence,

$$\lim_{N\to\infty}E(\mathbf{z}_{FE}) = -\frac{1}{T}\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\bar{\gamma}(h),$$

where $\bar{\gamma}(h)$ is given by (18). Therefore, we obtain

$$\lim_{N\to\infty}\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) = -\frac{1}{T}\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)^{-1}\bar{\boldsymbol{\chi}}_T,$$

as desired. To establish the order of the asymptotic small-T bias, note that under (10),

$$\|\bar{\boldsymbol{\chi}}_T\| \le \sum_{t=1}^{T-1}\|\bar{\gamma}(h)\| \le K\sum_{t=1}^{T-1}\rho^h = O(1).$$

In addition, $\bar{\boldsymbol{\Gamma}}(0)$ is nonsingular, and $\left\|\bar{\boldsymbol{\Psi}}_T\right\| < K$, which implies $\left\|\mathbf{Q}_T^{-1}\right\| = O(1)$,and therefore $Bias_T\left(\hat{\boldsymbol{\beta}}_{FE}\right) = O\left(T^{-1}\right)$, as required. ∎

Proof of Proposition 3. The exact analytical bias formula for the half-panel jackknife FE estimator is derived in the body of the paper; see the derivations preceding the result (36). We prove the order of the asymptotic bias in the general case next. First note that (using (15))

$$\begin{aligned} \bar{\boldsymbol{\Psi}}_T &= \bar{\boldsymbol{\Gamma}}(0)+\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\left[\bar{\boldsymbol{\Gamma}}(h)+\bar{\boldsymbol{\Gamma}}'(h)\right] \\ \bar{\boldsymbol{\Psi}}_{T/2} &= \bar{\boldsymbol{\Gamma}}(0)+\sum_{h=1}^{T/2-1}\left(1-\frac{2h}{T}\right)\left[\bar{\boldsymbol{\Gamma}}(h)+\bar{\boldsymbol{\Gamma}}'(h)\right], \end{aligned}$$

and

$$\begin{aligned} \bar{\boldsymbol{\Psi}}_T-\bar{\boldsymbol{\Psi}}_{T/2} &= \frac{1}{T}\sum_{h=1}^{T/2-1}h\left[\bar{\boldsymbol{\Gamma}}(h)+\bar{\boldsymbol{\Gamma}}'(h)\right] \\ &\quad +\sum_{h=T/2}^{T-1}\left(1-\frac{h}{T}\right)\left[\bar{\boldsymbol{\Gamma}}(h)+\bar{\boldsymbol{\Gamma}}'(h)\right]. \end{aligned}$$

Since $\left\|\bar{\boldsymbol{\Gamma}}(h)\right\| < K\rho^h$, then

$$\bar{\boldsymbol{\Psi}}_T-\bar{\boldsymbol{\Psi}}_{T/2} = O\left(\frac{1}{T}\right), \tag{A.6}$$

and[12]

$$
\begin{aligned}
Bias_T\left(\tilde{\boldsymbol{\beta}}_{FE}\right) &= -\frac{2}{T}\left[\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)^{-1}\bar{\boldsymbol{\chi}}_T-\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{2}{T}\bar{\boldsymbol{\Psi}}_{T/2}\right)^{-1}\bar{\boldsymbol{\chi}}_{T/2}\right] \\
&= -\frac{2}{T}\left\{\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)^{-1}\left(\bar{\boldsymbol{\chi}}_T-\bar{\boldsymbol{\chi}}_{T/2}\right)+\right. \\
&\quad \left.+\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)^{-1}\left(\frac{1}{T}\bar{\boldsymbol{\Psi}}_T-\frac{2}{T}\bar{\boldsymbol{\Psi}}_{T/2}\right)\left(\bar{\boldsymbol{\Gamma}}(0)-\frac{2}{T}\bar{\boldsymbol{\Psi}}_{T/2}\right)^{-1}\bar{\boldsymbol{\chi}}_{T/2}\right\}
\end{aligned}
$$

Now using (39) and (A.6), we have $Bias_T\left(\tilde{\boldsymbol{\beta}}_{FE}\right)=O\left(\frac{1}{T^2}\right)$, as required. ■

Proof of Proposition 4. The asymptotic variance is established in the body of the paper; see the derivations leading to (48). We establish the consistency of $\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE}\right)$ next. Let $\mathbf{h}_{it}=\mathbf{b}_{it}u_{it}$. Without the independence of $\mathbf{v}_{i,t+h}$ and u_{it} for all i,t and all $h\leq 0$, it is not guaranteed that

$$
\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}E\left(\mathbf{h}_{it}\mathbf{h}_{is}'\right)\rightarrow \mathbf{0}_{k\times k}, \tag{A.7}
$$

as $T,N\rightarrow\infty$ jointly such that $T=KN^{\epsilon}$, with $\epsilon>1/3$. But under the independence postulated in Assumption 4.b, we have

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\mathbf{h}_{it}\mathbf{h}_{is}'\right)\rightarrow\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\mathbf{h}_{it}\mathbf{h}_{it}'\right). \tag{A.8}
$$

To establish (A.8) first note that

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{b}_{it}\mathbf{b}_{is}'u_{it}u_{is} &= \sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{is}'u_{it}u_{is} \\
&\quad -\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\mathbf{q}}_{it}\bar{\mathbf{q}}_{is}'u_{it}u_{is},
\end{aligned} \tag{A.9}
$$

where $\bar{\mathbf{q}}_{it}=I\left(t\leq T/2\right)\bar{\mathbf{q}}_{ia}+I\left(t>T/2\right)\bar{\mathbf{q}}_{ib}$, $\bar{\mathbf{q}}_{ia}=2\bar{\boldsymbol{\omega}}_i-\bar{\boldsymbol{\omega}}_{ia}$, and $\bar{\mathbf{q}}_{ib}=2\bar{\boldsymbol{\omega}}_i-\bar{\boldsymbol{\omega}}_{ib}$. Consider the first term on the right side of (A.9). Under the independence of (current and past values of) regressors and future errors, we have, for $h=s-t>0$,

$$
E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{i,t+h}'u_{it}u_{i,t+h}\right)=\mathbf{0}_{k\times k}\text{ for all }i=1,2,...,N, t=1,2,...,T,\text{ and }h>0,
$$

and therefore

$$
\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{is}'u_{it}u_{is}\right)=\sum_{t=1}^{T}E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'u_{it}^2\right). \tag{A.10}
$$

Consider next the second term on the right side of (A.9) and note that

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\mathbf{q}}_{it}\bar{\mathbf{q}}_{is}'u_{it}u_{is} \\
=\;& \bar{\mathbf{q}}_{ia}\bar{\mathbf{q}}_{ia}'\sum_{t=1}^{T/2}\sum_{s=1}^{T/2}\frac{u_{it}u_{is}}{T}+\bar{\mathbf{q}}_{ia}\bar{\mathbf{q}}_{ib}'\frac{1}{T}\sum_{t=1}^{T/2}\sum_{s=T/2+1}^{T}u_{it}u_{is} \\
&+\bar{\mathbf{q}}_{ib}\bar{\mathbf{q}}_{ia}'\frac{1}{T}\sum_{t=T/2+1}^{T}\sum_{s=1}^{T/2}u_{it}u_{is}+\bar{\mathbf{q}}_{ib}\bar{\mathbf{q}}_{ib}'\frac{1}{T}\sum_{t=T/2+1}^{T}\sum_{s=T/2+1}^{T}u_{it}u_{is}.
\end{aligned}
$$

[12] We use the following identity: $\mathbf{A}^{-1}\mathbf{a}-\mathbf{B}^{-1}\mathbf{b}=\mathbf{A}^{-1}\left(\mathbf{a}-\mathbf{b}\right)+\mathbf{A}^{-1}\left(\mathbf{B}-\mathbf{A}\right)\mathbf{B}^{-1}\mathbf{b}$, for invertible $k\times k$ matrices $\mathbf{A}$, $\mathbf{B}$ and $k\times 1$ vectors $\mathbf{a},\mathbf{b}.$

But $\bar{\mathbf{q}}_{ia} = O_p\left(T^{-1/2}\right)$, $\bar{\mathbf{q}}'_{ib} = O_p\left(T^{-1/2}\right)$, and the double-sums involving the product of error terms $u_{it}u_{is}/T$ can be stochastically bounded as $O_p(1)$. This can be established by noting that

$$E\left[\left(\sum_{t=1}^{T/2}\sum_{s=1}^{T/2}\frac{u_{it}u_{is}}{T}\right)^2\right] = \frac{1}{T^2}\sum_{t=1}^{T/2}\sum_{s=1}^{T/2}\sum_{t'=1}^{T/2}\sum_{s'=1}^{T/2}E\left(u_{it}u_{is}u_{it'}u_{is'}\right),$$

but u_{it} and $u_{it'}$ are independent for any $t \neq t'$ and the fourth moments of u_{it} are uniformly bounded under Assumption 1. Hence,

$$\begin{aligned} & E\left[\left(\sum_{t=1}^{T/2}\sum_{s=1}^{T/2}\frac{u_{it}u_{is}}{T}\right)^2\right] \\ = \quad & \frac{1}{T^2}\sum_{t=1}^{T/2}E\left(u_{it}^4\right) + \frac{3}{T^2}\sum_{t=1}^{T/2}\sum_{s=1,s\neq t}^{T/2}E\left(u_{it}^2\right)E\left(u_{is}^2\right) < K, \end{aligned}$$

and the double-sums involving the product of error terms $u_{it}u_{is}/T$ are all $O_p(1)$. It now readily follows that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\mathbf{q}}_{it}\bar{\mathbf{q}}'_{is}u_{it}u_{is} = O_p\left(T^{-1}\right), \tag{A.11}$$

uniformly in i. Since, at the same time, $\left|\bar{\mathbf{q}}_{it}\bar{\mathbf{q}}'_{is}u_{it}u_{is}\right|$ is uniformly integrable, then (A.11) implies

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left|\bar{\mathbf{q}}_{it}\bar{\mathbf{q}}'_{is}u_{it}u_{is}\right| \rightarrow \mathbf{0}_{k\times k}, \tag{A.12}$$

uniformly in i, as $T \rightarrow \infty$. Results (A.10) and (A.12) in turn imply (A.8). Moreover, by independence of $\mathbf{h}_{it}$ across i, and by consistency of $\hat{u}_{it}$, we have

$$\hat{\mathbf{R}}_{FE} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\mathbf{h}_{it}\mathbf{h}'_{it}\right) \xrightarrow{p} \mathbf{0}_{k\times k},$$

as $N, T \rightarrow \infty$ jointly, which establishes the consistency of $\hat{\mathbf{R}}_{FE}$. As established in (44), $\mathbf{Q}$ can be consistently estimated using

$$\hat{\mathbf{Q}}_{FE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}\right)',$$

as $N, T \rightarrow \infty$ jointly. The consistency of $\hat{\mathbf{R}}_{FE}$ and $\hat{\mathbf{Q}}_{FE}$ proves the consistency of $AsyVar\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE}\right) = \hat{\mathbf{Q}}_{FE}^{-1}\hat{\mathbf{R}}_{FE}\hat{\mathbf{Q}}_{FE}^{-1}$, as required. ■

Proof of Proposition 5. Using $\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}} = \boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}$, $\hat{\mathbf{Q}}_{FE-TE}$ can be written as

$$\begin{aligned} \hat{\mathbf{Q}}_{FE-TE} &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}\right)\boldsymbol{\omega}'_{it} \\ &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{\omega}_{it}\boldsymbol{\omega}'_{it} - \frac{1}{N}\sum_{i=1}^{N}\bar{\boldsymbol{\omega}}_{i\cdot}\bar{\boldsymbol{\omega}}'_{i\cdot} - \frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}'_{\cdot t} + \bar{\boldsymbol{\omega}}\bar{\boldsymbol{\omega}}'. \end{aligned}$$

The last two terms are new, compared to the FE model analyzed in Proposition 2. Noting that the variance of $\boldsymbol{\omega}_{it}$ is uniformly bounded and $\boldsymbol{\omega}_{it}$ is independent across i, we obtain $\bar{\boldsymbol{\omega}}_{\cdot t} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\omega}_{it} = O_p\left(N^{-1/2}\right)$, and therefore, for a finite T, we have $T^{-1}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}'_{\cdot t} = O_p\left(N^{-1}\right)$, and

$$\bar{\boldsymbol{\omega}}\bar{\boldsymbol{\omega}}' = \left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\right)' = O_p\left(N^{-1}\right).$$

Hence, $\hat{\mathbf{Q}}_{FE-TE} \xrightarrow{p} \mathbf{Q}_T$, as T is fixed and $N \to \infty$, where $\mathbf{Q}_T$ is defined in (A.5). Consider next

$$\begin{aligned} \mathbf{z}_{FE-TE} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}\right) u_{it}, \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}\right) u_{it}. \end{aligned} \tag{A.13}$$

$\mathbf{z}_{FE-TE}$ consists of the following four terms:

$$\mathbf{z}_{FE-TE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\omega}_{it}u_{it} - \frac{1}{N}\sum_{i=1}^{N}\bar{\boldsymbol{\omega}}_{i\cdot}\bar{u}_{i\cdot} - \frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t} + \bar{\boldsymbol{\omega}}\bar{u},$$

where $\bar{u}_{i\cdot} = T^{-1}\sum_{t=1}^{T} u_{it}$, $\bar{u}_{\cdot t} = N^{-1}\sum_{i=1}^{N} u_{it}$, and $\bar{u} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}$. The last two terms are new compared to the earlier analysis. But $\bar{\boldsymbol{\omega}}_{\cdot t} = O_p\left(N^{-1/2}\right)$, and $\bar{u}_{\cdot t} = O_p\left(N^{-1/2}\right)$, and therefore $T^{-1}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t} = O_p\left(N^{-1}\right)$. Moreover,

$$\bar{\boldsymbol{\omega}}\bar{u} = \left(\frac{1}{T}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}\right) = O_p\left(N^{-1}\right).$$

Hence,

$$\lim_{N\to\infty} E\left(\mathbf{z}_{FE-TE}\right) = \lim_{N\to\infty} E\left(\mathbf{z}_{FE}\right) = -\frac{1}{T}\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\bar{\gamma}(h),$$

and, assuming $\left(\bar{\boldsymbol{\Gamma}}(0) - \frac{1}{T}\bar{\boldsymbol{\Psi}}_T\right)$ is invertible, the small-T bias of the FE-TE estimator $\hat{\boldsymbol{\beta}}_{FE-TE}$ in the model with fixed and time effects is the same as the bias of the FE estimator $\hat{\boldsymbol{\beta}}_{FE}$ in the model with FE. The analytical bias formula for the half-panel jackknife FE-TE estimator $\tilde{\boldsymbol{\beta}}_{FE-TE}$ defined in (60) will therefore be the same, given by (36). ■

Proof of Proposition 6. First we establish (66). Consider (63) and note that $\mathbf{d}_{it}^*$ is not cross-sectionally independent, but, using $\mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}} = \boldsymbol{\omega}_{it} - \bar{\boldsymbol{\omega}}_{i\cdot} - \bar{\boldsymbol{\omega}}_{\cdot t} + \bar{\boldsymbol{\omega}}$, $\mathbf{d}_{it}^*$ can be written as

$$\mathbf{d}_{it}^* = \mathbf{b}_{it} + \mathbf{c}_t, \tag{A.14}$$

where $\mathbf{b}_{it}$ is defined in (47),

$$\mathbf{c}_t = I\left(t \le T/2\right)\mathbf{c}_{ta} + I\left(t > T/2\right)\mathbf{c}_{tb},$$

and

$$\mathbf{c}_{ta} = -\bar{\boldsymbol{\omega}}_{\cdot t} + \left(2\bar{\boldsymbol{\omega}} - \bar{\boldsymbol{\omega}}_a\right), \ \mathbf{c}_{tb} = -\bar{\boldsymbol{\omega}}_{\cdot t} + \left(2\bar{\boldsymbol{\omega}} - \bar{\boldsymbol{\omega}}_b\right).$$

Hence,

$$\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}^* u_{it} = \mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{b}_{it}u_{it} + \mathbf{Q}^{-1}\sqrt{\frac{N}{T}}\sum_{t=1}^{T}\mathbf{c}_t\bar{u}_{\cdot t}. \tag{A.15}$$

Clearly, by independence of $\boldsymbol{\omega}_{it}$ across i, $\mathbf{b}_{it}$ is cross-sectionally independent. Using the same arguments as in the proof of Proposition 4, we obtain

$$AsyVar\left(\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{b}_{it}u_{it}\right) = \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}, \tag{A.16}$$

as $N, T \to \infty$ jointly such that $T = KN^{\epsilon}$, for some $0 < K < \infty$ and $\epsilon > 1/3$. Consider next the second term on the right side of (A.15). We have

$$\begin{aligned} \sqrt{\frac{N}{T}}\sum_{t=1}^{T}\mathbf{c}_t\bar{u}_{\cdot t} &= -\sqrt{\frac{N}{T}}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t} + 2\sqrt{NT}\bar{\boldsymbol{\omega}}\bar{u} \\ &\quad -\sqrt{\frac{N}{T}}\left(\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_a\bar{u}_{\cdot t} + \sum_{t=T/2+1}^{T}\bar{\boldsymbol{\omega}}_b\bar{u}_{\cdot t}\right). \end{aligned} \tag{A.17}$$

Note that $E\left(\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t}\right)=\mathbf{0}_{k\times 1}$ and

$$Var\left(\sqrt{\frac{N}{T}}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t}\right)=\frac{N}{T}\sum_{t=1}^{T}E\left(\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}'\right)E\left(\bar{u}_{\cdot t}^{2}\right).$$

But

$$\begin{aligned} E\left(\bar{\boldsymbol{\omega}}_{\cdot t}\bar{\boldsymbol{\omega}}_{\cdot t}'\right) &= E\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{jt}'\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'\right)=\frac{1}{N}\bar{\boldsymbol{\Gamma}}(0), \end{aligned}$$

and

$$E\left(\bar{u}_{\cdot t}^{2}\right)=\frac{1}{N^2}\sum_{i=1}^{N}E\left(u_{it}^{2}\right)=\frac{\sigma_{ui}^{2}}{N},$$

and therefore

$$\sqrt{\frac{N}{T}}\sum_{t=1}^{T}\bar{\boldsymbol{\omega}}_{\cdot t}\bar{u}_{\cdot t}\xrightarrow{p}\mathbf{0}_{k\times 1},$$

as $N,T\to\infty$ jointly. Consider next the term $2\sqrt{NT}\bar{\boldsymbol{\omega}}\bar{u}$. We have $\bar{\boldsymbol{\omega}}=O_p\left(N^{-1/2}T^{-1/2}\right)$ and $\bar{u}=O_p\left(N^{-1/2}T^{-1/2}\right)$, and hence

$$2\sqrt{NT}\bar{\boldsymbol{\omega}}\bar{u}\xrightarrow{p}\mathbf{0}_{k\times 1},$$

as $N,T\to\infty$ jointly. Consider next the terms in the second row on the right side of (A.17). We have

$$\begin{aligned} \sqrt{\frac{N}{T}}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{a}\bar{u}_{\cdot t} &= \sqrt{NT}\bar{\boldsymbol{\omega}}_{a}\frac{1}{2}\left(\frac{2}{T}\sum_{t=1}^{T/2}\bar{u}_{\cdot t}\right) \\ &= \frac{\sqrt{NT}}{2}\bar{\boldsymbol{\omega}}_{a}\bar{u}_{a}, \end{aligned}$$

where (same as before) $\bar{\boldsymbol{\omega}}_{a}=O_p\left(N^{-1/2}T^{-1/2}\right)$ and $\bar{u}_{a}=O_p\left(N^{-1/2}T^{-1/2}\right)$. Hence,

$$\sqrt{\frac{N}{T}}\sum_{t=1}^{T/2}\bar{\boldsymbol{\omega}}_{a}\bar{u}_{\cdot t}\xrightarrow{p}\mathbf{0}_{k\times 1},$$

as $N,T\to\infty$ jointly, and, using similar arguments,

$$\sqrt{\frac{N}{T}}\sum_{t=T/2+1}^{T}\bar{\boldsymbol{\omega}}_{b}\bar{u}_{\cdot t}\xrightarrow{p}\mathbf{0}_{k\times 1},$$

as $N,T\to\infty$ jointly. Overall, using these results in (A.17), we obtain

$$\sqrt{\frac{N}{T}}\sum_{t=1}^{T}\mathbf{c}_{t}\bar{u}_{\cdot t}\xrightarrow{p}\mathbf{0}_{k\times 1},$$

as $N,T\to\infty$ jointly. Hence,

$$\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{d}_{it}^{*}u_{it}\overset{d}{\sim}\mathbf{Q}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{b}_{it}u_{it},$$

and

$$\sqrt{NT}\left(\tilde{\boldsymbol{\beta}}_{FE-TE}-\boldsymbol{\beta}\right)\overset{d}{\sim}N\left(\mathbf{0},\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\right),$$

as desired. The consistency of $\widehat{AsyVar}\left(\sqrt{NT}\tilde{\boldsymbol{\beta}}_{FE-TE}\right)$ can now be established in the same way as in Proposition 4. ■

Proof of Proposition 7. Noting that

$$\hat{\mathbf{Q}}_{FE}=\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]',$$

where

$$\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]'=\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'-\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\bar{\boldsymbol{\omega}}_{i\cdot}'\left(T_i\right),$$

is independent across i, and $\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)=T_i^{-1}\sum_{t=T_{fi}}^{T_{li}}\boldsymbol{\omega}_{it}$, we have for a fixed $\{T_i,\ i=1,2,...,N\}$

$$\hat{\mathbf{Q}}_{FE}\overset{p}{\to}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\vartheta_i E\left(\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'-\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\bar{\boldsymbol{\omega}}_{i\cdot}'\left(T_i\right)\right).$$

But $T_i^{-1}\sum_{t=T_{fi}}^{T_{li}}E\left(\boldsymbol{\omega}_{it}\boldsymbol{\omega}_{it}'\right)=\boldsymbol{\Gamma}_i\left(0\right)$, and $E\left[\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\bar{\boldsymbol{\omega}}_{i\cdot}'\left(T_i\right)\right]=T_i^{-1}\boldsymbol{\Psi}_{iT_i}$, where

$$\boldsymbol{\Psi}_{iT_i}=\boldsymbol{\Gamma}_i\left(0\right)+\sum_{i=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\left[\boldsymbol{\Gamma}_i\left(h\right)+\boldsymbol{\Gamma}_i'\left(h\right)\right].$$

It is clear that $\boldsymbol{\Psi}_{iT_i}=O\left(1\right)$, and therefore $E\left[\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\bar{\boldsymbol{\omega}}_{i\cdot}'\left(T_i\right)\right]=O\left(T_i^{-1}\right)$. Hence,

$$\hat{\mathbf{Q}}_{FE}\overset{p}{\to}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\left(\boldsymbol{\Gamma}_i\left(0\right)-\frac{1}{T_i}\boldsymbol{\Psi}_{iT_i}\right)=\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}},\tag{A.18}$$

where $\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)$ is defined in (71) and $\bar{\boldsymbol{\Psi}}_{\{T_i\}}$ is defined in (72). $\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)=O\left(1\right)$ and it is nonsingular by assumption. Moreover, $\left\|\boldsymbol{\Psi}_{iT_i}\right\|<K$, and therefore

$$\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}}=O\left(\bar{T}^{-1}\right).\tag{A.19}$$

To obtain the small-T large-N bias of $\hat{\boldsymbol{\beta}}_{FE}$, note that $\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_{FE}=\hat{\mathbf{Q}}_{FE}^{-1}\mathbf{z}_{FE}$, where

$$\begin{aligned}\mathbf{z}_{FE}&=\frac{1}{\sum_{i=1}^{N}T_i}\sum_{i=1}^{N}\sum_{t=T_{fi}}^{T_{li}}\left[\mathbf{x}_{it}-\bar{\mathbf{x}}_{i\cdot}\left(T_i\right)\right]u_{it}\\&=\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\left[\boldsymbol{\omega}_{it}-\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\right]u_{it}.\end{aligned}$$

Using (A.18), we have

$$\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_{FE}\overset{d}{\sim}\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}}\right]^{-1}\mathbf{z}_{FE},$$

as $N\to\infty$ and $\{T_i\}$ are fixed. Consider $E\left(\mathbf{z}_{FE}\right)$ next.

$$E\left(\mathbf{z}_{FE}\right)=E\left[\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\left(\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}\boldsymbol{\omega}_{it}u_{it}-\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}u_{it}\right)\right].$$

But $E\left(\boldsymbol{\omega}_{it}u_{it}\right)=\mathbf{0}_{k\times 1}$, and

$$E\left(\bar{\boldsymbol{\omega}}_{i\cdot}\left(T_i\right)\frac{1}{T_i}\sum_{t=T_{fi}}^{T_{li}}u_{it}\right)=\frac{1}{T_i}\boldsymbol{\chi}_i,$$

where $\boldsymbol{\chi}_i = \sum_{i=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right) = O\left(1\right)$. Hence, $E\left(\mathbf{z}_{FE}\right) = N^{-1}\sum_{N=1}^{N}\frac{\vartheta_i}{T_i}\boldsymbol{\chi}_i$, and noting that $\left\|\boldsymbol{\chi}_i\right\| < K$, we have

$$\lim_{N\to\infty}\left\|E\left(\mathbf{z}_{FE}\right)\right\| \leq K\lim_{N\to\infty}\left(\frac{1}{N}\sum_{N=1}^{N}\frac{\vartheta_i}{T_i}\right) = O\left(\bar{T}^{-1}\right).$$

Therefore

$$\lim_{N\to\infty}\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) = \left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}}\right]^{-1}\frac{\bar{\boldsymbol{\chi}}_{\{T_i\}}}{\bar{T}} = O\left(\bar{T}^{-1}\right),$$

where $\bar{T}^{-1}\bar{\boldsymbol{\chi}}_{\{T_i\}} = \lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\frac{\vartheta_i}{T_i}\sum_{i=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right) = O\left(\bar{T}^{-1}\right)$, as required. ■

Proof of Proposition 8. The exact large-N small-T bias expression for $\tilde{\boldsymbol{\beta}}_{FE}$ directly follows from substituting the bias expressions into

$$\left(\tilde{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right) = 2\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}\right)-\frac{1}{2}\left[\left(\hat{\boldsymbol{\beta}}_{a,FE}-\boldsymbol{\beta}\right)+\left(\hat{\boldsymbol{\beta}}_{b,FE}-\boldsymbol{\beta}\right)\right].$$

We establish the order of the bias next. We have

$$\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right) = O\left(1\right), \text{ and } \bar{\boldsymbol{\Gamma}}_{\vartheta}^{-1}\left(0\right) = O\left(1\right), \tag{A.20}$$

by Assumption 5. Moreover,

$$\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}} = O\left(\bar{T}^{-1}\right), \tag{A.21}$$

see (A.19). Using the same arguments as in the derivation of (A.19), we also have

$$\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i/2\}} = O\left(\bar{T}^{-1}\right). \tag{A.22}$$

Consider the leading term $2\bar{\boldsymbol{\chi}}_{\{T_i\}}/\bar{T}-2\bar{\boldsymbol{\chi}}_{\{T_i/2\}}/\bar{T}$ next. We have

$$\left\|2\bar{\boldsymbol{\chi}}_{\{T_i\}}/\bar{T}-2\bar{\boldsymbol{\chi}}_{\{T_i/2\}}/\bar{T}\right\| = \left\|\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\frac{2}{T_i}\left(\sum_{t=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right)-\sum_{t=1}^{T_i/2-1}\left(1-\frac{h}{T_i/2}\right)\gamma_i\left(h\right)\right)\right\|. \tag{A.23}$$

But

$$\begin{aligned}\left\|\sum_{t=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right)-\sum_{t=1}^{T_i/2-1}\left(1-\frac{h}{T_i/2}\right)\gamma_i\left(h\right)\right\| &= \left\|\sum_{t=1}^{T_i/2-1}\frac{h}{T_i}\gamma_i\left(h\right)+\sum_{t=T/2}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right)\right\| \\ &\leq \frac{1}{T_i}\sum_{t=1}^{T_i/2-1}h\left\|\gamma_i\left(h\right)\right\|+\sum_{t=T/2}^{T_i-1}\left(1-\frac{h}{T_i}\right)\left\|\gamma_i\left(h\right)\right\|,\end{aligned}$$

and using $\left\|\gamma_i\left(h\right)\right\| < K\rho^h$, we obtain

$$\sum_{t=1}^{T_i/2-1}h\left\|\gamma_i\left(h\right)\right\| = O\left(1\right),\ \left(\sum_{i=T_i/2}^{T_i-1}\gamma_i\left(h\right)\right) = O\left(\rho^{T_i/2}\right),$$

and there exists $0 < K < \infty$ so that

$$\left\|\sum_{t=1}^{T_i-1}\left(1-\frac{h}{T_i}\right)\gamma_i\left(h\right)-\sum_{t=1}^{T_i/2-1}\left(1-\frac{h}{T_i/2}\right)\gamma_i\left(h\right)\right\| < \frac{K}{T_i}.$$

Using this result in (A.23), it follows that

$$\begin{aligned}\left\|2\bar{\boldsymbol{\chi}}_{\{T_i\}}/\bar{T}-2\bar{\boldsymbol{\chi}}_{\{T_i/2\}}/\bar{T}\right\| &\leq \left\|\frac{1}{N}\sum_{i=1}^{N}\vartheta_i\frac{2}{T_i}\left(\frac{K}{T_i}\right)\right\| \\ &\leq 2K\frac{1}{\frac{1}{N}\sum_{i=1}^{N}T_i}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i} \\ &= O\left(\bar{T}_N^{-1}\bar{T}_{h,N}^{-1}\right),\end{aligned}$$

where $\bar{T}_{h,N}$ is the harmonic mean of T_i given by (81). Overall, as $N\rightarrow\infty$,

$$\left\|2\bar{\boldsymbol{\chi}}_{\{T_i\}}/\bar{T}-2\bar{\boldsymbol{\chi}}_{\{T_i/2\}}/\bar{T}\right\| = O\left(\bar{T}^{-1}\bar{T}_h^{-1}\right),$$

and using also (A.20)-(A.22)

$$\lim_{N\rightarrow\infty}E\left(\tilde{\boldsymbol{\beta}}_{FE,a}-\boldsymbol{\beta}\right)=2\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i\}}\right]^{-1}\frac{\bar{\boldsymbol{\chi}}_{\{T_i\}}}{\bar{T}}-2\left[\bar{\boldsymbol{\Gamma}}_{\vartheta}\left(0\right)-\frac{1}{\bar{T}}\bar{\boldsymbol{\Psi}}_{\{T_i/2\}}\right]^{-1}\frac{\bar{\boldsymbol{\chi}}_{\{T_i/2\}}}{\bar{T}}=O\left(\bar{T}^{-1}\bar{T}_h^{-1}\right),$$

where $\bar{T}_h$ is given by (80). Using the arithmetic-harmonic mean inequality, we have $\bar{T}_h\leq\bar{T}$, and therefore $\bar{T}^{-1}\leq\bar{T}_h^{-1}$, and $\lim_{N\rightarrow\infty}E\left(\tilde{\boldsymbol{\beta}}_{FE,a}-\boldsymbol{\beta}\right)=O\left(\bar{T}_h^{-2}\right)$. Assuming

$$0<K_1<\vartheta_i<K_2<\infty, \tag{A.24}$$

for all i where K_1,K_2 do not change with sample size, we obtain $T_i^{-1}<\left(K_1\bar{T}_N\right)^{-1}$

$$\bar{T}_{h,N}^{-1}=N^{-1}\sum_{i=1}^{N}\frac{1}{T_i}<K_1^{-1}N^{-1}\sum_{i=1}^{N}\frac{1}{\bar{T}_N}=O\left(\bar{T}_N^{-1}\right),$$

$\bar{T}_h^{-1}=O\left(\bar{T}^{-1}\right)$, and

$$\lim_{N\rightarrow\infty}E\left(\tilde{\boldsymbol{\beta}}_{FE,a}-\boldsymbol{\beta}\right)=O\left(\bar{T}^{-2}\right), \tag{A.25}$$

as desired. When (A.24) does not hold for a finite subset of units, then $\bar{T}$ and $\bar{T}_h$ continue to be of the same order, and therefore (A.25) continues to hold. ■

Proof of Proposition 9. Proposition 9 can be established in the same way as Proposition 4. ■