Nickell Bias in Panel Local Projection: Financial Crises Are Worse Than You Think[†]

Ziwei Mei, Liugang Sheng, Zhentao Shi
The Chinese University of Hong Kong

First draft: February, 2023. This draft: October 2023.

Abstract

Local Projection is widely used for impulse response estimation, with the Fixed Effect (FE) estimator being the default for panel data. This paper highlights the presence of Nickell bias for all regressors in the FE estimator, even if lagged dependent variables are absent in the regression. This bias is the consequence of the inherent panel predictive specification. We recommend using the split-panel jackknife estimator to eliminate the asymptotic bias and restore the standard statistical inference. Revisiting three macro-finance studies on the linkage between financial crises and economic contraction, we find that the FE estimator substantially underestimates the post-crisis economic losses.

Key words: Bias correction, dynamic panel, impulse response, split-panel jackknife, macro-finance
JEL code: C33, C53, E44, G01

[†] Ziwei Mei, zwmei@link.cuhk.edu.hk. Liugang Sheng, lsheng@cuhk.edu.hk. Corresponding author: Zhentao Shi, zhentao.shi@cuhk.edu.hk, 9/F Esther Lee Building, Department of Economics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; Tel: (+852) 3943 8191. We thank the financial support from the Research Grants Council of the Hong Kong SAR (Project No.14501821). We thank Qiyu Dai and Shu Shen for excellent research assistance. We thank Greet Dhaene, Raffaella Giacomini, Frank Kleibergen, Xun Lu, Ryo Okui, and Liangjun Su for helpful comments.

1 Introduction

The impulse response function (IRF) characterizes the dynamic impact of external shocks on macroeconomic variables of interest. While the IRF is conventionally estimated by vector autoregression (VAR), the time series local projection (LP) method (Jordà, 2005) produces the IRF by regressing the future dependent variables at each horizon on the current independent variables via ordinary least squares (OLS). LP and its extensions have shaped a growing time series literature (Barnichon and Brownlees, 2019; Jordà, 2009; Xu, 2022), with systematic comparisons between LP and VAR theoretically (Montiel Olea and Plagborg-Møller, 2021; Plagborg-Møller and Wolf, 2021) and numerically (Li et al., 2022). Thanks to its simplicity, flexibility, and robustness to specifications, LP has been naturally carried over into panel regressions, where it is almost always estimated by the fixed effect (FE) method to control for unobserved individual-specific characteristics.

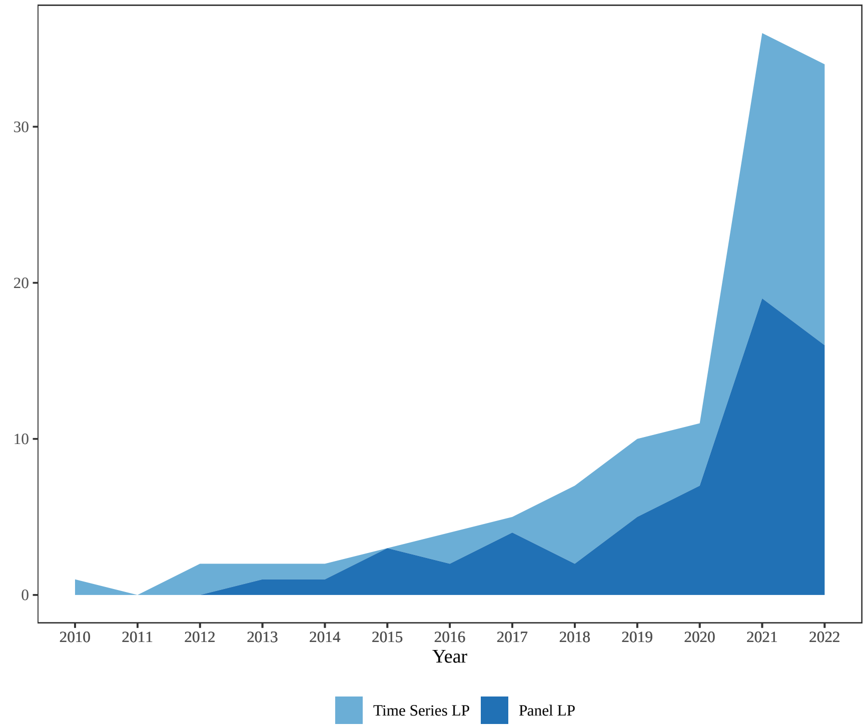
LP has been widely adopted in empirical macroeconomics recently, encompassing political economics, monetary and fiscal policies, financial crises, and pandemics (Acemoglu et al., 2019; Ramey and Zubairy, 2018; Ramey, 2016; Romer and Romer, 2017; Jordà et al., 2022). Figure 1 shows that the number of articles using LP in 11 leading economics journals has been growing rapidly from 2010 to 2022, and in total more than half of those 117 publications employed the panel LP method. One particularly exciting and active field is macro-finance, where the panel LP has played a methodological role in establishing the linkage between financial crises and economic contraction. For example, by applying the panel LP to various cross-country panel data, Romer and Romer (2017), Baron et al. (2021), and Mian et al. (2017) show that financial distress, banking crises, and household debts lead to severe economic contractions, respectively.[1]

In this paper, we focus on the Nickell bias of the classical FE estimators in the panel LP. To fix ideas, consider a simple h-period ahead panel LP model

$$y_{i,t+h} = \mu_i^{(h)y} + \beta^{(h)} x_{i,t} + e_{i,t+h}^{(h)}, \quad \text{for } t = 1, 2, \ldots, T-h, \text{ and } h = 0, 1, \ldots, H \tag{1}$$

where $x_{i,t}$ is a scalar predictor, $e_{i,t+h}^{(h)}$ is the error term uncorrelated with $x_{i,t}$, and $\mu_i^{(h)y}$ is the individual-specific heterogeneity, or the fixed effect. The IRF $(\beta^{(h)})_{h=0}^{H}$ is of central interest in learning an economic linkage between the two variables $x_{i,t}$ and $y_{i,t+h}$. For example, $y_{i,t+h}$ is the logarithm of real GDP and $x_{i,t}$ is some measure of financial crises. Despite the absence of a lagged dependent variable in (1), the FE estimator for this panel LP model is

[1] Many other studies including Jordà et al. (2013), Jordà et al. (2015), Jordà et al. (2016), Zeev (2017), Bhattarai et al. (2021) also adopt the panel LP method to examine the impacts of financial crises on economic contraction. Please see the timely survey of Sufi and Taylor (2021) for more discussions on the empirical methodology in this literature.



Note: All publications are counted from 2010 to 2022 in the following economics journals: *American Economic Review*, *Econometrica*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Review of Economic Studies*, *Journal of Monetary Economics*, *American Economic Journal: Macroeconomics*, *Review of Economics and Statistics*, *Economic Journal*, *International Economic Review*, and *Journal of International Economics*. See the `https://github.com/zhentaoshi/panel-local-projection/blob/main/Supplementary_Appendix_LP_Publications.pdf` for the full list of publications.

Figure 1: The Number of Papers Using LP in Leading Economics Journals

asymptotically biased when the number of cross-sectional units N and the time periods T are both large; to be precise, the leading case is $(N, T) \to \infty$ and $N/T \to c$ for some constant $c \in (0, \infty)$. This bias stems from the violation of *strict exogeneity* in the dynamic panel linear model, and therefore it can be viewed as a type of Nickell bias in multiple-equation panel VAR—the full dynamic model behind the single-equation LP. The Nickell bias in panel LP has a profound consequence to the practical estimation and inference. While standard statistical inference about the slope coefficient refers to the t-statistic in comparison with a critical value drawn from the standard normal distribution, this asymptotic bias will distort the test size based on the FE estimator, thereby leading to erroneous empirical findings.

To facilitate theoretical analysis of the bias, we first present a prototype model where the dependent variable $y_{i,t+1}$ is generated from

$$y_{i,t+1} = \mu_i^{(0)y} + \beta^{(0)} x_{i,t+1} + u_{i,t+1}^y, \tag{2}$$

and the variable of interest $x_{i,t}$ follows a panel AR(1) process

$$x_{i,t+1} = \mu_i^x + \rho x_{i,t} + u_{i,t+1}^x, \tag{3}$$

where $|\rho| < 1$ ensures stationarity. This simple model will allow us to derive the analytical expression of the bias from the underlying dynamic structure. We will show that this data generating process (DGP) implies that the error term $e_{i,t+h}^{(h)}$ in (1) violates strict exogeneity in general, which leads to an *implicit* Nickell bias, as on the right-hand side of (1) the lagged dependent variable $y_{i,t}$ does not *explicitly* show up.

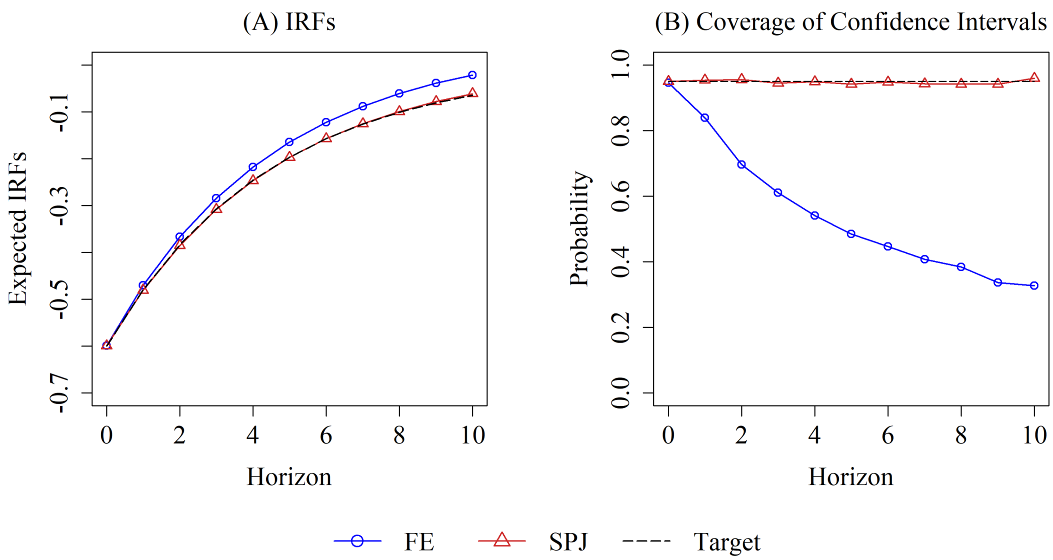
One of the key attractions of the LP method is that it allows applied researchers to concentrate on the specification of a main equation of interest that links $y_{i,t+h}$ and $x_{i,t}$, instead of modeling the entire VAR system. A desirable solution to the Nickell bias should preserve this advantage. To eliminate the first-order bias and thus restore valid statistical inference, we recommend using the split-panel jackknife (SPJ) estimator (Dhaene and Jochmans, 2015; Chudik et al., 2018) as a general solution to the panel LP single-equation regression. SPJ is easy to implement, with a simple formula

$$\widehat{\beta}^{(h)\text{spj}} = 2\widehat{\beta}^{(h)\text{fe}} - (\widehat{\beta}_a^{(h)\text{fe}} + \widehat{\beta}_b^{(h)\text{fe}})/2,$$

where $\widehat{\beta}^{(h)\text{fe}}$, $\widehat{\beta}_a^{(h)\text{fe}}$, and $\widehat{\beta}_b^{(h)\text{fe}}$ are the plain FE estimates from all the time periods, the first half ($t \leq T/2$), and the second half ($t > T/2$), respectively. Given that most LP empirical applications involve at least a moderate T, we show that the SPJ estimator $\widehat{\beta}^{(h)\text{spj}}$ is asymptotically unbiased and follows a zero-mean normal distribution when $N/T^3 \to 0$. Our formal theory about the SPJ is developed in a linear model that accommodates extra control variables. The theory can be further extended to entertain other popular settings in panel data (See Appendix B).

Figure 2 provides a quick snapshot of the performance of FE and SPJ estimation in a Monte Carlo simulation for the prototype model where the true IRF is $\beta^{(h)} = -0.6 \times 0.8^h$ in a panel setting with $N = 50$ and $T = 120$. In Panel (A), the SPJ estimator (red line) shows little bias as it overlaps with the true IRF (dashed line). The FE estimator (blue line), on the contrary, is biased toward 0 and the bias increases in the horizon h as the theory predicts. Striking is the coverage probability of the usual two-sided confidence interval constructed by inverting the t-statistic. As shown in Panel (B), the SPJ estimator aligns well with the nominal 95% coverage probability (dashed line), but the coverage probability of the FE's confidence interval deteriorates all the way down to 30–40% as the horizon h increases.

To explore the importance of the bias and the effectiveness of the SPJ correction, we revisit the empirical macro-finance question regarding the economic aftermath of financial



Note: The figure presents the Monte Carlo simulation results based on 1000 independent replications of the prototype model (2) and (3), where the key parameters are $\beta^{(0)} = -0.6$ and $\rho = 0.8$, and the sample size is $(N, T) = (50, 120)$. The full simulation design is detailed in Section 2.4. (A) presents the true (target) IRF and averaged IRFs using FE and SPJ, respectively, and (B) presents the targeted 95% nominal and the empirical coverage probabilities of the confidence intervals based on the t-statistic.

Figure 2: Expected IRFs and Coverage Probabilities from Simulation

crises. Economists have long been interested in the connection between finance and macroeconomics. Ben S. Bernanke, the 2022 Nobel laureate, demonstrated how bank runs played a decisive role in the Great Depression of the 1930s, the worst economic crisis in modern history (Bernanke et al., 1983). The Great Recession following the 2008 financial crisis sparked a renewed investigation into the relationship between financial shocks and economic recession. Several studies have shown that output and employment after financial crises are significantly lower than before, and recessions that accompany financial crises are more severe (Reinhart and Rogoff, 2009; Laeven and Valencia, 2013; Schularick and Taylor, 2012; Jordà et al., 2013).

Panel LP is commonly employed in this literature for assessing the economic loss of financial crises. Romer and Romer (2017) find that the output drops moderately following a typical financial distress shock, Baron et al. (2021) highlight the traditional role of banking crises and their economic consequences on output contraction and credit crunch, and Mian et al. (2017) underscore the role of household debt in generating recessions in the mid-run after the initial housing boom. Overall, our reexamination confirms their findings that financial shocks lead to lower future output (and exacerbate unemployment).

However, the FE estimates of economic contraction following financial distress, banking crises, and household debts as key results in those three studies respectively appear to substantially underestimate the output losses stemming from financial shocks in the mid- and long-run. The relative differences between the SPJ and FE estimates at the peak year when the decline in output is the largest range from 16% to 44% in the above three studies. For the case of Romer and Romer (2017), the SPJ estimate shows a peak decline in output of 6.3% three and a half years after a moderate financial crisis, whereas the FE estimate indicates a lower peak decline of 5.4%. Similarly, the FE estimates in Baron et al. (2021) suggest a 3.4% cumulative decline in GDP four years after the bank crisis, whereas SPJ widens the drop in GDP to about 4.2%. Lastly, a ten percentage point increase in the household debt to GDP ratio in Mian et al. (2017) is associated with a 3.9% decline in GDP by the FE estimates at the peak, versus a 5.6% decline by SPJ.

In summary, our analysis of these three studies reveals a consistent pattern of underestimation in the FE estimator when evaluating output loss following financial crises. Such underestimation echoes the bias in our prototype model which shifts the FE estimator toward zero. Our results carry important messages with policy implications. Since existing empirical strategies based on the FE estimator fall short of revealing the full magnitude of output losses, governments and policymakers must take financial crises more seriously and pay more attention to their prolonged adverse effects on economic growth when crafting monetary and fiscal stimulus policies.

Literature Review. Before we conclude this Introduction, we briefly discuss the econometric foundation on which this research stands. The finite sample bias of time series AR models is raised by Kendall (1954). The detrimental effect of such bias is amplified in panel data. Nickell (1981) showcases the treacherous nature of panel data: a seemingly innocuous procedure may face unexpected difficulty when a transformation takes care of the individual-specific heterogeneity embodied by the FE; see a recent survey by Okui (2021). When T is finite, Anderson and Hsiao (1982), and Arellano and Bond (1991), Arellano and Bover (1995), Hsiao et al. (2002) fix the Nickell bias based on instrument variables (IV) constructed from past observations. In recent years as more time periods are collected in panel data, researchers have taken advantage of the large-N-large-T asymptotic framework. Passing T to infinity opens many possibilities as well as challenges; see Fernández-Val and Weidner (2018). Kiviet (1995) and Bun and Carree (2005) show that the Nickell bias can be corrected by analytical formula without resorting to instruments. Dhaene and Jochmans (2015) tackle the Nickell bias by SPJ, which spares applied researchers from analytical derivations.

The Nickell bias in a single-equation dynamic panel model naturally extends to panel VAR. Holtz-Eakin et al. (1988) consider estimating the coefficients with the help of IV under finite T. Hahn and Kuersteiner (2002) start their analysis in a panel VAR and then narrow it down to a single equation with lagged dependent variables. Greenaway-McGrevy (2013) studies the bias-variance trade-off of multi-step prediction of panel VAR and follows Hahn and Kuersteiner (2002) for bias correction. Though the symptom of (implicit) Nickell bias arises from the VAR, LP focuses on a single equation of economic interest. As a result, in this paper we employ Dhaene and Jochmans (2015)'s single-equation SPJ estimator as an "automated" method, instead of SPJ's panel VAR version (Dhaene and Jochmans, 2016).

This paper is closely related to Chudik et al. (2018) and Herbst and Johannsen (2022). Chudik et al. (2018) explore the Nickell bias in a generic liner panel with weakly exogenous regressors and then extend SPJ to correct it in two-way FE models. This paper, on the other hand, is motivated by panel LP, and the solutions and empirical examples are provided for the panel LP. As panel LP involves a series of regressions, the same regressor $x_{i,t+1}$ satisfying strict exogeneity in (1) when $h = 0$ can violate it when $h \geq 1$. Therefore, weak exogeneity is deduced, rather than imposed, as an intrinsic feature of panel LP in learning the economic IRF. Moreover, we provide an in-depth study of panel VAR(∞) about when strict exogeneity is violated (see Appendix A.2) in panel LP. To further evaluate the practical relevance, we examine influential macro-finance studies under a unified theme of financial crises. Herbst and Johannsen (2022) investigate the finite sample bias in the time series LP model with and without (in their Appendix A.2) lagged dependent variables and then extend their discussion into panel data; they employ Bao and Ullah (2007) for analytical bias correction. This paper points out that all regressors in panel LP are subject to Nickell bias regardless of the linear specification, and provides a generic solution that is free of analytical bias formula.

This paper is the first to point out the omnipresence of Nickell bias in panel LP, to the best of our knowledge. We do not aim to introduce a new econometric method to resolve the Nickell bias. Instead, we expose the flawed common practice of FE estimation in the panel LP and offer an easy-to-implement solution. We attempt to draw the attention of applied macroeconomists toward this issue and strengthen the quality of economic discoveries with procedures corroborated by econometric theory.

Organization. The rest of the paper is organized as follows. In Section 2 we first use a simple two-equation model to demonstrate the presence of Nickell bias without lagged dependent variables and provide the analytical expression of this bias. We then show the Nickell bias is a generic phenomenon in panel LP and propose using the SPJ to restore asymptotic normality centered at zero. Moreover, Monte Carlo simulation is carried out to verify the theoretical predictions. Section 3 estimates the IRFs by FE and SPJ in the three empirical examples of macro-finance to evaluate the impact of financial crises. Proofs, theoretical extensions, and additional empirical outcomes are relegated to the Appendix.

2 Models and Theory

The FE estimator is the most popular method that estimates the slope coefficient in the linear panel regression and in the meantime controls the unobservable individual-specific heterogeneity. For simplicity let us consider the regression (1). Let $T_h = T - h$ be the effective sample size and $\mathcal{T}^h = [T_h] = \{1, 2, \ldots, T_h\}$ be the corresponding index set, where throughout the paper we use $[q] = \{1, 2, \cdots, q\}$ for a generic natural number q to denote the set of positive integers up to q. The FE estimator is

$$\widetilde{\beta}^{(h)\mathrm{fe}} = \sum_{i\in[N]}\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t}\widetilde{y}_{i,t+h} \Big/ \sum_{i\in[N]}\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t}^2, \tag{4}$$

where $\widetilde{y}_{i,t+h} = y_{i,t+h} - T_h^{-1}\sum_{t\in\mathcal{T}^h} y_{i,t+h}$ is the within-group demeaned dependent variable, and similarly within-group demeaned is the regressor $\widetilde{x}_{i,t}$. To conduct statistical inference about $\beta^{(h)}$, the standard deviation of $\widetilde{\beta}^{(h)\mathrm{fe}}$ is calculated as

$$\widetilde{s}^{(h)\mathrm{fe}} = (\sum_{i\in[N]}\sum_{t,s\in\mathcal{T}^h} \widetilde{x}_{i,t}\widetilde{x}_{i,s}\widetilde{e}_{i,t+h}^{(h)\mathrm{fe}}\widetilde{e}_{i,s+h}^{(h)\mathrm{fe}})^{1/2} / \sum_{i\in[N]}\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t}^2$$

where $\widetilde{e}_{i,t+h}^{(h)\mathrm{fe}} = \widetilde{y}_{i,t+h} - \widetilde{x}_{i,t}\widetilde{\beta}^{(h)\mathrm{fe}}$ is the estimation residual. We then construct the t-statistic

$$(\widetilde{\beta}^{(h)\mathrm{fe}} - \beta^{(h)\mathrm{null}})/\widetilde{s}^{(h)\mathrm{fe}},$$

where $\beta^{(h)\mathrm{null}}$ is a hypothesized value under a null of economics interest, and compare the value of the t-statistic with a critical value drawn from the standard normal distribution $\mathcal{N}(0,1)$. For example, for a two-sided test with size 5%, we reject the null if the absolute value of the t-statistic is larger than 1.96. Does the t-statistic based on $\widetilde{\beta}^{(h)\mathrm{fe}}$ provide valid statistical inference for the true IRF $\beta^{(h)}$? This question is complicated by the within-group transformation in the dynamic setting.

2.1 Implicit Nickell Bias in Panel LP

The procedure described above is the common practice based on FE. However, there is an intrinsic Nickell bias built into panel LP. Jordà (2005) constructs his time series LP by a VAR system. Here we use (2) and (3)—a stylized panel VAR(1)—as the true DGP for demonstration.

The dependent variable $y_{i,t+1}$ in (2) is linked to the regressor $x_{i,t+1}$ by the slope parameter $\beta^{(0)}$. In (3) $x_{i,t+1}$ follow one of the simplest time series models—a stationary AR(1) model.

Let $\mathbf{x}_i^t = (x_{i,0}, x_{i,1}, \ldots, x_{i,t})$ be the time series of the regressor from time 0 up to t. When assuming $u_{i,t+1}^x$ and $u_{i,t+1}^y$ independently and identically (i.i.d.) distributed across i and t, we have $\mathbb{E}\left[u_{i,t+1}^y | \mathbf{x}_i^T\right] = 0$ and thus in (2) strict exogeneity holds. The FE estimator is thus asymptotically unbiased in the regression (2) for $h = 0$.

LP is a series of linear regressions across different horizons. For $h = 1$ we substitute (3) into (2) and for $h \geq 2$ we repeat the substitution to produce (1) with the following closed-form expressions

$$\begin{aligned}
\beta^{(h)} &= \rho^h \beta^{(0)} \\
\mu_i^{(h)y} &= \mu_i^{(0)y} + \beta^{(0)} \mu_i^x \sum_{s=0}^{h-1} \rho^s \\
e_{i,t+h}^{(h)} &= u_{i,t+h}^y + \beta^{(0)} \sum_{s=0}^{h-1} \rho^s u_{i,t+h-s}^x .
\end{aligned}$$

The composite error term $e_{i,t+h}^{(h)}$ is weakly exogenous[2] in that

$$\mathbb{E}\left[e_{i,t+h}^{(h)} | \mathbf{x}_i^t\right] = \mathbb{E}\left[\beta^{(0)} \sum_{s=0}^{h-1} \rho^s u_{i,t+h-s}^x + u_{i,t+h}^y \middle| x_{i,0}, \mathbf{u}_i^{x,t}\right] = 0$$

where $\mathbf{u}_i^{x,t} = (u_{i,s}^x)_{s=1}^t$ but not strictly exogenous as

$$\mathbb{E}\left[e_{i,t+h}^{(h)} | \mathbf{x}_i^T\right] = \mathbb{E}\left[\beta^{(0)} \sum_{s=0}^{h-1} \rho^s u_{i,t+h-s}^x + u_{i,t+h}^y \middle| x_{i,0}, \mathbf{u}_i^{x,T}\right] = \beta^{(0)} \sum_{s=0}^{h-1} \rho^s u_{i,t+h-s}^x \neq 0$$

if $\beta^{(0)} \neq 0$.

Remark 1. For notational conciseness we set $0^0 = 1$ if $\rho = 0$. When $\rho = 0$, the regressor $x_{i,t}$ is independent across time, but strict exogeneity is still violated for all $h \geq 1$ due to $\beta^{(0)} \sum_{s=0}^{h-1} \rho^s u_{i,t+h-s}^x = \beta^{(0)} u_{i,t+h}^x \neq 0$. Strict exogeneity occurs in (1) if and only if $\beta^{(0)} = 0$, under which $(x_{i,t})$ and $(y_{i,t})$ are two autonomous time series with no connection.

At the first glance, (1) is a seemingly innocuous regression of $y_{i,t+h}$ on another variable $x_{i,t}$. It turns out that (2) and (3) consist of a two-equation panel *vector* AR model. When the within-group transformation is used to eliminate the fixed effects, the implicit Nickell bias is present even though the lagged y does not explicitly appear on the right-hand side of (1).

[2] Chudik et al. (2018) define weak exogeneity in terms of unconditional moments. We define weak exogeneity in terms of mean independence conditional on the past information, following Mikusheva and Sølvsten (2023).

The following proposition characterizes the bias in the asymptotic distribution.

Proposition 1. *Suppose the zero mean innovations* $u_{i,t}^y$ *and* $u_{i,t}^x$ *are independently and identically distributed across* i *and* t*. If*

$$s_x^2 := \frac{1}{NT_h} \sum_{i\in[N]} \sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t}^2 \xrightarrow{p} \sigma_x^2 > 0 \tag{5}$$

$$\frac{1}{\sqrt{NT_h}} \sum_{i\in[N]} \sum_{t\in\mathcal{T}^h} \left(\widetilde{x}_{i,t} e_{i,t+h}^{(h)} - \mathbb{E}\left[\widetilde{x}_{i,t} e_{i,t+h}^{(h)}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{xe,h}^2\right) \tag{6}$$

either as $N \to \infty$ *with a fixed* T*, or* $(N,T) \to \infty$ *jointly, then*

$$\sqrt{NT_h}\left(\widetilde{\beta}^{(h)\mathrm{fe}} - \beta^{(h)}\right) + \beta^{(0)} \cdot \frac{\sigma_{u_x}^2}{s_x^2} \sqrt{\frac{N}{T_h}} f_{T,h}(\rho) \xrightarrow{d} \mathcal{N}\left(0,\ \frac{\sigma_{xe,h}^2}{\sigma_x^4}\right), \tag{7}$$

where

$$f_{T,h}(\rho) = \frac{\left(1-\rho^h\right)}{\left(1-\rho\right)^2} - \frac{h}{\left(T-h\right)\left(1-\rho\right)^2} + \frac{\rho^{T-2h+1}\left(1-\rho^{2h}\right)}{\left(T-h\right)\left(1-\rho\right)^2\left(1-\rho^2\right)},$$

$\sigma_{u_x}^2 = \mathrm{var}\left[u_{i,t}^x\right]$, $\sigma_x^2 = \mathrm{var}\left[x_{i,t}\right]$ *and* $\mathcal{Z} \sim \mathcal{N}\left(0, \sigma_{xe,h}^2/\sigma_x^4\right)$ *is a zero-mean normal random variable.*

Remark 2. Simplifying assumptions are commonly employed in the large-N-large-T panel data literature to help make the key point clear. Here, the i.i.d. innovations keep concise the expressions of the bias and variance, and the high-level conditions in (5) and (6) avoid technical distractions highlighted in Phillips and Moon (1999)'s panel data joint (N,T) asymptotics.

Proposition 1 implies

$$\widetilde{\beta}^{(h)\mathrm{fe}} \overset{a}{\sim} \beta^{(h)} - \frac{\beta^{(0)}\sigma_{u_x}^2}{T_h s_x^2} f_{T,h}(\rho) + \frac{\mathcal{Z}}{\sqrt{NT_h}}, \tag{8}$$

where "$\overset{a}{\sim}$" signifies asymptotic similarity. The bias is of order $1/T$. If T is fixed, $\widetilde{\beta}^{(h)\mathrm{fe}}$ is inconsistent as $N \to \infty$. When $(N,T) \to \infty$ jointly with $N/T \to c$, the FE estimator $\widetilde{\beta}^{(h)\mathrm{fe}}$ is a consistent estimator of $\beta^{(h)}$, and hence ρ, $\sigma_{u_x}^2$, σ_x^2 and $\sigma_{xe,h}^2$ are also consistently estimable. However, the bias in (7)

$$-\beta^{(0)} \cdot \frac{\sigma_{u_x}^2}{s_x^2} \sqrt{\frac{N}{T_h}} f_{T,h}(\rho) \to -\beta^{(0)} \cdot \frac{\sigma_{u_x}^2}{\sigma_x^2} \sqrt{c} \cdot \frac{\left(1-\rho^h\right)}{\left(1-\rho\right)^2} \tag{9}$$

does not vanish asymptotically and it will distort the test size of the usual inference based on the t-statistic. Following Hahn and Kuersteiner (2002) and Okui (2010), one can correct the bias based on (7) for valid asymptotic inference.

Remark 3. Though it is possible to formulate Chudik et al. (2018)'s Eq.(1) by re-defining the regressors to incorporate our (1) as a special case, LP is a series of regressions where the direction and the magnitude of bias depend on many population parameters in the model. Consider $\rho > 0$ as in most economic autoregressive relationships, and we find the following features that are peculiar to the Nickell bias in panel LP.

1. The limit expression (9) implies that given all the parameters in the DGP, the bias worsens with large h, as $1 - \rho^h$ increases with h.

2. The bias shrinks the FE estimator toward zero as (8) becomes

$$\widetilde{\beta}^{(h)\text{fe}} \overset{a}{\sim} \beta^{(h)} \left(1 - \frac{1}{T_h} \cdot \frac{\sigma_{u_x}^2 f_{T,h}(\rho)}{\sigma_x^2 \rho^h}\right) + \frac{\mathcal{Z}}{\sqrt{NT_h}}$$

   by noticing the true IRF $\beta^{(h)} = \beta^{(0)}\rho^h$ and $0 < \sigma_{u_x}^2 f_{T,h}(\rho) / \left(\sigma_x^2 \rho^h\right) = O(1)$. No matter whether the true $\beta^{(h)}$ is positive or negative, the FE estimator exhibits an *attenuation bias* and underestimates $|\beta^{(h)}|$.

In simulation studies and empirical applications, we find that these intriguing phenomena of the FE estimator are common in the simple AR(1) specification as well as more general cases. Although the true DGPs in real data studies are unknown, in our empirical application in Section 3 we observe that the bias correction enlarges the magnitude of the IRFs and the biggest discrepancy often occurs on a relatively long horizon.

2.2 Main Equation based on Panel VAR

While we have demonstrated the bias in the FE estimation of (1) from the prototype DGP (2) and (3), the Nickell bias looms in general dynamic systems. In practical use of the panel AR model, researchers may want to include lagged dependent variables as well as other control variables. For example, Nickell (1981, p.1424) considers a panel ARX (in our notation) $y_{i,t+1} = \mu_i^y + \gamma y_{i,t} + \beta x_{i,t+1} + u_{i,t+1}^y$.

Though LP focuses on single-equation regressions instead of simultaneous-equation systems, it is helpful to present the underlying VAR system which implies the series of regressions. In this section, we consider data generated from a panel VAR. Let $\mathbf{x}_t$ be a K-dimensional random vector, and the observed data at time t is combined into a $(K+1)$-vector

$\mathbf{w}_{i,t} = (y_{i,t}, \mathbf{x}'_{i,t})'$. We write down a panel structural VAR(p) model

$$\mathbf{A}_0 \mathbf{w}_{i,t+1} = \boldsymbol{\mu}_i^{(0)} + \sum_{s=1}^{p} \mathbf{A}_s \mathbf{w}_{i,t+1-s} + \mathbf{u}_{i,t+1} \tag{10}$$

as the DGP, where $\mathbf{A}_s$, $s = 0, 1, \ldots, p$, are $(1+K)\times(1+K)$ coefficient matrices,[3] and $\boldsymbol{\mu}_i^{(0)}$ is the vector of individual-specific fixed effects. The Wold-causal order requests the left-bottom block of $\mathbf{A}_0$ to be a K-vector of zeros (Jordà, 2005), and we standardize its diagonal line as 1. We rewrite (10) as

$$\begin{pmatrix} 1 & -\mathbf{a}'_{0,yx} \\ \mathbf{0} & \mathbf{A}_{0,x} \end{pmatrix} \begin{pmatrix} y_{i,t+1} \\ \mathbf{x}_{i,t+1} \end{pmatrix} = \begin{pmatrix} \mu_i^{(0)y} \\ \boldsymbol{\mu}_i^{(0)x} \end{pmatrix} + \sum_{s=1}^{p} \begin{pmatrix} a_{s,y} & \mathbf{a}'_{s,yx} \\ \mathbf{a}_{s,xy} & \mathbf{A}_{s,x} \end{pmatrix} \begin{pmatrix} y_{i,t+1-s} \\ \mathbf{x}_{i,t+1-s} \end{pmatrix} + \begin{pmatrix} u^y_{i,t+1} \\ \mathbf{u}^x_{i,t+1} \end{pmatrix}, \tag{11}$$

where the matrices/vectors are partitioned in a compatible manner. The structural form of the first equation is

$$y_{i,t+1} = \mu_i^{(0)y} + \mathbf{a}'_{0,yx}\mathbf{x}_{i,t+1} + (1, \mathbf{0}') \times \sum_{s=1}^{p} \mathbf{A}_s \mathbf{w}_{i,t+1-s} + u^y_{i,t+1}. \tag{12}$$

As derived in Appendix A.1, the predictive equation for $y_{i,t+h}$ is

$$y_{i,t+h} = \mu_i^{(h)y} + \mathbf{W}'_{i,t}\boldsymbol{\theta}^{(h)} + e^{(h)}_{i,t+h}, \quad h = 1, 2, \ldots, H, \tag{13}$$

where $\mathbf{W}_{i,t} = (\mathbf{w}'_{i,t}, \mathbf{w}'_{i,t-1}, \ldots, \mathbf{w}'_{i,t-p+1})'$ is the $p(K+1)$ long vector of regressors, $\boldsymbol{\theta}^{(h)}$ is the corresponding coefficient vector, and the error term

$$e^{(h)}_{i,t+h} = (1, \mathbf{0}') \sum_{s=0}^{h-1} \mathbf{A}_0^{-s} \mathbf{A}_1^s \mathbf{A}_0^{-1} \mathbf{u}_{i,t+h-s}.$$

We impose the following assumption to ensure $e^{(h)}_{i,t+1}$ satisfies weak exogeneity.

Assumption 1. *(a) For each i, the times series $(\mathbf{u}_{i,t})_{t\in[T]}$ is a strictly stationary martingale difference sequence (m.d.s.) with respect to its natural filtration, and $\mathbb{E}[u^y_{i,t}|\mathbf{u}^x_{i,t}] = 0$. (b) All*

[3] Without loss of generality, here we write the numbers of the lags of all components in $\mathbf{w}_{i,t}$ being the same for all components. The researcher has the discretion to choose the numbers of lags for specific variables either by prior knowledge or by some information criterion, and in this case p here is viewed as the largest number of lags with the shorter lags being lengthened by zero coefficients.

roots of the determinant equation

$$g(z) = \det\left(\mathbf{A}_0 - \sum_{s=1}^{p} \mathbf{A}_s z^s\right) = 0$$

stay outside of the unit circle on the complex plane. (c) All individual time series are independent across i.

The m.d.s assumption and $\mathbb{E}[u_{i,t}^y | \mathbf{u}_{i,t}^x] = 0$ in Condition (a) guarantee that $u_{i,t+1}^y$ is mean independent of all the right-hand side regressors in (12). As $e_{i,t+h}^{(h)}$ is a linear combination of $(\mathbf{u}_{i,s})_{s=t+1}^{h}$, it is mean-independent of the past information. Condition (b) ensures that the observed variables $\mathbf{w}_{i,t}$ are strictly stationary over time, and Condition (c) rules out cross-sectional dependence for simplicity.

When $\mathbf{x}_{i,t}$ is multivariate, without loss of generality we can denote the variable of main economic interest as the first scalar $x_{i,t}^{\heartsuit}$, and the rest of the vector $\mathbf{x}_{i,t}^{\diamondsuit}$ as additional control variables to make more plausible the mean independence in (12) and weak exogeneity in (13).[4] In estimation, however, $x_{i,t}^{\heartsuit}$ and $\mathbf{x}_{i,t}^{\diamondsuit}$ are symmetric in that they share the same status as regressors accompanying the potential lagged dependent variables in the predictive equation (13). In the panel LP regression, all regressors in $\mathbf{w}_{i,t}$ incur the Nickell bias regardless of the numbers of lags, which is elaborated in Appendix A.2 for a VAR(∞) model.

Remark 4. This fact has profound implications to the IV method for the dynamic panel regression. The above discussion made clear that if we intend to seek IVs, we must prepare instruments not only for $(y_{i,t+1-s})_{s=1}^{p}$, but also for every variable in $(\mathbf{x}_{i,t+1-s})_{s=1}^{p}$. In practice, many IVs can bring about poor finite sample performance (Roodman, 2009), not to mention that weak IVs further complicate the econometrics (Andrews et al., 2019). Therefore, we would like to avoid using the IV approach in panel LP.

Closed-form expressions of the bias, as in Proposition 1 for the prototype model, are intractable in a full-scale dynamic system such as (13). We suggest an automated bias correction method in the next section.

2.3 Split-panel Jackknife

Panel LP requires a correct specification of the main equation of interest, but keeps an agnostic attitude toward the dynamics of $\mathbf{x}_{i,t}$ in the lower block of (11). Unlike the analytic bias correction, Dhaene and Jochmans (2015)'s SPJ is a data-driven method that suits well with the single-equation panel LP.

[4]For example, both Baron et al. (2021) and Mian et al. (2017) look at the effects of two regressors in x. When one of them is the focus of economic interpretation, the other serves as the respective control variable.

Denote the FE estimator of (13) as $\widehat{\boldsymbol{\theta}}^{(h)\text{fe}}$, and let $\widehat{\boldsymbol{\theta}}_a^{(h)\text{fe}}$ and $\widehat{\boldsymbol{\theta}}_b^{(h)\text{fe}}$ be the FE estimators using the first half of observations in the set $\mathcal{T}_a^h := \{t \le T_h/2\}$ and the second half of observations in the set $\mathcal{T}_b^h := \{T_h/2 < t \le T_h\}$, respectively. The SPJ estimator is defined as

$$\widetilde{\boldsymbol{\theta}}^{(h)\text{spj}} = 2\widehat{\boldsymbol{\theta}}^{(h)\text{fe}} - \frac{1}{2}\left(\widehat{\boldsymbol{\theta}}_a^{(h)\text{fe}} + \widehat{\boldsymbol{\theta}}_b^{(h)\text{fe}}\right).$$

Remark 5. There have been many proposed solutions in the literature of Nickell bias, and therefore it is not our intention to invent yet another new method. In an independent research from ours, Dube et al. (2023, April) is aware of the explicit Nickell bias in panel LP with the lagged dependent variables included and refers to Chen et al. (2019)'s SPJ as a potential solution. Our contribution here is to verify that the theory of SPJ goes through in panel LP with the series of predictive regressions. A key ingredient is our Lemma 1 in the Appendix, which establishes the order of $\sum_{t\in\mathcal{T}_a^h}\mathbb{E}\left[\bar{\mathbf{W}}_{i,b}e_{i,t+h}^{(h)}\right]$ that crosses the blocks $\mathcal{T}_a^h$ and $\mathcal{T}_b^h$. This term is peculiar to panel LP and its order depends on h, for $e_{i,t+h}^{(h)}$ is not an arbitrary exogenous shock but the error term yielded from iterating the first equation of the reduced-form VAR.

To establish the asymptotic properties of the SPJ estimator, we impose the following Assumption 2 that consists of high-level conditions of the law of large numbers and central limit theorem commonly seen in the literature of panel data, say Bai (2009). Define

$$\widehat{\mathbf{Q}}_k := \frac{1}{NT_h/2}\sum_{i\in[N]}\sum_{t\in\mathcal{T}_k^h}\widetilde{\mathbf{W}}_{i,t}\widetilde{\mathbf{W}}_{i,t}^\top,$$

for $k \in \{a, b\}$ associated with the two halves of the data over the time dimension, where $\widetilde{\mathbf{W}}_{i,t} = \mathbf{W}_{i,t} - T_h^{-1}\sum_{t\in\mathcal{T}^h}\mathbf{W}_{i,t}$. Let $\widehat{\mathbf{Q}} := (\widehat{\mathbf{Q}}_a + \widehat{\mathbf{Q}}_b)/2$. We use $\mathbf{1}\{\cdot\}$ to denote the indicator function.

Assumption 2. *There are positive-definite matrices* $\mathbf{Q}$ *and* $\mathbf{R}$ *such that*

$$\mathbf{Q} = \text{plim}_{(N,T)\to\infty}\widehat{\mathbf{Q}}_a = \text{plim}_{(N,T)\to\infty}\widehat{\mathbf{Q}}_b$$

and

$$\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\left[\mathbf{d}_{i,t}^*e_{i,t+h}^{(h)} - \mathbb{E}\left[\mathbf{d}_{i,t}^*e_{i,t+h}^{(h)}\right]\right] \xrightarrow{d} \mathcal{N}(0, \mathbf{R}) \tag{14}$$

as $(N,T)\to\infty$*, where*

$$\mathbf{R} = \lim_{(N,T)\to\infty}\frac{1}{N}\sum_{i\in[N]}\mathbb{E}\left[\frac{1}{T_h}\sum_{t,s\in\mathcal{T}^h}\mathbf{d}_{i,t}^*\mathbf{d}_{i,s}^{*\top}e_{i,t+h}^{(h)}e_{i,s+h}^{(h)}\right]$$

with $\mathbf{d}^*_{i,t} = (\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,b}) \cdot \mathbf{1}\left\{t \in \mathcal{T}^h_a\right\} + (\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,a}) \cdot \mathbf{1}\left\{t \in \mathcal{T}^h_b\right\}$ *and* $\bar{\mathbf{W}}_{i,k} = (T_h/2)^{-1} \sum_{t \in \mathcal{T}^h_k} \mathbf{W}_{i,t}$ *for* $k \in \{a, b\}$.

The SPJ estimator is asymptotically normal without a bias term if T is non-trivial relative to N in that $N/T^3 \to 0$.

Theorem 1. *If Assumptions 1 and 2 hold, then*

$$\sqrt{NT_h}\left(\widetilde{\boldsymbol{\theta}}^{(h)\mathrm{spj}} - \boldsymbol{\theta}^{(h)}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}) \tag{15}$$

as $(N, T) \to \infty$ *and* $N/T^3 \to 0$.

The asymptotic distribution centered around zero allows us to invoke the common procedure for inference. To construct a feasible t-statistic, we compute the

$$\widehat{\mathbf{R}} = (NT_h)^{-1} \sum_{i \in [N]} \sum_{t,s \in \mathcal{T}^h} \mathbf{d}^*_{i,t} \mathbf{d}^{*\top}_{is} \widetilde{e}^{(h)\mathrm{spj}}_{i,t+h} \widetilde{e}^{(h)\mathrm{spj}}_{i,s+h},$$

where $\widetilde{e}^{(h)\mathrm{spj}}_{i,t+h} = \widetilde{y}_{i,t+h} - \widetilde{\mathbf{W}}^{\top}_{i,t} \widetilde{\boldsymbol{\theta}}^{\mathrm{spj}}_h$ is the SPJ's estimated residual. Given some standard assumptions $\widehat{\mathbf{R}}$ consistently estimates the individual-clustered variance $\mathbf{R}$. The standard practice of statistical inference based on the t-statistic is asymptotically valid.

Remark 6. The theory in this section covers the case with cross-sectional FE and one-way clustered standard error only. There are many alternative specifications in empirical applications. For example, a researcher may want to allow cross-sectional correlation in the standard error, to add time fixed effects to control temporal heterogeneity, and to accommodate unbalanced panel data. To widen the applicability of the SPJ approach, Appendix B.2 discusses the two-way clustered standard error, the two-way fixed effects, and the observation-splitting procedure for unbalanced panels.

2.4 Simulations

To illustrate the presence of Nickel bias and the finite sample performance of SPJ, we conduct Monte Carlo simulation exercises based on the prototype model of the simple form as in Equations (2) and (3). We generate the innovations $\varepsilon_{i,t}$ and $u_{i,t}$ from independent $\mathcal{N}(0,1)$, and the fixed effect is generated by $\mu^{(0)y}_i = 0.2\sqrt{T}\bar{x}_i + \xi_i$ where $\xi_i \sim \mathcal{N}(0,1)$ is independent of all other variables. For the contemporaneous connection between $y_{i,t+1}$ and $x_{i,t+1}$, we specify $\beta^{(0)} = -0.6$, where the negative relationship is motivated by the impact of financial crises on the real economy. The AR(1) coefficient $\rho \in \{0, 0.2, 0.5, 0.8\}$ varies the persistence of the predictor. We consider the sample sizes $(N, T) \in \{(30, 60), (30, 120), (50, 120)\}$ in line with the empirical examples. All simulation results are produced by 1000 independent replications. Recall that Figure 2 in Introduction is a special case here with $\rho = 0.8$ and $(N, T) = (50, 120)$.

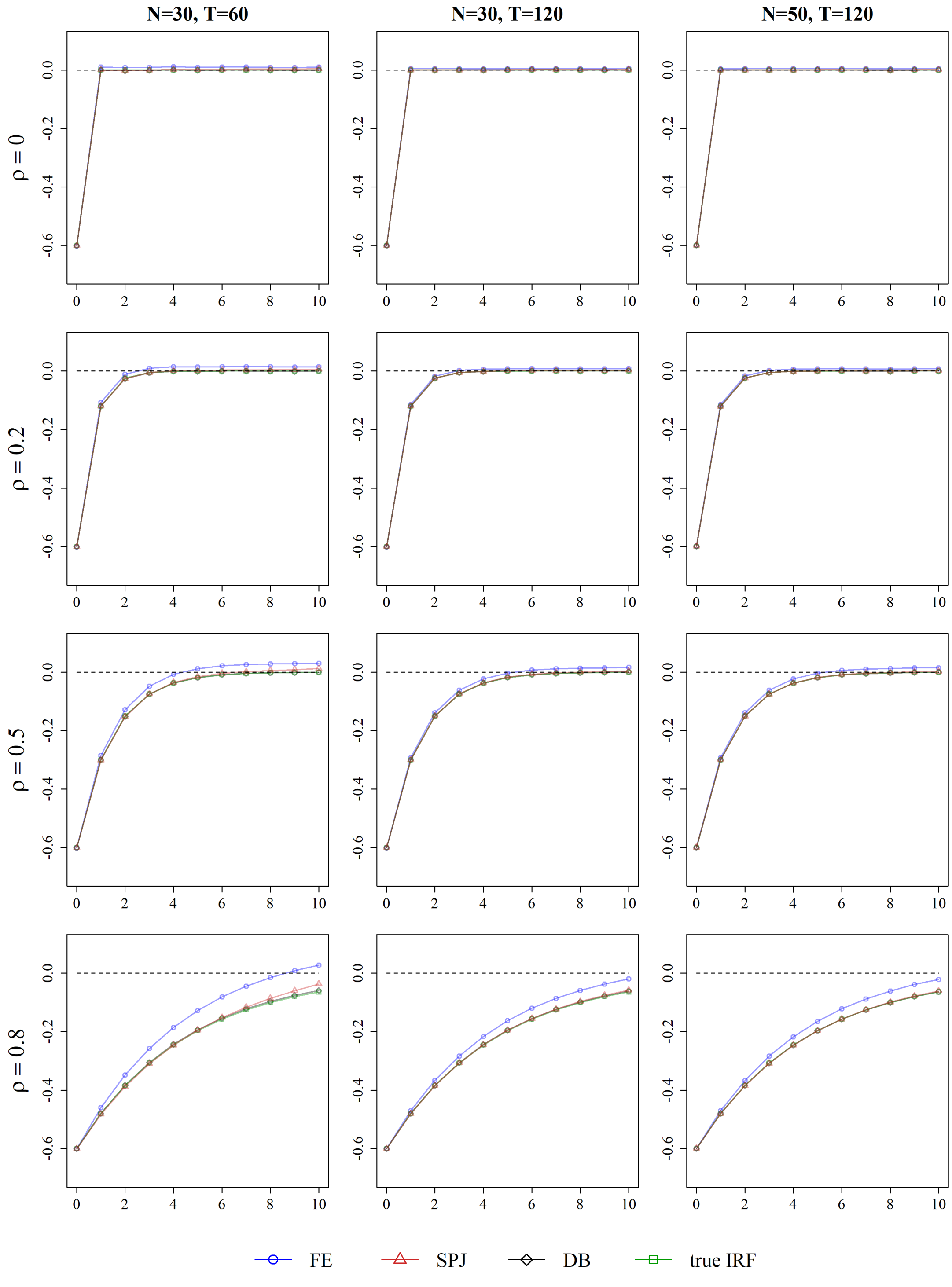
Proposition 1 offers the closed-form formula of the bias for the AR(1) specification of x. It provides an "oracle" debiased (DB) estimator (See Appendix A.4 for details) to take advantage of the closed-form formula, which is based on the oracle of correct specification of AR(1). It will be compared with the plain FE and the SPJ estimators that keep an agnostic attitude about the dynamics of x.

Figure 3 reports the average of estimated IRFs over the 1000 replications and compares them with the true IRFs. With $\beta^{(0)} < 0$, the true IRFs are negative and decay to zero exponentially fast. The FE estimator exhibits an upward bias, indicating a positive Nickell bias with an opposite sign of the true IRFs. In other words, the FE estimator underestimates the negative effects of $x_{i,t}$ on $y_{i,t+h}$. Moreover, the bias worsens as the horizon h increases. This finding is predicted by Proposition 1 and discussed in Remark 3.

Regarding the influence of ρ, FE's bias is relatively small when $\rho = 0$ or 0.2. As ρ increases to 0.5 and 0.8, the bias becomes more pronounced. The bias is mitigated when T grows from 60 to 120. When N grows from 30 to 50, the bias becomes larger. In comparison, the oracle DB estimator, as well as our recommended SPJ estimator, is very close to the true IRF.

When it comes to the estimation errors measured by root-mean-square error (RMSE), Figure 4 shows that the FE estimator produces the largest error among the three methods in all cases, while the oracle DB estimator yields the smallest RMSE. RMSEs of all the three estimators get larger as ρ gets closer to 1. The recommended method, SPJ, again exhibits similar performance to the oracle DB estimator in terms of RMSEs.

Figure 5 plots the empirical coverage probabilities of the confidence intervals (CIs) by inverting the t-statistic. The nominal 95% probability is marked by the horizontal dash line. For each (N, T) and ρ, the coverage rate of CIs from the FE estimator is close to the nominal probability when $h = 0$, while the bias emerges when $h > 0$ and the coverage rate falls short of 0.95 as h grows. Such Nickell bias is present even though $\rho = 0$ when the IRF equals zero as $h > 0$, and the performance further deteriorates as $x_{i,t}$ gets more persistent under a larger ρ. When $N = 30$, the Nickel bias is alleviated as T grows from 60 to 120, while it materializes again as N increases to 50. These numerical findings echo the closed-form form of the bias in Theorem 1. The CIs from the DB estimator, as an oracle estimator when the closeform of the Nickell bias is known, have coverage probabilities very close to 95% in all cases. The SPJ estimator closely resembles the oracle DB estimator, except for the slight deviation in the case of a small sample size $(N, T) = (30, 60)$ and $\rho = 0.8$; here SPJ still



Note: In this figure SPJ (red), DB (black) and the true IRF (green) mostly overlap.

Figure 3: Estimated IRFs Averaged Over Replications

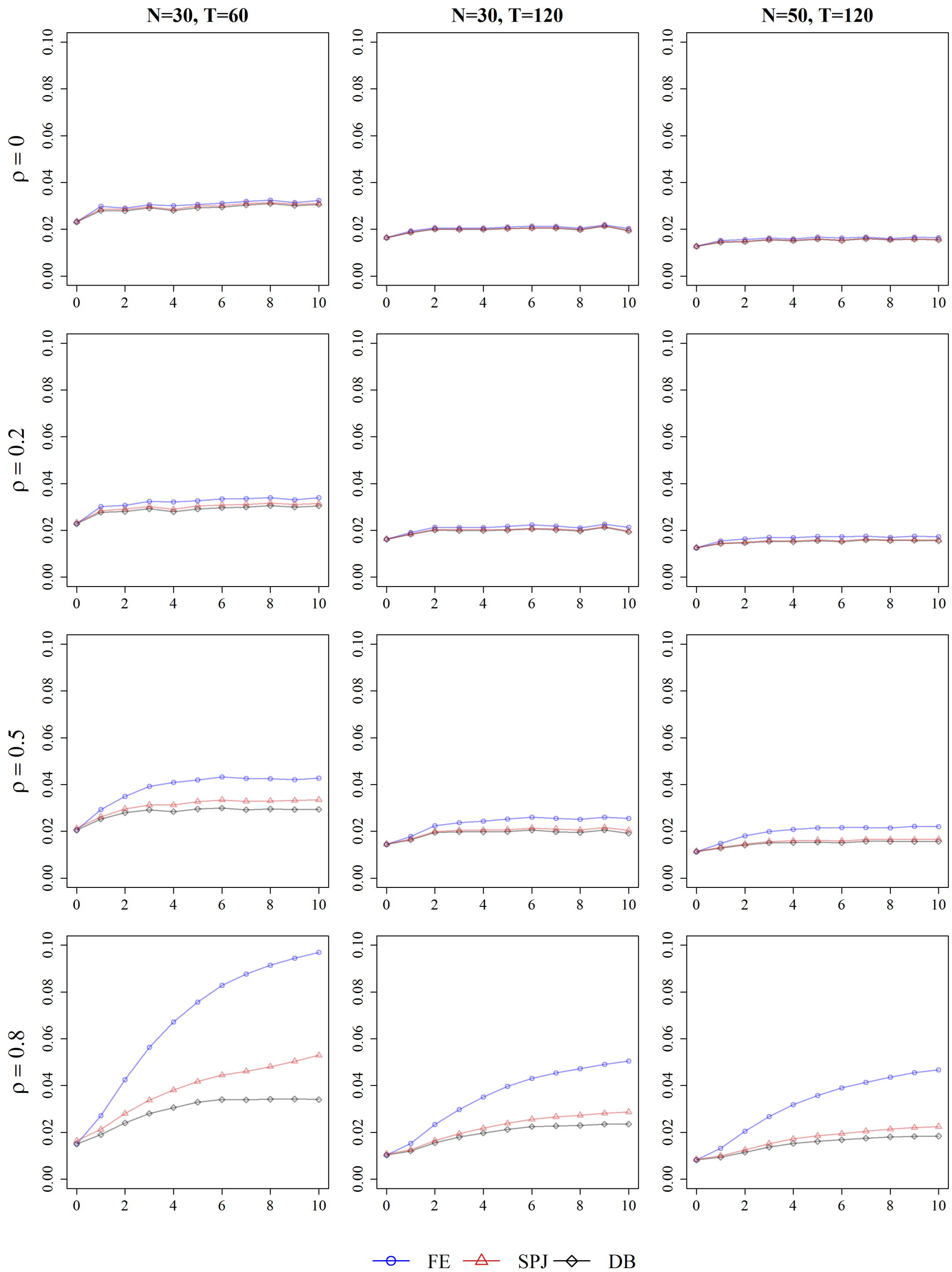
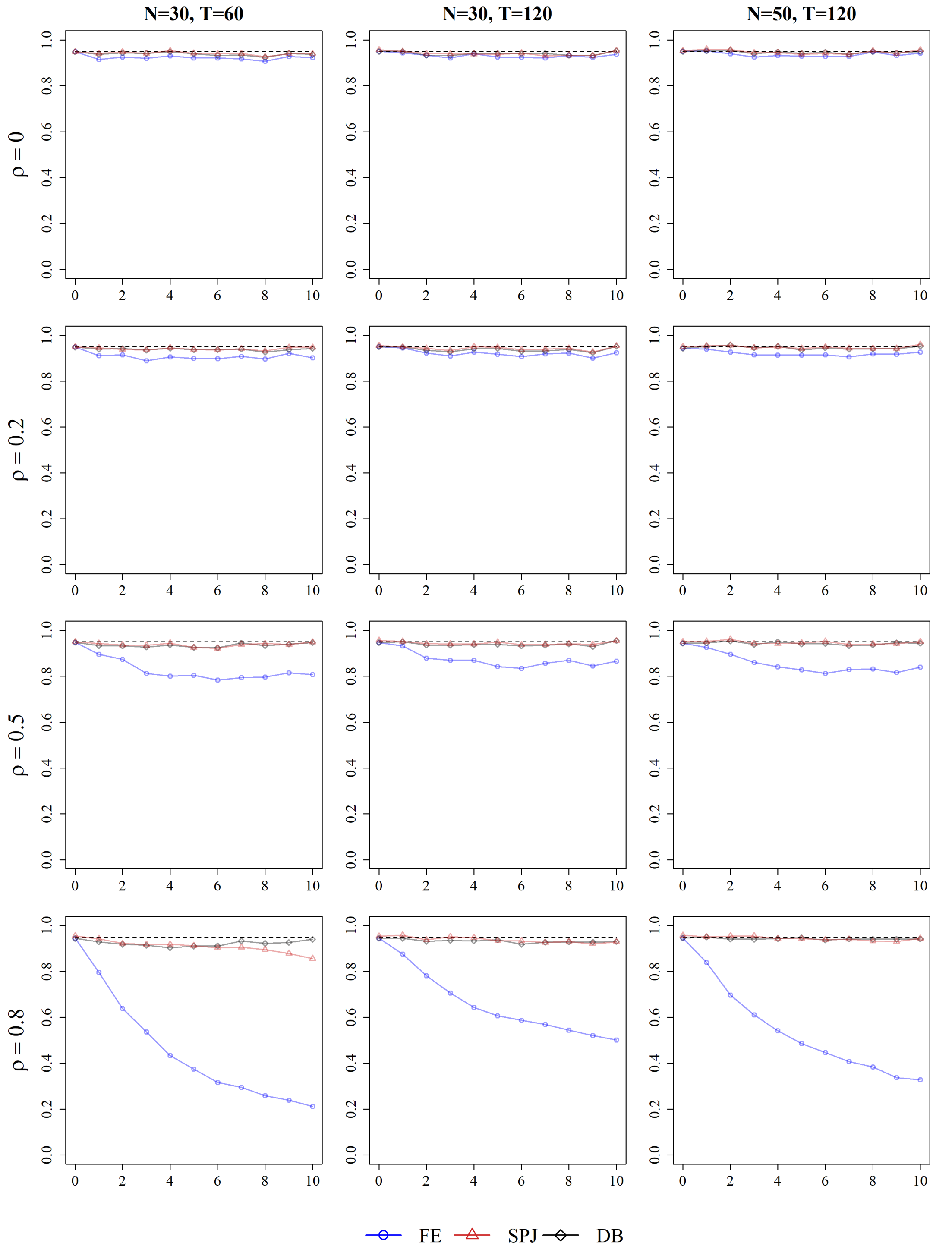


Figure 4: RMSEs of the Three Estimators



Note: The nominal level of 0.95 is marked by the horizontal dashed line.

Figure 5: Coverage Probability of the Confidence Interval Based on t-Statistic

substantially mitigates the severe bias from the FE.

To summarize, the simulation results provide a clear picture of the bias issue for the FE estimator in panel LP, and support the practical values of the SPJ estimator in bias correction. Besides the prototype model of Equations (2) and (3), real-data studies may include the lagged dependent variable as a control for the dynamics of $y_{i,t}$, which is used in all the three empirical examples in Section 3. To save space, we detail these additional simulation results in Appendix C with a model involving lagged dependent variables. As expected, the phenomena from the prototype model are observed again in the more general model.

3 Revisiting the Aftermath of Financial Crises

In this section, we apply our method to the three renowned empirical studies on macro-finance linkage as we mentioned in the Introduction. They all center on the aftermath of financial crises on output, but each of them places a distinct emphasis on different financial shocks. Romer and Romer (2017) design a new financial crisis chronology to discuss the responses of output and unemployment following the overall financial distress. Baron et al. (2021) explore the role of banking crises in the credit crunch and output reduction (Diamond and Dybvig, 1983; Bernanke, 2018). The third one and its related studies (Mian et al., 2017; Mian and Sufi, 2010) highlight the rising household debts in depressing output in the mid-run.

These studies all employ cross-country panel data and adopt the panel LP regression Eq.(13): $y_{i,t+h}$ denotes some measure of economic output, and the regressor vector $\mathbf{W}_{i,t}$ includes a key variable of interest $x_{i,t}^{\heartsuit}$—a certain measure of financial shocks. We revisit these works and contribute a methodological edge to the literature on financial crises.

3.1 Financial Distress: Romer and Romer (2017)

Crisis chronology can be traced back to the records of Caprio and Klingebiel (1996) about bank crisis events in the 1990s. Thereafter, Reinhart and Rogoff (2009), Schularick and Taylor (2012), Laeven and Valencia (2013), and Jordà et al. (2013) have adopted more precise subjective standards and quantitative data to measure financial crises and evaluate the associated economic losses. In a comprehensive study, Romer and Romer (2017) (RR, henceforth) create a narrative semiannual measure of financial distress for 24 advanced countries from 1967 to 2012, based on the contemporaneous narrative accounts of country conditions on the disruptions to credit supply in the OECD Economic Outlook. To capture the variation in financial disruption across countries and time periods, they classify financial distress into five categories: credit disruption, minor crisis, moderate crisis, big crisis, and extreme crisis, and each category is further broken into minus, regular, and plus. Thus, their new index of financial distress has a scale of 16, where 0 represents no financial distress, and a higher value indicates a more severe financial crisis. Based on this novel measure of general financial distress, they use FE to estimate panel LP and find that the average decline in output after a financial crisis is moderate in size, though persistent and statistically significant.

We revisit RR's open-access semi-annual dataset, an unbalanced panel of 24 OECD countries. RR construct the financial distress index from 1967 to 2012, but between 1967 and 1979 only Germany had a "Credit disruption (regular)" in 1974. Our analysis restricts the time period from 1980 to 2012; this sample adjustment barely changes their estimation results.

RR's panel LP specification is

$$y_{i,t+h} = \mu_i^{(h)y} + g_t^{(h)} + \beta^{(h)} x_{i,t}^{\mathrm{FD}} + \sum_{j=1}^{4} \tau_j^{(h)} x_{i,t-j}^{\mathrm{FD}} + \sum_{k=1}^{4} \eta_k^{(h)} y_{i,t-k} + e_{i,t+h}^{(h)}, \tag{16}$$

for $h = 0, ..., 10,$[5] where the dependent variable $y_{i,t+h}$ is the logarithm real GDP or the unemployment rate, and $x_{i,t}^{\mathrm{FD}}$ is their new index on financial distress. $\mu_i^{(h)y}$ and $g_t^{(h)}$ indicate country and time fixed effects for each horizon h. Moreover, four lags of the dependent variables and the financial distress index are included as control variables. The linear coefficient $\beta^{(h)}$ is the IRF of the economic activities in period h ahead of financial shocks occurring at time t, which corresponds to a minor level shift in financial distress as RR defines. To obtain a typical shock in financial distress, RR multiplies the estimates of $\beta^{(h)}$ by 7, which corresponds to a "Moderate crisis (minus)" financial crisis.

Figure 6's panels (A) and (B) show the IRFs of the real GDP and the unemployment rate to a typical financial shock, respectively, estimated by RR's FE and our recommended SPJ. SPJ confirms RR's finding that output falls and the unemployment rate rises significantly and persistently following the financial crisis, but its impulse responses exhibit considerable differences from those of the FE estimates, particularly for the long horizons.

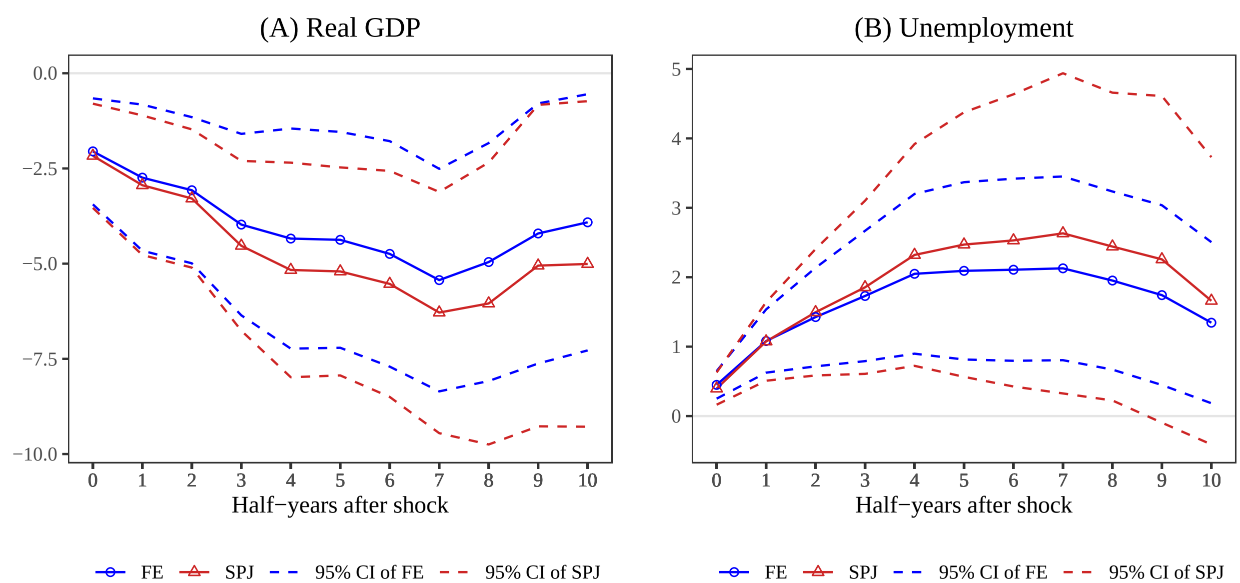
Panel (A) shows that the immediate aftermath of a moderate crisis is a fall in GDP of 2.2% based on the SPJ estimate, which is close to the FE estimate. However, the differences between the SPJ and the FE estimates become more pronounced as the horizon h rises. According to SPJ, the decline in output grows substantially 3.5 years after the financial shock and peaks at 6.3%, which is about 16% higher than the FE estimates. The differences

[5] The only difference from Eq.(13) is the presence of the time fixed-effect $g_t^{(h)y}$; See Sections B.2.2 and C for such an extension.

between the two estimates are also persistent. After 5 years ($h = 10$), the output fall based on SPJ returns to about 5%, which remains 28% larger in magnitude than the corresponding FE.

Consistent with the responses of real GDP, Panel (B) shows that the FE estimates also tend to underestimate the aftermath of a moderate financial crisis on the unemployment rate. A financial shock raises the unemployment rate steadily. As with GDP, the unemployment rate continued to increase through 3.5 years following the impulse, peaking at 2.6 percentage points based on the SPJ estimate (24% higher than the FE estimate), and after that, the differences between the two estimates remain.

Overall, FE tends to underestimate IRFs. The substantial shrinkage by FE may be partly caused by the persistence of financial distress. A simple FE estimation of a panel AR(1) model for the index of financial distress returns the autoregressive coefficient 0.836, which is comparable to the case of $\rho = 0.8$ in our simulation and echoes our Remark 3 concerning the attenuation effect of the asymptotic bias in the prototype model, even though (16) is a more general specification and the true dynamics of $x_{i,t}^{\mathrm{FD}}$ is unknown.



Note: Panel (A) and Panel (B) present the impulse responses of the logarithm real GDP and the unemployment rate to financial distress, respectively. The impulse responses are estimated from Model (16) and multiplied by 7, which corresponds to a "Moderate crisis (minus)" financial crisis. The blue solid lines are the FE estimates as in RR, while the red solid lines indicate the SPJ estimates. The dashed lines show the 95% confidence intervals based on standard errors clustered on country.

Figure 6: Impulse Responses of Real GDP and the Unemployment Rate to Financial Distress

3.2 Banking Crises: Baron et al. (2021)

The banking system is vital in the modern financial system, and a vast literature has studied the financial intermediary of banks and how banking crises led to a sustained decline in economic activities (Diamond and Dybvig, 1983; Calomiris and Mason, 2003; Gertler and Kiyotaki, 2010; Brunnermeier and Sannikov, 2014; Rampini and Viswanathan, 2019). Ben Bernake argues that bank runs played a critical role in the Great Depression of the 1930s, and the collapse of Lehman Brothers triggered global financial panics, which eventually turned into a worldwide financial crisis during 2007–2008 and the prolonged Great Recession (Bernanke et al., 1983; Bernanke, 2018). The recent meltdown of Silicon Valley Bank, the second largest failure of a financial institution in U.S. history, again spurred worries about the global financial markets.

Baron et al. (2021) (BVX, henceforth) study the aggregate output loss associated with bank crises. Different from the traditional approach that takes banking panics as banking crises, BVX argue that a large decline in bank equity return is the prerequisite for banking crises.[6] Specifically, they construct a new data set of bank equity return for 46 advanced and emerging economies over 1870–2016, and define a bank crisis as a bank equity crash of more than 30%. Their FE estimates find that bank equity crashes predict a sizable and long-lasting decline in future real GDP.

The baseline LP specification in BVX is

$$\begin{aligned}\Delta_h y_{i,t+h} = \mu_i^{(h)y} + \beta^{(h)\mathrm{B}} x_{i,t}^{\mathrm{B}} + \beta^{(h)\mathrm{N}} x_{i,t}^{\mathrm{N}} + \sum_{j=1}^{3} \left(\tau_j^{(h)\mathrm{B}} x_{i,t-j}^{\mathrm{B}} + \tau_j^{(h)\mathrm{N}} x_{i,t-j}^{\mathrm{N}} \right) \\ + \sum_{k=0}^{3} \eta_k^{(h)} \Delta y_{i,t-k} + \sum_{\ell=0}^{3} \zeta_\ell^{(h)} \Delta y_{i,t-\ell}^{\mathrm{credit}} + e_{i,t+h}^{(h)}\end{aligned} \tag{17}$$

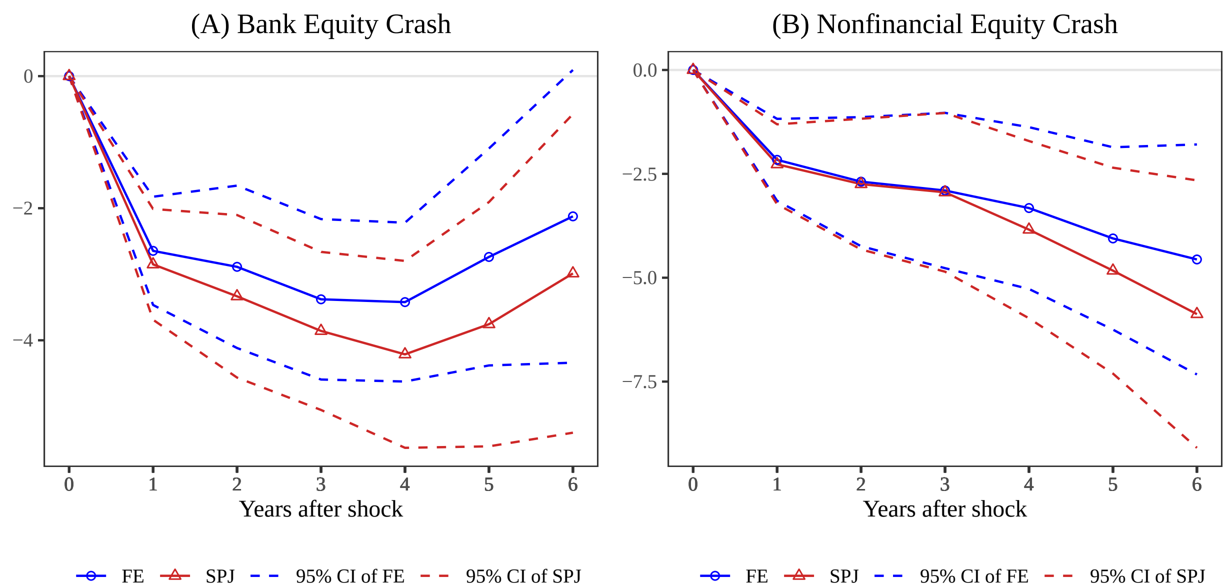
for $h = 1, ..., 6$, where the dependent variable $\Delta_h y_{i,t+h}$ is the change in the logarithm real GDP from year t to $t+h$; to be consistent with the other two empirical studies in this section, we multiply the change by 100 to represent percentages. The regressor $x_{i,t}^{\mathrm{B}} = \mathbf{1}\{r_{i,t}^{\mathrm{B}} \leq -30\%\}$ is a dummy variable for the bank equity crash when the bank equity index dropped more than 30%, and similarly $x_{i,t}^{\mathrm{N}} = \mathbf{1}\{r_{i,t}^{\mathrm{N}} \leq -30\%\}$ is a dummy for the nonfinancial equity crash.

[6] Many scholars have emphasized the criticality of bank equity. Gertler and Kiyotaki (2010) introduce financial intermediaries into their model, showing that disruptions in financial intermediation depress aggregate economic activity. He and Krishnamurthy (2013) find the asymmetry of capital market risk premiums during banking crises, that is, when bank equity is low, bank losses have a significant impact on the risk premium, but when bank equity is high, it has little effect. Brunnermeier and Sannikov (2014) suggest that a decline in bank assets through the price channel creates greater risk as the economy moves away from a steady state. Rampini and Viswanathan (2019) provide a dynamic model including bank net assets and find that low liquidity in banks will lead to more severe and prolonged recessions during banking crises.

The control variables include three lags of bank and nonfinancial equity crash dummies, and the contemporaneous and up to three-year lagged real GDP growth and change in credit-to-GDP.

The dependent variable here is the h-order difference $\Delta_h y_{i,t+h}$, and therefore the slope coefficients $\beta^{(h)\mathrm{B}}$ and $\beta^{(h)\mathrm{N}}$ represent cumulative responses. As detailed in Appendix B.1, this specification also suffers from the implicit Nickell bias, and the bias can be corrected by SPJ. Figure 7's Panel (A) shows the output declines significantly and persistently after a banking crisis. Compared with SPJ, FE underestimates the output depression, and the gap widens over time. The SPJ suggests that the cumulative decline in GDP peaks four years after the bank crisis at 4.2%, while the corresponding output gap is 3.4% by FE. Panel (B) reiterates the real output losses following a crash in nonfinancial equity. In the first three years after the nonfinancial stock crash, the IRF by SPJ is very close to the one by FE, whereas the differences in the two IRFs become visible afterward. FE again underestimates.

In the baseline specification, we do not control for time fixed effects as they may absorb the global impact of some big banking crises. BVX's FE estimates find that the inclusion of time fixed effects in the panel LP mitigates the output loss. We find similar results by SPJ, and thus the difference between them slightly declines. On the other hand, the gap between the IRFs of nonfinancial firms' equity crashes becomes more visible.



Note: Panel (A) and Panel (B) present the impulse responses of accumulative changes in the logarithm real GDP to bank and nonfinancial equity crashes respectively, which are estimated from the Model (17). The blue solid lines represent the FE estimates in BVX, while the red solid lines indicate our SPJ. The dashed lines represent 95% confidence intervals based on standard errors double-clustered on country and year.

Figure 7: Impulse Responses of Real GDP to Bank and Nonfinancial Equity Crashes

3.3 Household Debt: Mian et al. (2017)

Prior to 2007, the subprime mortgage market in the United States developed rapidly due to the continued prosperity of the U.S. housing market and the low interest rates. Household debt levels were on the rise, and mortgages were inversely proportional to household income (Mian and Sufi, 2009). The 2007–2008 global financial crisis reminds us of another important source of financial crisis: household debt. Expansion of household debt boosted consumer demand (Mian and Sufi, 2011), fueled housing speculation (Mian and Sufi, 2022), and raised house prices (Justiniano et al., 2019). A bubble cannot last forever. The cooling of the U.S. housing market, especially the hike in short-term interest rates, overloaded repayment burdens upon home buyers, leading to large-scale defaults that eventually triggered the subprime mortgage crisis.[7] The global financial crisis highlights that household debts would not only affect economic fluctuations in the short term but also hamper economic growth in the medium and long term.

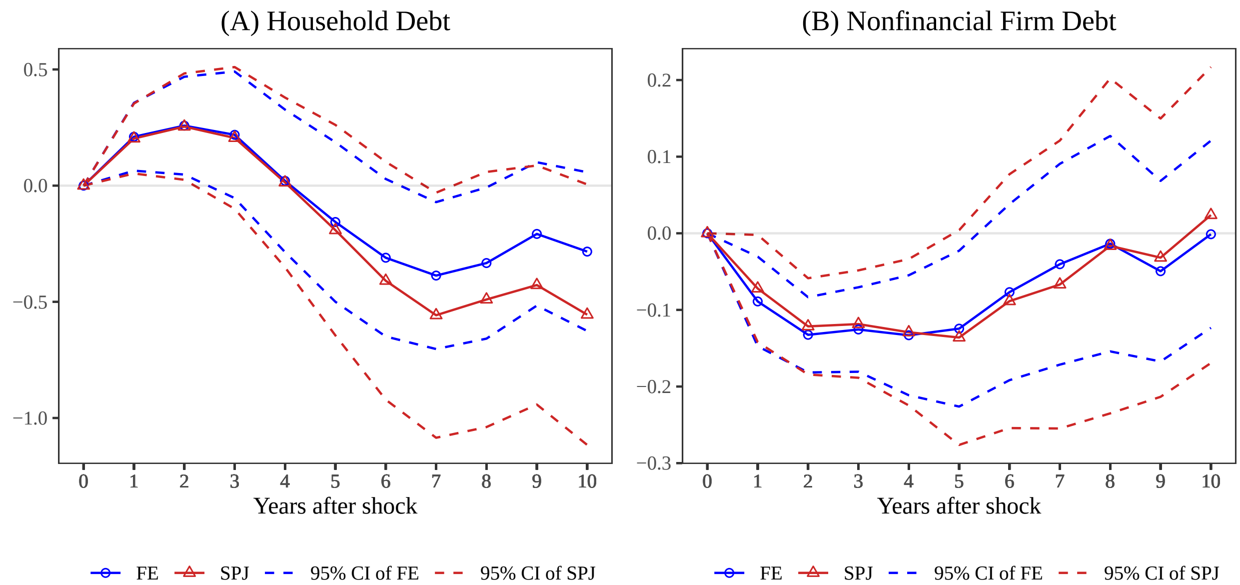
In an influential study, Mian et al. (2017) (MSV, henceforth) explore the dynamics between household debt, nonfinancial firms debt, and economic fluctuations, with a cross-country panel via both VAR and LP. They find that the household debt shock leads to a boom-recession cycle, that is, GDP first increases for 2-3 years but then decreases substantially. More precisely, they find that a rise in the household debt to GDP ratio from four years ago to last year predicts a substantial decline in subsequent real GDP growth from the current year onward. By contrast, the output declines immediately following firm debt shocks, whereas recovers gradually afterward. We will compare the baseline panel LP results using the FE estimation in MSV's Figure 2 (p.1770) with the results based on SPJ.

Collecting an unbalanced panel of 30 countries from 1960 to 2012, MSV specify the LP regression as

$$y_{i,t+h} = \mu_i^{(h)y} + \beta^{(h)\mathrm{HH}} x_{i,t}^{\mathrm{HH}} + \beta^{(h)\mathrm{F}} x_{i,t}^{\mathrm{F}} + \sum_{j=1}^{4} \left(\tau_j^{(h)\mathrm{HH}} x_{i,t-j}^{\mathrm{HH}} + \tau_j^{(h)\mathrm{F}} x_{i,t-j}^{\mathrm{F}} \right) + \sum_{k=0}^{4} \eta_k^{(h)} y_{i,t-k} + e_{i,t+h} \tag{18}$$

for $h = 1, ..., 10$, where $y_{i,t+h}$ is the logarithm real GDP. The key explanatory variables of interest are the household debt to GDP ratio $x_{i,t}^{\mathrm{HH}}$ and nonfinancial firm debt to GDP ratio $x_{i,t}^{\mathrm{F}}$. Control variables include four lags of household debt to GDP ratio and nonfinancial firm debt to GDP ratio, and contemporaneous and up to four-year lagged logarithm real GDP.

[7]In addition, the expansion of household debt in most developed economies such as the United States mainly affects the economy through the channel of consumption demand (Mian and Sufi, 2018; Mian et al., 2020). The higher the debt leverage, the greater the responsiveness of household consumption propensity to changes in housing wealth (Mian et al., 2013).



Note: Panel (A) and Panel (B) present the impulse responses of the logarithm real GDP to household debt and nonfinancial firm debt, respectively, which are estimated from Model (18). The blue solid lines represent the FE estimates in MSV, while the red solid lines indicate our SPJ. The dashed lines represent 95% confidence intervals based on standard errors double-clustered on country and year.

Figure 8: Impulse Responses of Real GDP to Household Debt and Nonfinancial Firm Debt

Our empirical analysis supports and reinforces MSV's main findings that household debt booms predict a sustained decline in output after a temporary rise. SPJ's IRFs in Figure 8 maintain the boom-recession cycle, with long-term real GDP declines greater than MSV's FE. In Panel (A) SPJ resembles FE in the first to three years, indicating a temporary rise in GDP due to the housing market boom. The output has started to decline persistently since the fourth year. The trough is observed in the seventh year, where FE estimates a ten percentage point shock in the household debt to GDP ratio is associated with a 3.9% drop, in contrast with SPJ's 5.6% drop. This is consistent with existing literature that finds debt booms may distort resource allocation and human capital accumulation, and thus reduce long-run growth (Gopinath et al., 2017; Charles et al., 2018; Borio et al., 2016). As a comparison, Panel (B) shows that the IRFs of output to firm debt shocks are close.

3.4 Summary

Our empirical exercises reexamine the aftermath of financial shocks and compare our SPJ estimates with FE. Qualitatively, we confirm that financial shocks yield significant negative impacts on output and employment, reinforcing the consensus that financial crisis recessions are deeper and longer than regular recessions. Quantitatively, SPJ indicates that the magnitudes of the effects tend to be larger than the results suggested by the conventional FE estimation, and the bias of the FE estimates tends to worsen as the horizon increases. Table 1 documents the absolute relative differences between the two estimates at the peak period of the output declines following financial shocks. Column (6) suggests that the FE estimates of economic contraction following financial distress, banking crises, and household debts as key results in those three studies suffer from substantial underestimation, ranging from 16% to 44%. Moreover, our results also suggest that the economic recovery from financial crises is slower than the FE estimates suggest. This helps to explain why the recovery from the global financial crisis in 2008 has been weak and slow (IMF, 2018). Thus, our study has important implications for policymakers who design financial surveillance systems and macro-prudential policies in targeting financial stability and economic growth.

Note that the estimated quantitative effects of financial crises on output vary widely in the existing literature for various reasons such as the definition of financial crisis, the time period, the sample of countries, and the measurement issues of GDP (Sufi and Taylor, 2021). Thus, researchers may not put too much emphasis on the exact magnitude. However, our motivation to re-examine the link between financial crisis and output loss is to showcase the substantial estimation bias arising from the common econometric method used in the literature, and to provide better econometric tools for practitioners to achieve more accurate estimates of financial crises on output loss.

Table 1: Relative Difference in the FE and SPJ estimates of IRFs

Paper and Model	Dep. Variable (1)	Indep. Variable (2)	h^{peak} (3)	$\widehat{\beta}^{(h)\text{spj}}$ (4)	$\widehat{\beta}^{(h)\text{fe}}$ (5)	Relative Diff. (6)
Romer and Romer (2017):						
Model (16)	Real GDP	Financial distress	7	-6.285	-5.432	15.69%
	Unemployment		7	2.632	2.128	23.69%
Baron et al. (2021):						
Model (17)	Real GDP growth	Bank equity crash	4	-4.213	-3.422	23.13%
		Nonfinancial equity crash	6	-5.876	-4.561	28.84%
Mian et al. (2017):						
Model (18)	Real GDP	Household debt	7	-0.558	-0.387	44.08%
		Nonfinancial firm debt	5	-0.136	-0.124	9.30%

Note: Column (3) is the peak period when the economic activities contract the most, and Columns (4) and (5) present the SPJ and FE estimates at this horizon. The period is half-year in RR and yearly in the other two studies. The relative difference, defined as $|\widehat{\beta}^{(h)\text{spj}}/\widehat{\beta}^{(h)\text{fe}} - 1| \times 100\%$, is shown in Column (6).

4 Conclusion

The Nickell bias for the FE estimator is well-known in linear dynamic panel data models where lagged dependent variables serve as regressors. Yet, it has not received a thorough examination and treatment in the panel LP, despite that recently this approach has been widely used in empirical macroeconomic studies for the estimation of impulse responses. A few empirical papers using panel LP were aware of the explicit Nickell bias and they suggested or used lagged dependent variables as IVs (Chong et al., 2012, Choi et al., 2018; Hobijn et al., 2021). We show that in the panel LP, asymptotic bias also emerges in the FE estimator even if no lagged dependent variables are present, which is employed in several empirical applications of the panel LP, for example, see Chodorow-Reich et al. (2019), Ottonello and Winberry (2020), and Bahaj et al. (2022). Furthermore, in the general specification, we illustrate that all predetermined regressors incur Nickell bias. It is imperative to correct this bias for valid asymptotic inference.

The SPJ method is capable of correcting the Nickell bias for general panel LP models. After bias correction, the standard statistical inference based on the t-statistic remains valid. By applying our bias correction procedure to three empirical studies on assessing the economic loss associated with financial crises, we show that FE tends to substantially underestimate the economic contraction following financial crises in the mid- and long- runs.

SPJ is easy to implement and can be extended to many variants of specifications. To facilitate the implementation, we provide an open-source R package named `pLP` (`https://github.com/zhentaoshi/panel-local-projection`). Given input data and the option `method = "SPJ"` or `method = "FE"`, a single-line command of the main function will automatically compute the IRF, standard error, and the corresponding 95% CI.

References

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019). Democracy does cause growth. *Journal of Political Economy 127*(1), 47–100.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*(1), 47–82.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*, 727–753.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*(1), 29–51.

Bahaj, S., A. Foulis, G. Pinter, and P. Surico (2022). Employment and the residential collateral channel of monetary policy. *Journal of Monetary Economics 131*, 26–44.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bao, Y. and A. Ullah (2007). The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics 140*(2), 650–669.

Barnichon, R. and C. Brownlees (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics 101*(3), 522–530.

Baron, M., E. Verner, and W. Xiong (2021). Banking crises without panics. *Quarterly Journal of Economics 136*(1), 51–113.

Bernanke, B. S. (2018). The real effects of disrupted credit: evidence from the global financial crisis. *Brookings Papers on Economic Activity 2018*(2), 251–342.

Bernanke, B. S. et al. (1983). Nonmonetary effects of the financial crisis in propagation of the great depression. *American Economic Review 73*(3), 257–276.

Bhattarai, S., F. Schwartzman, and C. Yang (2021). Local scars of the US housing crisis. *Journal of Monetary Economics 119*, 40–57.

Borio, C., E. Kharroubi, C. Upper, and F. Zampolli (2016). Labour reallocation and productivity dynamics: financial causes, real consequences. BIS working papers, Bank for International Settlements.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Bun, M. J. and M. A. Carree (2005). Bias-corrected estimation in dynamic panel data models. *Journal of Business & Economic Statistics 23*(2), 200–210.

Calomiris, C. W. and J. R. Mason (2003). Fundamentals, panics, and bank distress during the depression. *American Economic Review 93*(5), 1615–1647.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Caprio, G. J. and D. Klingebiel (1996). Bank insolvencies: cross-country experience. Policy research working paper series, World Bank.

Charles, K. K., E. Hurst, and M. J. Notowidigdo (2018). Housing booms and busts, labor market opportunities, and college attendance. *American Economic Review 108*(10), 2947–2994.

Chen, S., V. Chernozhukov, and I. Fernández-Val (2019). Mastering panel metrics: causal impact of democracy on growth. *American Economic Association Papers and Proceedings 109*, 77–82.

Chodorow-Reich, G., J. Coglianese, and L. Karabarbounis (2019). The macro effects of unemployment benefit extensions: a measurement error approach. *Quarterly Journal of Economics 134*(1), 227–279.

Choi, S., D. Furceri, P. Loungani, S. Mishra, and M. Poplawski-Ribeiro (2018). Oil prices and inflation dynamics: evidence from advanced and developing economies. *Journal of International Money and Finance 82*, 71–96.

Chong, Y., Ò. Jordà, and A. M. Taylor (2012). The Harrod-Balassa-Samuelson hypothesis: real exchange rates and their long-run equilibrium. *International Economic Review 53*(2), 609–634.

Chudik, A., M. H. Pesaran, and J.-C. Yang (2018). Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics 33*(6), 816–836.

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

Dhaene, G. and K. Jochmans (2016). Bias-corrected estimation of panel vector autoregressions. *Economics Letters 145*, 98–103.

Diamond, D. W. and P. H. Dybvig (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy 91*(3), 401–419.

Dube, A., D. Girardi, O. Jorda, and A. M. Taylor (2023). A local projections approach to difference-in-differences event studies. Technical report, National Bureau of Economic Research.

Fernández-Val, I. and M. Weidner (2018). Fixed effects estimation of large-T panel data models. *Annual Review of Economics 10*, 109–138.

Gertler, M. and N. Kiyotaki (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of Monetary Economics*, Volume 3, pp. 547–599. Elsevier.

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital allocation and productivity in south europe. *Quarterly Journal of Economics 132*(4), 1915–1967.

Greenaway-McGrevy, R. (2013). Multistep prediction of panel vector autoregressive processes. *Econometric Theory 29*(4), 699–734.

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

He, Z. and A. Krishnamurthy (2013). Intermediary asset pricing. *American Economic Review 103*(2), 732–70.

Herbst, E. and B. K. Johannsen (2022). Bias in local projections. Technical Report 2020-010R1, Board of Governors of the Federal Reserve System (U.S.).

Hobijn, B., F. Nechio, and A. H. Shapiro (2021). Using Brexit to identify the nature of price rigidities. *Journal of International Economics 130*, 103448.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*(6), 1371–1395.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

IMF (2018). *World Economic Outlook. Challenges to Steady Growth.* Washingtion DC: International Monetary Fund.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Jordà, Ò. (2009). Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics 91*(3), 629–647.

Jordà, Ò., M. Schularick, and A. M. Taylor (2013). When credit bites back. *Journal of Money, Credit and Banking 45*(s2), 3–28.

Jordà, Ò., M. Schularick, and A. M. Taylor (2015). Leveraged bubbles. *Journal of Monetary Economics 76*, S1–S20.

Jordà, Ò., M. Schularick, and A. M. Taylor (2016). The great mortgaging: housing finance, crises and business cycles. *Economic Policy 31*(85), 107–152.

Jordà, Ò., S. R. Singh, and A. M. Taylor (2022). Longer-run economic consequences of pandemics. *Review of Economics and Statistics 104*(1), 166–175.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2019). Credit supply and the housing boom. *Journal of Political Economy 127*(3), 1317–1350.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika 41*(3-4), 403–404.

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

Laeven, L. and F. Valencia (2013). Systemic banking crises database. *IMF Economic Review 61*(2), 225–270.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2022). Local projections vs. VARs: lessons from thousands of DGPs. Working paper, National Bureau of Economic Research.

Mian, A., K. Rao, and A. Sufi (2013). Household balance sheets, consumption, and the economic slump. *Quarterly Journal of Economics 128*(4), 1687–1726.

Mian, A. and A. Sufi (2009). The consequences of mortgage credit expansion: evidence from the US mortgage default crisis. *Quarterly Jurnal of Economics 124*(4), 1449–1496.

Mian, A. and A. Sufi (2010). Household leverage and the recession of 2007–09. *IMF Economic Review 58*(1), 74–117.

Mian, A. and A. Sufi (2011). House prices, home equity-based borrowing, and the US household leverage crisis. *American Economic Review 101*(5), 2132–56.

Mian, A. and A. Sufi (2018). Finance and business cycles: the credit-driven household demand channel. *Journal of Economic Perspectives 32*(3), 31–58.

Mian, A. and A. Sufi (2022). Credit supply and housing speculation. *Review of Financial Studies 35*(2), 680–719.

Mian, A., A. Sufi, and E. Verner (2017). Household debt and business cycles worldwide. *Quarterly Journal of Economics 132*(4), 1755–1817.

Mian, A., A. Sufi, and E. Verner (2020). How does credit supply expansion affect the real economy? the productive capacity and household demand channels. *Journal of Finance 75*(2), 949–994.

Mikusheva, A. and M. Sølvsten (2023). Linear regression with weak exogeneity. *arXiv preprint arXiv:2308.08958*.

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*(4), 1789–1823.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–1426.

Okui, R. (2010). Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data. *Econometric Theory 26*(5), 1263–1304.

Okui, R. (2021). Linear dynamic panel data models. *Handbook of Research Methods and Applications in Empirical Microeconomics*, 2–22.

Ottonello, P. and T. Winberry (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica 88*(6), 2473–2502.

Phillips, P. C. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), 1057–1111.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and VARs estimate the same impulse responses. *Econometrica 89*(2), 955–980.

Ramey, V. (2016). Macroeconomic shocks and their propagation. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, pp. 71–162. Elsevier.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy 126*(2), 850–901.

Rampini, A. A. and S. Viswanathan (2019). Financial intermediary capital. *Review of Economic Studies 86*(1), 413–455.

Reinhart, C. M. and K. S. Rogoff (2009). The aftermath of financial crises. *American Economic Review 99*(2), 466–472.

Romer, C. D. and D. H. Romer (2017). New evidence on the aftermath of financial crises in advanced countries. *American Economic Review 107*(10), 3072–3118.

Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and statistics 71*(1), 135–158.

Schularick, M. and A. M. Taylor (2012). Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review 102*(2), 1029–1061.

Sufi, A. and A. M. Taylor (2021). Financial crises: a survey. Working paper, National Bureau of Economic Research.

Xu, K.-L. (2022). On local projection based inference. CAEPR working papers, Center for Applied Economics and Policy Research, Department of Economics, Indiana University Bloomington.

Zeev, N. B. (2017). Capital controls as shock absorbers. *Journal of International Economics 109*, 43–67.

Appendix for "Implicit Nickell Bias in Panel Local Projection: Financial Crises Are Worse Than You Think"

This appendix consists of three sections. Section A provides derivations and proofs of the theoretical statements in the main text. Section B enriches the estimation and inference procedures to allow alternative specifications of the regressions and dependence structures of the error terms. Section C gives additional simulation results of panel local projection (LP) with time fixed effects and lagged dependent variables.

A Technical Appendix

A.1 Derivation of Eq.(13)

Given the structural equation (12), elementary linear algebra leads to the predictive regression for $y_{i,t+h}$. Notice that the identification of the panel VAR system demands that $\mathbf{A}_{0,x}$, the $K \times K$ right-bottom block of $\mathbf{A}_0$, must be of full rank; otherwise the regressors $\mathbf{x}_{i,t}$ would be perfectly collinear and the coefficients in (12) fail be well-defined. The block upper triangular feature of $\mathbf{A}_0$ allows us to write down its inverse $\mathbf{A}_0^{-1} = \begin{pmatrix} 1 & \mathbf{a}_{0,yx}' \mathbf{A}_{0,x}^{-1} \\ \mathbf{0} & \mathbf{A}_{0,x}^{-1} \end{pmatrix}$ and therefore the reduced-form VAR is

$$\mathbf{w}_{i,t+1} = \mathbf{A}_0^{-1} \left(\boldsymbol{\mu}_i^{(0)} + \sum_{s=1}^{p} \mathbf{A}_s \mathbf{w}_{i,t+1-s} + \mathbf{u}_{i,t+1} \right) = \boldsymbol{\mu}_i^{(1)} + \sum_{s=1}^{p} \mathbf{B}_s^{(1)} \mathbf{w}_{i,t+1-s} + \mathbf{e}_{i,t+1}^{(1)},$$

where $\boldsymbol{\mu}_i^{(1)} = \mathbf{A}_0^{-1} \boldsymbol{\mu}_i^{(0)}$, $\mathbf{B}_s^{(1)} = \mathbf{A}_0^{-1} \mathbf{A}_s$, and $\mathbf{e}_{i,t+1}^{(1)} = \mathbf{A}_0^{-1} \mathbf{u}_{i,t+1}$. The above expression is the predictive regression for $h = 1$. Assumption 1 (b) implies that all the roots of the determinant equation

$$\det \left(\mathbf{I}_{K+1} - \sum_{s=1}^{p} \mathbf{B}_s^{(1)} z^s \right) = 0$$

stay outside of the unit circle on the complex plane to rule out unit roots and explosive roots. For $h \geq 2$, we have

$$\mathbf{w}_{i,t+h} = \boldsymbol{\mu}_i^{(h)} + \sum_{s=1}^{p} \mathbf{B}_s^{(h)} \mathbf{w}_{i,t+1-s} + \mathbf{e}_{i,t+h}^{(h)}$$

where the slope coefficients are defined recursively as

$$\mathbf{B}_s^{(h)} = \mathbf{B}_1^{(h-1)}\mathbf{B}_s^{(1)} + \mathbf{B}_{s+1}^{(h-1)} \cdot \mathbf{1}\left\{s \leq p-1\right\},$$

and the intercept and error term can be written in closed-forms as

$$\boldsymbol{\mu}_i^{(h)} = \sum_{s=0}^{h-1} \mathbf{B}_1^{(s)} \boldsymbol{\mu}_i^{(1)}, \quad \mathbf{e}_{i,t+h}^{(h)} = \sum_{s=0}^{h-1} \mathbf{B}_1^{(s)} \mathbf{e}_{i,t+h-s}^{(1)} \tag{A1}$$

with $\mathbf{B}_1^{(0)} = \mathbf{I}$. For panel LP, we are interested in the first equation of the h-period-ahead predictive regression

$$y_{i,t+h} = (1, \mathbf{0}')\mathbf{w}_{i,t+h} = \mu_i^{(h)y} + \sum_{s=1}^{p} \boldsymbol{\theta}_s^{(h)\prime}\mathbf{w}_{i,t+1-s} + e_{i,t+h}^{(h)}$$

where $\boldsymbol{\theta}_s^{(h)\prime} = (1, \mathbf{0}')\mathbf{B}_s^{(h)}$ is the first row of $\mathbf{B}_s^{(h)}$, and $e_{i,t+h}^{(h)} = (1, \mathbf{0}')\mathbf{e}_{i,t+h}^{(h)} = \sum_{s=0}^{h-1} \boldsymbol{\theta}_s^{(h)\prime}\mathbf{e}_{i,t+h-s}^{(1)}$ is the first element of the vector $\mathbf{e}_{i,t+h}^{(h)}$.

A.2 Violation of Strict Exogeneity in Panel VAR

The VAR(∞) considered by Plagborg-Møller and Wolf (2021) incorporates as a special case of the VAR(p) in the main text. VAR(∞) is convenient for the discussion of the population model. Let $\mathbf{A}(z) = \sum_{s=0}^{\infty} \mathbf{A}_s z^s$ be an infinite-order polynomial for $(1+K) \times (1+K)$ matrices $\mathbf{A}_s$, and the diagonal of $\mathbf{A}_0$ is standardized as 1. Following the notations in (10), we write the VAR(∞) for a representative individual i as

$$\mathbf{A}(\mathbb{L})\mathbf{w}_{i,t} = \boldsymbol{\mu}_i^{(0)} + \mathbf{u}_{i,t},$$

where $\mathbb{L}$ is the lag operator.

Assume all roots of the determinant equation $\det(\mathbf{A}(z)) = 0$ are outside of the unit cycle, and thus there exists $\mathbf{C}(\mathbb{L}) = \mathbf{A}(\mathbb{L})^{-1}$, where $\mathbf{C}(z) = \sum_{s=0}^{\infty} \mathbf{C}_s z^s$ is another infinite-order polynomial with $c_{y,0} = 1$. It transforms the VAR(∞) system into a corresponding vector moving average (VMA) system

$$\mathbf{w}_{i,t} = \mathbf{C}(\mathbb{L})\left(\boldsymbol{\mu}_i^{(0)} + \mathbf{u}_{i,t}\right) = \boldsymbol{\delta}_i + \sum_{s=0}^{\infty} \mathbf{C}_s \mathbf{u}_{i,t-s}$$

where $\boldsymbol{\delta}_i = \mathbf{C}(1)\boldsymbol{\mu}_i^{(0)}$ is the unconditional mean vector. Partition every matrix

$$\mathbf{A}_s = \begin{pmatrix} a_{y,s} & \mathbf{a}_{yx,s}^\top \\ \mathbf{a}_{xy,s} & \mathbf{A}_{x,s} \end{pmatrix}, \quad \mathbf{C}_s = \begin{pmatrix} c_{y,s} & \mathbf{c}_{yx,s}^\top \\ \mathbf{c}_{xy,s} & \mathbf{C}_{x,s} \end{pmatrix}$$

in a compatible manner with $(u_{i,t+1}^y, \mathbf{u}_{i,t+1}^{x\prime})'$. Parallel to (11), the Wold-causal order implies $\mathbf{a}_{xy,0} = \mathbf{0}$ and leads to $\mathbf{c}_{xy,0} = \mathbf{0}$ as well. The first equation of the VMA(∞) is

$$y_{i,t+h} = \delta_i^y + \sum_{s=0}^{\infty} c_{y,s} u_{i,t+h-s}^y + \sum_{s=0}^{\infty} \mathbf{c}_{yx,s}^\top \mathbf{u}_{i,t+h-s}^x$$

for $h \geq 2$.

Let $\mathcal{F}^{i,t} = \sigma\left((\mathbf{w}_{i,s})_{s=-\infty}^t\right) = \sigma((\mathbf{u}_{i,s})_{s=-\infty}^t)$ be the information set from the infinite past up to time t, where $\sigma(\cdot)$ is the sigma-field generated by the corresponding random variables. Since $y_{i,t+h}$ is a linear combination of the innovations, the conditional mean model is

$$\mathbb{E}\left[y_{i,t+h}|\mathcal{F}^{i,t}\right] = \delta_i^y + \sum_{s=h}^{\infty} c_{y,s} u_{i,t+h-s}^y + \sum_{s=h}^{\infty} \mathbf{c}_{yx,s}^\top \mathbf{u}_{i,t+h-s}^x. \tag{A2}$$

It follows that the error term against the information up to time t is

$$e_{i,t+h}^{(h)} = y_{i,t+h} - \mathbb{E}\left[y_{i,t+h}|\mathcal{F}^{i,t}\right] = \sum_{s=0}^{h-1} c_{y,s} u_{i,t+h-s}^y + \sum_{s=0}^{h-1} \mathbf{c}_{yx,s}^\top \mathbf{u}_{i,t+h-s}^x. \tag{A3}$$

Let $\mathcal{G}$ by the information set of all included regressors, to be discussed below. Notice that $c_{y,0} = 1$ by normalization. Whether the local projection regression violates strict exogeneity is equivalent to check if

$$\mathbb{E}[e_{i,t+h}^{(h)}|\mathcal{G}] = \mathbb{E}\left[u_{i,t+h}^y|\mathcal{G}\right] + \sum_{s=1}^{h-1} c_{y,s}\mathbb{E}\left[u_{i,t+h-s}^y|\mathcal{G}\right] + \sum_{s=0}^{h-1} \mathbf{c}_{yx,s}^\top \mathbb{E}\left[\mathbf{u}_{i,t+h-s}^x|\mathcal{G}\right] \tag{A4}$$

equals 0 or not.

1. If $c_{y,s} \neq 0$ for some $s \geq 1$ ($c_{y,0} = 1$ by standardization), the researcher includes lagged dependent variables in the panel LP regression we have $\sigma\left((y_{i,s})_{s=-\infty}^{T-h}\right) \subseteq \mathcal{G}$. As a result, the first term in (A4) is $\mathbb{E}\left[u_{i,t+h}^y|\mathcal{G}\right] = u_{i,t+h}^y \neq 0$ whenever $t \leq T - 2h$, and strict exogeneity must be violated. This is the explicit Nickell bias due to lagged dependent variables.

2. If the researcher includes $\mathbf{x}_{i,t}$ in the regression and thus $\sigma\left((\mathbf{u}_{i,s}^x)_{s=-\infty}^{T-h}\right) \subseteq \mathcal{G}$. There are

two cases:

(a) If in the true DGP $\mathbf{c}_{yx,s} \neq 0$ for some $s \geq 0$, then the third term in the right-most expression of (A4) whenever $t \leq T - 2h$ becomes

$$\sum_{s=0}^{h-1} \mathbf{c}_{yx,s}^{\top} \mathbb{E}\left[\mathbf{u}_{i,t+h-s}^{x} | \mathcal{G}\right] = \sum_{s=0}^{h-1} \mathbf{c}_{yx,s}^{\top} \mathbf{u}_{i,t+h-s}^{x} \neq 0.$$

(b) If $\mathbf{c}_{xy,s} \neq \mathbf{0}$ for some $s \geq 1$ ($\mathbf{c}_{xy,0} = \mathbf{0}$ is the implication of Wold-causality), then for some $\tau < t$ the lagged $y_{i,\tau}$ Granger-causes $x_{i,t}$. In this case, again in (A4) the first term $\mathbb{E}\left[u_{i,t+h}^{y} | \mathcal{G}\right] \neq 0$ whenever $t \leq T - 2h$ violates strict exogeneity.

In summary, Point 1 translates to the familiar explicit Nickell bias. Point 2(a) has $\mathbf{x}_{i,t}$ as regressors and it incurs Nickell bias as well; the prototype model in Section 2.1 is a special case with $c_{y,s} = 0$ but $\mathbf{c}_{yx,0} \neq \mathbf{0}$, leading to implicit Nickell bias despite the absence of the lagged dependent variables. The information of $y_{i,\tau}$ is leaked via $\mathbf{c}_{xy,s} \neq 0$ into the future $\mathbf{x}_{i,t}$ in Point 2(b); the "static model" in Example 2 of Chudik et al. (2018)'s Online Appendix is a special case of it.

When $\{c_{y,s} = 0\}_{s=1}^{H-1}$, $\{\mathbf{c}_{xy,s} = \mathbf{0}\}_{s=1}^{H-1}$, and $\{\mathbf{c}_{yx,s} = \mathbf{0}\}_{s=0}^{H-1}$, the true DGP becomes a void model $y_{i,t+h} = \delta_i^y + u_{i,t+h}^y$ and it is completely unpredictable by the past information. But if this happens, a researcher shall exclude all regressors from the linear regression in the first place. Given that she has no prior knowledge about the true regression coefficients, a researcher should worry that each potentially relevant regressor based on her economic reasoning is subject to Nickell bias in panel LP. As a consequence, it is imperative for her to conduct bias correction for asymptotically valid inference.

A.3 Proof of Proposition 1

We begin with the following decomposition

$$\begin{aligned}\sqrt{NT_h}\left(\widetilde{\beta}^{(h)\mathrm{fe}} - \widetilde{\beta}^{(h)}\right) &= \frac{1}{\sqrt{NT_h} s_x^2} \sum_{i \in [N]} \sum_{t \in \mathcal{T}^h} \widetilde{x}_{i,t} e_{i,t+h}^{(h)} \\ &= \frac{1}{\sqrt{N} s_x^2} \sum_{i \in [N]} \left(\frac{1}{\sqrt{T_h}} \sum_{t \in \mathcal{T}^h} \widetilde{x}_{i,t} u_{i,t+h}^{y} + \frac{1}{\sqrt{T_h}} \beta^{(0)} \sum_{t \in \mathcal{T}^h} \sum_{s=0}^{h-1} \rho^s \widetilde{x}_{i,t} u_{i,t+h-s}^{x} \right).\end{aligned}$$

The independence between $u_{i,t}^y$ and $u_{i,t}^x$ immediately implies the first term $\mathbb{E}\left[T_h^{-1/2} \sum_{t \in \mathcal{T}^h} \widetilde{x}_{i,t} u_{i,t+h}^{y}\right] = 0$, and next we focus on the second term $T_h^{-1/2} \beta^{(0)} \sum_{t \in \mathcal{T}^h} \sum_{s=0}^{h-1} \rho^s \widetilde{x}_{i,t} u_{i,t+h-s}^{x}$. Notice that

$\mathbb{E}[x_{i,t}u^x_{i,t+h-s}] = 0$ for any $0 \le s \le h-1$ and hence

$$\sum_{t\in\mathcal{T}^h} \mathbb{E}\left[x_{i,t}\sum_{s=0}^{h-1}\rho^s u^x_{i,t+h-s}\right] = 0. \tag{A5}$$

By the AR(1) model (3), we have $x_{i,t} - \delta_{0i}\sum_{\ell=0}^{t}\rho^\ell = \sum_{\ell=0}^{t}\rho^\ell u^x_{i,t-\ell}$, which yields

$$\bar{x}_i = \frac{1}{T_h}\sum_{t\in\mathcal{T}^h} x_{i,t} = \frac{1}{T_h}\sum_{r=1}^{T_h}\sum_{\ell=0}^{r}\rho^\ell u^x_{i,r-\ell} + \delta_{0i}\frac{1}{T_h}\sum_{r=1}^{T_h}\sum_{\ell=0}^{r}\rho^\ell.$$

For any $s \in [h]$, the independence of $u^x_{i,t}$ across i and t gives

$$\begin{aligned}
&\sum_{t=0}^{T_h}\mathbb{E}\left[\bar{x}_i\rho^s u^x_{i,t+h-s}\right] \\
=\ & \frac{\rho^s}{T_h}\sum_{t=0}^{T_h}\sum_{r=1}^{T_h}\sum_{\ell=0}^{r}\mathbb{E}\left[u^x_{i,t+h-s}\cdot\rho^\ell u^x_{i,r-\ell}\right] = \frac{\rho^s}{T_h}\sum_{t=0}^{T_h}\sum_{r=1}^{T_h}\sum_{\ell=0}^{r}\rho^{r-\ell}\mathbb{E}\left[u^x_{i,t+h-s}u^x_{i,\ell}\right] \\
=\ & \frac{\rho^s}{T_h}\sum_{t=h-s}^{T_h}\sum_{r=1}^{T_h}\sum_{\ell=0}^{r}\rho^{r-\ell}\mathbb{E}\left[u^x_{i,t}u^x_{i,\ell}\right] = \sigma^2_{u_x}\frac{\rho^s}{T_h}\sum_{t=h-s}^{T_h}\left(T_h - t\right)\rho^{t-(h-s)}
\end{aligned}$$

where the last equality follows as $E\left[u^x_{i,t}u^x_{i,\ell}\right] = \sigma^2_{u_x}\mathbf{1}\{t=\ell\}$. Define

$$\begin{aligned}
f_{T,h}(\rho) &:= \sum_{s=0}^{h-1}\sum_{t\in\mathcal{T}^h}\mathbb{E}\left[\bar{x}_i\rho^s u^x_{i,t+h-s}\right] = \sum_{s=0}^{h-1}\sum_{t=h-s}^{T_h}\left(1-\frac{t}{T_h}\right)\rho^{t-h+2s} \\
&= \frac{(T-2h)\left(1-\rho^2\right) - (T-h)\rho^h\left(1-\rho^2\right) + \rho^{T-2h+1}\left(1-\rho^{2h}\right)}{(T-h)\left(1-\rho\right)^2\left(1-\rho^2\right)} \\
&= \frac{\left(1-\rho^h\right)}{\left(1-\rho\right)^2} - \frac{h}{(T-h)\left(1-\rho\right)^2} + \frac{\rho^{T-2h+1}\left(1-\rho^{2h}\right)}{(T-h)\left(1-\rho\right)^2}
\end{aligned}$$

after long but elementary calculations for the third equality. We deduce

$$\frac{1}{\sqrt{T_h}}\sum_{s=0}^{h-1}\sum_{t\in\mathcal{T}^h}\rho^s\mathbb{E}\left[\widetilde{x}_{i,t}u^x_{i,t+h-s}\right] = -\frac{1}{\sqrt{T_h}}f_{T,h}(\rho).$$

As a result,

$$\mathbb{E}\left[\frac{1}{\sqrt{T_h}}\sum_{t\in\mathcal{T}^h}\widetilde{x}_{i,t}e^{(h)}_{i,t+h} + \frac{\beta^{(0)}}{\sqrt{T_h}}f_{T,h}(\rho)\right] = 0 \tag{A6}$$

for all $i \in [N]$. The asymptotic normality assume in (6) gives

$$\begin{aligned} & \frac{1}{\sigma_{xe,h}\sqrt{NT_h}} \sum_{i\in[N]} \sum_{t\in\mathcal{T}^h} \left(\widetilde{x}_{i,t} e^{(h)}_{i,t+h} + \frac{\beta^{(0)}}{\sqrt{T_h}} f_{T,h}(\rho) \right) \\ = \ & \frac{1}{\sigma_{xe,h}\sqrt{NT_h}} \sum_{i\in[N]} \sum_{t\in\mathcal{T}^h} \left(\widetilde{x}_{i,t} e^{(h)}_{i,t+h} - \mathbb{E}\left(\widetilde{x}_{i,t} e^{(h)}_{i,t+h} \right) \right) \xrightarrow{d} \mathcal{N}(0,1). \end{aligned}$$

Furthermore, the law of large numbers assumed in (5) yields

$$\sqrt{NT_h}\left(\widetilde{\beta}^{(h)\text{fe}} - \beta^{(h)}\right) + \frac{\beta^{(0)}\sigma^2_{u_x}}{s^2_x}\sqrt{\frac{N}{T_h}} f_{T,h}(\rho) \xrightarrow{d} \mathcal{N}\left(0,\ \frac{\sigma^2_{xe,h}}{\sigma^4_x}\right)$$

as stated in the theorem.

A.4 Oracle Debiased Estimator

Theorem 1 shows that under (2) and (3), the Nickel bias of $\widetilde{\beta}^{(h)\text{fe}}$ is $-\frac{\beta^{(0)}\sigma^2_{u_x}}{s^2_x}\sqrt{\frac{N}{T_h}} f_{T,h}(\rho)$. We thus define the following debiased FE estimator

$$\widetilde{\beta}^{(h)\text{db}} = \widetilde{\beta}^{(h)\text{fe}} + \frac{\widehat{\beta}^{(0)}}{T_h s^2_x} \cdot \widehat{\sigma}^2_{u_x} \cdot f_{T,h}(\widehat{\rho})$$

where $\widehat{\beta}^{(0)}$ is the FE estimator for (2), $\widehat{\rho}$ is the FE estimator of the AR(1) model (3). For the variances, $\widehat{\sigma}^2_{u_x}$ is the sum of squared residuals of the AR(1) regression, and s^2_x is defined as in (5).

We need to further estimate $\sigma^2_{xe,h}$ for inference. A consistent estimator is given by

$$\widehat{\sigma}^2_{xe,h} := \frac{1}{NT_h} \sum_{i\in[N]} \left[\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} \widehat{e}^{(h)\text{db}}_{i,t+h} \right]^2 \tag{A7}$$

where $\widehat{e}^{(h)\text{db}}_{i,t+h} = y_{i,t+h} - \widehat{\beta}^{(h)\text{db}} x_{i,t}$. Under standard regularity conditions we can show

$$\widehat{\sigma}^2_{xe,h} - \frac{1}{NT_h} \sum_{i\in[N]} \mathbb{E}\left[\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} e^{(h)}_{i,t+h} \right]^2 = o_p(1).$$

Besides, we have deduced by (A6) that

$$\frac{1}{NT_h} \sum_{i\in[N]} \left[\mathbb{E} \sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} e^{(h)}_{i,t+h} \right]^2 = \frac{(\beta^{(0)})^2 \sigma^4_{u_x} f^2_{T,h}(\rho)}{T_h} = o_p(1)$$

as $T_h \to \infty$. As a result $\widehat{\sigma}_{xe,h}^2 - \sigma_{xe,h}^2 = o_p(1)$ by the fact that

$$\begin{aligned}\sigma_{xe,h}^2 &= \lim_{(N,T)\to\infty} \frac{1}{NT_h} \sum_{i\in[N]} \operatorname{var}\left[\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} e_{i,t+h}^{(h)}\right] \\ &= \lim_{(N,T)\to\infty} \frac{1}{NT_h} \sum_{i\in[N]} \left\{ \mathbb{E}\left[\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} e_{i,t+h}^{(h)}\right]^2 - \left[\mathbb{E}\sum_{t\in\mathcal{T}^h} \widetilde{x}_{i,t} e_{i,t+h}^{(h)}\right]^2 \right\}.\end{aligned}$$

Familiar two-sided, symmetric (around the point estimate) confidence intervals can be constructed given $\widehat{s}^{(h)\text{db}} = \dfrac{\widehat{\sigma}_{xe,h}}{s_x\sqrt{NT_h}}$.

A.5 Proof of Theorem 1

We first present a useful lemma.

Lemma 1. *If Assumption 1 holds, then*

$$\sup_{i\in[N]} \left\| \sum_{t\in\mathcal{T}_a^h} \mathbb{E}\left(\bar{\mathbf{W}}_{i,b} e_{i,t+h}^{(h)}\right) \right\| = O\left(\frac{h}{T_h}\right)$$

where $e_{i,t+h}^{(h)} = \sum_{s=0}^{h-1} \boldsymbol{\theta}_s^{(h)\prime} \mathbf{e}_{i,t+h-s}^{(1)}$ *is the first entry of* $\mathbf{e}_{i,t+h}^{(h)}$ *defined in (A1).*

Proof of Lemma 1. Recall that $\mathbf{W}_{i,t} = (\mathbf{w}_{i,t}', \mathbf{w}_{i,t-1}', \ldots, \mathbf{w}_{i,t-p+1}')'$. It suffices to show that $\sup_{i\in[N]} \left\|\mathbb{E}\left[\sum_{t\in\mathcal{T}_a^h} \bar{\mathbf{w}}_{i,b}^j e_{i,t+h}^{(h)}\right]\right\| = O(h/T_h)$ where $\bar{\mathbf{w}}_{i,b}^j = (T_h/2)^{-1}\sum_{t\in\mathcal{T}_b^h} \mathbf{w}_{i,t-j}$ for $j = 0, 1, \cdots, p-1$. The VAR(p) process $\mathbf{w}_{i,t}$ has the representation $\mathbf{w}_{i,t} = \boldsymbol{\delta}_i + \sum_{d=0}^{\infty} \boldsymbol{\Psi}_d \mathbf{e}_{i,t-d}^{(1)}$ with the coefficient matrices $\|\boldsymbol{\Psi}_d\| \leq C\mathrm{e}^{-cd}$ for some absolute constants $C, c > 0$ and the unconditional mean $\boldsymbol{\delta}_i$. Besides, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t\in\mathcal{T}_a^h} \bar{\mathbf{w}}_{i,b}^j e_{i,t+h}^{(h)}\right] &= \frac{1}{T_h/2} \sum_{t\in\mathcal{T}_a^h} \sum_{r\in\mathcal{T}_b^h} \mathbb{E}\left[\mathbf{w}_{i,r-j}^j e_{i,t+h}^{(h)}\right] \\ &= \frac{1}{T_h/2} \sum_{r\in\mathcal{T}_b^h} \sum_{t\in\mathcal{T}_a^h} \sum_{s=0}^{h-1} \sum_{d=0}^{\infty} \mathbb{E}\left[\boldsymbol{\Psi}_d \mathbf{e}_{i,r-j-d}^{(1)} \cdot \boldsymbol{\theta}_s^{(h)\prime} \mathbf{e}_{i,t+h-s}^{(1)}\right]\end{aligned}$$

where

$$\mathbb{E}\left[\boldsymbol{\Psi}_d \mathbf{e}_{i,r-j-d}^{(1)} \cdot \boldsymbol{\theta}_s^{(h)\prime} \mathbf{e}_{i,t+h-s}^{(1)}\right] = \begin{cases} \boldsymbol{\Psi}_d \mathbf{A}_0^{-1} \boldsymbol{\Sigma}_u (\mathbf{A}_0^{-1})' \boldsymbol{\theta}_s^{(h)}, & d = r - t - j - h + s \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{A8}$$

where $\boldsymbol{\Sigma}_u = \mathbb{E}\left(\mathbf{u}_{i,t}\mathbf{u}_{i,t}'\right)$. By (A8) and the fact that $j \in [p], 0 \leq s \leq h-1$, we have

$$\left\| \sum_{d=0}^{\infty} \mathbb{E}\left[\boldsymbol{\Psi}_d \mathbf{e}_{i,r-j-d}^{(1)} \cdot \boldsymbol{\theta}_s^{(h)\prime} \mathbf{e}_{i,t+h-s}^{(1)} \right] \right\| \leq C_1 \cdot C\mathrm{e}^{-c(r-t)} \cdot \mathrm{e}^{c(p+h)}$$

where $C_1 = \|\mathbf{A}_0^{-1}\|^2 \cdot \|\boldsymbol{\Sigma}_u\| \cdot \sup_{0 \leq s \leq h-1} \|\boldsymbol{\theta}_s^{(h)}\|$. Hence,

$$\begin{aligned}
\left\| \mathbb{E}\left[\sum_{t \in \mathcal{T}_a^h} \bar{\mathbf{w}}_{i,b}^{j} e_{i,t+h}^{(h)} \right] \right\| &\leq \frac{C_1 \cdot C\mathrm{e}^{c(p+h)}}{T_h/2} \sum_{r \in \mathcal{T}_b^h} \sum_{t \in \mathcal{T}_a^h} \sum_{s=0}^{h-1} \mathrm{e}^{-c(r-t)} \\
&= \frac{C_2 \cdot h}{T_h} \sum_{r=T_h/2+1}^{T_h} \mathrm{e}^{-cr} \sum_{t=1}^{T_h/2} \mathrm{e}^{ct} \\
&= \frac{C_2 \cdot h}{T_h} \cdot \frac{\mathrm{e}^{-c(T_h/2+1)}(1-\mathrm{e}^{-cT_h/2})}{1-\mathrm{e}^{-c}} \cdot \frac{\mathrm{e}^{cT_h/2}-1}{1-\mathrm{e}^{-c}} \\
&\leq \frac{C_3 \cdot h}{T_h} \mathrm{e}^{-c(T_h/2+1)} \cdot \mathrm{e}^{cT_h/2} \leq \frac{C_3 \cdot h}{T_h}
\end{aligned}$$

where $C_2 = 2C_1 \cdot C\mathrm{e}^{c(p+h)}$ and $C_3 = \frac{C_2}{(1-\mathrm{e}^{-c})^2}$. This non-asymptotic bound holds uniformly for all $i \in [N]$. □

We proceed with the main proof. The FE estimator is

$$\widetilde{\boldsymbol{\theta}}^{(h)\mathrm{fe}} = \widehat{\mathbf{Q}}^{-1} \frac{1}{NT_h} \sum_{i \in [N]} \sum_{t \in \mathcal{T}^h} \widetilde{\mathbf{W}}_{i,t} y_{i,t+h}$$

and thus

$$\widetilde{\boldsymbol{\theta}}^{(h)\mathrm{fe}} - \boldsymbol{\theta}^{(h)} = \widehat{\mathbf{Q}}^{-1} \boldsymbol{\zeta}$$

where $\boldsymbol{\zeta} = \dfrac{1}{NT_h} \sum_{i \in [N]} \sum_{t \in \mathcal{T}^h} \widetilde{\mathbf{W}}_{i,t} e_{i,t+h}^{(h)}$. The estimation error can be decomposed as

$$\widetilde{\boldsymbol{\theta}}^{(h)\mathrm{spj}} - \boldsymbol{\theta}^{(h)} = \widehat{\mathbf{Q}}^{-1}\left(2\boldsymbol{\zeta} - \frac{1}{2}\left(\boldsymbol{\zeta}_a + \boldsymbol{\zeta}_b\right)\right) + \left(\widehat{\mathbf{Q}}^{-1} - \widehat{\mathbf{Q}}_a^{-1}\right)\frac{\boldsymbol{\zeta}_a}{2} + \left(\widehat{\mathbf{Q}}^{-1} - \widehat{\mathbf{Q}}_b^{-1}\right)\frac{\boldsymbol{\zeta}_b}{2} \tag{A9}$$

where $\boldsymbol{\zeta}_k = \dfrac{2}{NT_h} \sum_{i \in [N]} \sum_{t \in \mathcal{T}_k^h} \left(\mathbf{w}_{i,t} - \bar{\mathbf{w}}_{i,k}\right) e_{i,t+h}^{(h)}$ for $k \in \{a, b\}$.

Multiply $\sqrt{NT_h}$ on both side of (A9). We focus on

$$\sqrt{NT_h}\left(2\boldsymbol{\zeta} - \frac{1}{2}\left(\boldsymbol{\zeta}_a + \boldsymbol{\zeta}_b\right)\right) = \frac{1}{\sqrt{NT_h}} \sum_{i \in [N]} \sum_{t \in \mathcal{T}^h} \mathbf{d}_{i,t}^{*} e_{i,t+h}^{(h)}.$$

The summand $\left\{T_h^{-1/2}\sum_{t\in\mathcal{T}^h}\left[\mathbf{d}^*_{i,t}e_{i,t+h}-\mathbb{E}\left(\mathbf{d}^*_{i,t}e_{i,t+h}\right)\right]\right\}_{i=1}^N$ is independent across $i\in[N]$ under Assumption 1. The definition of $\mathbf{d}^*_{i,t}$ yields the following decomposition

$$\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}=\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h_a}(\mathbf{W}_{i,t}-\bar{\mathbf{W}}_{i,b})e^{(h)}_{i,t+h}+\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h_b}(\mathbf{W}_{i,t}-\bar{\mathbf{W}}_{i,a})e^{(h)}_{i,t+h}.$$

Recall that $e^{(h)}_{i,t+h}$ is the first entry of $\mathbf{e}^{(h)}_{i,t+h}=\sum_{s=0}^{h-1}\mathbf{B}_1^{(s)}\mathbf{e}^{(1)}_{i,t+h-s}$ defined in (A1), a linear combination of $\left(\mathbf{e}^{(1)}_{i,t+\ell}\right)_{\ell=1}^h$. By construction, $\mathbb{E}\left[\mathbf{W}_{i,t}e^{(h)}_{i,t+h}\right]=0$ for all $t\geq 1$, and $\mathbb{E}\left[\bar{\mathbf{W}}_{i,a}e^{(h)}_{i,t+h}\right]=0$ for all $t\in\mathcal{T}^h_b$, since $\mathbb{E}\left[\mathbf{W}_{i,s}e^{(h)}_{i,t+h}\right]=0$ whenever $s<t$. Lemma 1 then gives

$$\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\mathbb{E}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}\right]=O\left(\frac{1}{\sqrt{NT_h}}\cdot N\frac{h}{T_h}\right)=O\left(h\frac{N^{1/2}}{T_h^{3/2}}\right)\to 0 \tag{A10}$$

under the condition $N/T^3\to 0$ with a fixed h.

Parallel calculation delivers $\sqrt{NT_h}\boldsymbol{\zeta}_a=O_p(1)$ and $\sqrt{NT_h}\boldsymbol{\zeta}_b=O_p(1)$. We therefore conclude

$$\begin{aligned}
&\sqrt{NT_h}\left(\widetilde{\boldsymbol{\theta}}^{(h)\mathrm{spj}}-\boldsymbol{\theta}^{(h)}\right)\\
=\;&\widehat{\mathbf{Q}}^{-1}\left(2\boldsymbol{\zeta}-\frac{\boldsymbol{\zeta}_a}{2}-\frac{\boldsymbol{\zeta}_b}{2}\right)+\left(\widehat{\mathbf{Q}}^{-1}-\widehat{\mathbf{Q}}_a^{-1}\right)\cdot O_p(1)+\left(\widehat{\mathbf{Q}}^{-1}-\widehat{\mathbf{Q}}_b^{-1}\right)\cdot O_p(1)\\
=\;&\widehat{\mathbf{Q}}^{-1}\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}-\mathbb{E}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}\right]\right]+\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\mathbb{E}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}\right]+o_p(1)\\
=\;&\widehat{\mathbf{Q}}^{-1}\frac{1}{\sqrt{NT_h}}\sum_{i\in[N]}\sum_{t\in\mathcal{T}^h}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}-\mathbb{E}\left[\mathbf{d}^*_{i,t}e^{(h)}_{i,t+h}\right]\right]+o(1)+o_p(1)\\
\stackrel{d}{\to}\;&\mathcal{N}\left(\mathbf{0},\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}^{-1}\right)
\end{aligned}$$

where the second equality follows by Assumption 2 as $\mathbf{Q}=\mathrm{plim}\widehat{\mathbf{Q}}_a=\mathrm{plim}\widehat{\mathbf{Q}}_b$ and so does $\mathbf{Q}=\mathrm{plim}\widehat{\mathbf{Q}}$, the third equality from (A10), and the limiting distribution follows by (14).

B Extensions for Alternative Specifications

B.1 Differenced Dependent Variable

Some empirical works, for example Baron et al. (2021), specifies the IRF in the following regression form

$$\Delta_h y_{i,t+h}=\mu_i^{(h)}+b^{(h)}x_{i,t}+e^{(h)}_{i,t+h} \tag{A11}$$

where $\Delta_h y_{i,t+h} := y_{i,t+h} - y_{i,t}$. This specification does not circumvent the implicit Nickell bias. Suppose that the true DGP now becomes

$$\Delta y_{i,t} = \mu_i^{(0)y} + \beta^{(0)} x_{i,t} + u_{i,t}^y, \tag{A12}$$

and the dynamic of $x_{i,t}$ still follows (3). Similar derivations as in Section 2 yield

$$\Delta y_{i,t+r} = \mu_i^{(r)} + \beta^{(r)} x_{i,t} + e_{i,t+r}^{(r)} \tag{A13}$$

for all $r \geq 1$, where $\beta^{(r)} = \rho^r \beta^{(0)}$, $\mu_i^{(r)} = \mu_i^{(0)} + \beta^{(0)} \mu_i^x \sum_{s=0}^{r-1} \rho^s$, and $e_{i,t+r}^{(r)} = u_{i,t+r}^y + \beta^{(0)} \sum_{s=0}^{r-1} \rho^s u_{i,t+r-s}^x$. The components in (A12) can be found by taking r from 1 to h and summing up (A13):

$$\begin{aligned}
b^{(h)} &= \sum_{r=1}^{h} \rho^r \beta^{(0)} = \frac{\beta^{(0)} \rho (1-\rho^h)}{1-\rho}; \\
\mu_i^{(h)} &= h\mu_i^{(0)y} + \beta^{(0)} \mu_i^x \sum_{r=1}^{h} \sum_{s=0}^{r-1} \rho^s = h\mu_i^{(0)y} + \frac{\beta^{(0)} \mu_i^x [h(1-\rho) - \rho(1-\rho^h)]}{(1-\rho)^2}; \\
e_{i,t+h}^{(h)} &= \sum_{r=1}^{h} u_{i,t+r}^y + \beta^{(0)} \sum_{r=1}^{h} \sum_{s=0}^{r-1} \rho^s u_{i,t+r-s}^x.
\end{aligned}$$

The Nickell bias of (A11) stems from the second term $\beta^{(0)} \sum_{r=1}^{h} \sum_{s=0}^{r-1} \rho^s u_{i,t+r-s}^x$ of $e_{i,t+h}^{(h)}$ by parallel arguments as in Section 2.

B.2 Estimation and Inference with Split-Panel Jackknife

In the main text we present the theory in simple forms to highlight that the Nickell bias in LP can be effectively resolved by the split-panel jackknife (SPJ) estimator. In practice, applied researchers may attempt alternative specifications in both the predictors and the correlation structure in the error terms. These extensions can be accommodated by adapting existing econometric methods.

B.2.1 Two-way Clustered Standard Error

We first provide the formula for two-way clustered standard error of the SPJ estimator. Recall that the asymptotic variance estimator $\widehat{\mathbf{V}}_N := \widehat{\mathbf{Q}}^{-1} \widehat{\mathbf{R}}_N \widehat{\mathbf{Q}}^{-1}$ below Theorem 1 is clustered over time, where

$$\widehat{\mathbf{R}}_N := \widehat{\mathbf{R}} = (NT_h)^{-1} \sum_{i\in[N]} \sum_{t,s\in\mathcal{T}^h} \mathbf{d}^*_{i,t}\mathbf{d}^{*\top}_{i,s} \widetilde{e}^{(h)\mathrm{spj}}_{i,t+h} \widetilde{e}^{(h)\mathrm{spj}}_{i,s+h}.$$

Similarly, we can construct an asymptotic variance estimator $\widehat{\mathbf{V}}_T := \widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{R}}_T\widehat{\mathbf{Q}}^{-1}$ with

$$\widehat{\mathbf{R}}_T := \widehat{\mathbf{R}} = (NT_h)^{-1} \sum_{t\in\mathcal{T}^h} \sum_{i,j\in[N]} \mathbf{d}^*_{i,t}\mathbf{d}^{*\top}_{j,t} \widetilde{e}^{(h)\mathrm{spj}}_{i,t+h} \widetilde{e}^{(h)\mathrm{spj}}_{j,t+h}$$

to cluster over individuals. Finally, the robust variance estimator (or White estimator) is $\widehat{\mathbf{V}}_{NT} := \widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{R}}_{NT}\widehat{\mathbf{Q}}^{-1}$ with

$$\widehat{\mathbf{R}}_{NT} := \widehat{\mathbf{R}} = (NT_h)^{-1} \sum_{t\in\mathcal{T}^h} \sum_{i\in[N]} \mathbf{d}^*_{i,t}\mathbf{d}^{*\top}_{i,t} (\widetilde{e}^{(h)\mathrm{spj}}_{i,t+h})^2.$$

Following Cameron et al. (2011), we can construct the two-way clustered variance estimator as $\widehat{\mathbf{V}}_{\mathrm{TW}} = \widehat{\mathbf{V}}_N + \widehat{\mathbf{V}}_T - \widehat{\mathbf{V}}_{NT}$.

B.2.2 Time Fixed Effect

In the main text we only discuss the cross-sectional fixed effects. Time fixed effects are ubiquitous in panel data empirical applications. The two-way FE estimator captures potential time FE along with the individual FE. Chudik et al. (2018) study SPJ for a generic two-way FE regression. We borrow their approach in panel LP. Specifically, we can redefine the demeaned variable

$$\widetilde{\mathbf{W}}_{i,t} = \mathbf{W}_{i,t} - T_h^{-1} \sum_{t\in\mathcal{T}^h} \mathbf{W}_{i,t} - N^{-1} \sum_{i\in[N]} \mathbf{W}_{i,t} + (NT_h)^{-1} \sum_{t\in\mathcal{T}^h} \sum_{i\in[N]} \mathbf{W}_{i,t}$$

and similar two-way demeaning also applies to other observable variables. All other steps will follow the formulae for the one-way fixed effect. As for the standard error, we update the expression for $\mathbf{d}^*_{i,t}$ in Assumption 2 given as

$$\begin{aligned}\mathbf{d}^*_{i,t} = &\left(\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,b} - \frac{1}{T_h} \sum_{t\in\mathcal{T}^h} (\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,b})\right) \cdot \mathbf{1}\left\{t \in \mathcal{T}^h_a\right\} \\ &+ \left(\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,a} - \frac{1}{T_h} \sum_{t\in\mathcal{T}^h} (\mathbf{W}_{i,t} - \bar{\mathbf{W}}_{i,a})\right) \cdot \mathbf{1}\left\{t \in \mathcal{T}^h_b\right\}\end{aligned}$$

where $\bar{\mathbf{W}}_{i,k} = (T_h/2)^{-1} \sum_{t\in\mathcal{T}^h_k} \mathbf{W}_{i,t}$ for $k \in \{a, b\}$.

B.2.3 Unbalanced Panel

While the balanced panel facilitates theoretical analysis, in practice unbalanced panels are the norm rather than the exception. We follow Section A7 of Chudik et al. (2018) to handle unbalanced panels for SPJ with the observation-splitting procedure. Suppose that for the cross-sectional unit i we observe all variables over $t = T_{f_i}, T_{f_i+1}, \cdots, T_{l_i}$, where T_{f_i} and T_{l_i} are the first and last time periods, respectively. Without loss of generality we assume $T_i = T_{l_i} - T_{f_i} + 1$ is an even number for every i. We implement SPJ by dividing the T_i observations into two sub-samples. The first sub-sample, denoted by subscript a, contains the first $T_i/2$ observations; the second half, denoted by subscript b, consists of the remaining $T_i/2$ observations.

C Additional Simulation Results

We consider a more general DGP with a lagged dependent variable given as

$$\begin{aligned} y_{i,t} &= \mu_i^{(0)y} + g_t^{(0)y} + \tau y_{i,t-1} + \beta^{(0)} x_{i,t} + u_{i,t}^y, \\ x_{i,t} &= \mu_i^x + \kappa y_{i,t-1} + \rho x_{i,t-1} + u_{i,t}^x. \end{aligned} \tag{A14}$$

The time fixed effect $g_t^{(0)y}$ as well as the lagged dependent variable $y_{i,t-1}$ are added to mimic common practice in applied works. In model (A14), the true impulse response function for Period h becomes the $(2,2)$th element of the matrix $\mathbf{P}^h$ where $\mathbf{P} = \big((\tau,\kappa)', (\beta^{(0)},\rho)'\big)$. Based on the insights from Section 2.1, this panel VAR(1) model also incurs the Nickell bias issue, though the closed-form bias formula for this setting is too complicated to be deduced and coded. Therefore in our simulation we consider the FE estimator and the SPJ estimator, but omit the DB estimator.

In our setting, $\mu_i^{(0)y}$, μ_i^x, $u_{i,t}^y$, $u_{i,t}^x$ and the sample sizes follow the simulation setting of simple models (2) and (3). The time effect is set as $g_t^{(0)} := 0.025t + 0.001t^2$. We fix $\beta^{(0)} = -0.25$, $\kappa = -0.5$, $\tau = 0.5$ and vary $\rho \in \{0, 0.2, 0.4, 0.5\}$; to ensure that the VAR(1) system is stationary (net of the deterministic trend) we cannot choose $\rho = 0.8$ as in Section 2.4. We slightly abuse the abbreviation "FE" to denote the (two-way) individual-time fixed effect estimator. The SPJ estimator that allows time effects is introduced in Appendix B.2.2. Following the spirit of LP we estimate the model

$$y_{i,t+h} = \mu_i^{(h)y} + g_t^{(h)y} + \tau^{(h)} y_{i,t} + \beta^{(h)} x_{i,t} + e_{i,t+h}^{(h)}$$

by FE and SPJ estimators, and the IRFs are estimated by the estimators of $\beta^{(h)}$. Simulation results are shown in Figures A1-A3 below and the key findings are similar to those in Section 2.4, including the attenuation bias of FE and its substantial under-coverage of the confidence interval.

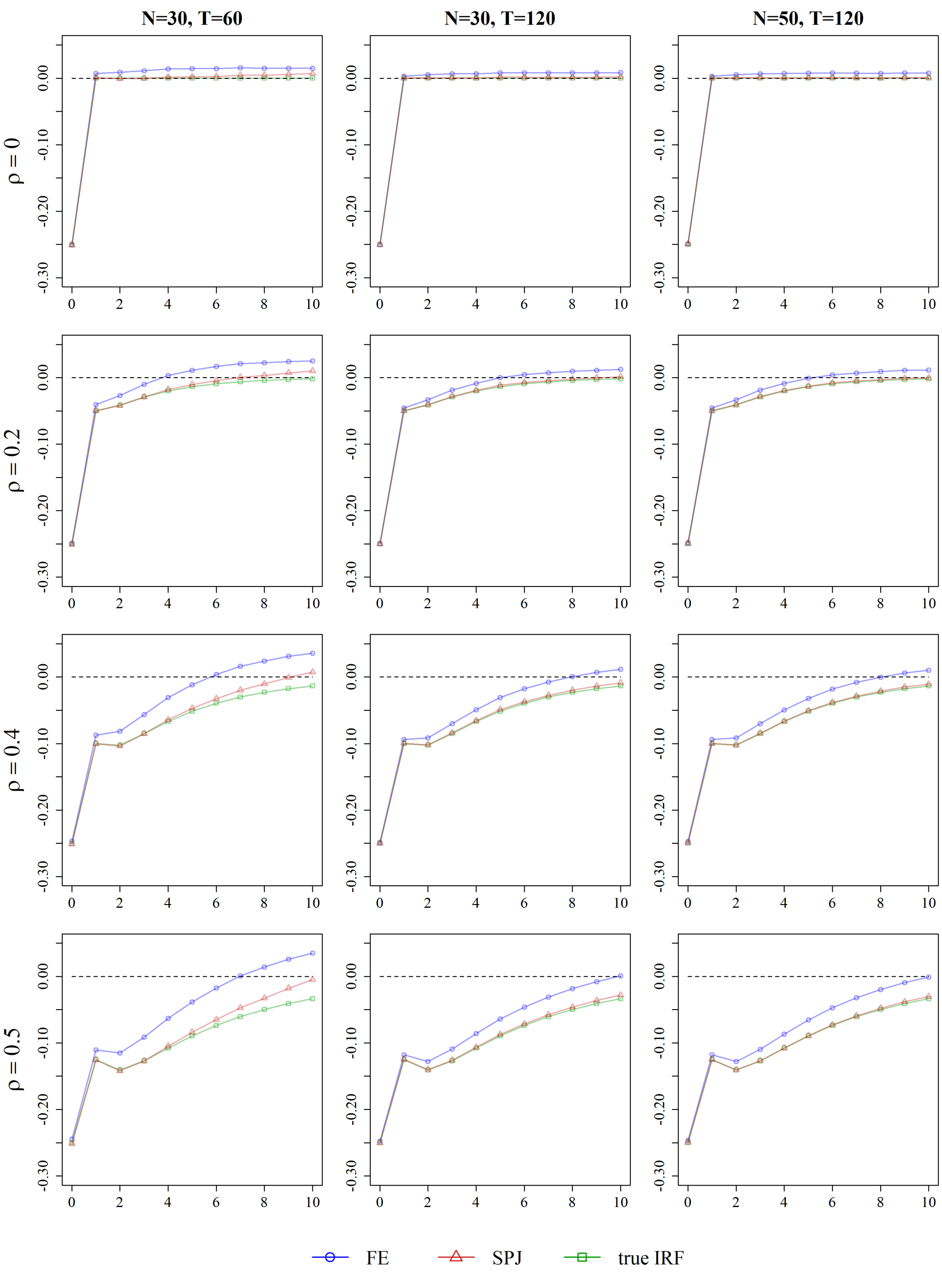


Figure A1: Estimated IRFs Averaged Over Replications for Model (A14)

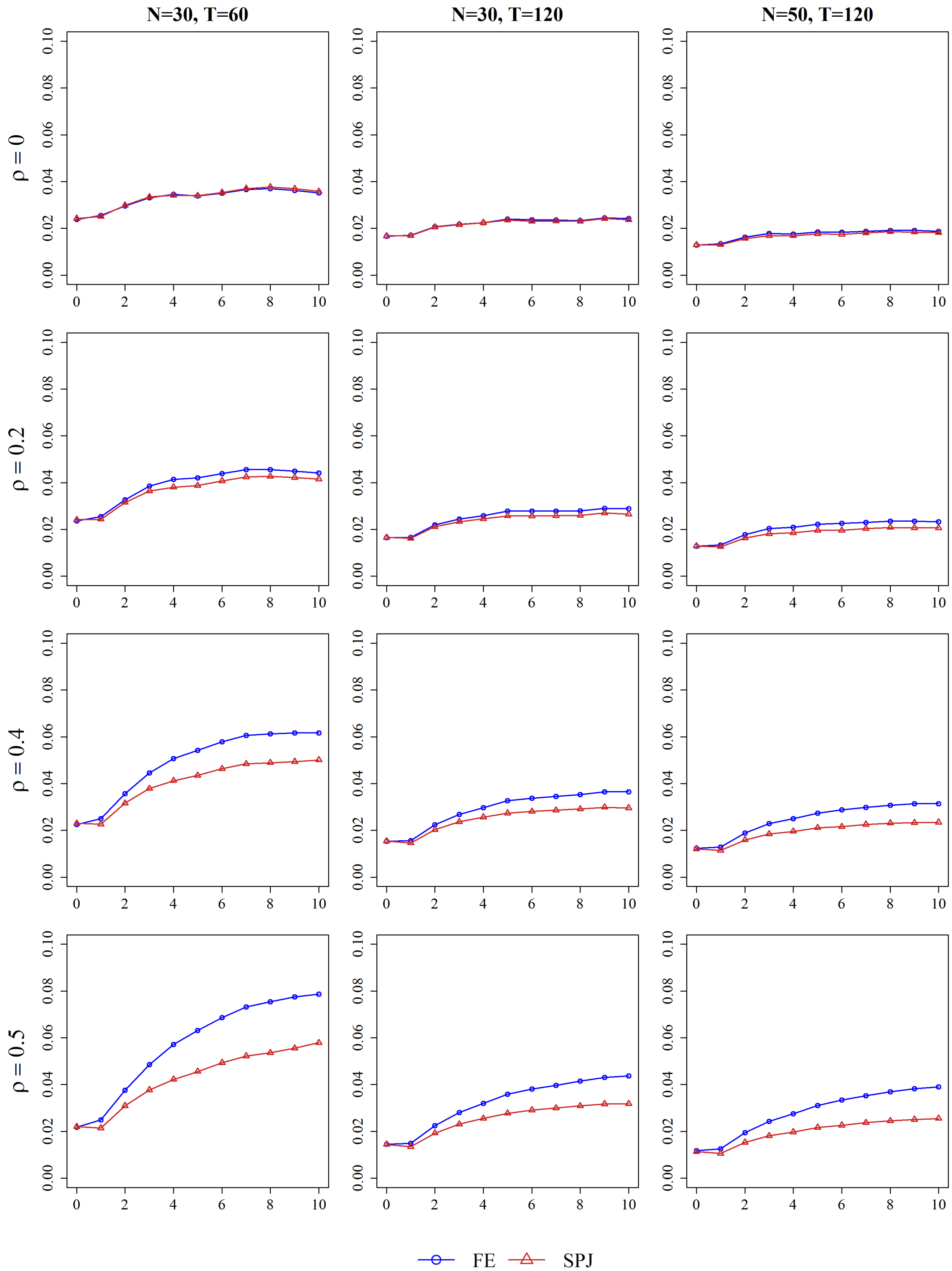
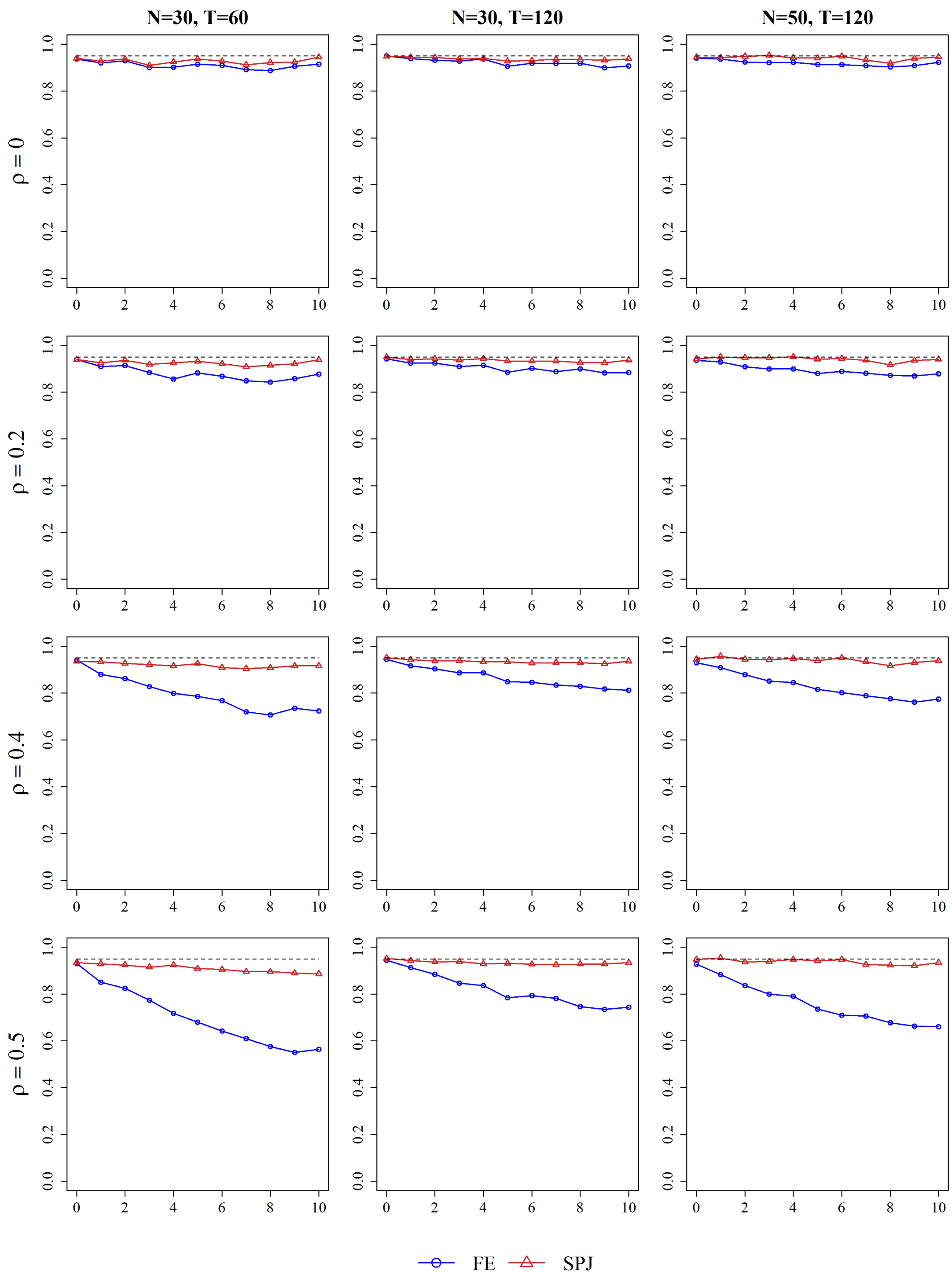


Figure A2: RMSEs from Model (A14)



Note: The nominal level of 0.95 is marked by the horizontal dashed line.

Figure A3: Coverage Probability of Confidence Interval Based on t-Statistic for Model (A14)