STAFF
REPORTS

NO. 1133
OCTOBER 2024

REVISED
JANUARY 2025

A Jackknife Variance Estimator for Panel Regressions

Richard K. Crump | Nikolay Gospodinov | Ignacio Lopez Gaffney

FEDERAL RESERVE BANK *of* NEW YORK

A Jackknife Variance Estimator for Panel Regressions

Richard K. Crump, Nikolay Gospodinov, and Ignacio Lopez Gaffney

Federal Reserve Bank of New York Staff Reports, no. 1133
October 2024; revised January 2025
https://doi.org/10.59576/sr.1133

Abstract

We introduce a new jackknife variance estimator for panel-data regressions. Our variance estimator can be motivated as the conventional leave-one-out jackknife variance estimator on a transformed space of the regressors and residuals using orthonormal trigonometric basis functions. We prove the asymptotic validity of our variance estimator and demonstrate desirable finite-sample properties in a series of simulation experiments. We also illustrate how our method can be used for jackknife bias-correction in a variety of time-series settings.

JEL classification: C12, C13, C22, C23
Key words: leave-one-out jackknife, panel data models, strong time-series and cross-sectional dependence, cluster-robust variance estimation, trigonometric basis functions

---

Crump: Federal Reserve Bank of New York (email: richard.crump@ny.frb.org). Gospodinov: Federal Reserve Bank of Atlanta (email: nikolay.gospodinov@atl.frb.org). Gaffney: Central Bank of the Republic of Argentina (email: iml2114@columbia.edu). The authors would like to thank Martín Almuzara, Matias Cattaneo, Bruce Hansen, Tassos Magdalinos, Ulrich Müller, Katerina Petrova, Mikkel Plagborg-Møller, and Mark Watson for helpful comments and discussions.

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the author(s) and do not necessarily reflect the position of the Central Bank of the Republic of Argentina, Federal Reserve Bank of Atlanta, Federal Reserve Bank of New York, or the Federal Reserve System. Any errors or omissions are the responsibility of the author(s).

To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr1133.html.

1 Introduction

Panel data models are often characterized by strong cross-sectional *and* time-series dependence. In particular, typical applications, such as panels of states or countries, feature time series behavior which is, at least partially, driven by common components (e.g., the state of the business cycle). This adds an additional layer of complexity as it induces a more intricate dependence structure across individuals and time. However, the standard variance estimators in panel regression models are not well-suited to accommodate this complicated two-way dependence structure which can result in substantial size distortions.

In this paper, we endeavor to address this challenge by proposing a jackknife variance estimator for panel regressions. The proposed estimator is fully agnostic about the source of the serial dependence in the common time-series behavior, is tuning-parameter-free, and offers substantial improvements over existing methods, especially when the serial correlation is particularly strong. While this paper is primarily concerned with panel data models, the proposed variance estimator naturally specializes to time-series models.

The validity of the standard (leave-one-out) jackknife approach relies on the lack of serial correlation in the data and is not directly applicable in a setting with time dependence. To accommodate this case, the proposed variance estimator is founded on a data transformation that renders the time dependence asymptotically negligible without affecting the OLS estimate. More specifically, we rotate the data using orthonormal trigonometric functions so that the transformed observations are now associated with particular frequencies that, under certain conditions, are asymptotically orthogonal to each other.[1] It is important to note that this property obviates the need for HAC/HAR adjustment in the variance matrix estimation.

We capitalize on this insight and construct a leave-one-out jackknife variance estimator on the space of the rotated variables. The variance estimator takes the usual sandwich form, but on the transformed variables, can account for clustering, and is weighted by the leverage. The advantages of the proposed approach are particularly notable under strong time dependence but it continues to offer size improvements even in the case of weak serial correlation. The proposed jackknife variance estimator can be nested in a more general framework which includes jackknife bias correction.[2]

[1] These basis functions may be generated as the eigenvectors of the variance matrix of a random walk process.

[2] In Appendix A.3, we show that such a bias-correction approach shows substantial promise. We simulate data from a persistent AR(1) and a standard predictive regression model and show that jackknife- and bootstrap-based

We assess the finite-sample properties of our proposed variance estimator in an extensive simulation experiment relying on data-generating processes already considered by Chiang et al. (2024), Chen and Vogelsang (2023), and Hidalgo and Schafgans (2021). We show that our jackknife variance estimator tightly controls empirical size in all of these designs, which feature strong cross-sectional or time-series dependence (or both). In contrast, all of the existing procedures show substantial size distortions. Furthermore, after size adjustment, the jackknife variance estimator exhibits comparable power properties to these alternatives. The simulation experiments also show that our jackknife approach is particularly effective when the degree of time-series dependence is high. This aligns with previous work using these basis functions. Crump and Gospodinov (2021) show that these basis functions provide close approximations to the eigenvectors of a persistent spatial AR(1) process. Crump, Gospodinov, and Lopez Gaffney (2024b), in a companion paper, show that these basis functions (asymptotically) orthogonalize a wide class of time-series processes.

Our paper is related to the growing literature studying cluster-robust variance estimation in panel data settings. The form of our jackknife variance estimator may be interpreted as clustering by individual unit whereas the orthogonalization property of our data transformation obviates the need for clustering over time. In contrast, recent contributions focus on variance estimators which explicitly accommodate different clustering structures. Cameron et al. (2011) were the first to propose variance estimators for general multi-way clustering (see also Davezies et al., 2021) whereas Thompson (2011) studied the two-cluster setting with temporal dependence that is nonzero for a finite lag length and absent otherwise. Menzel (2021) and MacKinnon et al. (2021) propose bootstrap-based inference procedures in multi-way clustering setups without serial dependence. More recently, Chiang et al. (2024) endeavor to accommodate general forms of unknown serial correlation in two-way clustered standard errors. They add a Newey-West type autocorrelation correction to the standard two-way clustered standard errors. Chen and Vogelsang (2023) modify the approach of Chiang et al. (2024) by considering a bias-corrected version of their standard errors and fixed-b asymptotic approximations to the associated t-statistic (see also Vogelsang, 2012). We share the motivation of this strand of the literature to conduct trustworthy inference across a range of different time-series behavior. However, our approach is fundamentally different as we remove, rather than accommodate, serial correlation before constructing our variance estimator.

Our paper is most closely related to Hidalgo and Schafgans (2021) who first transform the data

(on the rotated space) corrections largely eliminate the well documented OLS bias in these settings.

using the discrete Fourier transform (DFT) and then apply the Eicker-Huber-White variance estimator on the transformed data. There are three important differences between our approach and that of Hidalgo and Schafgans (2021). First, we rely on a different set of basis functions which have fundamentally different properties. In particular, our basis functions are well-equipped to accommodate both weak dependence but also much more persistent processes (Crump, Gospodinov, and Lopez Gaffney, 2024b). Since, in practice, the degree of persistence is unknown, this robustness is an important property as it substantially widens the set of applications where it can be used. Second, we propose a jackknife variance estimator which corresponds to HC3 rather than HC0 (in the parlance of MacKinnon and White, 1985), and, in different contexts has been shown to have better performance in theory and practice (see, e.g., Hansen, 2023; MacKinnon et al., 2023a,b). Furthermore, the DFT-based transformation does not naturally lend itself to leave-one-out estimation as the resultant regression equation is in terms of complex random variables. Third, in a series of simulation experiments using the simulation designs from Chiang et al. (2024), Chen and Vogelsang (2023), and Hidalgo and Schafgans (2021), we show that our variance estimator uniformly controls size, without compromising power, whereas the Hidalgo and Schafgans (2021) approach can exhibit size distortions.

Finally, since our variance estimator retains its validity when $N = 1$ our paper is related to the vast literature on HAC/HAR estimation. See Lazarus et al. (2018), Baillie et al. (2024) among others for recent contributions and a broad discussion of the existing literature. Furthermore, our results on the jackknife bias correction are related to the literature studying the finite-sample bias of the OLS estimator in nonstandard settings such as persistent autoregressions (e.g., Kendall, 1954) or predictive regressions (e.g., Stambaugh, 1999).

The paper is organized as follows. Section 2 introduces the relevant notation and provides a heuristic motivation for our jackknife variance estimator. Section 3 formally introduces the procedure and states all of the main theoretical results. Extensive simulation evidence on the finite-sample properties of the jackknife estimator, as compared to alternative procedures available in the literature, is presented in Section 4. Section 5 concludes. Appendix A.1 contains the proofs of the main results. Appendix A.2 provides supplemental simulation results while Appendix A.3 demonstrates the appealing finite-sample properties of a jackknife bias correction in autoregressive and predictive regression models.

Notation All limits are taken as $N, T \to \infty$. For sequences of numbers or random variables, we use $\mathsf{a}_{N,T} \lesssim \mathsf{b}_{N,T}$ to denote that $\limsup_{N,T} |\mathsf{a}_{N,T}/\mathsf{b}_{N,T}|$ is finite, $\mathsf{a}_{N,T} \lesssim_p \mathsf{b}_{N,T}$ or $\mathsf{a}_{N,T} = O_p(\mathsf{b}_{N,T})$ to denote $\limsup_{\varepsilon\to\infty} \limsup_{N,T} \mathbb{P}[|\mathsf{a}_{N,T}/\mathsf{b}_{N,T}| \geq \varepsilon] = 0$, $\mathsf{a}_{N,T} = o(\mathsf{b}_{N,T})$ implies $\mathsf{a}_{N,T}/\mathsf{b}_{N,T} \to 0$, and $\mathsf{a}_{N,T} = o_p(\mathsf{b}_{N,T})$ implies that $\mathsf{a}_{N,T}/\mathsf{b}_{N,T} \to_p 0$, where $\to_p$ denotes convergence in probability. In addition, let $\to_d$ denote convergence in distribution. Finally, let $\|\cdot\|$ denote the Euclidean norm such that $\|A\|^2 = \text{tr}(A'A)$ for a real-valued matrix, A.

2 Motivation for Variance Estimator

Our variance estimator relies on leave-one-out jackknife estimation with transformed dependent and independent variables. We use a particular choice of orthonormal trigonometric basis functions which have desirable features. In this section, we provide background and motivation for our approach before we introduce our formal results in the next section.

2.1 Jackknife Variance Estimation

Consider the linear regression model:

$$y_i = \alpha + x_i'\beta + \epsilon_i, \qquad i = 1, \ldots, N, \tag{1}$$

where $\mathbb{E}[x_i\epsilon_i] = 0$. We can stack this model as

$$y = \mathbf{X}\theta + \epsilon, \tag{2}$$

where $\theta = \left(\alpha,\ \beta'\right)'$, $\mathbf{X} = [\iota_N,\ x]$, where ι_N is $N \times 1$ vector of ones and x is a matrix with ith row equal to x_i. Let $\widehat{\theta}$ be the OLS estimator of equation (2). Then, the jackknife variance estimator of $\widehat{\theta}$ is

$$\widehat{\mathsf{V}}_\theta = \sum_{j=1}^{N} (\widehat{\theta}_{(-j)} - \widehat{\theta})(\widehat{\theta}_{(-j)} - \widehat{\theta})', \tag{3}$$

where $\widehat{\theta}_{(-j)}$ is the OLS estimator of (2) for the sample which excludes the jth observation (see Shao and Tu, 1996, for an introduction).[3] MacKinnon and White (1985) clarify the relationship between the variance estimator in equation (3) and the more familiar Eicker-White-Huber robust variance estimator. We have that,

$$\widehat{\mathbb{V}}_\theta = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{i=1}^{N} (1 - p_{ii})^{-2} x_i x_i' \, \hat{\epsilon}_i^2 \right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}, \tag{4}$$

where $x_i = (x_{1,i}, x_{2,i}')'$ and $p_{ii} = x_i' \left(\mathbf{X}'\mathbf{X}\right)^{-1} x_i$. The presence of the leverage term, $(1 - p_{ii})^{-2}$ introduces a more conservative variance estimator than the conventional heteroskedasticity-robust formulation which has been found to have desirable properties both in theory and in practice (see Hansen, 2023; MacKinnon et al., 2023a,b,c, and references therein).

2.2 Rotating the Data

The appealing properties of the jackknife variance estimator we have just introduced rely on an assumption of independence across units and so does not carry over immediately to the time-series setting. To address this issue, we rotate the data using a particular choice of basis functions which has the effect of strongly diminishing the covariation across observations. We now define the orthogonal trigonometric basis. Let $\psi_j \equiv (\psi_{1,j}, ..., \psi_{T,j})'$, where

$$\psi_{h,j} \equiv \frac{2}{\sqrt{2T+1}} \sin\left(\frac{h(2j-1)\pi}{2T+1} \right). \tag{5}$$

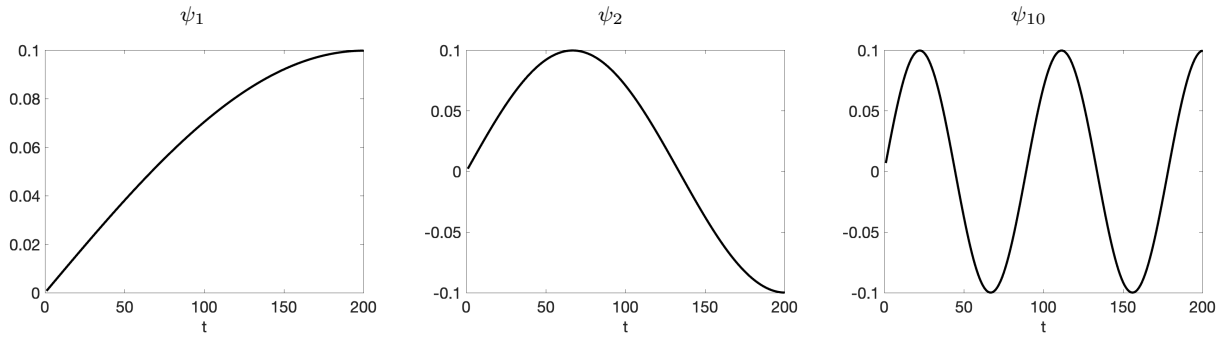
Each ψ_j is an eigenvector of the variance matrix of a random walk and so satisfies $\Psi'\Psi = \Psi\Psi' = \mathsf{I}_T$. Figure 1 shows the first, second and tenth basis functions as examples. We can observe that ψ_1 is the lowest frequency basis function and as j increases, the periodicity of ψ_j shortens.

Crump, Gospodinov, and Lopez Gaffney (2024b) prove that for a large class of time-series processes we have that

$$\text{Corr}(\psi_j' y, \psi_k' y) = o(1), \qquad \forall j \neq k, \tag{6}$$

[3] Jackknife variance estimators have also been proposed where the centering occurs at $N^{-1} \sum_{j=1}^{N} \widehat{\theta}_{(-j)}$ instead of $\widehat{\theta}$. In practice, the results are very similar and so we focus on the latter version of the jackknife variance estimator in this paper.

Figure 1. Examples of Basis Functions. This figure shows selected basis functions based on the choice of Ψ, as given by equation (5), for $T = 200$.



as $T \to \infty$, where $y = (y_1, \ldots, y_T)$. Even in cases where this result does not hold, Crump, Gospodinov, and Lopez Gaffney (2024b) show that there is very little residual correlation across observations when this transformation is applied, even in small samples.

Consider the following time series model:

$$y_t = \alpha + x_t'\beta + \epsilon_t, \qquad i = 1, \ldots, T, \tag{7}$$

which we stack similarly to obtain

$$y = \mathbf{X}\vartheta + \epsilon, \tag{8}$$

where we use the notation $\vartheta = (\alpha, \beta')'$ to distinguish from the cross-sectional case above. We can limit the degree of correlation across observations by rotating the data using Ψ. To do so, we pre-multiply to obtain

$$\Psi' y = \Psi'\mathbf{X}\vartheta + \Psi'\epsilon. \tag{9}$$

By the orthonormal property of Ψ, we can interpret $\Psi' y$ as the OLS regression of y on each of the T basis functions (and similarly for each column of $\mathbf{X}$) since

$$w = (\Psi'\Psi)^{-1}\Psi' y = \Psi' y. \tag{10}$$

Thus, w collects these coefficient estimates, where w_j is associated with a particular frequency of the data. In our case, w_1 is the loading on the lowest frequency basis function, w_2 is the loading

on the next lowest, and so on.

Importantly, this transformation has no effect on the OLS estimator as

$$\widehat{\vartheta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = (\mathbf{X}'\Psi\Psi'\mathbf{X})^{-1}\mathbf{X}'\Psi\Psi'y = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'w, \tag{11}$$

where $\mathbf{Z} = \Psi'\mathbf{X}$ and $w = \Psi'y$. Although the OLS estimator is invariant to the transformation Ψ, when we omit the jth observation of w and $\mathbf{Z}$ we obtain an OLS estimate which is different from a leave-one-out estimate in the time domain (i.e., the un-rotated data). This leads to the modified jackknife estimation approach we consider herein

$$\widehat{\mathsf{V}}_\vartheta = \sum_{j=1}^{T}(\widehat{\vartheta}_{(-j)} - \widehat{\vartheta})(\widehat{\vartheta}_{(-j)} - \widehat{\vartheta})', \tag{12}$$

where $\widehat{\vartheta}_{(-j)}$ is the estimate constructed by removing the jth observation (i.e., omitting a "frequency"). In the next section, we will generalize this approach to panel data and prove its asymptotic validity for inference.

Remark 1 (Bias Correction)**.** The focus of this paper is on jackknife variance estimation but, by analogy, we can also use our jackknife approach to bias correct the OLS estimator,

$$\widehat{\vartheta}_{\text{bc}} = T\widehat{\vartheta} - \frac{(T-1)}{T}\sum_{j=1}^{T}\widehat{\vartheta}_{(-j)}. \tag{13}$$

Furthermore, we can also consider the pairs bootstrap on the transformed data, $\{(w_j, \mathbf{Z}_j') : j = 1, \ldots, T\}$, to construct an alternative bias correction. In Appendix A.3, we show extensive simulation evidence suggesting that both of these bias correction methods perform well in challenging estimation settings. □

3 Main Results

We focus on the following linear panel-data model:

$$y_{it} = \alpha_i + x_{1,it}\beta_1 + x_{2,it}'\beta_2 + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{14}$$

where $x_{1,it} \in \mathbb{R}$, $x_{2,it} \in \mathbb{R}^K$. Without loss of generality we assume that β_1 is the parameter of interest. We can stack the model as

$$Y_i = \alpha_i \iota_T + X_{1,i}\beta_1 + X_{2,i}\beta_2 + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{15}$$

where $Y_i = (y_{i1,\ldots,}y_{iT})'$ and similarly for $X_{1,i}$, $X_{2,i}$, and ε_i. As in the previous section, we can transform this equation using our basis functions Ψ which yields

$$W_i = \alpha_i \zeta + Z_{1,i}\beta_1 + Z_{2,i}\beta_2 + u_i, \qquad i = 1, \ldots, N, \tag{16}$$

where $W_i = \Psi' Y_i$, $Z_{1,i} = \Psi' X_{1,i}$, $Z_{2,i} = \Psi' X_{2,i}$, and $\zeta = \Psi' \iota_T$. We will construct our variance estimator as a leave-out jackknife of equation (16). Unlike the previous section, however, we will omit N observations (rather than one) when we construct our leave-out OLS estimator. Let w_{ij} be the jth element of W_i and similarly for ζ_j, $z_{1,ij}$, $z_{2,ij}$, and u_{ij}. We omit the jth observation for each i from $i = 1, \ldots, N$ and, using the remaining data (sample of size $(T-1)N$), we calculate the OLS estimator, $\widehat{\beta}_{(-j)}$. Let $\widehat{\beta}_{1,(-j)}$ be the component of this OLS estimator corresponding to β_1. Then, our jackknife (JN) variance estimator is

$$\widehat{\mathsf{V}}_1 = \sum_{j=1}^{T} (\widehat{\beta}_{1,(-j)} - \widehat{\beta}_1)^2. \tag{17}$$

We can rewrite $\widehat{\mathsf{V}}_1$ in a more familiar form. To do so, let us define the following OLS estimator from the stacked panel regression of $x_{1,it}$ on $x_{2,it}$, and individual fixed effects,

$$\widehat{\lambda} = \Big(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{2,it} - \widehat{\mu}_{2,i})(x_{2,it} - \widehat{\mu}_{2,i})' \Big)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{2,it} - \widehat{\mu}_{2,i}) x_{1,it} \tag{18}$$

along with

$$\widehat{\mu}_{1,i} = T^{-1} \sum_{t=1}^{T} x_{1,it}, \qquad \widehat{\mu}_{2k,i} = T^{-1} \sum_{t=1}^{T} x_{2k,it}, \tag{19}$$

and $\widehat{\mu}_{2,i} = (\widehat{\mu}_{21,i}, \ldots, \widehat{\mu}_{2K,i})'$. Then, since we focus on inference on β_1, the relevant (scaled) Gram matrix is

$$\widehat{\omega}_1 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{1,it} - \widehat{\mu}_{1,i} - (x_{2,it} - \widehat{\mu}_{2,i})' \widehat{\lambda} \right)^2 . \tag{20}$$

Finally, let $\widehat{\varepsilon}_i$ be the OLS residual from equation (15) and define $\widehat{u}_{ij} = \psi_j' \widehat{\varepsilon}_i$. Also define,

$$\widehat{v}_{ij} = \psi_j' \big(X_{1,i} - \widehat{\mu}_{1,i} \iota_T - (X_{2,i} - \iota_T \widehat{\mu}_{2,i}') \widehat{\lambda} \big), \tag{21}$$

and stack these objects across i as $\widehat{u}^j = (\widehat{u}_{1j}, \ldots, \widehat{u}_{Nj})'$ and $\widehat{V}^j = (\widehat{v}_{1j}, \ldots, \widehat{v}_{Nj})'$. We can then re-express $\widehat{\mathsf{V}}_1$ as

$$\widehat{\mathsf{V}}_1 = \frac{1}{NT} \times \widehat{\omega}_1^{-1} \Big(\frac{1}{NT} \sum_{j=1}^{T} \widehat{V}^{j\prime} \left(\mathsf{I}_N - \mathsf{P}_{jj} \right)^{-1} \widehat{u}^j \widehat{u}^{j\prime} \left(\mathsf{I}_N - \mathsf{P}_{jj} \right)^{-1} \widehat{V}^j \Big) \widehat{\omega}_1^{-1}, \tag{22}$$

where $\mathsf{P}_{jj} = \mathbf{Z}^j \big(\mathbf{X}'\mathbf{X} \big)^{-1} \mathbf{Z}^j$, $\mathbf{Z}^j = (\mathsf{I}_N \otimes \psi_j')\mathbf{X}$, and $\mathbf{X}$ is the $NT \times (N + K + 1)$ matrix of right-hand side variables (including dummy variables for the fixed effects) from equation (14).[4] Although $\widehat{\omega}_1$ is a scalar, we write the expression in equation (22) in "sandwich" form for clarity. $\widehat{\omega}_1^{-1}$ represents the inverse of the Gram matrix and would generally be present with any variance estimator. The middle of the sandwich, however, is fundamentally different from the usual formulation. In the absence of the matrix $\left(\mathsf{I}_N - \mathsf{P}_{jj} \right)^{-1}$, the estimator is of Eicker-Huber-White form (HC0 or HC1) but, importantly, on the transformed variables $\widehat{v}_{ij}$ and $\widehat{u}_{ij}$. Instead, the presence of $\left(\mathsf{I}_N - \mathsf{P}_{jj} \right)^{-1}$ lends the interpretation of a weighted version of this estimator (HC3); see MacKinnon and White (1985).

To prove the validity of our variance estimator we require some assumptions on the data generating process.

Assumption 1 (Model)**.** *We observe* (y_{it}, x_{it}') *which satisfies equation* (14) *and assume the following:*

[4] In setups where there are natural clusters in the cross-sectional dimension (e.g., states, when cities are the observational units), and where we assume, e.g., independence across clusters (as in Hansen, 2023, 2024), we can implement a specialized version of our variance estimator, $\widehat{\mathsf{V}}_1 = \frac{1}{NT} \times \widehat{\omega}_1^{-1} \Big(\frac{1}{NT} \sum_{j=1}^{T} \sum_{g=1}^{G} \widehat{V}_g^{j\prime} \left(\mathsf{I}_{N_g} - \mathsf{P}_{jj,g} \right)^{-1} \widehat{u}_g^j \widehat{u}_g^{j\prime} \left(\mathsf{I}_{N_g} - \mathsf{P}_{jj,g} \right)^{-1} \widehat{V}_g^j \Big) \widehat{\omega}_1^{-1}$, where g indexes the N_g observations in each of the G clusters.

(*i*) *For each* i, (y_{it}, x_{it}') *is strictly stationary and ergodic with* $\mu_{1,i} := \mathbb{E}[x_{1,it}]$ *and* $\mu_{2k,i} := \mathbb{E}[x_{2k,it}]$ *for* $k = 1, \ldots, K$;

(*ii*) $\sqrt{NT}(\widehat{\lambda} - \lambda) = O_p(1)$;

(*iii*) *Let* $\omega_1 = \text{plim}_{N,T\to\infty}\widehat{\omega}_1$ *and* $\gamma_1 = \lim_{N,T\to\infty}\mathbb{E}\Big[\big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{1,it} - x_{2,it}'\lambda)\varepsilon_{it}\big)^2\Big]$. *For some* $c > 0$, $c < \omega_1 \lesssim 1$, $c < \gamma_1 \lesssim 1$ *and* $\sqrt{NT}\Big(\hat{\beta}_1 - \beta_1\Big) \to_d \mathcal{N}\left(0, \omega_1^{-2}\gamma_1\right)$;

(*iv*) $\sqrt{NT}\Big(\hat{\beta}_2 - \beta_2\Big) = O_p(1)$.

Assumption 1 imposes relatively modest restrictions on the data-generating process. We assume that the time-series dependence is well behaved and that OLS estimation of β_1 with the associated standard limiting distribution holds in our model. In addition to Assumption 1, we also assume the following:

Assumption 2 (Moments and Dependence)**.** *Assume that:*

(*i*) $NT^{-1}\log(T) \to 0$;

(*ii*) $\text{ALRV}(\varepsilon_{it}) \lesssim 1$, $\text{ALRV}(x_{1,it}) \lesssim 1$, *and* $\text{ALRV}(x_{2k,it}) \lesssim 1$ *for* $k = 1, \ldots, K$ *where* $\text{ALRV}(\mathsf{x}_{it}) := \lim_{N,T\to\infty} N^{-1}\sum_{i=1}^{N}\mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathsf{x}_{it}\right)$;

(*iii*) $\max_{1\le j\le T}\max_{1\le i\le N}\mathbb{E}\left[|u_{ij}|^4\right] \lesssim 1$, $\max_{1\le j\le T}\max_{1\le i\le N}\mathbb{E}\left[|z_{1,ij}|^4\right] \lesssim 1$, *and* $\max_{1\le j\le T}\max_{1\le i\le N}\mathbb{E}\left[|z_{2k,ij}|^4\right] \lesssim 1$ *for* $k = 1, \ldots, K$;

(*iv*) *Define* $\widetilde{v}_{ij} = \psi_j'\big(x_{1,it} - \mu_{1,i} - (x_{2,it} - \mu_{2,i})'\lambda\big)$, $\widetilde{V}^j = (\widetilde{v}_{1j}, \ldots, \widetilde{v}_{Nj})'$, *and* $u^j = (u_{1j}, \ldots, u_{Nj})'$. *Then,* $(NT)^{-1}\sum_{j=1}^{T}\Big\{\big(\widetilde{V}^{j\prime}u^j\big)^2 - \mathbb{E}\big[\big(\widetilde{V}^{j\prime}u^j\big)^2\big]\Big\} = o_p(1)$;

(*v*) $(NT)^{-1}\sum_{j=1}^{T}\mathbb{E}\Big[\big(\widetilde{V}^{j\prime}u^j\big)^2\Big] - (NT)^{-1}\mathbb{E}\Big[\big(\sum_{j=1}^{T}\widetilde{V}^{j\prime}u^j\big)^2\Big] = o(1)$.

Assumption 2 makes further restrictions on how the data are generated. Assumption 2(*i*) imposes that T must grow sufficiently fast relative to N. Assumption 2(*ii*) ensures that the average long-run variances (ALRV) of both the regressors and the regression errors exist and are finite. Assumption 2(*iii*) $-$ (*v*) regulate the degree of time-series and cross-sectional dependence in the observed data. Assumption 2(*iii*) restricts the higher-order covariances of the transformed regressors and regression errors to diminish as observations become increasingly far apart. Meanwhile, Assumptions 2(*iv*) $-$ (*v*) place joint restrictions on the degree of dependence across units and time.

These latter two assumptions play a key role in establishing consistency of the variance estimator. Assumption 2(v) is weaker than assuming independence of the regressors and errors (as assumed by Hidalgo and Schafgans, 2021) and is similar to assumptions commonly made in the panel data literature (e.g., Gonçalves, 2011). Section 3.1 discusses the case when Assumption 2(v) fails.

We can now state our main result.

Theorem 1. *Let Assumptions 1 and 2 hold. Then,*

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{\mathsf{V}}_1^{1/2}} \to_d \mathcal{N}(0,1) \tag{23}$$

as $N, T \to \infty$.

Theorem 1 shows that valid asymptotic inference can be conducted using the jackknife variance estimator we introduce. Because there are no additional tuning parameters necessary, our estimator can be implemented easily and in a computationally-effective manner.

Theorem 1 proves validity of the t-statistic for a test of significance of the marginal coefficient β_1. Our results do not immediately carry over to general tests of linear restrictions for the coefficient vector β. To accommodate Wald-type tests, the variance estimator must be modified to adjust for the fact that the basis functions only marginally orthogonalize each variable.

There are several other remarks on our proposed method which are warranted:

Remark 2 (Unbalanced Panels)**.** In practice, panels may not be balanced. We can accommodate the unbalanced panel case by defining a selection matrix $S \in \mathbb{R}^{\bar{N} \times (N \cdot T)}$, where $\bar{N}$ is the number of observations in the unbalanced panel, and S is a matrix of ones and zeros corresponding to the observations available in the sample. In order to calculate the jackknife estimator in this unbalanced case, define

$$\Psi_{NT,(-j)} = S(\mathsf{I}_N \otimes \Psi_{(-j)}),$$

where $\Psi_{(-j)}$ is the matrix Ψ with the jth column removed. Then, we can calculate $\widehat{\beta}_{(-j)}$ through a regression of $\Psi'_{NT,(-j)}\mathbf{Y}$ on $\tilde{\Psi}'_{NT,(-j)}\mathbf{X}$, where $\mathbf{X}$ is the $\bar{N} \times (N + K + 1)$ vector of right-hand side variables and $\mathbf{Y}$ is the corresponding $\bar{N} \times 1$ vector of left-hand side variables. □

Remark 3 (Spatial Models)**.** We can also consider more general forms of dependence. It is straightforward to tailor our procedure to models of spatial dependence, i.e., $X \in \mathbb{R}^{N^2}$ which follows a $d = 2$ Levy-Brownian motion as in Müller and Watson (2024). Here, N is the cardinality of the sets $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^N$. Correspondingly, the index set $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$ has cardinality N^2 where $\mathcal{C}(i)$ defines the location or coordinate where X_i is observed for all $i \in \{1, ..., N^2\}$. Here, we assume a single cross-section for simplicity but similar steps as below can be followed in the panel spatial case.

Given that X follows a Levy-Brownian motion, we can describe its covariance matrix $\Sigma \in \mathbb{R}^{N^2 \times N^2}$ as

$$\Sigma_{i,j} = \frac{1}{2}(|\mathcal{C}(i)| + |\mathcal{C}(j)| - |\mathcal{C}(i) - \mathcal{C}(j)|),$$

where $||x|| = \sqrt{x'x}$. An eigendecomposition of Σ produces a Ψ which satisfies

$$\Psi' \Sigma \Psi = \Lambda,$$

where Λ is a diagonal matrix collecting the eigenvalues. Although, in this case, Ψ is not generally available in closed-form, it can be easily computed numerically.

In practice, we may have that $\mathcal{C} \neq \mathcal{X} \times \mathcal{Y}$ (i.e., the data are irregularly sampled over a grid). To accommodate this case, we can define a set $\tilde{\mathcal{C}} \equiv \mathcal{X} \times \mathcal{Y}$ and a selection matrix $S \in \mathbb{R}^{(N_x \cdot N_y) \times \bar{N}}$, where $\bar{N}$ is the cardinality of $\mathcal{C}$. Then, we can form the matrix $\tilde{\Sigma} \in \mathbb{R}^{(N_x \cdot N_y) \times (N_x \cdot N_y)}$ as above and then perform an eigendecomposition on $\tilde{\Sigma}$ to obtain $\tilde{\Psi}$. We then construct $\Psi = S'\tilde{\Psi}$, and utilize the transformation $\Psi' X$.

To generalize to the regression setting, consider the spatial regression model:

$$y = \alpha \iota_{\bar{N}} + X_1 \beta_1 + X_2 \beta_2 + \epsilon,$$

where y is an $\bar{N} \times 1$ vector. We can pre-multiply by Ψ' as before and obtain the spatial jackknife variance estimator,

$$\widehat{\mathsf{V}}_1 = \sum_{j=1}^{(N_x N_y)} (\widehat{\beta}_{1,(-j)} - \widehat{\beta}_1)^2. \tag{24}$$

□

3.1 Properties under Weak Exogeneity of the Regressors

Assumption 2(v) may be too strict for some applications, such as panel local projections. When Assumption 2(v) fails, the JN variance estimator is no longer consistent. However, in many cases the JN variance estimator will still produce asymptotically valid, albeit conservative, inference. Below we characterize the conditions which ensure this validity.

We start with the case of $N = 1$ for simplicity and to cement intuition. Let us simplify further and assume that $\{(y_t, x_t') : t = 1, \ldots, T\}$ is a jointly Gaussian time series satisfying equation (14) with a corresponding $T \times T$ "endogeneity" matrix, $\Gamma = \mathbb{E}\left[(X_1 - X_2\lambda)\,\varepsilon'\right]$. We assume that the regressors are pre-determined so that Γ is a lower triangular matrix with diagonal elements all equal to zero. Then, it can be shown that

$$\hat{\mathsf{V}}_1 - \mathsf{V}_1 = \omega_1^{-2} \cdot \frac{2}{T} \sum_{j=1}^{T} \left(\psi_j' \Gamma \psi_j\right)^2 - \omega_1^{-2} \cdot \frac{1}{T} \sum_{j \neq s} \psi_j' \Sigma^x \psi_s \psi_j' \Sigma^\varepsilon \psi_s + o_p\left(1\right), \tag{25}$$

where $\Sigma^x = \mathrm{Cov}\left(X_1 - X_2\lambda, X_1 - X_2\lambda\right)$ and $\Sigma^\varepsilon = \mathbb{E}\left[\varepsilon\varepsilon'\right]$. The second term can be shown to be $o(1)$ under mild restrictions on the dependence structure of the data using results in Crump, Gospodinov, and Lopez Gaffney (2024b). As a simple example, if $\Sigma^\varepsilon \propto \mathsf{I}_T$ then this term is identically zero by the orthonormality of the basis functions. Since the first term in equation (25) is always positive, then the JN variance estimator is guaranteed to be, at worst, conservative as $T \to \infty$.

To generalize this result to wider settings including with $N > 1$, we make the following assumption.

Assumption 3 (Asymptotic Conservativeness)**.** *Assume that:*

(*i*) $NT^{-1} \cdot \mathrm{tr}(\bar{\Gamma})^2 = o_p(1)$ *if* $N > 1$ *or* $\bar{\Gamma} = 0$ *if* $N = 1$, *where* $\bar{\Gamma} = N^{-1}\sum_{i=1}^{N} \Gamma_{ii}$ *and* $\Gamma_{ii} := \mathbb{E}\left[(X_{1,i} - X_{2,i}\lambda)\varepsilon_i'\right]$;

(*ii*) $(NT)^{-1}\sum_{i=1}^{N}\sum_{\ell=1}^{N} \mathrm{tr}(\Gamma_{i\ell}\Gamma_{\ell i}) = o_p(1)$ *if* $N > 1$ *and* $\sum_{i=1}^{N}\sum_{\ell=1}^{N} \mathrm{tr}(\Gamma_{i\ell}\Gamma_{\ell i}) = 0$ *if* $N = 1$, *where* $\Gamma_{i\ell} := \mathbb{E}\left[(X_{1,i} - X_{2,i}\lambda)\varepsilon_\ell'\right]$;

(*iii*) $(NT)^{-1}\left|\kappa_{NT}\right| = o_p(1)$, *where*

$$\kappa_{NT} := \sum_{j \neq s} \sum_{i,\ell=1}^{N} \left(\mathbb{E}\left[\widetilde{v}_{ij} u_{ij} \widetilde{v}_{\ell s} u_{\ell s}\right] - \mathbb{E}\left[\widetilde{v}_{ij}\widetilde{v}_{\ell s}\right]\mathbb{E}\left[u_{ij}u_{\ell s}\right] - \mathbb{E}\left[\widetilde{v}_{ij}u_{ij}\right]\mathbb{E}\left[\widetilde{v}_{\ell s}u_{\ell s}\right] - \mathbb{E}\left[\widetilde{v}_{ij}u_{\ell s}\right]\mathbb{E}\left[\widetilde{v}_{\ell s}u_{ij}\right]\right);$$

(iv) $(NT)^{-1}\,|\varkappa_{NT}| = o(1)$, *where*

$$\varkappa_{NT} := \sum_{i,\ell=1}^{N} \sum_{j\neq s} \psi_j'\big(\Sigma_{i\ell}^{x} + \Sigma_{i\ell}^{x\prime}\big)\psi_s\psi_j'\big(\Sigma_{i\ell}^{\varepsilon} + \Sigma_{i\ell}^{\varepsilon\prime}\big)\psi_s$$

and $\Sigma_{i\ell}^{x} = \text{Cov}(X_{1,i} - X_{2,i}\lambda, X_{1,\ell} - X_{2,\ell}\lambda)$ *and* $\Sigma_{i\ell}^{\varepsilon} = \mathbb{E}[\varepsilon_i\varepsilon_\ell']$.

Assumption 3(*i*) is innocuous and is, in fact, implied by the assumption of the consistency of the OLS estimator (Assumption 1). For completeness, we list it here to enable comparisons with Assumption 3(*ii*). A simple sufficient condition for Assumption 3(*ii*) to hold is that the regressors are pre-determined both within the same unit and across units. Assumption 3(*iii*) imposes that $(\widetilde{v}_{ij}, u_{ij})$ across i and j have fourth-order cross-moments that behave, as if, these random variables were Gaussian. A sufficient condition for Assumption 3(*iii*) to hold is if the joint distribution of the regressors and the errors is conditionally Gaussian for each (N, T). This would accommodate, for example, general forms of stochastic volatility (e.g., GARCH) if, conditional on the time-varying volatility component, the errors are Gaussian. Moreover, Assumption 3(*iii*) holds more generally so long as the joint convergence of $(\widetilde{v}_{ij}, u_{ij})$ to a Gaussian limit occurs sufficiently fast. For reference, with j fixed, the results in Abadir et al. (2014) can be used to show that $(\widetilde{v}_{ij}, u_{ij})$ are jointly asymptotically normal when $\{(x_t', \varepsilon_t) : t = 1, \ldots, T\}$ follow a linear time series process with martingale difference innovations. Finally, sufficient conditions for Assumption 3(*iv*) to hold are mild restrictions on the dependence structure of the data (Crump, Gospodinov, and Lopez Gaffney, 2024b).

We can now characterize the general result.

Theorem 2. *Let Assumption 1, Assumption 2(i)–(iv), and Assumption 3 hold. Then,*

$$\widehat{\mathsf{V}}_1 - \mathsf{V}_1 = \omega_1^{-2} \cdot \frac{N}{T} \sum_{j=1}^{T} \left(\psi_j'\bar{\Gamma}\psi_j\right)^2 + \omega_1^{-2} \cdot \frac{1}{NT} \sum_{j=1}^{T} \left\| (\mathsf{I}_N \otimes \psi_j)'\Gamma_{NT}(\mathsf{I}_N \otimes \psi_j) \right\|^2 + o_p(1), \tag{26}$$

where Γ_{NT} *is a block matrix with* (i, ℓ) *block equal to* $\Gamma_{i\ell}$.

Theorem 2 implies that $\widehat{\mathsf{V}}_1 \to_p \mathsf{V}_1^*$, where $\mathsf{V}_1^* \geq \mathsf{V}_1$, and so we can still rely on the JN variance estimator to conduct inference on the regression coefficient of interest. As an alternative, we can regain consistency of the variance estimator by correcting $\widehat{\mathsf{V}}_1$ to remove the two terms on the right-hand side of equation (26). Then, under regularity conditions, and if $N\|\widehat{\bar{\Gamma}} - \bar{\Gamma}\| = o_p(1)$ and

$N^{-1}\sum_{i,\ell}\|\widehat{\Gamma}_{i\ell}-\Gamma_{i\ell}\|=o_p(1)$, it can be shown that

$$\widehat{\mathsf{V}}_1-\mathsf{V}_1-\omega_1^{-2}\cdot\frac{N}{T}\sum_{j=1}^{T}\left(\psi_j'\widehat{\bar{\Gamma}}\psi_j\right)^2-\omega_1^{-2}\cdot\frac{1}{NT}\sum_{j=1}^{T}\left\|(\mathsf{I}_N\otimes\psi_j)'\widehat{\Gamma}_{NT}(\mathsf{I}_N\otimes\psi_j)\right\|^2=o_p(1).$$

4 Simulation Evidence

In this section, we present simulation evidence on the finite-sample properties of our JN estimator relative to a number of other procedures available in the literature. For our simulation experiments, we use the exact designs utilized in Chiang et al. (2024), Chen and Vogelsang (2023), and Hidalgo and Schafgans (2021). This has several advantages. First, we can cover a wide range of different data-generating mechanisms. Second, these are established designs and so are not tailored to our approach. Finally, this allows us to (implicitly) make comparisons to an even wider range of methods since these papers include computationally-intensive bootstrap methods which we avoid for simplicity (e.g., Chiang et al., 2024; MacKinnon et al., 2021; Hidalgo and Schafgans, 2021) and variance estimators which are nested in other procedures (e.g., Thompson, 2011).

We consider the following three data-generating processes (DGPs):

CHS This design is from Chen and Vogelsang (2023) but was originally considered in Chiang et al. (2024). The data follow the linear model

$$Y_{it}=\beta_0+X_{it}\beta_1+U_{it},$$

with $(\beta_0,\beta_1)=(1,1)$, where (X_{it},U_{it}) are generated as

$$X_{it}=w_\alpha\alpha_i^x+w_\gamma\gamma_t^x+w_\epsilon\epsilon_{it}^x,\qquad U_{it}=w_\alpha\alpha_i^u+w_\gamma\gamma_t^u+w_\epsilon\epsilon_{it}^u,$$
$$\gamma_t^x=\rho\gamma_{t-1}^x+\left(\sqrt{1-\rho^2}\right)\tilde{\gamma}_t^x,\qquad \gamma_t^u=\rho\gamma_{t-1}^x+\left(\sqrt{1-\rho^2}\right)\tilde{\gamma}_t^u,$$

and $(w_\alpha^x,w_\gamma^x,w_\epsilon^x)=(w_\alpha^u,w_\gamma^u,w_\epsilon^u)=(.25,.5,.25)$. Finally, $(\alpha_i^x,\alpha_i^u,\epsilon_{it}^x,\epsilon_{it}^u,\tilde{\gamma}_t^x,\tilde{\gamma}_t^u)$ along with the initial conditions for $\tilde{\gamma}_t^x$ and $\tilde{\gamma}_t^u$, are mutually independent standard Gaussian random variables.

CHS-NL This design is a non-linear version of the CHS design and is taken from Chen and Vogelsang

(2023). They replace X_{it} and U_{it} in the CHS design above with

$$X_{it} = \Phi\left(w_\alpha \alpha_i^x + w_\gamma \gamma_t^x + w_\epsilon \epsilon_{it}^x\right), \qquad U_{it} = \Phi\left(w_\alpha \alpha_i^u + w_\gamma \gamma_t^u + w_\epsilon \epsilon_{it}^u\right),$$

where $\Phi(\cdot)$ is the CDF of a standard Gaussian random variable.

HS This design is taken from Hidalgo and Schafgans (2021). The data follow the linear model

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

where α_t and η_i are mutually independent Gaussian random variables with unit mean and variance and are held fixed across simulations. We follow Hidalgo and Schafgans (2021) and set $\beta = 0$. To generate spatial dependence, we draw s_i as an independent uniform random variable on $[0, N]$ for $i = 1, \ldots, N$. These are the locations of the units. Next, define the spatially-dependent error process as

$$u_{it} = \rho u_{i(t-1)} + \sqrt{1-\rho^2}\epsilon_{it}^u,$$

where

$$\epsilon_{it}^u = \sigma_u \sum_{\ell=1}^{N} c_\ell(i) e_{\ell t}, \qquad c_\ell(i) = (1 + |s_\ell - s_i|)^{-0.7}.$$

Here $e_{\ell t}$ are *i.i.d.* standard Gaussian random variables, σ_u is chosen to ensure that ϵ_{it}^u has unit variance, and the weighting function $c_\ell(i)$ generates strong spatial dependence.[5] The x_{it} are generated in the exact same way and then transformed as $x_{it} \mapsto x_{it} + \mu_t$, where μ_t are mutually independent Gaussian random variables with unit mean and variance.

We present results for our JN variance estimator along with the following alternatives:

- [OLS] Conventional variance estimator based on the assumption of a spherical error variance.
- [EHW] Eicker-Huber-White heteroskedasticity robust estimator.

[5] We also considered the weak spatial dependence design of Hidalgo and Schafgans (2021) and the JN variance estimator performed similarly as in the strong spatial dependence case. We omit these results here for brevity but are available upon request.

- [Ci] Cluster-robust variance estimator within i.
- [Ct] Cluster-robust variance estimator within t.
- [DK] The Driscoll and Kraay (1998) variance estimator.
- [CGM] Two-way cluster-robust variance estimator as in Cameron et al. (2011).
- [CHS] CHS variance estimator as in Chiang et al. (2024).
- [CHSbc] bias-corrected CHS variance estimator as in Chen and Vogelsang (2023).
- [DKA] Driscoll-Kray and Arrelano variance estimator as in Chen and Vogelsang (2023).
- [HS] HS variance estimator as in Hidalgo and Schafgans (2021).

All procedures rely on an asymptotic standard Gaussian limiting distribution.[6] All simulations set $N = 50$ and vary T and ρ, and are based on 5,000 replications. The nominal size is set equal to 5%.

We start with the linear CHS design and present results for $\rho \in \{0.2, 0.5, 0.9, 0.95\}$ and $T \in \{25, 75, 125\}$. The empirical size of each procedure is presented in Table 1. The OLS, EHW, and Ci methods lead to severe size distortions, as they are not designed to accommodate time dependence. The Ct, DK, CGM, and CHS methods exhibit somewhat lower over-rejections for larger T and smaller ρ but their empirical size uniformly exceeds the nominal size and grows markedly as the degree of persistence increases. The CHSbc and DKA methods modestly improve size control relative to CHS but are still over-sized throughout, especially for larger ρ. The HS procedure outperforms all the other alternative methods although it can be over-sized for small T and also for larger values of ρ. For example, when $\rho = 0.9$ and $T = 25$, the empirical size is over 10%.

In contrast to all of these approaches, the JN variance estimator has empirical size close to nominal size for all values of T and ρ. In fact, the empirical size resides between 4.5% and 6.2% for all specifications. Importantly, the excellent size control demonstrated by the JN estimator does not come at the expense of lower power. Table 2 presents size-adjusted power for this design.

[6] For more direct comparison and computational ease, we do not explore bootstrap-based procedures here. However, Chiang et al. (2024) include the approaches of Menzel (2021) and MacKinnon et al. (2021) in their simulation results while Hidalgo and Schafgans (2021) also include a bootstrap version of their method. Based on the simulation results presented in these papers, we can conclude that the JN variance estimator controls size better than these resampling-based methods.

Broadly speaking, the power is similar across different approaches so we can conclude that the JN variance estimator has comparable (size-adjusted) power properties. Notably, when ρ moves toward unity, both the HS and JN approaches have higher power relative to the other procedures although the power of the JN method tends to exceed that of the HS method.

In Tables 3 and 4 we present the corresponding empirical size and size-adjusted empirical power for the CHS-NL design. The results are very similar as those presented in Tables 1 and 2 highlighting the robustness of the JN variance estimator to a nonlinear components structure. In Appendix A.2, we report the corresponding results for the CHS and CHS-NL design when individual fixed effects are included in all procedures. This is the more empirically relevant setting. Tables A.1 and A.3 show that although all procedures have worse size control in the presence of fixed effects, the deterioration in the performance of the JN variance estimator is extremely modest. The empirical size remains tightly concentrated around the nominal size, ranging from 3.3% to 6.7%. In terms of power, Tables A.2 and A.4 continue to show that the JN method does not suffer from meaningful power losses relative to procedures which fail to control size.

We next consider the DGP utilized in HS. Because of the presence of time effects, we modify the implementation of the JN variance estimator to accommodate them. In particular, we work with the transformed variables $\dot{y}_{it}$ and $\dot{x}_{it}$, defined as

$$\dot{y}_{it} = y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it}, \qquad \dot{x}_{it} = x_{it} - \frac{1}{N}\sum_{i=1}^{N} x_{it}.$$

We can make such a transformation as the presence of time effects is unaffected by whether the data have been rotated or not. Once we obtain $\dot{y}_{it}$ and $\dot{x}_{it}$, we proceed as in Section 3. In Table 5, we present the empirical size results. As in the previous two designs, the JN variance estimator has excellent size control even when T is small or ρ is high and despite the very strong cross-sectional dependence. For the other methods, when $T = 128$ and $\rho = 0.7$, CHS, CHSbc, DKA, and HS perform reasonably well, but these procedures exhibit more substantial size distortions in all other specifications. The remaining alternatives all fail to control size across all specifications.

Table 6 presents the corresponding size-adjusted power for the HS design. Again, the excellent size control does not come at the expense of lower power. It may be worth noting that OLS and Ct generally have the highest size-adjusted power but this power is practically unattainable since, as shown in Table 5, they have empirical size which is highly distorted (at least an order of magnitude

greater than the nominal size). Compared to the other competing procedures, the JN variance estimator appears to exhibit higher power as ρ increases.

In the next set of simulation experiments, we decouple the degree of persistence of the regressor from the error. We designate the autoregressive parameter used to generate the regressors as ρ_X and the autoregressive parameter used to generate the errors as ρ_U.[7] We fix $\rho_X = 0.9$ and vary the value of ρ_U. In Table 7, we report the ρ_X-fixed case for the CHS design. Importantly, this mismatch between persistence of the regressors and the error does not compromise the ability of the JN variance estimator to control size. In contrast, all of the other procedures continue to have material size distortions. Furthermore, Table 8 shows that the power from using the JN variance estimator continues to be comparable, on a size-adjusted basis, to the other computing procedures. Tables 9 and 10 (and Tables A.5 and A.6 in Appendix A.2) show that the same pattern emerges in the HS and CHS-NL DGPs. For example, in the ρ_X-fixed HS design, the empirical size of the HS procedure is 23.4% when $T = 16$ as compared to 5.4% for the JN estimator. Even the best performing alternative (CHSbc) has an empirical size of 16.9% in this case. Finally, in unreported results, we confirm that the JN variance estimator continues to control size when the Gaussian innovations used in the CHS, CHS-NL and HS designs are replaced by heavy-tailed or asymmetric distributions and in the presence of conditional heteroskedasticity using a GARCH(1,1) model.

Taken in sum, these simulation results highlight the highly desirable finite-sample properties of the JN variance estimator introduced in Section 3.

5 Conclusion

In this paper, we propose a novel jackknife variance estimator for panel-data (and time-series) regressions. The procedure is particularly simple to implement, requiring no additional tuning parameters, and can be characterized as leave-one-out jackknife variance estimation on a rotated data space. The variance estimator performs well for varying degrees of persistence but especially outperforms alternative approaches when the degree of persistence in the data is high. We prove asymptotic validity of our approach and demonstrate excellent finite-sample properties in a series of simulation experiments using designs that have previously been investigated in the literature.

[7] We also consider decoupling the persistence across individuals instead. In particular, for each i, we impose that γ_i^x and γ_i^u is a Gaussian AR(1) with ρ of 0.2 or 0.95, each with equal probability of occurrence. Tables A.7 and A.8 in the Appendix report these results for the CHS DGP which are qualitatively similar to all other simulation results.

We also show that our jackknife approach leads to a broader framework including jackknife bias correction. This naturally leads to consideration of a pairs bootstrap on the rotated data. In Appendix A.3, we show that such an approach provides effective bias adjustment but lays promise for more general improvements in inference which are currently under study by the authors.

Our proposed jackknife variance estimator is also well suited to panel data models with more structure imposed, such as the reduced-rank regression models prevalent in empirical finance (e.g., Adrian, Crump, and Moench, 2015). This setting is studied in the companion paper, Crump, Gospodinov, and Lopez Gaffney (2024a).

References

Abadir, K. M., Distaso, W., Giraitis, L., Koul, H. L., 2014. Asymptotic normality for weighted sums of linear processes. Econometric Theory 30, 252–284.

Adrian, T., Crump, R. K., Moench, E., 2015. Regression-based estimation of dynamic asset pricing models. Journal of Financial Economics 118, 211–244.

Baillie, R. T., Diebold, F. X., Kapetanios, G., Kim, K. H., Mora, A., 2024. On robust inference in time series regression.

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

Campbell, J. Y., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Chen, K., Vogelsang, T. J., 2023. Fixed-b asymptotics for panel models with two-way clustering, working Paper.

Chiang, H. D., Hansen, B. E., Sasaki, Y., 2024. Standard errors for two-way clustering with serially correlated time effects. Review of Economics and Statistics (forthcoming) .

Crump, R. K., Gospodinov, N., 2021. On the factor structure of bond returns. Econometrica 9, 295–314.

Crump, R. K., Gospodinov, N., Lopez Gaffney, I., 2024a. Robust empirical asset pricing, working paper.

Crump, R. K., Gospodinov, N., Lopez Gaffney, I., 2024b. A simple diagnostic for time-series and panel-data regressions. Staff Report 1132, Federal Reserve Bank of New York.

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

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

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

Hansen, B. E., 2023. Jackknife standard errors for clustered regression, working paper.

Hansen, B. E., 2024. Standard errors for difference-in-difference regression. Journal of Applied Econometrics (forthcoming).

Hidalgo, J., Schafgans, M., 2021. Inference without smoothing for large panels with cross-sectional and temporal dependence. Journal of Econometrics 223, 125–160.

Kendall, M. G., 1954. Note on bias in the estimation of autocorrelation. Biometrika 41, 403–404.

Kirch, C., Politis, D. N., 2011. TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. Annals of Statistics 39, 1427–1470.

Lazarus, E., Lewis, D. J., Stock, J. H., Watson, M. W., 2018. HAR inference: Recommendations for practice. Journal of Business & Economic Statistics 36, 541–559.

MacKinnon, J. G., Nielsen, M. Ø., Webb, M. D., 2021. Wild bootstrap and asymptotic inference with multiway clustering. Journal of Business & Economic Statistics 39, 505–519.

MacKinnon, J. G., Nielsen, M. Ø., Webb, M. D., 2023a. Cluster-robust inference: A guide to empirical practice. Journal of Econometrics 232, 272–299.

MacKinnon, J. G., Nielsen, M. Ø., Webb, M. D., 2023b. Fast and reliable jackknife and bootstrap methods for cluster-robust inference. Journal of Applied Econometrics 38, 671–694.

MacKinnon, J. G., Nielsen, M. Ø., Webb, M. D., 2023c. Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust. Stata Journal 23, 942–982.

MacKinnon, J. G., White, H., 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics 29, 305–325.

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

Müller, U., Watson, M., 2024. Spatial unit roots and spurious regression. Econometrica 92, 1661–1695.

Shao, J., Tu, D., 1996. The Jackknife and Bootstrap. Springer, New York, NY.

Shevelev, V., Moses, P. J. C., 2014. Tangent power sums and their applications. Integers 14, 1–15.

Stambaugh, R. F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Thompson, S. B., 2011. Simple formulas for standard errors that cluster by both firm and time. Journal of Financial Economics 99, 1–10.

Vogelsang, T. J., 2012. Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects. Journal of Econometrics 166, 303–319.

Table 1. Empirical Size (CHS). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.704	0.71	0.694	0.115	0.127	0.108	0.123	0.114	0.104	0.071	0.056
0.2	75	0.716	0.718	0.58	0.091	0.089	0.08	0.077	0.073	0.069	0.063	0.062
0.2	125	0.713	0.716	0.49	0.087	0.085	0.071	0.071	0.068	0.063	0.063	0.06
0.5	25	0.736	0.742	0.713	0.185	0.184	0.18	0.175	0.152	0.141	0.082	0.053
0.5	75	0.759	0.763	0.626	0.162	0.124	0.147	0.112	0.099	0.094	0.069	0.057
0.5	125	0.759	0.759	0.557	0.161	0.113	0.137	0.094	0.085	0.082	0.065	0.058
0.9	25	0.773	0.775	0.648	0.453	0.401	0.397	0.346	0.282	0.274	0.118	0.046
0.9	75	0.872	0.874	0.739	0.53	0.298	0.483	0.274	0.232	0.23	0.112	0.052
0.9	125	0.873	0.873	0.723	0.531	0.244	0.485	0.224	0.19	0.189	0.095	0.059
0.95	25	0.728	0.731	0.526	0.503	0.452	0.374	0.332	0.248	0.247	0.138	0.046
0.95	75	0.871	0.873	0.7	0.626	0.375	0.535	0.332	0.266	0.268	0.121	0.045
0.95	125	0.892	0.892	0.734	0.645	0.325	0.58	0.297	0.248	0.249	0.109	0.058

Table 2. Size-Adjusted Empirical Power (CHS). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.461	0.447	0.417	0.4	0.396	0.406	0.4	0.402	0.407	0.454	0.407
0.2	75	0.899	0.898	0.842	0.878	0.868	0.886	0.871	0.873	0.874	0.895	0.884
0.2	125	0.983	0.982	0.962	0.978	0.976	0.979	0.978	0.978	0.978	0.979	0.977
0.5	25	0.345	0.326	0.286	0.296	0.299	0.304	0.3	0.298	0.299	0.383	0.345
0.5	75	0.745	0.749	0.668	0.737	0.691	0.741	0.697	0.697	0.706	0.747	0.715
0.5	125	0.924	0.923	0.863	0.912	0.891	0.914	0.899	0.899	0.901	0.907	0.896
0.9	25	0.2	0.181	0.145	0.228	0.206	0.215	0.203	0.226	0.219	0.327	0.319
0.9	75	0.246	0.242	0.197	0.253	0.258	0.25	0.27	0.264	0.267	0.33	0.305
0.9	125	0.34	0.326	0.255	0.32	0.305	0.326	0.317	0.312	0.312	0.387	0.373
0.95	25	0.266	0.237	0.173	0.346	0.281	0.287	0.289	0.326	0.309	0.426	0.434
0.95	75	0.208	0.2	0.158	0.243	0.232	0.204	0.234	0.234	0.23	0.32	0.35
0.95	125	0.232	0.216	0.162	0.229	0.243	0.214	0.236	0.234	0.236	0.329	0.326

Table 3. Empirical Size (CHS-NL). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS-NL DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.702	0.708	0.695	0.117	0.129	0.11	0.124	0.116	0.105	0.071	0.056
0.2	75	0.715	0.719	0.578	0.093	0.091	0.081	0.079	0.077	0.069	0.062	0.064
0.2	125	0.712	0.716	0.492	0.088	0.087	0.071	0.071	0.068	0.064	0.063	0.061
0.5	25	0.737	0.742	0.71	0.187	0.186	0.18	0.175	0.154	0.142	0.083	0.055
0.5	75	0.758	0.763	0.626	0.161	0.124	0.148	0.111	0.101	0.094	0.068	0.057
0.5	125	0.758	0.761	0.557	0.161	0.113	0.138	0.093	0.086	0.081	0.064	0.059
0.9	25	0.772	0.774	0.646	0.454	0.403	0.398	0.346	0.282	0.274	0.117	0.046
0.9	75	0.873	0.875	0.737	0.531	0.301	0.483	0.275	0.231	0.231	0.112	0.052
0.9	125	0.873	0.875	0.723	0.531	0.247	0.485	0.225	0.193	0.191	0.096	0.059
0.95	25	0.729	0.729	0.524	0.501	0.453	0.373	0.331	0.247	0.248	0.138	0.047
0.95	75	0.87	0.871	0.7	0.625	0.376	0.533	0.334	0.267	0.27	0.121	0.045
0.95	125	0.892	0.895	0.731	0.646	0.328	0.579	0.298	0.249	0.251	0.108	0.058

Table 4. Size-Adjusted Empirical Power (CHS-NL). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS-NL DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.468	0.46	0.43	0.408	0.401	0.416	0.409	0.408	0.415	0.457	0.411
0.2	75	0.9	0.895	0.844	0.879	0.864	0.884	0.87	0.87	0.872	0.892	0.883
0.2	125	0.983	0.982	0.961	0.976	0.976	0.978	0.976	0.976	0.977	0.98	0.975
0.5	25	0.351	0.331	0.299	0.304	0.305	0.316	0.304	0.306	0.31	0.386	0.352
0.5	75	0.747	0.751	0.67	0.734	0.693	0.746	0.708	0.704	0.706	0.751	0.721
0.5	125	0.924	0.924	0.865	0.909	0.884	0.915	0.9	0.899	0.9	0.909	0.901
0.9	25	0.207	0.191	0.149	0.235	0.203	0.221	0.217	0.234	0.229	0.335	0.322
0.9	75	0.252	0.255	0.21	0.265	0.261	0.262	0.269	0.265	0.266	0.336	0.308
0.9	125	0.346	0.342	0.264	0.326	0.314	0.337	0.325	0.319	0.321	0.395	0.377
0.95	25	0.27	0.252	0.184	0.351	0.27	0.293	0.294	0.32	0.307	0.429	0.436
0.95	75	0.217	0.216	0.171	0.251	0.242	0.214	0.244	0.239	0.238	0.326	0.357
0.95	125	0.245	0.228	0.17	0.233	0.252	0.231	0.244	0.237	0.239	0.335	0.327

Table 5. Empirical Size (HS). This table presents empirical size for the t-test of the null hypothesis that $\beta = 0$ for the HS DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.7	16	0.67	0.667	0.56	0.276	0.376	0.15	0.242	0.145	0.209	0.181	0.056
0.7	32	0.689	0.687	0.566	0.288	0.253	0.144	0.158	0.111	0.159	0.132	0.059
0.7	64	0.698	0.696	0.556	0.263	0.165	0.119	0.088	0.07	0.11	0.087	0.051
0.7	128	0.685	0.683	0.544	0.256	0.115	0.103	0.055	0.046	0.078	0.063	0.043
0.9	16	0.738	0.732	0.556	0.44	0.606	0.259	0.343	0.164	0.329	0.295	0.047
0.9	32	0.786	0.783	0.566	0.508	0.503	0.289	0.296	0.162	0.279	0.246	0.053
0.9	64	0.81	0.81	0.558	0.528	0.345	0.301	0.215	0.143	0.195	0.176	0.055
0.9	128	0.815	0.813	0.542	0.52	0.214	0.301	0.143	0.107	0.133	0.111	0.044

Table 6. Size-Adjusted Empirical Power (HS). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta = 0.3$ for the HS DGP. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.7	16	0.362	0.34	0.323	0.298	0.187	0.309	0.258	0.265	0.266	0.28	0.294
0.7	32	0.548	0.54	0.501	0.535	0.436	0.508	0.48	0.479	0.483	0.481	0.483
0.7	64	0.881	0.876	0.808	0.856	0.762	0.843	0.813	0.811	0.791	0.82	0.816
0.7	128	0.994	0.994	0.982	0.993	0.972	0.992	0.985	0.985	0.979	0.984	0.978
0.9	16	0.212	0.198	0.187	0.172	0.101	0.176	0.18	0.166	0.165	0.158	0.202
0.9	32	0.267	0.252	0.225	0.254	0.132	0.243	0.231	0.229	0.206	0.209	0.254
0.9	64	0.424	0.427	0.372	0.401	0.28	0.409	0.373	0.379	0.354	0.369	0.405
0.9	128	0.688	0.673	0.621	0.66	0.548	0.685	0.632	0.629	0.609	0.607	0.645

Table 7. Empirical Size (CHS; ρ_x Fixed). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ_U	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.791	0.797	0.883	0.151	0.157	0.153	0.161	0.149	0.145	0.141	0.068
0.2	75	0.788	0.789	0.892	0.115	0.101	0.118	0.103	0.101	0.097	0.106	0.062
0.2	125	0.786	0.79	0.896	0.114	0.091	0.117	0.094	0.092	0.088	0.091	0.066
0.5	25	0.829	0.835	0.9	0.267	0.243	0.27	0.245	0.224	0.219	0.16	0.075
0.5	75	0.836	0.84	0.917	0.244	0.155	0.248	0.158	0.146	0.142	0.119	0.068
0.5	125	0.838	0.84	0.922	0.24	0.128	0.244	0.131	0.124	0.12	0.102	0.069
0.9	25	0.829	0.833	0.858	0.454	0.377	0.452	0.372	0.326	0.321	0.183	0.046
0.9	75	0.895	0.896	0.923	0.536	0.289	0.538	0.289	0.268	0.266	0.146	0.067
0.9	125	0.898	0.901	0.934	0.528	0.236	0.532	0.237	0.22	0.218	0.116	0.066
0.95	25	0.784	0.787	0.8	0.452	0.371	0.446	0.361	0.318	0.309	0.159	0.032
0.95	75	0.886	0.888	0.91	0.578	0.313	0.578	0.312	0.286	0.282	0.122	0.039
0.95	125	0.909	0.91	0.928	0.58	0.263	0.581	0.263	0.242	0.241	0.105	0.048

Table 8. Size-Adjusted Empirical Power (CHS; ρ_x Fixed). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ_U	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.424	0.409	0.425	0.374	0.358	0.373	0.357	0.36	0.363	0.292	0.281
0.2	75	0.953	0.951	0.944	0.949	0.932	0.948	0.931	0.93	0.93	0.888	0.848
0.2	125	0.999	0.998	0.996	0.998	0.996	0.998	0.996	0.996	0.996	0.993	0.982
0.5	25	0.251	0.244	0.25	0.224	0.213	0.223	0.212	0.208	0.213	0.18	0.174
0.5	75	0.738	0.718	0.724	0.699	0.638	0.699	0.638	0.634	0.636	0.623	0.572
0.5	125	0.938	0.933	0.916	0.928	0.906	0.928	0.905	0.905	0.906	0.875	0.833
0.9	25	0.129	0.118	0.116	0.105	0.121	0.105	0.12	0.12	0.119	0.173	0.201
0.9	75	0.163	0.161	0.156	0.158	0.16	0.158	0.161	0.159	0.16	0.181	0.203
0.9	125	0.244	0.225	0.218	0.212	0.196	0.21	0.196	0.194	0.194	0.24	0.25
0.95	25	0.159	0.138	0.123	0.168	0.192	0.165	0.193	0.204	0.2	0.282	0.362
0.95	75	0.163	0.146	0.136	0.151	0.18	0.15	0.178	0.177	0.177	0.233	0.29
0.95	125	0.19	0.175	0.159	0.172	0.171	0.173	0.171	0.17	0.171	0.242	0.265

Table 9. Empirical Size (HS; ρ_x Fixed). This table presents empirical size for the t-test of the null hypothesis that $\beta = 0$ for the CHS DGP. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ_U	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.7	16	0.69	0.688	0.548	0.338	0.415	0.201	0.263	0.169	0.237	0.234	0.054
0.7	32	0.728	0.729	0.559	0.363	0.293	0.204	0.191	0.142	0.185	0.192	0.061
0.7	64	0.744	0.741	0.562	0.364	0.201	0.183	0.117	0.091	0.132	0.136	0.059
0.7	128	0.75	0.746	0.546	0.347	0.138	0.17	0.078	0.064	0.098	0.091	0.044
0.9	16	0.738	0.732	0.556	0.44	0.606	0.259	0.343	0.164	0.329	0.295	0.047
0.9	32	0.786	0.783	0.566	0.508	0.503	0.289	0.296	0.162	0.279	0.246	0.053
0.9	64	0.81	0.81	0.558	0.528	0.345	0.301	0.215	0.143	0.195	0.176	0.055
0.9	128	0.815	0.813	0.542	0.52	0.214	0.301	0.143	0.107	0.133	0.111	0.044

Table 10. Size-Adjusted Empirical Power (HS; ρ_x Fixed). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta = 0.3$ for the HS DGP. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ_U	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.7	16	0.153	0.141	0.14	0.14	0.085	0.154	0.13	0.127	0.122	0.116	0.136
0.7	32	0.308	0.308	0.269	0.295	0.223	0.282	0.253	0.252	0.256	0.227	0.218
0.7	64	0.618	0.61	0.54	0.591	0.496	0.581	0.538	0.538	0.546	0.503	0.485
0.7	128	0.921	0.918	0.857	0.913	0.863	0.904	0.889	0.887	0.874	0.859	0.856
0.9	16	0.212	0.198	0.187	0.172	0.101	0.176	0.18	0.166	0.165	0.158	0.202
0.9	32	0.267	0.252	0.225	0.254	0.132	0.243	0.231	0.229	0.206	0.209	0.254
0.9	64	0.424	0.427	0.372	0.401	0.28	0.409	0.373	0.379	0.354	0.369	0.405
0.9	128	0.688	0.673	0.621	0.66	0.548	0.685	0.632	0.629	0.609	0.607	0.645

Appendix

A.1 Proofs

Before proceeding to the main proof, we introduce the following lemmas which we rely on repeatedly in the proof of Theorem 1.

Lemma A.1. *Let* $\zeta_j = \psi_j' \iota_T$. *Then,*

(ii) $\zeta_j > 0$;

(ii) $\frac{\zeta_j}{\sqrt{T}} \to \bar{\zeta}_j = \frac{2\sqrt{2}}{\pi(2j-1)}$;

(iii) $\sum_{j=1}^{T} \zeta_j = O\left(\log(T)\sqrt{T}\right)$;

(iv) $\sum_{j=1}^{T} \zeta_j^m = O\left(T^{m/2}\right)$ *for* $m \in \mathbb{N}$ *with* $m \geq 2$.

Proof of Lemma A.1. For (i), we have that

$$\begin{aligned}
\zeta_j &= \frac{2}{\sqrt{2T+1}} \sum_{t=1}^{T} \sin\left(\frac{\pi t(2j-1)}{2T+1}\right) \\
&= \frac{-1}{\sqrt{2T+1}} \csc\left(\frac{\pi(2j-1)}{4T+2}\right) \sin\left(\frac{\pi(j-1-T)}{2T+1}\right) \\
&= \frac{-1}{\sqrt{2T+1}} \tan\left(\frac{\pi(j+T)}{2T+1}\right).
\end{aligned}$$

As $j \in \{1, ..., T\}$, we know $\frac{\pi(j+T)}{2T+1} \in (\frac{\pi}{2}, \pi)$ which in turn implies $\tan\left(\frac{\pi(j+T)}{2T+1}\right) < 0$ as $\tan(x) < 0$ $\forall\ x \in (\frac{\pi}{2} + \pi a, \pi + \pi a)$ and all $a \in \mathbb{Z}$, the result follows.

For (ii), by a series expansion at $T = \infty$ of ζ_j we have that

$$\zeta_j = \frac{2\sqrt{2}\sqrt{T}}{\pi(2j-1)} + O(T^{-1/2})$$

and the result follows. For (iii), note that from the above we have that

$$\zeta_j = \frac{-1}{\sqrt{2T+1}} \tan\left(\frac{\pi j}{2T+1} + \frac{\pi T}{2T+1}\right)$$

so that

$$\begin{aligned}\sum_j \zeta_j &= \left(\frac{-1}{2\sqrt{2T+1}}\sum_{j=1}^{T}\tan\left(\frac{\pi j}{2T+1}+\frac{\pi T}{2T+1}\right)\right)\\ &= \left(\frac{-T}{2\sqrt{2T+1}}\frac{1}{T}\sum_{j=1}^{T}\tan\left(\frac{\pi Tj/T}{2T+1}+\frac{\pi T}{2T+1}\right)\right)\\ &\leq \left(\frac{-T}{2\sqrt{2T+1}}\int_{1/T}^{1}\tan\left(\frac{\pi T(j+1)}{2T+1}\right)dj\right)[1+O(T^{-1})].\end{aligned}$$

Evaluating the integral, we obtain

$$\begin{aligned}\int_{1/T}^{1}\tan\left(\frac{\pi T(j+1)}{2T+1}\right)dj &= -\frac{2T+1}{\pi T}\ln\left(\cos\left(\frac{\pi T(j+1)}{2T+1}\right)\right)|_{1/T}^{1}\\ &= -\frac{2T+1}{\pi T}\left[\ln\left(\cos\left(\frac{2T\pi}{2T+1}\right)\right)-\ln\left(\cos\left(\frac{T\pi(1+T^{-1})}{2T+1}\right)\right)\right]\\ &= -\frac{2T+1}{\pi T}\ln\left(\frac{\cos\left(\frac{2T\pi}{2T+1}\right)}{\cos\left(\frac{T\pi(1+T^{-1})}{2T+1}\right)}\right).\end{aligned}$$

From a series expansion at $T=\infty$, we then have that

$$\begin{aligned}\int_{1/T}^{1}\tan\left(\frac{\pi T(j+1)}{2T+1}\right)dj &= -\frac{2T+1}{\pi T}\ln\left(\frac{4}{\pi}T+\frac{2}{\pi}-O(T^{-1})\right)\\ &= -\frac{2T+1}{\pi T}O(\ln(T))\\ &= -O(\ln(T))\end{aligned}$$

and the result follows. Finally, for (iv), we have that

$$\begin{aligned}\sum_{j=1}^{T}\zeta_i^{2p} &= \sum_{j=1}^{T}\left(\frac{1}{\sqrt{2T+1}}\sum_{t=1}^{T}\sin\left(\frac{\pi t(2j-1)}{2T+1}\right)\right)^{2p}\\ &\lesssim T^{-p}\sum_{j=1}^{T}\left[\frac{-1}{2}\csc\left(\frac{\pi(2j-1)}{4T+2}\right)\left(\sin(\pi j)+\sin\left(\frac{\pi(j-1-T)}{2T+1}\right)\right)\right]^{2p}.\end{aligned}$$

Since $0<\csc^{2p}(\frac{\pi(2j-1)}{4T+2})<\infty\ \forall\ j\in\{1,..,T\}$ and $\sin(\pi j)=0\ \forall\ j\in\mathbb{Z}$,

$$\sum_{j=1}^{T}\zeta_i^{2p}\lesssim T^{-p}\sum_{j=1}^{T}\left[\frac{-1}{2}\csc\left(\frac{\pi(2j-1)}{4T+2}\right)\sin\left(\frac{\pi(j-1-T)}{2T+1}\right)\right]^{2p}$$

$$= T^{-p} \sum_{j=1}^{T} \tan^{2p}\left(\frac{\pi(j+T)}{2T+1}\right)$$
$$= T^{-p} \sum_{j=1}^{T} \tan^{2p}\left(\frac{\pi j}{2T+1}\right).$$

By Theorem 2 of Shevelev and Moses (2014), we have

$$\sum_{j=1}^{T} \tan^{2p}\left(\frac{\pi j}{2T+1}\right) = \frac{2^{2p-1}(2^{2p-1}-1)}{(2p)!}|B_{2p}|T^{2p},$$

where B_{2p} is a Bernoulli number and so the result holds for even m. For odd values of m, we can use the Cauchy–Schwarz inequality and the result for even m to obtain the general result. □

Lemma A.2. *Let Assumptions 1 and 2 hold. Then,* $\sqrt{\frac{T}{N}}\|\widehat{\mu}_1 - \mu_1\| = O_p(1)$ *and* $\sqrt{\frac{T}{N}}\|\widehat{\mu}_2 - \mu_2\| = O_p(1)$. *Let* $\widehat{\mu}_e$ *be the* $N \times 1$ *vector with* ℓ*th element equal to* $\widehat{\mu}_e$ *is* $\frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$, *then* $\sqrt{\frac{T}{N}}\|\widehat{\mu}_e\| = O_p(1)$.

Proof of Lemma A.2. For the first result, note that

$$\begin{aligned}
\frac{T}{N}\mathbb{E}\left[\|\widehat{\mu}_1 - \mu_1\|^2\right] &= \frac{T}{N}\sum_{i=1}^{N}\mathbb{E}\left[(\widehat{\mu}_{1,i} - \mu_{1,i})^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_{1,it} - \mathbb{E}[x_{1,it}])\right)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{1,it}\right),
\end{aligned}$$

so that if $\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{it}\right) \lesssim 1$, then by Markov's inequality we have that $\sqrt{\frac{T}{N}}\|\widehat{\mu}_1 - \mu_1\| = O_p(1)$. The other results follow by similar steps. □

We can now proceed to the proof of Theorem 1.

Proof of Theorem 1. By equation (22) and Assumption 1, we need only to show that

$$(NT)^{-1}\sum_{j=1}^{T}\widehat{V}^{j}\widehat{M}_j^{-1}\widehat{u}^{j}\widehat{u}^{j\prime}\widehat{M}_j^{-1}\widehat{V}^{j\prime} = \gamma_1 + o_p(1), \tag{A.1}$$

where $\widehat{M}_j^{-1} = (\mathsf{I}_N - \mathsf{P}_{jj})^{-1}$. First, note that using the results Lemma A.1 it can be shown that

$$\left|(NT)^{-1}\sum_{j=1}^{T}\widehat{V}^j\widehat{M}_j^{-1}\widehat{u}^j\widehat{u}^{j\prime}\widehat{M}_j^{-1}\widehat{V}^{j\prime} - (NT)^{-1}\sum_{j=1}^{T}\widehat{V}^j\widehat{u}^j\widehat{u}^{j\prime}\widehat{V}^{j\prime}\right| = o_p(1), \tag{A.2}$$

since for sufficiently large j, $\left|\widehat{M}_j^{-1} - \mathsf{I}_n\right| = o_p(1)$. Then, we have that

$$\left|(NT)^{-1}\sum_{j=1}^{T}\widehat{V}^j\widehat{u}^j\widehat{u}^{j\prime}\widehat{V}^{j\prime} - \gamma_1\right| \leq \mathfrak{R}_1 + \mathfrak{R}_2 + \mathfrak{R}_3, \tag{A.3}$$

where

$$\mathfrak{R}_1 = \left|(NT)^{-1}\sum_{j=1}^{T}\widehat{V}^j\widehat{u}^j\widehat{u}^{j\prime}\widehat{V}^{j\prime} - (NT)^{-1}\sum_{j=1}^{T}\widetilde{V}^j u^j u^{j\prime}\widetilde{V}^{j\prime}\right|, \tag{A.4}$$

$$\mathfrak{R}_2 = \left|(NT)^{-1}\sum_{j=1}^{T}\widetilde{V}^j u^j u^{j\prime}\widetilde{V}^{j\prime} - (NT)^{-1}\sum_{j=1}^{T}\mathbb{E}\left[\widetilde{V}^j u^j u^{j\prime}\widetilde{V}^{j\prime}\right]\right|, \tag{A.5}$$

$$\mathfrak{R}_3 = \left|(NT)^{-1}\sum_{j=1}^{T}\mathbb{E}\left[\widetilde{V}^j u^j u^{j\prime}\widetilde{V}^{j\prime}\right] - \gamma_1\right|. \tag{A.6}$$

Thus, it is sufficient to show that $\mathfrak{R}_i = o_p(1)$ for $i \in \{1,2,3\}$. The results for $\mathfrak{R}_2$ and $\mathfrak{R}_3$ follow by Assumption 2. To show $\mathfrak{R}_1$, first note that $\mathfrak{R}_1 = \sum_{\ell=1}^{9}\mathfrak{R}_{1\ell}$, where

$$\mathfrak{R}_{11} = \frac{1}{NT}\sum_{j=1}^{T}\delta_{V,j}'\delta_{u,j}\delta_{u,j}'\delta_{V,j},$$

$$\mathfrak{R}_{12} = \frac{2}{NT}\sum_{j=1}^{T}\delta_{V,j}'\delta_{u,j}u_j'\delta_{V,j},$$

$$\mathfrak{R}_{13} = \frac{2}{NT}\sum_{j=1}^{T}\delta_{V,j}'\delta_{u,j}\delta_{u,j}'\widetilde{V}^j,$$

$$\mathfrak{R}_{14} = \frac{2}{NT}\sum_{j=1}^{T}\delta_{V,j}'\delta_{u,j}u_j'\widetilde{V}^j,$$

$$\mathfrak{R}_{15} = \frac{1}{NT}\sum_{j=1}^{T}\delta_{V,j}'u_j u_j'\delta_{V,j},$$

$$\mathfrak{R}_{16} = \frac{2}{NT}\sum_{j=1}^{T}\delta_{V,j}'u_j\delta_{u,j}'\widetilde{V}^j,$$

$$\mathfrak{R}_{17} = \frac{2}{NT}\sum_{j=1}^{T}\delta_{V,j}'u_j u_j'\widetilde{V}^j,$$

$$\mathfrak{R}_{18} = \frac{1}{NT}\sum_{j=1}^{T} \widetilde{V}^{j\prime}\delta_{u,j}\delta_{u,j}'\widetilde{V}^{j},$$
$$\mathfrak{R}_{19} = \frac{2}{NT}\sum_{j=1}^{T} \widetilde{V}^{j\prime}\delta_{u,j}u_j'\widetilde{V}^{j}.$$

Here,

$$\begin{aligned}
\delta_{u,j} &= \widetilde{Z}_1^j\left(\beta_1 - \widehat{\beta}_1\right) + \widetilde{Z}_2^j\left(\beta_2 - \widehat{\beta}_2\right) + \zeta_j\Gamma_u,\\
\Gamma_u &= (\widehat{\mu}_1 - \mu_1)\left(\widehat{\beta}_1 - \beta_1\right) + (\widehat{\mu}_2 - \mu_2)\left(\widehat{\beta}_2 - \beta_2\right) - \hat{\mu}_e,\\
\delta_{V,j} &= \zeta_j\Gamma_V + \widetilde{Z}_2^j(\lambda - \hat{\lambda}),\\
\Gamma_V &= (\mu_1 - \widehat{\mu}_1) + (\mu_2 - \widehat{\mu}_2)(\widehat{\lambda} - \lambda) + (\mu_2 - \widehat{\mu}_2)\lambda,
\end{aligned}$$

where $\widetilde{Z}_1^j = Z_1^j - \zeta_j\mu_1$ and $Z_1^j = (z_{1,1j},\ldots,z_{1,Nj})'$ (and similarly for $\widetilde{Z}_2^j$). We can show that $\mathfrak{R}_{11}$, $\ldots$, $\mathfrak{R}_{19}$ are $o_p(1)$ repeatedly using elementary bounds and Assumption 2. To see this, consider $\mathfrak{R}_{11}$. We have

$$\begin{aligned}
\mathfrak{R}_{11} &= \frac{1}{NT}\sum_{j=1}^{T}\delta_{V,j}'\delta_{u,j}\delta_{u,j}'\delta_{V,j}\\
&\lesssim \frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j\Gamma_V'\widetilde{Z}_1^j\left(\beta_1 - \widehat{\beta}_1\right)\right|^2 + \frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j\Gamma_V'\widetilde{Z}_2^j\left(\beta_2 - \widehat{\beta}_2\right)\right|\\
&+ \frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j^2\Gamma_V'\Gamma_u\right|^2 + \frac{1}{NT}\sum_{j=1}^{T}\left|\left(\lambda - \widehat{\lambda}\right)'\widetilde{Z}_2^{j\prime}\widetilde{Z}_1^j\left(\beta_1 - \widehat{\beta}_1\right)\right|^2\\
&+ \frac{1}{NT}\sum_{j=1}^{T}\left|\left(\lambda - \widehat{\lambda}\right)'\widetilde{Z}_2^{j\prime}\widetilde{Z}_2^j\left(\beta_2 - \widehat{\beta}_2\right)\right|^2 + \frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j\left(\lambda - \widehat{\lambda}\right)'\widetilde{Z}_2^{j\prime}\Gamma_u\right|^2.
\end{aligned}$$

The first term satisfies,

$$\frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j\Gamma_V'\widetilde{Z}_1^j\left(\beta_1 - \widehat{\beta}_1\right)\right|^2 \leq NT\left|\beta_1 - \widehat{\beta}_1\right|^2 \times \frac{T}{N}\left\|\Gamma_V\right\|^2 \times \frac{1}{NT^3}\sum_{j=1}^{T}\zeta_j^2\left\|\widetilde{Z}_1^j\right\|^2.$$

The first two factors are $O_p(1)$ by Assumption 1 and Lemma A.2 and the last factor is $O_p(T^{-2})$ using Lemma A.1. Thus, the first term is $o_p(1)$. The second term follows by similar steps. The

third term is

$$\frac{1}{NT}\sum_{j=1}^{T}\left|\zeta_j^2\Gamma_V'\Gamma_u\right|^2 \leq \frac{T}{N}\|\Gamma_V\|^2 \times \frac{T}{N}\|\Gamma_u\|^2 \times \frac{N}{T^3}\sum_{j=1}^{T}\zeta_j^4.$$

The first two factors are $O_p(1)$ by Assumption 1 and Lemma A.2 and the last factor is $O_p(NT^{-1})$ using Lemma A.1. Thus, the third term is $o_p(1)$. The fourth term is,

$$\frac{1}{NT}\sum_{j=1}^{T}\left|\left(\lambda-\widehat{\lambda}\right)'\widetilde{Z}_2^{j\prime}\widetilde{Z}_1^{j}\left(\beta_1-\widehat{\beta}_1\right)\right|^2 \leq NT\|\lambda-\widehat{\lambda}\|^2 \times NT|\beta_1-\widehat{\beta}_1|^2 \times \frac{1}{N^3T^3}\sum_{j=1}^{T}\|\widetilde{Z}_2^{j\prime}\widetilde{Z}_1^{j}\|^2.$$

The first two factors are $O_p(1)$ by Assumption 1 and Lemma A.2 and the last factor is $O_p(N^{-1}T^{-2})$. Thus, the fourth term is $o_p(1)$. The fifth term follows by similar steps. Finally, the sixth term follows by similar steps as for the first and second terms. The bounds for the terms $\mathfrak{R}_{12}, \ldots, \mathfrak{R}_{19}$ can then be obtained by similar steps.

□

Proof of Theorem 2. In order to show the result we need only show that when Assumption 2(v) fails, the middle of the sandwich of the variance estimator has an additional term which is asymptotically non-negative. Without Assumption 2(v), we have the additional term

$$\begin{aligned}
&(NT)^{-1}\sum_{j=1}^{T}\mathbb{E}\Big[\big(\widetilde{V}^{j\prime}u^j\big)^2\Big] - (NT)^{-1}\mathbb{E}\Big[\big(\sum_{j=1}^{T}\widetilde{V}^{j\prime}u^j\big)^2\Big] \\
&\quad= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\mathbb{E}\big[\widetilde{v}_{ij}u_{ij}\widetilde{v}_{\ell s}u_{\ell s}\big] \\
&\quad= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\mathbb{E}\big[\widetilde{v}_{ij}\widetilde{v}_{\ell s}\big]\mathbb{E}\big[u_{ij}u_{\ell s}\big] + \mathbb{E}\big[\widetilde{v}_{ij}u_{ij}\big]\mathbb{E}\big[\widetilde{v}_{\ell s}u_{\ell s}\big] + \mathbb{E}\big[\widetilde{v}_{ij}u_{\ell s}\big]\mathbb{E}\big[\widetilde{v}_{\ell s}u_{ij}\big] + o_p(1) \\
&\quad= \mathfrak{R}_A + \mathfrak{R}_B + \mathfrak{R}_C + o_p(1),
\end{aligned}$$

where the penultimate line follows by Assumption 3(iii). We start with $\mathfrak{R}_B$.

$$\begin{aligned}
\mathfrak{R}_B &= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\mathbb{E}\big[\widetilde{v}_{ij}u_{ij}\big]\mathbb{E}\big[\widetilde{v}_{\ell s}u_{\ell s}\big] \\
&= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\psi_j'\Gamma_{ii}\psi_j\psi_s'\Gamma_{\ell\ell}\psi_s
\end{aligned}$$

$$= \frac{N}{T}\sum_{j=1}^{T}\left(\psi_j'\bar{\Gamma}\psi_j\right)^2 - \frac{N}{T}\left[\mathrm{tr}\left(\bar{\Gamma}\right)^2\right]$$
$$= \frac{N}{T}\sum_{j=1}^{T}\left(\psi_j'\bar{\Gamma}\psi_j\right)^2 + o_p(1),$$

where the last line follows by Assumption 3(i). Next, consider $\mathfrak{R}_C$.

$$\begin{aligned}
\mathfrak{R}_C &= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\mathbb{E}\left[\widetilde{v}_{ij}u_{\ell s}\right]\mathbb{E}\left[\widetilde{v}_{\ell s}u_{ij}\right] \\
&= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\psi_j'\Gamma_{i\ell}\psi_s\psi_s'\Gamma_{\ell i}\psi_j \\
&= -\frac{1}{NT}\sum_{j,s=1}^{T}\sum_{i,\ell=1}^{N}\psi_j'\Gamma_{i\ell}\psi_s\psi_s'\Gamma_{\ell i}\psi_j + \frac{1}{NT}\sum_{j=1}^{T}\sum_{i,\ell=1}^{N}\psi_j'\Gamma_{i\ell}\psi_j\psi_j'\Gamma_{\ell i}\psi_j \\
&= \mathfrak{R}_{C,1} + \mathfrak{R}_{C,2}.
\end{aligned}$$

First, we have that

$$\mathfrak{R}_{C,1} = -\frac{1}{NT}\sum_{i,\ell=1}^{N}\sum_{j=1}^{T}\psi_j'\Gamma_{i\ell}\Gamma_{\ell i}\psi_j = -\frac{1}{NT}\sum_{i,\ell=1}^{N}\mathrm{tr}\left(\Gamma_{i\ell}\Gamma_{\ell i}\right) = o_p(1),$$

by Assumption 3(ii) and next we have that

$$\mathfrak{R}_{C,2} = \frac{1}{NT}\sum_{j=1}^{T}\sum_{i,\ell=1}^{N}\psi_j'\Gamma_{i\ell}\psi_j\psi_j'\Gamma_{\ell i}\psi_j = \frac{1}{NT}\sum_{j=1}^{T}\mathrm{tr}\left(\mathbf{\Psi}_j'\Gamma_{NT}\mathbf{\Psi}_j\mathbf{\Psi}_j'\Gamma_{NT}'\mathbf{\Psi}_j\right)$$

where $\mathbf{\Psi}_j = (\mathsf{I}_N \otimes \psi_j)$ so that $\mathfrak{R}_C = \frac{1}{NT}\sum_{j=1}^{T}\left\|\mathbf{\Psi}_j'\Gamma_{NT}\mathbf{\Psi}_j\right\|^2 + o_p(1)$. Finally, consider $\mathfrak{R}_A$,

$$\begin{aligned}
\mathfrak{R}_A &= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\mathbb{E}\left[\widetilde{v}_{ij}\widetilde{v}_{\ell s}\right]\mathbb{E}\left[u_{ij}u_{\ell s}\right] \\
&= -\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\psi_j'\Sigma_{i\ell}^{x}\psi_s\psi_j'\Sigma_{i\ell}^{\varepsilon}\psi_s \\
&= -\frac{1}{NT}\sum_{j\neq s}\mathrm{tr}\left(\mathbf{\Psi}_j'\Sigma_{NT}^{x}\mathbf{\Psi}_s\mathbf{\Psi}_j'\Sigma_{NT}^{\varepsilon}\mathbf{\Psi}_s\right),
\end{aligned}$$

where Σ_{NT}^{x} is a block matrix with (i,ℓ) block equal to $\Sigma_{i\ell}^{x}$ and similarly for $\Sigma_{NT}^{\varepsilon}$. Since Σ_{NT}^{x} and $\Sigma_{NT}^{\varepsilon}$ are symmetric then we can see that the j and s indices in the summand of $\mathfrak{R}_A$ are

interchangeable so that

$$\mathfrak{R}_A = -\frac{1}{4}\frac{1}{NT}\sum_{j\neq s}\sum_{i,\ell=1}^{N}\psi_j'\big(\Sigma_{i\ell}^{x}+\Sigma_{i\ell}^{x\prime}\big)\psi_s\psi_j'\big(\Sigma_{i\ell}^{\varepsilon}+\Sigma_{i\ell}^{\varepsilon\prime}\big)\psi_s = o(1),$$

by Assumption 3(iv). □

A.2 Additional Simulation Results

Table A.1. Empirical Size (CHS). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.755	0.763	0.846	0.113	0.13	0.115	0.133	0.12	0.114	0.087	0.065
0.2	75	0.746	0.751	0.851	0.082	0.081	0.084	0.083	0.078	0.074	0.064	0.059
0.2	125	0.742	0.745	0.843	0.069	0.066	0.071	0.069	0.065	0.062	0.054	0.051
0.5	25	0.788	0.794	0.861	0.188	0.187	0.191	0.188	0.168	0.162	0.106	0.063
0.5	75	0.793	0.796	0.872	0.148	0.113	0.151	0.115	0.105	0.101	0.08	0.062
0.5	125	0.788	0.789	0.872	0.144	0.093	0.147	0.095	0.09	0.086	0.06	0.052
0.9	25	0.829	0.833	0.858	0.454	0.377	0.452	0.372	0.326	0.321	0.183	0.046
0.9	75	0.895	0.896	0.923	0.536	0.289	0.538	0.289	0.268	0.266	0.146	0.067
0.9	125	0.898	0.901	0.934	0.528	0.236	0.532	0.237	0.22	0.218	0.116	0.066
0.95	25	0.8	0.803	0.816	0.483	0.394	0.476	0.384	0.341	0.334	0.191	0.033
0.95	75	0.9	0.905	0.922	0.619	0.361	0.621	0.36	0.336	0.333	0.171	0.05
0.95	125	0.918	0.92	0.94	0.646	0.32	0.647	0.32	0.298	0.296	0.146	0.058

Table A.2. Size-Adjusted Empirical Power (CHS). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.314	0.306	0.311	0.271	0.273	0.272	0.274	0.279	0.277	0.288	0.25
0.2	75	0.799	0.788	0.777	0.77	0.747	0.77	0.741	0.745	0.748	0.77	0.755
0.2	125	0.957	0.952	0.948	0.947	0.943	0.947	0.943	0.943	0.944	0.954	0.945
0.5	25	0.245	0.222	0.239	0.196	0.191	0.197	0.193	0.189	0.192	0.217	0.214
0.5	75	0.612	0.602	0.585	0.577	0.54	0.578	0.541	0.539	0.541	0.587	0.559
0.5	125	0.845	0.838	0.826	0.821	0.806	0.821	0.804	0.802	0.803	0.822	0.805
0.9	25	0.129	0.118	0.116	0.105	0.121	0.105	0.12	0.12	0.119	0.173	0.201
0.9	75	0.163	0.161	0.156	0.158	0.16	0.158	0.161	0.159	0.16	0.181	0.203
0.9	125	0.244	0.225	0.218	0.212	0.196	0.21	0.196	0.194	0.194	0.24	0.25
0.95	25	0.148	0.134	0.118	0.151	0.163	0.146	0.166	0.175	0.168	0.244	0.335
0.95	75	0.149	0.134	0.131	0.144	0.136	0.142	0.134	0.135	0.137	0.175	0.23
0.95	125	0.166	0.147	0.152	0.138	0.133	0.138	0.134	0.137	0.134	0.162	0.214

Table A.3. Empirical Size (CHS-NL). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS-NL DGP where each procedure includes individual fixed effects in the estimation. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.756	0.766	0.847	0.114	0.132	0.116	0.135	0.121	0.116	0.089	0.067
0.2	75	0.749	0.753	0.85	0.083	0.082	0.086	0.085	0.081	0.075	0.064	0.06
0.2	125	0.743	0.746	0.84	0.068	0.067	0.07	0.069	0.067	0.063	0.054	0.05
0.5	25	0.786	0.796	0.861	0.191	0.19	0.193	0.191	0.17	0.164	0.103	0.063
0.5	75	0.791	0.797	0.872	0.149	0.113	0.151	0.116	0.106	0.102	0.079	0.062
0.5	125	0.788	0.79	0.871	0.142	0.093	0.145	0.096	0.09	0.087	0.06	0.053
0.9	25	0.83	0.833	0.858	0.452	0.377	0.451	0.373	0.327	0.322	0.182	0.047
0.9	75	0.893	0.895	0.92	0.536	0.292	0.539	0.292	0.268	0.265	0.145	0.065
0.9	125	0.899	0.901	0.931	0.53	0.24	0.532	0.241	0.223	0.221	0.115	0.066
0.95	25	0.798	0.801	0.816	0.483	0.393	0.476	0.383	0.342	0.335	0.189	0.033
0.95	75	0.901	0.905	0.922	0.618	0.362	0.619	0.361	0.338	0.334	0.171	0.05
0.95	125	0.919	0.921	0.941	0.645	0.322	0.647	0.322	0.299	0.298	0.145	0.059

Table A.4. Size-Adjusted Empirical Power (CHS-NL). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS-NL DGP where each procedure includes individual fixed effects in the estimation. The results are reported for a nominal level of 5%, $N = 50$, and different values of T and ρ. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.319	0.315	0.327	0.275	0.281	0.276	0.283	0.286	0.286	0.287	0.249
0.2	75	0.798	0.789	0.78	0.769	0.745	0.769	0.746	0.746	0.747	0.771	0.754
0.2	125	0.957	0.952	0.948	0.948	0.943	0.948	0.943	0.943	0.944	0.954	0.945
0.5	25	0.249	0.237	0.25	0.204	0.199	0.205	0.199	0.199	0.199	0.218	0.21
0.5	75	0.611	0.61	0.596	0.583	0.55	0.582	0.553	0.553	0.552	0.589	0.563
0.5	125	0.846	0.838	0.83	0.823	0.809	0.823	0.808	0.805	0.807	0.822	0.809
0.9	25	0.133	0.124	0.123	0.108	0.123	0.109	0.122	0.122	0.123	0.178	0.21
0.9	75	0.171	0.173	0.18	0.169	0.162	0.169	0.162	0.162	0.162	0.185	0.206
0.9	125	0.249	0.24	0.246	0.22	0.202	0.219	0.202	0.199	0.199	0.239	0.257
0.95	25	0.153	0.139	0.132	0.151	0.166	0.152	0.171	0.178	0.168	0.251	0.345
0.95	75	0.151	0.144	0.151	0.148	0.138	0.151	0.138	0.138	0.139	0.182	0.235
0.95	125	0.169	0.158	0.165	0.146	0.14	0.148	0.139	0.142	0.141	0.164	0.219

Table A.5. Empirical Size (CHS-NL; ρ_x Fixed). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS-NL DGP where each procedure includes individual fixed effects in the estimation. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.793	0.8	0.878	0.153	0.161	0.155	0.164	0.155	0.151	0.139	0.069
0.2	75	0.784	0.787	0.895	0.119	0.105	0.121	0.108	0.104	0.102	0.105	0.061
0.2	125	0.787	0.79	0.894	0.113	0.091	0.119	0.094	0.091	0.088	0.091	0.066
0.5	25	0.831	0.838	0.899	0.268	0.242	0.269	0.244	0.224	0.219	0.159	0.076
0.5	75	0.839	0.844	0.918	0.246	0.16	0.249	0.163	0.151	0.147	0.118	0.068
0.5	125	0.838	0.842	0.919	0.24	0.131	0.246	0.133	0.126	0.123	0.1	0.069
0.9	25	0.83	0.833	0.858	0.452	0.377	0.451	0.373	0.327	0.322	0.182	0.047
0.9	75	0.893	0.895	0.92	0.536	0.292	0.539	0.292	0.268	0.265	0.145	0.065
0.9	125	0.899	0.901	0.931	0.53	0.24	0.532	0.241	0.223	0.221	0.115	0.066
0.95	25	0.784	0.787	0.8	0.451	0.37	0.446	0.36	0.318	0.309	0.158	0.032
0.95	75	0.885	0.888	0.908	0.578	0.314	0.578	0.314	0.286	0.284	0.122	0.04
0.95	125	0.906	0.907	0.926	0.58	0.265	0.581	0.265	0.243	0.241	0.104	0.048

Table A.6. Size-Adjusted Empirical Power (CHS-NL; ρ_x Fixed). This table presents size-adjusted empirical power for the t-test of the alternative hypothesis that $\beta_1 = 0$ for the CHS-NL DGP where each procedure includes individual fixed effects in the estimation. The results are reported for $\rho_X = 0.9$, a nominal level of 5%, $N = 50$, and different values of T and ρ_U. Results are based on 5,000 simulations.

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
0.2	25	0.49	0.485	0.495	0.445	0.421	0.446	0.419	0.424	0.427	0.348	0.322
0.2	75	0.97	0.967	0.963	0.964	0.953	0.964	0.952	0.95	0.952	0.928	0.898
0.2	125	0.999	0.999	0.997	0.999	0.998	0.999	0.998	0.998	0.998	0.996	0.991
0.5	25	0.29	0.288	0.317	0.261	0.247	0.264	0.248	0.248	0.248	0.21	0.196
0.5	75	0.788	0.774	0.789	0.756	0.701	0.756	0.699	0.696	0.697	0.693	0.645
0.5	125	0.957	0.952	0.946	0.95	0.929	0.95	0.929	0.929	0.929	0.908	0.877
0.9	25	0.133	0.124	0.123	0.108	0.123	0.109	0.122	0.122	0.123	0.178	0.21
0.9	75	0.171	0.173	0.18	0.169	0.162	0.169	0.162	0.162	0.162	0.185	0.206
0.9	125	0.249	0.24	0.246	0.22	0.202	0.219	0.202	0.199	0.199	0.239	0.257
0.95	25	0.157	0.142	0.127	0.169	0.197	0.164	0.195	0.202	0.2	0.282	0.361
0.95	75	0.167	0.149	0.148	0.149	0.177	0.149	0.175	0.174	0.174	0.232	0.288
0.95	125	0.19	0.179	0.172	0.173	0.168	0.173	0.167	0.17	0.17	0.239	0.261

Table A.7. Empirical Size (CHS w/ heterogeneous ρ). This table presents empirical size for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The degree of persistence for the regressor and the error are heterogeneous: for each i we have that γ_t^x and γ_t^u are a Gaussian AR(1) with ρ of 0.2 or 0.95, each with equal probability of occurrence. The results are reported for a nominal level of 5% and different values of N and T. Results are based on 5,000 simulations.

N	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
25	25	0.595	0.577	0.641	0.16	0.186	0.159	0.186	0.154	0.134	0.1	0.054
25	75	0.711	0.705	0.704	0.301	0.225	0.281	0.206	0.189	0.175	0.092	0.045
25	125	0.75	0.747	0.732	0.373	0.22	0.341	0.195	0.177	0.166	0.08	0.04
50	25	0.707	0.692	0.74	0.151	0.188	0.151	0.184	0.156	0.144	0.095	0.053
50	75	0.794	0.788	0.791	0.304	0.224	0.288	0.218	0.193	0.185	0.091	0.046
50	125	0.824	0.821	0.812	0.386	0.225	0.367	0.213	0.195	0.189	0.081	0.037
75	25	0.756	0.744	0.783	0.157	0.194	0.155	0.191	0.161	0.153	0.098	0.057
75	75	0.824	0.821	0.822	0.301	0.223	0.292	0.217	0.195	0.191	0.093	0.043
75	125	0.85	0.85	0.841	0.385	0.218	0.37	0.21	0.194	0.192	0.085	0.039
100	25	0.792	0.778	0.813	0.156	0.195	0.154	0.193	0.161	0.154	0.095	0.053
100	75	0.847	0.843	0.846	0.3	0.22	0.293	0.216	0.194	0.191	0.089	0.044
100	125	0.871	0.871	0.865	0.377	0.226	0.371	0.217	0.197	0.195	0.08	0.038

Table A.8. Empirical Power (CHS w/ heterogeneous ρ). This table presents size-adjusted empirical power for the t-test of the null hypothesis that $\beta_1 = 1$ for the CHS DGP where each procedure includes individual fixed effects in the estimation. The degree of persistence for the regressor and the error are heterogeneous: for each i we have that γ_t^x and γ_t^u are a Gaussian AR(1) with ρ of 0.2 or 0.95, each with equal probability of occurrence. The results are reported for a nominal level of 5% and different values of N and T. Results are based on 5,000 simulations.

N	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN
25	25	0.399	0.373	0.371	0.33	0.291	0.353	0.308	0.31	0.32	0.363	0.376
25	75	0.539	0.517	0.556	0.436	0.533	0.571	0.594	0.591	0.594	0.651	0.648
25	125	0.585	0.572	0.658	0.525	0.617	0.697	0.682	0.681	0.68	0.717	0.708
50	25	0.435	0.401	0.415	0.327	0.31	0.337	0.322	0.328	0.327	0.38	0.409
50	75	0.506	0.493	0.588	0.454	0.535	0.535	0.577	0.565	0.571	0.652	0.659
50	125	0.569	0.56	0.666	0.53	0.609	0.657	0.66	0.66	0.658	0.721	0.713
75	25	0.434	0.404	0.415	0.347	0.307	0.351	0.318	0.323	0.326	0.391	0.387
75	75	0.542	0.508	0.595	0.457	0.524	0.525	0.562	0.555	0.557	0.653	0.654
75	125	0.601	0.59	0.673	0.562	0.619	0.641	0.649	0.647	0.646	0.718	0.712
100	25	0.432	0.405	0.42	0.345	0.308	0.348	0.313	0.325	0.326	0.394	0.4
100	75	0.553	0.53	0.602	0.481	0.546	0.531	0.561	0.558	0.558	0.656	0.657
100	125	0.61	0.6	0.68	0.563	0.628	0.645	0.648	0.646	0.648	0.711	0.71

A.3 Jackknife- and Bootstrap-Based Bias Correction

In this section, we propose to use our framework to construct bias-corrected estimators in time series and panel data settings. Consider the linear time series model[8]

$$y_t = \alpha + \beta' x_t + \varepsilon_t, \qquad t = 1, \ldots, T,$$

which we transform, as in the main text, to obtain $\{w_j, z_j\}_{j=1}^T$. Let $\hat{\beta}$ be the OLS estimator and $\hat{\beta}_{(-\ell)}$ the OLS estimator after leaving out $\{w_\ell, z_\ell\}$ from the sample. We can then construct a JN bias corrected estimator on the rotated space of the variables by following the standard prescription (see Shao and Tu, 1996),

$$\hat{\beta}^*_{\mathsf{JN}} = \hat{\beta} - \frac{T-1}{T} \sum_{t=1}^{T} \left(\hat{\beta}_{(-j)} - \hat{\beta} \right). \tag{A.7}$$

We can generalize this bias correction by adding triangular weights $\{\omega_{j,T}\}_{j=1}^T$ which satisfy $\sum_{j=1}^T \omega_{j,T} = 1$ for all T,

$$\hat{\beta}^*_{\mathsf{JN},\omega} = \hat{\beta} - \frac{T-1}{T} \sum_{t=1}^{T} \omega_{j,T} \left(\hat{\beta}_{(-j)} - \hat{\beta} \right). \tag{A.8}$$

In our simulation experiments below, we choose $\omega_{j,T} = (z_j'(\mathbf{X}'\mathbf{X})z_j)^{-\zeta_T}$ where $\zeta_T = T^{1/3}$. Finally, since the (instantaneous) jackknife can be interpreted as a first-order approximation to the bootstrap, it is natural to also consider a bootstrap-based bias correction. Since we work with the rotated data, which are heteroskedastic in general, we utilize the pairs bootstrap by resampling (with replacement) from $\{w_j, z_j\}_{j=1}^T$.[9] Let $\{w^*_{j,b}, z^*_{j,b}\}_{j=1}^T$ be the bth bootstrap sample with corresponding OLS estimator $\hat{\beta}^*_b$. Then,

$$\hat{\beta}^*_{\mathsf{boot}} = 2\hat{\beta} - \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}^*_b. \tag{A.9}$$

To assess the properties of these bias-corrected estimators, we consider two simulation designs featuring data generating processes which are well known to produce severely biased OLS estimates.

[8] We can follow the same steps as below in the panel data setting. We focus on the case where $N = 1$ for notational simplicity.

[9] See Kirch and Politis (2011) for an alternative bootstrap procedure relying on the DFT.

The first design is an autoregression of order one,

$$y_t = \beta y_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim_{iid} \mathcal{N}(0, 1-\beta^2)$ and we focus on values of β near unity.[10] The second design is a predictive regression model,

$$y_t = \beta x_{t-1} + \varepsilon_{y,t}$$
$$x_t = \phi x_{t-1} + \varepsilon_{x,t},$$

where $(\varepsilon_{y,t}, \varepsilon_{x,t})$ are multivariate Gaussian random variables with variance matrix equal to

$$\begin{bmatrix} 1 & \rho\sqrt{0.1} \\ \rho\sqrt{0.1} & 0.1 \end{bmatrix}.$$

In the predictive regression model, we generate data under the null hypothesis that $\beta = 0$ (see, e.g., Stambaugh, 1999; Campbell and Yogo, 2006).

For all simulations, we consider sample sizes of $T \in \{50, 100, 200\}$ and report average and median bias along with the mean-square error (MSE). Initial conditions are set equal to zero and there is a burn-in period of 100 observations. All results are based on 5,000 simulations.

We begin with the AR(1) model. Following common practice, we consider two specifications for the deterministic component. The first includes just a constant in the OLS estimation of the autoregressive parameter whereas the second includes both a constant and a linear time trend. We consider values of $\beta \in \{0.3, 0.9, 0.95, 0.99\}$. We include $\beta = 0.3$ in the specification with a constant only to show how the bias-correction procedures perform when the OLS estimator has little bias.[11]

The results are presented in Table A.9. The OLS estimator of β exhibits a strong downward bias in small samples and as β approaches unity. As documented in the existing literature, the downward bias is exacerbated when a linear time trend is included in the estimation. For $T = 100$ and $\beta = 0.99$ in the specification with a time trend, for example, the mean and the median bias of the OLS estimator is -0.092 and -0.081, respectively. The bias of the OLS estimator is greatly

[10] All results are robust to using standard Gaussian innovations rather than $\varepsilon_t \sim_{iid} \mathcal{N}(0, 1-\beta^2)$.

[11] We do not report results for $\beta = 0.3$ in the second specification because it is unlikely that a linear time trend will be included in the model when the underlying process is strongly mean reverting.

reduced using our jackknife and bootstrap bias-correction procedures. The JN method effectively eliminates the downward bias for $T > 50$ but this comes at the cost of increased variance and MSE. The weighted JN also manages to materially reduce the downward bias of the OLS estimator but without elevating the variability of the bias-adjusted estimate. As a result, the MSE for the weighted JN procedure is substantially below the MSE of the OLS estimator for these higher values of β. The best performing method is the bootstrap on the rotated space of the data. This method almost fully corrects the bias of the OLS estimator for $T = 100$ and $T = 200$ with an MSE that is materially lower (often twice as low) than that of OLS. This correction can be of great importance for computing accurate half-lives, impulse responses, and for long-horizon forecasting.

We now move on to the predictive regression model. In this setting, the OLS bias rises as a function of both ϕ and $|\rho|$. As such, we consider $\phi \in \{0.9, 0.95, 0.99\}$ and $\rho \in \{0, 0.7, -0.95\}$. The value of $\rho = -0.95$ mimics the properties of equity return predictive regressions when the regressor is the dividend yield or dividend-price ratio (e.g., Campbell and Yogo, 2006). In this case, it is known that the OLS estimator of β is upwardly biased. In contrast, when $\rho = 0.7$, the OLS estimator will be downward biased. Finally, we include the case of $\rho = 0$ to assess the robustness properties of the bias correction when the bias is small in magnitude.

Table A.10 reports the results for the predictive regression model as we vary T, ϕ and ρ. When the degree of endogeneity is large ($\rho = -0.95$ or 0.7), the OLS estimator is characterized by a large bias which tends to increase as the persistence parameter ϕ approaches one. The jackknife and bootstrap methods prove to be very effective in reducing this bias which becomes rather negligible for the larger sample sizes. While the JN method achieves this bias reduction at the expense of increased variability (reflected in its larger MSE), the weighted JN and, especially, the bootstrap bias-correction methods have a MSE that is lower than that of OLS. For example, the reduction of MSE of the bootstrap relative to OLS is over 30% for the empirically relevant case of $T = 200$ and $\phi = 0.99$. The corresponding mean (median) bias in this case is 0.074 (0.059) for the OLS estimator and 0.005 (-0.005) for the bootstrap-based bias correction. The performance of the jackknife and bootstrap methods remains satisfactory even when bias correction is unnecessary which is the case of $\rho = 0$. We should note that the proposed methods can also straightforwardly accommodate multiple predictors with varying degrees of persistence – a setting that cannot be readily handled by many of the existing methods.

Table A.9. Jackknife Bias Correction of AR(1). This table presents mean and median bias, along with the (scaled) mean-square error of four different estimators of the autoregressive coefficient in an AR(1) model: (1) the OLS estimator; the jackknife bias-corrected estimator as in equation (A.7); (3) the weighted jackknife bias-corrected estimator as in equation (A.8); the bootstrap-based bias corrected estimator as in equation (A.9). The top panel reports results for estimators with only a constant included as the deterministic term whereas the bottom panel reports results when both a constant and a linear time trend are included. All results are based on 5,000 simulations.

Deterministic Component: Constant Only

		OLS			JN			Weighted JN			Bootstrap		
		Mean	Med.	MSE	Mean	Med.	MSE	Mean	Med.	MSE	Mean	Med.	MSE
T	β	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$
50	0.30	-0.0381	-0.0313	2.0052	0.0048	0.0075	2.3333	-0.0113	-0.0074	2.1331	-0.0067	-0.0017	2.1481
50	0.90	-0.0819	-0.0661	1.5577	0.0263	0.0202	2.0989	-0.0195	-0.0115	1.3020	-0.0097	0.0012	1.2155
50	0.95	-0.0876	-0.0705	1.5079	0.0299	0.0200	2.0872	-0.0232	-0.0168	1.1567	-0.0114	-0.0006	1.0692
50	0.99	-0.0979	-0.0815	1.6238	0.0299	0.0244	2.1839	-0.0322	-0.0237	1.1409	-0.0185	-0.0065	1.0271
100	0.30	-0.0192	-0.0172	0.9219	0.0028	0.0043	0.9973	-0.0038	-0.0023	0.9573	-0.0021	-0.0001	0.9586
100	0.90	-0.0389	-0.0297	0.4647	0.0137	0.0151	0.5614	-0.0034	0.0024	0.4126	-0.0007	0.0070	0.3840
100	0.95	-0.0431	-0.0340	0.4389	0.0164	0.0132	0.5735	-0.0046	-0.0000	0.3654	-0.0016	0.0051	0.3275
100	0.99	-0.0499	-0.0397	0.4492	0.0182	0.0141	0.6377	-0.0090	-0.0052	0.3467	-0.0056	0.0009	0.2925
200	0.30	-0.0104	-0.0100	0.4800	0.0007	0.0017	0.4988	-0.0019	-0.0009	0.4905	-0.0015	-0.0009	0.4900
200	0.90	-0.0189	-0.0143	0.1594	0.0063	0.0092	0.1711	0.0001	0.0037	0.1469	0.0006	0.0052	0.1380
200	0.95	-0.0200	-0.0156	0.1241	0.0082	0.0093	0.1471	0.0008	0.0037	0.1119	0.0010	0.0050	0.0987
200	0.99	-0.0242	-0.0189	0.1157	0.0100	0.0069	0.1678	-0.0009	-0.0003	0.0995	-0.0011	0.0015	0.0789

Deterministic Component: Constant and Linear Time Trend

		OLS			JN			Weighted JN			Bootstrap		
		Mean	Med.	MSE	Mean	Med.	MSE	Mean	Med.	MSE	Mean	Med.	MSE
T	β	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$	Bias	Bias	$\times 100$
50	0.90	-0.1393	-0.1231	3.1341	0.0336	0.0095	4.1863	-0.0557	-0.0468	2.1418	-0.0305	-0.0196	1.9550
50	0.95	-0.1554	-0.1376	3.4747	0.0263	-0.0009	4.3289	-0.0699	-0.0638	2.2084	-0.0422	-0.0320	1.9366
50	0.99	-0.1795	-0.1624	4.2903	0.0141	-0.0154	4.6869	-0.0914	-0.0882	2.5973	-0.0612	-0.0517	2.1888
100	0.90	-0.0641	-0.0550	0.8064	0.0189	0.0116	1.0698	-0.0158	-0.0120	0.6087	-0.0070	-0.0015	0.5537
100	0.95	-0.0741	-0.0646	0.8919	0.0205	0.0048	1.1826	-0.0213	-0.0191	0.6071	-0.0114	-0.0055	0.5275
100	0.99	-0.0916	-0.0810	1.1588	0.0140	-0.0022	1.3474	-0.0355	-0.0341	0.6939	-0.0248	-0.0185	0.5858
200	0.90	-0.0303	-0.0249	0.2397	0.0079	0.0078	0.2710	-0.0047	-0.0010	0.1959	-0.0020	0.0025	0.1790
200	0.95	-0.0343	-0.0292	0.2250	0.0096	0.0063	0.2698	-0.0058	-0.0035	0.1639	-0.0032	0.0004	0.1425
200	0.99	-0.0426	-0.0372	0.2644	0.0112	0.0030	0.3406	-0.0102	-0.0107	0.1699	-0.0074	-0.0045	0.1378

Table A.10. Jackknife Bias Correction of the Predictive Regression Model. This table presents mean and median bias, along with the (scaled) mean-square error of four different estimators of the slope coefficient in the standard predictive regression model: (1) the OLS estimator; the jackknife bias-corrected estimator as in equation (A.7); (3) the weighted jackknife bias-corrected estimator as in equation (A.8); the bootstrap-based bias corrected estimator as in equation (A.9). The three panels report results for different degrees of endogeneity: $\rho = 0.7$, $\rho = 0$, and $\rho = -0.95$, respectively. All results are based on 5,000 simulations.

Degree of Endogeneity: $\rho = 0.7$

		OLS			JN			Weighted JN			Bootstrap		
T	β	Mean Bias	Med. Bias	MSE ×100	Mean Bias	Med. Bias	MSE ×100	Mean Bias	Med. Bias	MSE ×100	Mean Bias	Med. Bias	MSE ×100
50	0.90	-0.1809	-0.1424	10.8375	0.0576	0.0497	14.9891	-0.0437	-0.0229	9.9847	-0.0210	0.0075	9.4721
50	0.95	-0.2007	-0.1572	10.7000	0.0567	0.0446	15.4931	-0.0589	-0.0369	9.4868	-0.0326	-0.0018	8.8782
50	0.99	-0.2170	-0.1718	10.3651	0.0648	0.0500	15.4823	-0.0724	-0.0447	8.5658	-0.0405	-0.0036	7.9525
100	0.90	-0.0874	-0.0669	3.6264	0.0313	0.0359	4.4449	-0.0073	0.0076	3.4883	-0.0018	0.0172	3.2766
100	0.95	-0.0954	-0.0721	3.0379	0.0340	0.0317	4.0741	-0.0113	0.0019	2.8437	-0.0046	0.0143	2.5997
100	0.99	-0.1090	-0.0823	2.8909	0.0415	0.0353	4.4934	-0.0188	-0.0050	2.6276	-0.0115	0.0079	2.2918
200	0.90	-0.0387	-0.0287	1.3447	0.0184	0.0233	1.4940	0.0044	0.0115	1.3358	0.0053	0.0135	1.2754
200	0.95	-0.0455	-0.0345	0.9635	0.0175	0.0204	1.1624	0.0008	0.0082	0.9404	0.0011	0.0113	0.8564
200	0.99	-0.0528	-0.0409	0.7231	0.0228	0.0190	1.1183	-0.0013	0.0028	0.7162	-0.0019	0.0063	0.5864

Degree of Endogeneity: $\rho = 0$

50	0.90	-0.0070	-0.0081	7.3627	-0.0032	-0.0003	10.3906	-0.0047	-0.0024	8.2579	-0.0058	-0.0051	8.0228
50	0.95	-0.0026	-0.0022	5.9564	-0.0010	0.0019	9.9525	-0.0020	-0.0011	7.0244	-0.0023	-0.0018	6.7370
50	0.99	-0.0015	-0.0021	4.5649	-0.0062	-0.0081	9.3018	-0.0038	-0.0044	5.6666	-0.0035	-0.0048	5.4618
100	0.90	-0.0006	-0.0002	2.6811	0.0017	-0.0011	3.3358	0.0009	-0.0015	2.9374	0.0006	-0.0012	2.8337
100	0.95	-0.0004	0.0007	1.8085	-0.0008	0.0016	2.6933	-0.0006	0.0027	2.1411	-0.0006	0.0019	2.0061
100	0.99	0.0020	0.0019	1.1922	0.0008	0.0013	2.1938	0.0014	0.0014	1.5176	0.0015	0.0016	1.3871
200	0.90	-0.0021	-0.0005	1.0713	-0.0018	-0.0006	1.2206	-0.0019	-0.0004	1.1484	-0.0021	0.0002	1.1145
200	0.95	-0.0009	-0.0023	0.6916	-0.0019	-0.0021	0.8864	-0.0016	-0.0018	0.7898	-0.0014	-0.0017	0.7435
200	0.99	0.0001	-0.0004	0.3526	0.0000	0.0011	0.6362	0.0001	-0.0001	0.4779	0.0001	0.0001	0.4165

Degree of Endogeneity: $\rho = -0.95$

50	0.90	0.2416	0.1901	14.2368	-0.0796	-0.0667	19.4009	0.0567	0.0252	12.1099	0.0264	-0.0144	11.3643
50	0.95	0.2720	0.2243	14.7666	-0.0824	-0.0482	20.6774	0.0769	0.0548	11.6859	0.0415	0.0080	10.7064
50	0.99	0.2962	0.2420	15.4953	-0.0872	-0.0715	20.9835	0.0988	0.0719	11.2862	0.0560	0.0168	10.1439
100	0.90	0.1188	0.0908	4.6373	-0.0399	-0.0465	5.4714	0.0114	-0.0071	4.1236	0.0033	-0.0200	3.8671
100	0.95	0.1275	0.1004	3.9868	-0.0520	-0.0430	5.4968	0.0116	-0.0031	3.4299	0.0029	-0.0195	3.0496
100	0.99	0.1468	0.1207	3.9699	-0.0580	-0.0442	5.5855	0.0238	0.0145	3.0160	0.0136	-0.0033	2.5915
200	0.90	0.0594	0.0462	1.5888	-0.0162	-0.0245	1.6301	0.0022	-0.0089	1.4315	0.0004	-0.0114	1.3541
200	0.95	0.0658	0.0513	1.2940	-0.0186	-0.0202	1.4726	0.0036	-0.0047	1.1539	0.0029	-0.0092	1.0242
200	0.99	0.0740	0.0590	1.0868	-0.0291	-0.0195	1.5608	0.0036	0.0018	0.9475	0.0047	-0.0050	0.7458