A New Jackknife Variance Estimator for Time Series and Panel Regressions[†]

Richard K. Crump
Federal Reserve Bank of New York

Nikolay Gospodinov
Federal Reserve Bank of Atlanta

Ignacio Lopez Gaffney
Yale University

Venice Time Series Workshop (22-23 April, 2022)

[†] The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of New York or the Federal Reserve System.

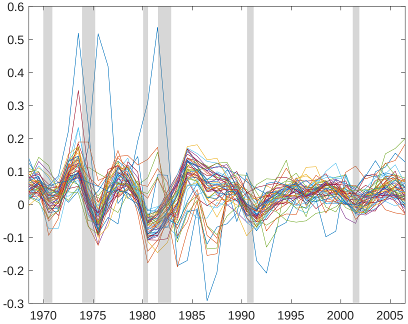
Challenges for Inference: Dependence

- Strong dependence makes accurate inference challenging
- Becomes increasingly difficult as dependence strengthens to extreme cases:
  - *Time-series dependence*: e.g., unit roots and classical spurious regression problem (e.g., Granger & Newbold, 1974; Phillips, 1987)
  - *Cross-sectional dependence*: e.g., spatial unit roots and new type of spurious regression problem (e.g., Müller & Watson, 2024)
- This (joint) dependence structure makes accurate inference and assessing robustness a difficult task
  - Reliable inference in time-series regression is, by now, a well-known challenge
  - Panel data are often characterized by strong cross-sectional <u>and</u> time-series dependence which further complicates the issue
- Inference procedures that diminish sensitivity to exact form of dependence are desirable

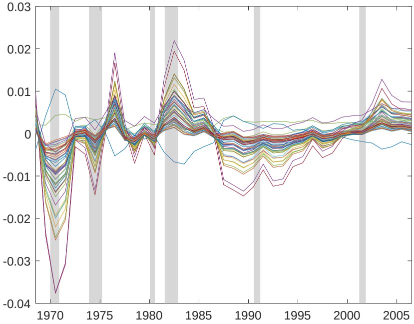
Challenges for Inference: Dependence

Motivating Example: Nakamura and Steinsson (2014, AER)

(a) *LHS Variable*



(b) *RHS Variable*



This Paper

- Propose a new jackknife variance estimator for time-series/panel-data regressions
  - Rotate ("orthogonalize") the data using a specific choice of orthonormal trigonometric basis functions
  - Then perform leave-one-out jackknife variance estimation using the rotated data
- The estimator is:
  - Easy to implement
  - Tuning-parameter-free
  - Agnostic about source of serial dependence in common time-series behavior
  - Performs very well in finite samples, especially under strong dependence
  - Panel Data: Allows for arbitrary correlations across units
- Setting: t-statistic based inference (i.e., one-dimensional linear restrictions)
  - Prove variance estimator enables asymptotically valid inference in range of settings

Conventional Jackknife Variance Estimator

- Consider standard cross-sectional linear regression setup:

$$y_i = \alpha + x_i'\beta + \epsilon_i, \qquad \mathbb{E}[x_i \epsilon_i] = 0 \qquad i = 1, \ldots, N.$$

- We can stack as

$$y = \mathbf{X}\theta + \epsilon, \qquad \theta = \left(\beta', \alpha\right)', \qquad \mathbf{X} = \left[X \quad \iota_N\right].$$

- Let $\widehat{\theta}$ be OLS estimator of full sample, and $\widehat{\theta}_{(-i)}$ be OLS estimator for sample excluding the ith observation

- Recall the leave-one-out formula:

$$\widehat{\theta} - \widehat{\theta}_{(-j)} = \frac{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i\left(y_i - \mathbf{x}_i'\widehat{\theta}\right)}{1 - \mathbf{x}_i'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i}, \quad \text{where } \mathbf{x}_i = (x_i, 1)'.$$

- Here $\mathbf{x}_i'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i \in (0, 1)$ is the “leverage” term with larger leverage values implying “unusual” $\mathbf{x}_i$ relative to rest of sample

Conventional Jackknife Variance Estimator (cont'd)

- Jackknife (JN) variance estimator of $\widehat{\theta}$ is

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

- We can then obtain the HC3 variance estimator

$$\begin{aligned}\widehat{\mathsf{V}}_{\theta} &= \sum_{j=1}^{N} (\widehat{\theta}_{(-j)} - \widehat{\theta})(\widehat{\theta}_{(-j)} - \widehat{\theta})' \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left[\sum_{i=1}^{N} \frac{1}{\left(1 - \mathbf{x}_i'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_i\right)^2} \mathbf{x}_i \mathbf{x}_i' \left(y_i - \mathbf{x}_i'\widehat{\theta}\right)^2\right] \left(\mathbf{X}'\mathbf{X}\right)^{-1}\end{aligned}$$

- Without term in red just Eicker-Huber-White variance estimator

- Leverage weights are asymptotically negligible...but lead to better size control in finite samples

Orthogonalizing Dependent Processes

- How can we map standard jackknife estimator to dependent case?
- We want to somehow make corr$(y_i, y_j) \approx 0$ and corr$(x_i, x_j) \approx 0$
- But we want to avoid assuming a specific parametric data-generating process to avoid misspecification
- Instead, we apply a linear transformation to the data using the functions:

$$\Psi = [\psi_1, ..., \psi_T], \qquad \psi_j = (\psi_{1,j}, ..., \psi_{T,j})',$$

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

  - Ψ is orthonormal; i.e., $\Psi\Psi' = \Psi'\Psi = I_T$
  - ψ_j are trigonometric basis functions of different frequencies
  - We order Ψ so that ψ_1 is the lowest frequency and ψ_T the highest

What is Ψ?

- Ψ are the eigenvectors of the variance matrix of a random walk (RW), i.e.

  $$\Sigma = \Psi \Lambda_{\mathsf{RW}} \Psi',$$

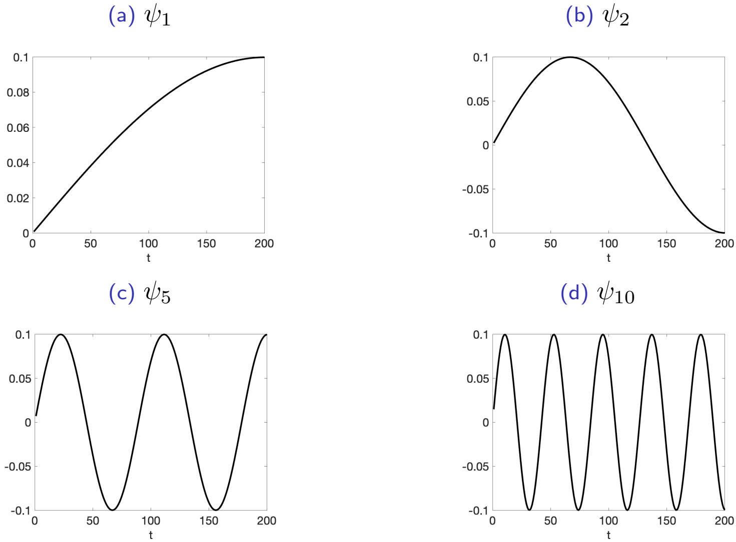
  where $\left[\Sigma_{\mathsf{RW}}\right]_{ij} = \min(i,j)$ and Λ_{RW} is diagonal with rth diagonal element

  $$\left[\Lambda_{\mathsf{RW}}\right]_{rr} = \frac{\sigma^2}{2 - 2\cos\left(\frac{(2r-1)\pi}{2T+1}\right)}$$

- By orthonormality, Ψ are also eigenvectors of a white noise (WN) process.
- More generally, corr$(\psi_i' x, \psi_j' x) = o(1)$ for a wide class of time series processes, with rate of convergence determined by their dependence properties.
- <u>Not</u> the same basis functions as:
  - Discrete Fourier transform (e.g., Hidalgo & Schafgans, 2021)
  - Discrete cosine transform (e.g., Müller & Watson, 2008)

What is Ψ? (cont'd)

Example: $T = 200$



Orthogonalizing Dependent Processes (cont'd)

- Let x be a RW. Define $z = \Psi' x$ with $z_i = \psi_i' x$. Then,

$$\mathbb{V}[z] = \mathbb{V}[\Psi' x] = \Psi' \underbrace{\left(\Psi \Lambda \Psi'\right)}_{\Sigma_{\mathsf{RW}}} \Psi = \Lambda_{\mathsf{RW}} \implies \mathsf{corr}(z_i, z_j) = 0, \; i \neq j$$

- z are OLS coefficients from regressing x on each ψ_j since

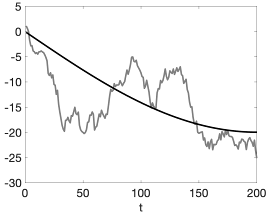
$$z = \Psi' x = \underbrace{\left(\Psi' \Psi\right)^{-1} \Psi' x}_{\text{OLS estimator}} \implies x = \Psi z = \sum_{j=1}^{T} \psi_j z_j$$

- <ins>Intuition</ins>: Represent Gaussian RW as linear combinations of (deterministic) eigenvectors, where weights are independent (not *i.i.d.*) Gaussian variables, i.e.

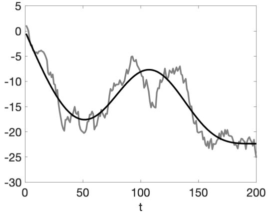
$$x \equiv (x_1, x_2, ..., x_T)' =_d \sum_{j=1}^{T} \psi_j \eta_j, \qquad \eta_j \sim \mathcal{N}\Big(0, \left[\Lambda_{\mathsf{RW}}\right]_{jj}\Big).$$

Orthogonalizing Dependent Processes (cont'd)

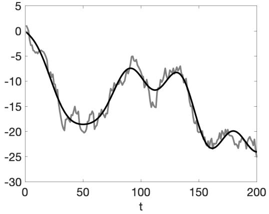
(a) $\eta_1\psi_1$



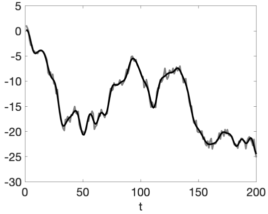
(b) $\sum_{j=1}^{5}\eta_j\psi_j$



(c) $\sum_{j=1}^{10}\eta_j\psi_j$



(d) $\sum_{j=1}^{50}\eta_j\psi_j$



Combining Orthogonalization and Jackknife

- Consider a standard time-series regression,

$$y = \alpha \cdot \iota_T + X\beta + \varepsilon,$$

where $y = (y_1, \ldots, y_T)'$, $X = (x_1, \ldots, x_T)'$, and ι_T is $T \times 1$ vector of ones

- Transform the data by multiplying through by Ψ'

$$\Psi' y = \alpha \cdot \Psi' \iota_T + \Psi' X \beta + \Psi' \varepsilon \iff w = \alpha \cdot \zeta_T + Z\beta + u$$

- OLS estimator is invariant to this transformation since $\Psi\Psi' = \mathrm{I}_T$

$$\begin{bmatrix} \widehat{\beta} \\ \widehat{\alpha} \end{bmatrix} = \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'w = \left(\mathbf{X}'\Psi\Psi'\mathbf{X}\right)^{-1}\mathbf{X}'\Psi\Psi' y = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' y$$

where $\mathbf{Z} = \left[Z \ \zeta_T\right]$ and $\mathbf{X} = \left[X \ \iota_T\right]$

Combining Orthogonalization and Jackknife (cont'd)

- Leave-one-out (LOO) OLS estimator is <u>not</u> invariant to transformation

- Let $\widehat{\beta}_{(-j)}$ be the LOO estimator, i.e.,

$$\begin{bmatrix} \widehat{\beta}_{(-j)} \\ \widehat{\alpha}_{(-j)} \end{bmatrix} = \left(\mathbf{Z}'\mathbf{Z} - \mathbf{Z}_j'\mathbf{Z}_j\right)^{-1}\left(\mathbf{Z}'w - \mathbf{Z}_j'w_j\right)$$

  where $\mathbf{Z}_j$ and w_j are the jth rows of $\mathbf{Z}$ and w, respectively.

- The JN variance estimator is simply,

$$\begin{aligned} \mathsf{V}_{\hat{\beta}} &= \sum_{j=1}^{T}\left(\widehat{\beta}_{(-j)} - \widehat{\beta}\right)^2 \\ &= \mathsf{e}_1'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left[\sum_{j=1}^{T}\frac{1}{\left(1 - \mathbf{Z}_j'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{Z}_j\right)^2}\,\widehat{u}_j^2 \cdot \mathbf{Z}_j\mathbf{Z}_j'\right]\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathsf{e}_1 \end{aligned}$$

  where $\mathsf{e}_1 = (1, 0, \ldots, 0)'$

Intuition: Why Might This Work?

- We have that $(\mathbf{Z}'\mathbf{Z})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$ because $\Psi\Psi' = \mathsf{I}_T$
- Simple case: no constant, one regressor, and strict exogeneity (drop leverage):

$$\begin{aligned}
\frac{1}{T}\sum_{j=1}^{T} \widehat{u}_j^2 \cdot z_j^2 &= \frac{1}{T}\sum_{j=1}^{T} \mathbb{E}\left[u_j^2 \cdot z_j^2\right] + o_p(1) \\
&= \frac{1}{T}\sum_{j=1}^{T}\sum_{\ell=1}^{T} \mathbb{E}\left[u_j u_\ell z_j z_\ell\right] - \frac{1}{T}\sum_{j\neq\ell} \mathbb{E}\left[u_j u_\ell z_j z_\ell\right] + o_p(1) \\
&= \underbrace{\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} \mathbb{E}\left[\varepsilon_t\varepsilon_s x_t x_s\right]}_{\text{Long-run variance as } T\to\infty} - \frac{1}{T}\sum_{j\neq\ell} \psi_j'\mathbb{E}\left[\varepsilon\varepsilon'\right]\psi_\ell \cdot \psi_j'\mathbb{E}\left[xx'\right]\psi_\ell + o_p(1)
\end{aligned}$$

where the third equality follows since $z'u = x'\Psi'\Psi\varepsilon = x'\varepsilon$

- The second term is $o(1)$ under regularity conditions and the orthogonalization properties of Ψ $\Rightarrow$ consistency of the JN variance estimator

Illustrative Example

- Simulate data from the model (Müller, 2014)

$$
\begin{aligned}
y_t &= x_t'\beta + \varepsilon_t \\
x_t &= \rho x_{t-1} + v_t^x, \qquad \varepsilon_t = \rho\varepsilon_t + v_t^\varepsilon,
\end{aligned}
$$

  where $|\rho_x| < 1, |\rho_\varepsilon| < 1$, and $\{v_t^x\}$ and $\{v_t^\varepsilon\}$ are mutually independent, *i.i.d.*$N(0,1)$ innovations.

- Vary the sample size and the value of ρ; 50,000 replications
- 2 cases: (i) single regressor and (ii) four regressors
- Empirical size of t-test of $\mathbb{H}_0 : \beta = \beta_0$ based on
  1. the proposed jackknife variance estimator (Ψ-JN);
  2. the NW HAC variance estimator (with Andrews' data-driven implementation).
- Both procedures use standard normal critical values at 5% nominal level.

Simulation Evidence: Empirical Size

	$T = 50$		$T = 100$		$T = 200$		$T = 500$	
	Ψ-JN	NW	Ψ-JN	NW	Ψ-JN	NW	Ψ-JN	NW
Scalar Regressor								
$\rho = 0.00$	0.057	0.087	0.054	0.070	0.051	0.059	0.049	0.052
$\rho = 0.70$	0.066	0.191	0.063	0.145	0.057	0.109	0.054	0.084
$\rho = 0.90$	0.070	0.340	0.073	0.266	0.068	0.193	0.060	0.125
$\rho = 0.95$	0.066	0.422	0.071	0.351	0.072	0.273	0.066	0.173
$\rho = 0.98$	0.063	0.487	0.068	0.444	0.073	0.380	0.070	0.266
$\rho = 0.999$	0.064	0.537	0.066	0.535	0.068	0.526	0.069	0.500
Four Regressors								
$\rho = 0.00$	0.049	0.097	0.051	0.073	0.049	0.061	0.049	0.053
$\rho = 0.70$	0.053	0.224	0.052	0.163	0.054	0.120	0.052	0.085
$\rho = 0.90$	0.048	0.376	0.051	0.307	0.054	0.220	0.053	0.140
$\rho = 0.95$	0.047	0.432	0.049	0.388	0.052	0.307	0.053	0.201
$\rho = 0.98$	0.045	0.464	0.047	0.455	0.048	0.408	0.050	0.314
$\rho = 0.999$	0.044	0.479	0.045	0.486	0.045	0.483	0.045	0.476

Sampling Assumptions

Assumption TS-1

Assume that (y_t, x_t') are strictly stationary and ergodic satisfying

$$y_t = \alpha + \beta x_t + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with:

(i) $\mathbb{E}[\varepsilon_t \,|\, x_t, x_{t-1}, \ldots] = 0$;

(ii) $\big(\mathbf{X}'\mathbf{X}/T\big) \to_p \Omega$ and $\lambda_{\min}(\Omega)$ is bounded away from zero;

(iii) $T^{-1/2}\mathbf{X}'\varepsilon \to_d \mathcal{N}(0, \Sigma)$, where $\Sigma = \lim_{T\to\infty} T^{-1}\mathbb{E}\big[\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}\big]$ is the long-run variance.

- Throughout we will present results for a constant and single regressor to simplify notation; paper covers general case
- Let V_1 be the $(1,1)$ element of $\Omega^{-1}\Sigma\Omega^{-1}$

Sampling Assumptions

Moments and Dependence

Assumption TS-2

Assume that:

(i) $\mathbb{C}\big(v_{t_1}v_{t_2}, \varepsilon_{t_1}\varepsilon_{t_2}\big) = 0$ for $t_1 \neq t_2$;

(ii) $T^{-2}\sum_{t_1,t_2,t_3,t_4} |\kappa_4(v_{t_1}, \varepsilon_{t_2}, v_{t_3}, \varepsilon_{t_4})| = o(1)$, where κ_4 is the fourth-order joint cumulant;

(iii) $\max_{1\leq j\leq T} \mathbb{E}\left[|u_j|^4\right] \lesssim 1$, $\max_{1\leq j\leq T} \mathbb{E}\left[\left\|\psi_j'([X_1 \;\, X_2] - \iota_T\mu')\right\|^4\right] \lesssim 1$ uniformly for all T sufficiently large;

(iv) $T^{-1}\sum_{j=1}^{T}\left(q_j^2u_j^2 - \mathbb{E}\big[q_j^2u_j^2\big]\right) = o_p(1)$, where $q_j = \psi_j'v$.

- (i) holds, e.g., errors are independent of regressors or if ε_t is MDS with respect to information set at time t
- (ii)–(iv) hold under restrictions on the dependence structure, e.g., mixing or cumulant summability conditions

Asymptotic Expansion of $\widehat{\mathsf{V}}_1$

Proposition TS-1

Under Assumptions TS-1 and TS-2,

$$T \cdot \left(\widehat{\mathsf{V}}_1 - \mathsf{V}_1\right) = \mathbb{V}(x_t)^{-2} \cdot \Bigg[\frac{2}{T}\sum_{j=1}^{T}\left(\psi_j'\mathbb{E}\left[x\varepsilon'\right]\psi_j\right)^2 - \mathbb{C}(x_t^2, \varepsilon_t^2) - \frac{1}{T}\sum_{j \neq s}\psi_j'\mathbb{V}\left(x\right)\psi_s\psi_j'\mathbb{E}\left[\varepsilon\varepsilon'\right]\psi_s\Bigg] + o_p(1),$$

where $x = (x_1, \ldots, x_T)'$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$.

- First Term: always non-negative
- Second Term: can be straightforwardly estimated
- Third Term: can be handled by orthogonalization properties of Ψ

Sampling Assumptions

Orthogonalization of Variance Matrices

Assumption TS-3

Assume that

$$T^{-1} \sum_{j \neq s} \psi_j' \mathbb{V}\left[x\right] \psi_s \psi_j' \mathbb{E}\left[\varepsilon\varepsilon'\right] \psi_s = o(1).$$

- Simple example: $\mathbb{E}\left[\varepsilon\varepsilon'\right] \propto \mathsf{I}_T$, then term is zero
- Crump, Gospodinov, and Lopez Gaffney (2025) show that Ψ orthogonalizes a wide variety of time series processes
  - For example, for a stationary ARMA(p, q), $\max_{1 \leq j,s \leq T} \psi_j' \mathbb{E}\left[xx'\right] \psi_s = O(T^{-1})$
  - For stronger dependence (e.g., some long memory processes) rate slows to $O(T^{-\nu})$ for $\nu < 1$

Consistency of $\widehat{\mathsf{V}}_1$

Proposition TS-2

Let Assumptions TS-1–TS-3 hold. If $\sup_{T \geq 1} \lambda_{\max}(\mathbb{E}[x\varepsilon'] + \mathbb{E}[\varepsilon x']) = o_p(1)$ and $\mathbb{C}(x_t^2, \varepsilon_t^2) = 0$, then $\widehat{\mathsf{V}}_1 \to_p \mathsf{V}_1$.

- Comparing restrictions to strict exogeneity
  - Do not require that $\mathbb{E}[\,\varepsilon_t \,|\, x_1, \ldots, x_T] = 0$
  - But does require $\mathbb{C}(x_t^2, \varepsilon_t^2) = 0$
- Can circumvent this latter requirement by modifying the variance estimator as

$$\widehat{\mathsf{V}}_{1,\mathrm{bc}} = \widehat{\mathsf{V}}_1 + \frac{\widehat{\mathbb{C}}(x_t^2, \widehat{\varepsilon}_t^2)}{T \cdot \widehat{\mathbb{V}}(x_t)^2}$$

Asymptotic Validity of $\widehat{\mathsf{V}}_1$

Proposition TS-3 (Asymptotic Conservativeness)

Let Assumptions TS-1–TS-3 hold and $\widehat{\mathbb{C}}(x_t^2, \widehat{\varepsilon}_t^2) = \mathbb{C}(x_t^2, \varepsilon_t^2) + o_p(1)$. Then,

$$T\big(\widehat{\mathsf{V}}_{1,\mathrm{bc}} - \mathsf{V}_1\big) = 2\omega_1^{-2} T^{-1} \sum_{j=1}^{T} \big(\psi_j' \mathbb{E}\big[v\varepsilon'\big]\psi_j\big)^2 + o_p(1),$$

where ω_1^{-1} is the $(1,1)$ element of Ω^{-1}. Furthermore, if $\mathbb{C}(v_t^2, \varepsilon_t^2) = 0$, then

$$T\big(\widehat{\mathsf{V}}_1 - \mathsf{V}_1\big) = 2\omega_1^{-2} T^{-1} \sum_{j=1}^{T} \big(\psi_j' \mathbb{E}\big[v\varepsilon'\big]\psi_j\big)^2 + o_p(1).$$

- This proposition shows that (the bias-corrected) jackknife variance estimator may not consistently estimate the long-run variance, but the gap is guaranteed to be positive asymptotically.

Asymptotic Validity of $\widehat{\mathsf{V}}_1$ (cont'd)

Corollary TS-1

Let Assumptions TS-1–TS-3 hold. Then for $a \in (0, 1)$,

$$\limsup_{T\to\infty} \mathbb{P}\left(\left|\frac{\widehat{\beta}_1 - \beta_1}{\widehat{\mathsf{V}}_{1,\mathrm{bc}}}\right| \geq \Phi^{-1}\left(1 - \frac{a}{2}\right)\right) \leq a,$$

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

- t-statistic based inference, using the bias-corrected jackknife variance estimator and $N(0, 1)$ critical values, provides valid inference under general assumptions.

General Bias Correction

Predictive Regressions, Local Projections

- In some cases such as predictive regressions and local projections we need to change the bias corrected estimator

- Recall that predictive regressions/local projections for horizon h have $MA(h-1)$ structure

- We work with a modified variance estimator of the form

$$\widehat{\mathsf{V}}_{1,\mathrm{bc}}^{(h)} = \widehat{\mathsf{V}}_{1,\mathrm{bc}} + \frac{2\sum_{r=1}^{h-1} \widehat{\mathbb{C}}(x_t x_{t-r}, \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-r})}{T \cdot \widehat{\mathbb{V}}(x_t)^2}$$

  to accommodate additional moving average structure

- In simulation results, we use $\widehat{\mathsf{V}}_{1,\mathrm{bc}}$ or $\widehat{\mathsf{V}}_{1,\mathrm{bc}}^{(h)}$ as appropriate

Extensive Simulation Evidence

- h-period predictive regressions with multiple predictors, strong endogeneity and persistence, and GARCH errors.
- Local projections with GARCH and non-Gaussian errors.
- The LLSW (2018) set of simulation experiments which includes regression specifications covering distributed lag models, forecasting regressions, and local projections. Using a dynamic factor model, LLSW randomly select 828 individual data-generating processes for each of 23 different regression specifications and choices of innovation distributions.
- We compare our jackknife-based inference to HAR estimators (EWC, NW), lag augmentation, DFT- and DCT-based jackknife methods.
- We show that the jackknife approach using Ψ has excellent finite-sample (size and power) properties in these challenging environments and (almost) always outperforms the alternative procedures.

Panel-Data Regressions: General Setup

- Standard linear panel-data model:

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

  - Assume β_1 is the parameter of interest
  - $x_{1,it} \in \mathbb{R}$, $x_{2,it} \in \mathbb{R}^K$

- We can stack the model as

$$\underbrace{Y_i}_{T \times 1} = \alpha_i \iota_T + X_{1,i}\beta_1 + X_{2,i}\beta_2 + \varepsilon_i, \qquad i = 1, \ldots, N,$$

- We will proceed with FEs only (without $x_{2,it}$) but paper covers general case

- Unbalanced panel case also covered in paper

Jackknife Variance Estimator

- Stacked model is

$$\underbrace{Y_i}_{T\times 1} = \alpha_i \iota_T + X_i \beta_1 + \varepsilon_i, \qquad i = 1, \ldots, N,$$

- We can rotate data for each i via

$$\begin{aligned} \Psi' Y_i &= \alpha_i \cdot \Psi' \iota_T + \Psi' X_i \beta_1 + \Psi' \varepsilon_i \\ \iff W_i &= \alpha_i \cdot \zeta_T \quad + Z_i \beta_1 \quad + u_i \end{aligned}$$

- OLS estimator is still invariant to the transformation using $\Psi \Rightarrow$ can continue using OLS estimator as basis of comparison

- In panel setting, we form JN variance estimator by
  - removing jth observation from each $i = 1, \ldots, N$
  - using $(T-1)N$ observations for each JN estimate of β_1

Jackknife Variance Estimator (cont'd)

- Let $\widehat{\beta}_{1,(-j)}$ be the OLS estimator of β_1 with the jth observation removed for each $i = 1, \dots, N$; collect $\widehat{\beta}_{1,(-j)}$ for $j = 1, \dots, T$

- JN variance estimator is then

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

- Continues to have an HC3 interpretation but now with a leverage *matrix*

- Not as computationally efficient when $N > 1$ (can no longer vectorize estimator) but still easy to implement

- Can be modified straightforwardly to cluster at group level rather than individual level

Asymptotic Expansion of $\widehat{\mathsf{V}}_1$

Proposition PD-1

Under regularity conditions and if $NT^{-1}\log(T) \to 0$,

$$NT \cdot \left(\widehat{\mathsf{V}}_1 - \mathsf{V}_1\right) = \omega_1^{-2}\Bigg[\frac{N}{T}\sum_{j=1}^{T}\left(\psi_j'\bar{\Gamma}\psi_j\right)^2 + \frac{1}{N}\sum_{i,\ell=1}^{N}\mathbb{C}(x_{1,it}x_{1,\ell t},\varepsilon_{it}\varepsilon_{\ell t})$$

$$+ \frac{1}{NT}\sum_{j=1}^{T}\left\|(\mathsf{I}_N \otimes \psi_j)'\Gamma_{NT}(\mathsf{I}_N \otimes \psi_j)\right\|^2$$

$$+ \frac{1}{NT}\sum_{i,\ell=1}^{N}\sum_{j\neq s}\psi_j'\left(\Sigma_{i\ell}^{x} + \Sigma_{i\ell}^{x\prime}\right)\psi_s\psi_j'\left(\Sigma_{i\ell}^{\varepsilon} + \Sigma_{i\ell}^{\varepsilon\prime}\right)\psi_s\Bigg] + o_p(1)$$

where $\bar{\Gamma} = N^{-1}\sum_{i=1}^{N}\mathbb{E}[X_i\varepsilon_i']$, Γ_{NT} is a block matrix with (i,ℓ) block equal to $\mathbb{E}[X_i\varepsilon_\ell']$, ω^{-1} is the $(1,1)$ element of $\mathrm{plim}_{N,T\to\infty}(\mathbf{X}'\mathbf{X}/NT)^{-1}$, and

$$\Sigma_{i\ell}^{x} = \mathbb{C}(X_i, X_\ell), \qquad \Sigma_{i\ell}^{\varepsilon} = \mathbb{C}(\varepsilon_i, \varepsilon_\ell).$$

Asymptotic Validity of $\widehat{\mathsf{V}}_1$

Corollary PD-1

Under regularity conditions and if $NT^{-1}\log(T) \to 0$. Then for $a \in (0,1)$,

$$\limsup_{T\to\infty} \mathbb{P}\left(\left| \frac{\widehat{\beta}_1 - \beta_1}{\widehat{\mathsf{V}}_{1,\mathrm{bc}}} \right| \geq \Phi^{-1}\left(1 - \frac{a}{2}\right) \right) \leq a,$$

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

- t-statistic based inference, using the bias-corrected jackknife variance estimator and $N(0,1)$ critical values, provides valid inference under general assumptions

- In the panel case,

$$\widehat{\mathsf{V}}_{1,\mathrm{bc}} = \widehat{\mathsf{V}}_1 + \frac{\frac{1}{N}\sum_{i,\ell=1}^{N} \widehat{\mathbb{C}}(x_{1,it}x_{1,\ell t}, \widehat{\varepsilon}_{it}\widehat{\varepsilon}_{\ell t})}{NT \cdot \widehat{\omega}_1^2}.$$

Simulation Evidence

- We use simulation designs from Hidalgo and Schafgans (2021), Chiang, Hansen, and Sasaki (2024), and Chen and Vogelsang (2023)
- All three designs feature cross-sectional and time-series dependence
  - We fix $N = 50$ and vary T
- Compare jackknife variance estimator to following alternatives:
  - [OLS] Conventional variance estimator
  - [EHW] Eicker-Huber-White heteroskedasticity-robust estimator
  - [Ci] Cluster-robust variance estimator within i
  - [Ct] Cluster-robust variance estimator within t
  - [DK] The Driscoll-Kraay (1998) variance estimator
  - [CGM] Two-way cluster-robust variance (Cameron, Gelbach and Miller, 2011)
  - [CHS] Chiang, Hansen, and Sasaki (2024) variance estimator
  - [CHSbc] bias-corrected CHS variance estimator (Chen and Vogelsang, 2023)
  - [DKA] Driscoll-Kray and Arrelano variance estimator (Chen and Vogelsang, 2023)
  - [HS] Hidalgo and Schafgans (2021) variance estimator

Simulation Design: Hidalgo and Schafgans (2021)

- Data follow the linear model:

$$y_{it} = \alpha_t + \eta_i + \beta x_{it} + \xi_{it}, \qquad \beta = 0$$

  where α_t and η_i are mutually independent Gaussian random variables with unit mean and variance and are held fixed across simulations

- Draw location of units, s_i, as independent uniform random variable on $[0, N]$

- Define spatially-dependent error process as

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

  where

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

  and $e_{\ell t}$ are *i.i.d.* standard Gaussian random variables, $\mathbb{V}\left(\epsilon_{it}^{\xi}\right) = 1$

- x_{it} are generated same as ξ_{it}, transformed as $x_{it} \mapsto x_{it} + \mu_t$, where μ_t are mutually independent Gaussian random variables with unit mean and variance

Simulation Design: Hidalgo and Schafgans (2021)

Empirical Size (5% Nominal Size)

ρ	T	OLS	EHW	Ci	Ct	DK	CGM	CHS	CHSbc	DKA	HS	JN	JNBC
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	0.073
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	0.060
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	0.057
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	0.053
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	0.063
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	0.063
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	0.061
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	0.061

- To handle time effects, we use the transformed variables $\dot{y}_{it}$ and $\dot{x}_{it}$,

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

- Presence of time effects is unaffected by whether data have been rotated or not
- Empirical size close to nominal size across designs

Further Extensions Under Study

- **Instrumental Variables**
  - Can modify approach for IV/2SLS settings: rotate data and remove jth observation from first- <u>and</u> second-stage regressions
- **Spatial Regressions**
  - Utilize Müller & Watson (2024) framework and use as basis functions the eigenvectors of variance matrix of Levy-Brownian motion
- **Bias Correcting Estimators**
  - In paper we provide simulation evidence that JN bias correction,

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

    performs well in AR(1), standard predictive regression model
- **General Framework of Resampling "Rotated" Data**
  - Pairs bootstrap of $\{w_j, Z_j\}_{j=1}^{T}$ by sampling with replacement
  - This "rotated" bootstrap bias correction performs well for AR(1), standard predictive regression model (simulation evidence in paper)

Conclusion

- We propose a novel JN variance estimator for time-series/panel-data regressions and prove asymptotic validity of our approach
- We demonstrate excellent finite-sample properties in a series of simulation experiments using designs that have previously been investigated in the literature
- Also show that our JN approach leads to a broader framework
  - Regression diagnostic (Crump, Gospodinov, and Lopez Gaffney 2025)
  - JN bias correction (in this paper)
  - Naturally leads to consideration of pairs bootstrap on rotated data (under study)
- Our proposed JN variance estimator is also well suited to panel data models with more structure imposed
  - e.g., reduced-rank regression models prevalent in empirical finance (Crump, Gospodinov, and Lopez Gaffney 2026)

APPENDIX

Additional Simulation Evidence: Chiang, Hansen and Sasaki (2024)

Simulation Design: Chiang, Hansen and Sasaki (2024)

- Data follow the linear model:

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

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

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

and $(w_\alpha^x, w_\gamma^x, w_\epsilon^x) = (w_\alpha^\xi, w_\gamma^\xi, w_\epsilon^\xi) = (.25, .5, .25)$.

- $(\alpha_i^x, \alpha_i^\xi, \epsilon_{it}^x, \epsilon_{it}^\xi, \tilde{\gamma}_t^x, \tilde{\gamma}_t^\xi)$ and initial conditions for $\tilde{\gamma}_t^x$ and $\tilde{\gamma}_t^\xi$, are mutually independent standard Gaussian random variables

- Note that γ_t^x and γ_t^ξ generate cross-sectional and time-series dependence

Simulation Design: Chiang, Hansen and Sasaki (2024)

Empirical Size (Nominal Size: 5%)

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

- JN variance estimator: empirical size close to nominal size across all designs
  - Between 4.5% and 6.2% for all specifications
  - All other approaches exhibit size distortions – especially for large ρ