ARTICLE IN PRESS

Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.sciencedirect.com/journal/econometrics-and-statistics

Spectral Omnibus test for cross-sectional dependence in panel models

Marcell T. Kurbucz [a,*], Betsabé Pérez Garrido [b], Antal Jakovác [c,d]

[a] Institute for Global Prosperity, The Bartlett, University College London, 9–11 Endsleigh Gardens, London, WC1H 0EH, United Kingdom
[b] Department of Computer Science, Corvinus University of Budapest, 8 Fővám Square, Budapest, 1093, Hungary
[c] Department of Computational Sciences, HUN-REN Wigner Research Centre for Physics, 29–33 Konkoly–Thege Miklós Street, Budapest, 1121, Hungary
[d] Department of Statistics, Corvinus University of Budapest, 8 Fővám Square, Budapest, 1093, Hungary

ARTICLE INFO

Article history:
Received 7 November 2025
Received in revised form 28 March 2026
Accepted 29 March 2026

Keywords:
Panel data
Residual diagnostics
Cross-sectional dependence
Spectral methods
Eigenvalues
Random matrix theory

ABSTRACT

The Spectral Omnibus test (SPECO) is introduced as a diagnostic for assessing departures from cross-sectional independence in panel model residuals. SPECO operates on the eigenvalue spectrum of the residual correlation matrix and aggregates six complementary spectral indicators — capturing dominance, separation, concentration, and disorder — into a single omnibus decision. For each indicator, empirical significance values are obtained from a Monte Carlo null cache indexed by panel dimension and combined using the Cauchy method, yielding reliable finite-sample inference without relying on large-sample edge approximations. Extended simulations spanning global (linear and nonlinear), structured (sparse and block), and robustness (temporal and non-Gaussian) dependence structures show that all procedures achieve nominal size after empirical calibration. In power comparisons, SPECO attains near-unit power under linear and monotonic dependence and delivers substantial gains under oscillatory, sign-varying alternatives, where standard moment-based and pairwise diagnostics can exhibit substantially reduced power. SPECO also remains stable under heavy-tailed errors, Gaussian mixtures, heterogeneous panels, and moderate temporal dependence. Overall, SPECO provides a computationally efficient, broadly applicable diagnostic when the form of cross-sectional dependence is unknown.

1. Introduction

Panel data models, which integrate both cross-sectional and time-series information, play a central role in empirical economics and applied statistics. Their main advantage lies in the ability to control for unobserved heterogeneity across entities while exploiting temporal variation within them, thereby improving efficiency and reducing omitted-variable bias. However, the validity of these models depends critically on several assumptions about the behavior of the residuals—notably, cross-sectional independence, homoscedasticity, and the absence of serial correlation or other systematic dependence (Meijer et al., 2017). When these assumptions are violated, parameter estimates may become biased or inconsistent, and standard inferential procedures can yield misleading conclusions (Garba et al., 2013).

* Corresponding author.
E-mail address: m.kurbucz@ucl.ac.uk (M.T. Kurbucz).

https://doi.org/10.1016/j.ecosta.2026.03.002

Please cite this article as: M.T. Kurbucz, B.P. Garrido and A. Jakovác, Spectral Omnibus test for cross-sectional dependence in panel models. Econometrics and Statistics (2026), https://doi.org/10.1016/j.ecosta.2026.03.002.

To safeguard against such misspecifications, numerous diagnostic tests have been proposed. Cross-sectional dependence (CSD) is commonly assessed using the Lagrange Multiplier (LM) test of Breusch and Pagan (1980) and its large-sample extensions by Pesaran (2004, 2021), which evaluate the average pairwise correlation among residuals. Rank-based and nonparametric alternatives such as those of Frees (1995) and Feng et al. (2021) have been developed to detect monotonic or nonlinear dependence. Additional diagnostics — including the Breusch–Pagan and White tests for heteroscedasticity (Breusch and Pagan, 1979; White, 1980), and the Jarque–Bera test for normality (Jarque and Bera, 1980) — are widely used to ensure model adequacy. Stationarity checks via panel unit-root and stationarity tests (Levin et al., 2002; Im et al., 2003; Hadri and Rao, 2008) also remain standard practice.

While these diagnostics have become routine, most focus on a single property of residuals — such as correlation, variance, or distributional shape — and are primarily sensitive to linear or monotonic forms of dependence. Empirical data often exhibit complex, nonlinear, or oscillatory dependence structures that cannot be adequately captured by moment-based or pairwise methods alone. Recent developments in random matrix theory (RMT) and spectral analysis extend this perspective: Feng et al. (2022) proposed dual max–sum tests exploiting both the largest and aggregate residual correlations, while Huang et al. (2025) established a unified RMT-based omnibus framework for cross-sectional independence in large panels. These studies demonstrate that spectral measures — such as eigenvalue traces, entropy, or extreme eigenvalues — can characterize both weak and strong forms of residual dependence.

The Spectral Omnibus test (SPECO) is a finite-sample diagnostic for cross-sectional independence based on the eigenvalue spectrum of the residual correlation matrix. The test is designed to be sensitive to both linear and nonlinear forms of dependence by summarizing six spectral characteristics — including the leading eigenvalue, eigenvalue gaps, partial trace of the top-k eigenvalues, the linear spectral statistic (LSS), and spectral entropy (top-k entropy is used by default for numerical stability and to focus on the leading spectral mass; sensitivity to k is examined in Section 3.7) — into a single omnibus statistic. Each component is benchmarked against its empirical null distribution, and the resulting p-values are aggregated using the Cauchy combination method of Liu and Xie (2020), extended by Chen (2022) to ensure validity under dependence among components. This yields reliable finite-sample inference via empirical calibration, without relying on large-sample edge approximations, and exhibits stable performance under heavy-tailed and non-Gaussian errors in the simulations presented in Section 3.

Throughout, the target hypothesis is *cross-sectional independence of residuals*. Although robustness under moderate temporal dependence is also reported, SPECO is not intended as a dedicated serial-correlation test. When serial dependence is a primary concern, standard time-series diagnostics (e.g., Wooldridge, 2010) remain the appropriate first-line tools; SPECO is then applied to assess remaining cross-sectional structure in the (appropriately filtered) residuals.

The remainder of the article is organized as follows. Section 2 presents the methodological framework, detailing the construction of SPECO, benchmark diagnostics, and the extended Monte Carlo design. Section 3 presents empirical results across linear, nonlinear, sparse, dense, temporal, and non-Gaussian alternatives. Section 4 concludes with practical recommendations and directions for future research.

2. Methodology and experimental settings

2.1. Model framework and assumptions

A general panel data model of the form

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta}_i + u_{it}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{1}$$

is considered, where $\boldsymbol{\beta}_i$ may be unit-specific (heterogeneous panel) or common across units (homogeneous panel). After estimation via OLS, within-group, or other panel estimators, residuals $\hat{u}_{it}$ are obtained and row-standardized to form the $N \times T$ matrix $\mathbf{E}$, where each row is centered and scaled to unit variance.

Asymptotic approximations for CSD diagnostics depend on the panel regime and the dependence structure. The Breusch–Pagan LM test, for instance, is derived for fixed N and large T (Breusch and Pagan, 1980), while large-panel variants and bias corrections target settings with $(N, T) \to \infty$ (Pesaran, 2004; Pesaran et al., 2008; Pesaran, 2021); moreover, correlation-based limits can change under latent factor structures (Xie and Pesaran, 2022). Empirical calibration is therefore adopted: for each (N, T) configuration, a Monte Carlo null cache (Section 2.3) is constructed and all procedures are calibrated to the same finite-sample null regime. For SPECO, Rank/RankAbs, MaxSum, and the RMT benchmark, empirical p-values are computed from the cache. For CD and LM the cache is used to standardize the statistics (finite-sample mean and variance) and two-sided Gaussian-reference p-values are reported; size matching is verified in Table 4.

2.2. Spectral Omnibus Test (SPECO)

Let $\mathbf{E} \in \mathbb{R}^{N \times T}$ denote the matrix of row-standardized residuals. The unnormalized residual covariance matrix is defined as

$$\mathbf{S} = \frac{1}{T}\mathbf{E}\mathbf{E}^{\top}, \tag{2}$$

Table 1
Spectral indicators used by SPECO. Each feature targets a distinct aspect of the eigenvalue spectrum, ensuring complementary coverage of dependence structures.

Indicator	Definition	Interpretation	Detects
Edge	λ_1	Dominant eigenvalue	Pervasive factor strength
First gap	$\lambda_1 - \lambda_2$	Separation between top eigenmodes	Spiked vs. diffuse spectrum
Gap ratio	$\lambda_1/(\lambda_2 + \varepsilon)$	Relative dominance of λ_1	Single vs. multi-factor structure
Partial trace	$\sum_{i=1}^{k} \lambda_i$	Leading spectral mass (top-k)	Distributed dependencies
LSS	$\sum_{i=1}^{k} \lambda_i^2$	Concentration/inequality	Few strong vs. many weak modes
Entropy	$-\sum_{i=1}^{k} w_i \log w_i$	Spectral disorder (top-k)	Uniform vs. uneven spectrum

the diagonal scaling matrix (with a small ridge term $\varepsilon \approx 10^{-12}$ for numerical stability) as

$$\mathbf{D} = \mathrm{diag}(\mathbf{S}) + \varepsilon\, \mathbf{I}_N, \tag{3}$$

and the residual correlation matrix as

$$\mathbf{R} = \mathbf{D}^{-1/2}\, \mathbf{S}\, \mathbf{D}^{-1/2}. \tag{4}$$

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ denote the eigenvalues of $\mathbf{R}$. SPECO combines six complementary spectral indicators (Table 1) to capture different aspects of dependence.

The six indicators are chosen for their complementary geometric signatures in the eigenvalue distribution. The first three — the edge eigenvalue, the first gap, and the gap ratio — are designed to detect dominant factor structures such as linear cross-sectional dependence and block patterns. In contrast, the partial trace and the linear spectral statistic capture distributed spectral mass, making them sensitive to sparse dependencies and diffuse patterns where no single eigenvalue dominates. Finally, spectral entropy responds to disorder in the eigenvalue distribution, and can be effective at detecting oscillatory and sign-varying residual patterns.

Moment-based diagnostics such as CD and LM are deliberately excluded from the spectral aggregation. These tests operate on correlation-moment summaries rather than on the eigenvalue decomposition, and combining moment- and spectrum-based statistics would require modeling their joint null distribution under dependence. Restricting attention to spectral features allows application of the Cauchy combination with minimal additional assumptions, relying on Liu and Xie (2020) for validity under arbitrary dependence among the component statistics.

Entropy and trace-based statistics are computed from the top k eigenvalues, with $k = 5$ as the default throughout. Section 3.7 reports sensitivity analyses showing that power is stable for $k \in \{3, 5, 7\}$ and increases modestly at $k = 10$. Accordingly, $k = 5$ is recommended as a conservative default that balances signal capture against noise inclusion. For panels with fewer than 50 units, or when only one or two factors are suspected, $k = 3$ may reduce dilution from noise eigenvalues; conversely, for high-dimensional panels with $N \geq 100$, larger values such as $k \in \{7, 10\}$ may capture additional distributional features without excessive contamination. Entropy is computed from normalized top-k masses:

$$w_i = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j}, \qquad H = -\sum_{i=1}^{k} w_i \log w_i. \tag{5}$$

2.3. Empirical calibration and null cache

To control finite-sample size at each (N, T), an empirical null cache is precomputed by simulating B Gaussian panels with i.i.d. $\mathcal{N}(0, 1)$ entries and row-standardizing the resulting residual matrices. For each simulated matrix, the six spectral statistics are recorded. The empirical p-value for an observed statistic $\mathcal{T}$ is

$$p = \Pr_0(\mathcal{T} \geq t) \approx \frac{\#\{\mathcal{T}^{(b)} \geq t\} + 1}{B + 1}, \tag{6}$$

for right-tailed indicators (edge, gaps, trace, LSS), and

$$p = \Pr_0(H \leq h) \approx \frac{\#\{H^{(b)} \leq h\} + 1}{B + 1}, \tag{7}$$

for entropy (left-tailed, as lower entropy indicates stronger structure).

The six component p-values $(p_1, \ldots, p_6)$ are combined via the Cauchy Combination Test (Liu and Xie, 2020):

$$p_{\mathrm{CCT}} = \tfrac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{1}{6} \sum_{j=1}^{6} \tan[(0.5 - p_j)\pi] \right). \tag{8}$$

The resulting omnibus p-value is denoted $p_{\mathrm{SPECO}} \equiv p_{\mathrm{CCT}}$.

Motivated by Chen (2022), two hybrid aggregation schemes are also considered, designed to improve power when components exhibit heterogeneous signal strengths. The Max–Mean Combination (MCM) applies both a Bonferroni-corrected minimum and the Cauchy mean on the same scale, taking $T_{\mathrm{MCM}} = \max(T_{\mathrm{mean}}, T_{\mathrm{min}})$ where T_{mean} is the mean of the Cauchy-transformed p-values and $T_{\mathrm{min}} = \tan[(0.5 - p_{\mathrm{min}} \cdot m)\pi]$, with $p_{\mathrm{MCM}} = 1 - F_C(T_{\mathrm{MCM}})$. The Cauchy–Max Combination (CMC) uses a weighted mixture, $T_{\mathrm{CMC}} = (1 - w)\,T_{\mathrm{mean}} + w\,T_{\mathrm{min}}$, with $w = 1/(1 + \sqrt{m})$, where $m = 6$ is the number of features. Section 3.8 compares all three approaches. The standard Cauchy is retained as the default for comparability with prior work and computational simplicity.

The null of cross-sectional independence is rejected when $p_{\mathrm{SPECO}} < \alpha$.

2.4. Benchmark tests

SPECO is compared with six established diagnostics. The Pesaran CD test (Pesaran, 2004) targets average pairwise correlation via the statistic $\mathrm{CD} = \sqrt{2T/[N(N-1)]}\sum_{i<j} r_{ij}$, while the Breusch–Pagan LM test (Breusch and Pagan, 1980) focuses on squared correlations through $\mathrm{LM} = T\sum_{i<j} r_{ij}^2$, making it sensitive to pervasive weak dependence. Rank-based Kendall diagnostics (Feng et al., 2021) compute signed and absolute sums of the upper-triangular entries of the Kendall correlation matrix, capturing monotonic associations that may be missed by moment-based methods.

Recent random matrix theory benchmarks are also included. The Max–Sum test of Feng et al. (2022) exploits both the maximum and the sum of absolute pairwise correlations. Specifically, $\mathrm{MaxSum}_{\max} = \max_{i<j}|r_{ij}|$ and $\mathrm{MaxSum}_{\sum} = \sum_{i<j} r_{ij}^2$ are computed, and their empirical p-values are aggregated via Bonferroni correction, $p_{\mathrm{MaxSum}} = \min\{1,\ 2\min(p_{\max}, p_{\sum})\}$. Additionally, a simplified version of the Unified RMT test proposed by Huang et al. (2025) is implemented, using a linear spectral statistic given by the sum of squared eigenvalues of the residual correlation matrix, $\sum_{i=1}^{N}\lambda_i^2$, as the test statistic.

For all benchmark procedures, the same (N, T)-indexed Monte Carlo null cache is used to ensure a common finite-sample calibration regime. For SPECO, Rank/RankAbs, MaxSum, and the RMT benchmark, p-values are computed empirically from the cache using the appropriate tail(s). For CD and LM, the cache is instead used for finite-sample z-calibration (null mean and variance), i.e. $z = (S - \mu_0)/\sigma_0$, and two-sided Gaussian-reference p-values $p = 2\{1 - \Phi(|z|)\}$ are reported. Size matching under this mixed calibration is verified in Table 4, and all subsequent power comparisons are therefore size-matched within each (N, T) configuration.

2.5. Interpreting nonlinear residual dependence

The detection of nonlinear residual dependence raises natural questions regarding its economic and practical relevance. While many panel specifications incorporate linear structure through regressors or factors, several mechanisms can generate residual patterns that are not well described by linear dependence alone. Omitted threshold or regime effects — such as crisis versus normal periods in macroeconomic panels or discrete growth regimes — may induce discontinuous dependence structures that escape linear specifications. Likewise, unmodeled spatial or network spillovers can produce sign-varying or state-dependent interactions that are not well captured by conventional linear spatial models. Time-varying volatility clustering, common in financial panels but absent from static factor structures, may also induce nonlinear residual dependence. In addition, measurement error combined with heteroskedastic disturbances can generate complex cross-sectional dependence patterns in standardized residuals.

When such structure is detected, further specification analysis may be warranted. Researchers may investigate whether nonlinear terms, interaction effects, or regime-switching dynamics account for the observed dependence. Threshold models in the spirit of Hansen (1999) or factor models with interaction effects provide natural extensions. Robustness checks — including estimation with panel-robust standard errors or tests for time-varying parameters — can help determine whether the detected dependence materially affects inference. If the evidence suggests structural nonlinearity, the underlying linear specification may require augmentation, for example through GARCH-type dynamics, enriched spatial lag structures, or regime-dependent factor loadings.

2.6. Monte Carlo simulation design

Finite-sample size and power are evaluated through an extended Monte Carlo design with panels of dimension $N \times T$, where $T = 50$ and $N \in \{30, 50, 100\}$. Table 2 summarizes the configuration.

Implementation is carried out in `R`, using `RSpectra` (v0.16-2) for eigenvalue computations, `Matrix` (v1.7-3) for sparse algebra, and `ggplot2` (v3.5.2) for visualization. All seeds, parameters, and configurations are logged for reproducibility.

2.6.1. Data-generating processes (DGPs)

Thirteen DGPs are considered, grouped into three classes: global cross-sectional dependence designs, structured dependence designs (sparse or clustered), and robustness designs incorporating temporal and distributional deviations (Table 3).

This tripartite classification (global, structured, robustness) provides a transparent taxonomy for interpreting test performance across fundamentally different dependence mechanisms.

Table 2
Monte Carlo configuration for extended experiments.

Parameter	Description	Value(s)
N	Cross-sectional units	{30, 50, 100}
T	Time periods	50
n_{rep} (null)	Replications for size estimation	500
n_{rep} (alt.)	Replications for power estimation	200
B (null cache)	Simulated null draws (size/power)	800/500
α	Nominal significance level	0.05
Calibration	Empirical standardization	Yes (all tests)
k values tested	Eigenvalue truncation	{3, 5, 7, 10}
Combination methods	Aggregation schemes	Cauchy, MCM, CMC

Table 3
Extended data-generating processes (DGPs). Signal parameter θ or $\rho \in \{0.2, 0.5, 0.8\}$ unless stated.

Type	DGP	Description	Purpose
Global	Null	i.i.d. $\mathcal{N}(0, 1)$ noise	Size validation
	Linear	Common factor: $a\mathbf{1}_N F_t + \eta_{it}$, $a = \sqrt{\rho/(1-\rho)}$	Pervasive linear CSD
	Monotonic (scale)	Time-varying scale: $\sigma_t = \exp(3\theta F_t)$	Variance-based dependence
	Oscillatory	$g_\theta(F_t) = \theta[\sin(F_t) + 0.3\cos(F_t^2)]$	Smooth non-monotonic
	Adversarial mix	$\theta F_t \cdot \text{sgn}(\sin(3F_t))$	Sharp sign inversions
	Time warp	Deterministic $F(t)$ with $\sin(2F(t))$ warp	Phase-shifted oscillations
	Checkerboard	Alternating ± 1 segments in t	Block-structured oscillations
Structured	Sparse	20% of units correlated via aF_t	Diluted dependencies
	Block (dense)	3 blocks with intra-block factor	Clustered structure
Robustness	Temporal	AR(1) in t: $e_{it} = \rho_{\text{AR}} e_{i,t-1} + \epsilon_{it}$	Serial dependence robustness
	Heavy-tailed	Linear factor with t_5 errors	Robustness to heavy tails
	Mixture	80% $\mathcal{N}(0, 1)$ + 20% $\mathcal{N}(0, 4)$	Non-Gaussianity
	Heterogeneous	Unit-specific loadings: $\beta_i \sim \mathcal{N}(\theta, 0.5\theta)$	Estimation heterogeneity

Table 4
Empirical size at $\alpha = 0.05$ under the Null DGP ($T = 50$, $n_{\text{rep}} = 500$).

N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
30	0.044	0.044	0.050	0.044	0.046	0.042	0.044
50	0.042	0.042	0.052	0.056	0.052	0.044	0.052
100	0.060	0.062	0.048	0.038	0.072	0.070	0.064

3. Results and discussion

Size control under the null hypothesis is verified first, followed by power across the global dependence designs (see Table 3). Structured dependence designs (sparse and block patterns) and robustness scenarios are then studied, and the section concludes with sensitivity analyses for k and the aggregation rule.

3.1. Size control under the null

Table 4 reports empirical size at $\alpha = 0.05$ after calibration. All tests maintain size $\in [0.038, 0.072]$, confirming that empirical standardization successfully equalizes rejection rates across (N, T) regimes. SPECO specifically achieves size $\in [0.046, 0.072]$ across $N \in \{30, 50, 100\}$.

At $N = 100$, SPECO and MaxSum are mildly liberal, with empirical sizes of 0.072 and 0.070 at $\alpha = 0.05$. Given $n_{\text{rep}} = 500$, the Monte Carlo standard error at $p \approx 0.05$ is about $\sqrt{0.05 \cdot 0.95/500} \approx 0.010$, so deviations of this magnitude are consistent with simulation noise and finite-cache variability. Importantly, all methods are evaluated under the same (N, T)-indexed empirical calibration, so the ensuing power comparisons remain size-matched and are not driven by regime-specific reference distributions.

3.2. Power across global scenarios

Fig. 1 summarizes power for the six original DGPs (Linear, Monotonic, Oscillatory, Adversarial, Warp, Checkerboard). Detailed results are in Appendix Tables A.1–A.6.

Under linear dependence (Table A.1), all tests achieve 100% power at $\rho \geq 0.2$, confirming that spectral aggregation does not sacrifice efficiency when dependence follows a simple common-factor structure. This result is reassuring, as it demonstrates that the multi-feature approach does not dilute power against the baseline alternative for which traditional tests were specifically designed.

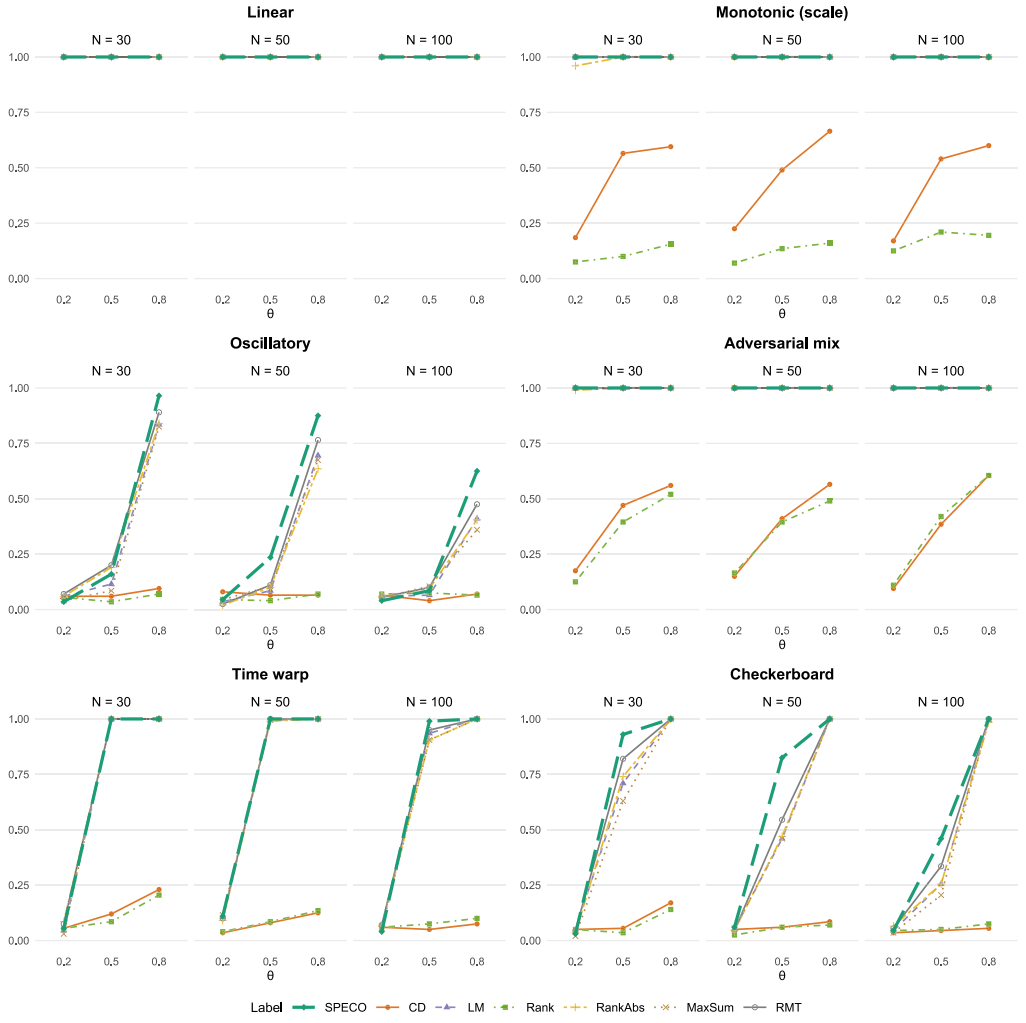

Fig. 1. Empirical power of SPECO and benchmark diagnostics across six global DGPs. Panels facet by $N \in \{30, 50, 100\}$; $\alpha = 0.05$, $n_{\text{rep}} = 200$. SPECO consistently matches or exceeds benchmarks, with largest gains under oscillatory patterns.

For monotonic scale dependence (Table A.2), SPECO, LM, MaxSum, and RMT all achieve 100% power across all configurations, reflecting their shared sensitivity to variance-based patterns. In contrast, the CD test underperforms substantially (17%–67%), as it targets mean correlations rather than variance shifts. The rank-based signed diagnostic similarly lags (7%–21%), consistent with scale-driven dependence being more naturally detected through second-moment and spectral inflation than through signed rank association.

The most notable differences emerge under oscillatory dependence (Table A.3). At strong signal strength ($\theta = 0.8$, $N = 30$), SPECO achieves 96.5% power, outperforming RMT (89%), LM (83.5%), and RankAbs (84%). Meanwhile, MaxSum reaches 82.5% and CD only 9.5%, reflecting the difficulty of detecting sign-varying patterns via pairwise correlation averages. As panel size increases to $N = 100$, SPECO maintains 62.5% power while MaxSum drops to 36%, RMT to 47.5%, and CD to 7%. This performance gap underscores SPECO's advantage: by aggregating multiple spectral features sensitive to different forms of structure, it avoids the power loss that affects tests relying on a single summary statistic.

Similar patterns persist across the remaining nonlinear scenarios. Under adversarial mix dependence (Table A.4), LM, SPECO, RMT, RankAbs, and MaxSum all achieve 99%–100% power across configurations. The time warp design (Table A.5) reveals SPECO achieving 100% at $\theta = 0.8$ across all N values, while the checkerboard alternative (Table A.6) produces SPECO power of 82.5%–100% at $\theta \geq 0.5$, substantially exceeding CD (4.5%–17%) and maintaining a consistent edge over LM and RMT.

3.3. Sparse and block dependence

Table 5 summarizes results for sparse (20% correlated units) and block (3-block dense) structures at weak signal strength.

Table 5
Power under sparse and dense CSD at $\theta = 0.2$ (weak signal). Full results in Tables A.7–A.8.

DGP	N	CD	LM	SPECO	MaxSum	RMT
Sparse	30	0.14	0.08	0.15	0.09	0.08
	50	0.22	0.17	0.28	0.20	0.20
	100	0.65	0.40	0.73	0.40	0.47
Block (dense)	30	0.84	0.30	0.43	0.31	0.36
	50	0.99	0.59	0.66	0.60	0.67
	100	1.00	0.97	0.99	0.96	0.99

Table 6
Power under temporal dependence (AR(1) structure, $N = 50$, $T = 50$). Full results in Table A.9.

ρ_{AR}	CD	LM	SPECO	MaxSum	RMT
0.2	0.06	0.46	0.38	0.38	0.59
0.5	0.15	1.00	1.00	1.00	1.00
0.8	0.28	1.00	1.00	1.00	1.00

Table 7
Power under non-Gaussian errors at $\theta = 0.2$ ($N = 50$, $T = 50$). Full results in Tables A.10–A.11.

Error type	CD	LM	SPECO	MaxSum	RMT
Heavy-tailed (t_5)	1.00	1.00	1.00	1.00	1.00
Mixture (80%–20%)	1.00	0.89	0.99	0.89	0.92

Under sparse CSD, SPECO achieves 15%–73% power compared to 14%–65% for CD and 8%–40% for LM, demonstrating superior sensitivity to diluted dependencies where only a minority of units exhibit correlation. This advantage stems from SPECO's ability to detect subtle shifts in leading spectral mass and entropy measures that remain informative even when average pairwise correlations are weak. Conversely, under block-structured dense CSD, the CD test dominates with 84%–100% power, aligning with its design for detecting pervasive average correlation. SPECO achieves 43%–99% in this setting, converging toward CD as panel size increases and the block structure becomes more evident in the leading eigenvalues.

At moderate-to-strong signals ($\theta \geq 0.5$), all tests converge to near-perfect power regardless of structure (Tables A.7–A.8), confirming that test differences primarily manifest at the weak-signal boundary where practitioners face the greatest diagnostic uncertainty.

3.4. Temporal dependence

Table 6 reports power under AR(1) temporal structure: $e_{it} = \rho_{AR} e_{i,t-1} + \sqrt{1 - \rho_{AR}^2}\,\epsilon_{it}$.

SPECO maintains competitive power (38%–100%) under temporal dependence, comparable to LM and MaxSum. The RMT test achieves the highest power at weak temporal signals ($\rho_{AR} = 0.2$: 59% vs. 38% for SPECO), reflecting the sensitivity of linear spectral statistics to eigenvalue inflation induced by serial dependence. The CD test underperforms (6%–28%), as serial autocorrelation need not manifest as strong cross-sectional correlation in row-standardized residuals. An important caveat applies: SPECO is not specifically designed for serial correlation detection, and practitioners requiring joint assessment of temporal and cross-sectional structure should apply dedicated serial correlation diagnostics (Wooldridge, 2010) before employing SPECO for cross-sectional patterns.

3.5. Robustness to non-Gaussian errors

Table 7 reports power under heavy-tailed (t_5) and Gaussian mixture errors at weak signal strength.

Under heavy-tailed errors following a t_5 distribution, all tests achieve perfect or near-perfect power (100%), confirming that row-standardization effectively mitigates marginal heavy tails and preserves practical detectability in this design. For Gaussian mixture errors (80% standard normal, 20% high-variance normal), SPECO achieves 98.5% power at $N = 50$, trailing the CD test (100%) but outperforming LM (88.5%) and MaxSum (89%). This pattern suggests that SPECO maintains robust performance across distributional perturbations while CD's focus on pairwise correlation remains particularly stable under this mixture setting.

When additional dataset-specific calibration is required, a parametric bootstrap procedure is recommended (Appendix A.13). Simulations confirm that this procedure maintains nominal size but incurs substantial computational cost—approximately 20 times slower than cache-based inference. Consequently, it should be reserved for cases with strong evidence of extreme distributional departures and adequate sample size ($N, T \geq 50$) to support the additional computational burden.

Table 8
Power under heterogeneous panel at $\theta = 0.2$ ($T = 50$). Full results in Table A.12.

N	CD	LM	SPECO	MaxSum	RMT
30	0.93	0.24	0.51	0.28	0.32
50	1.00	0.56	0.81	0.51	0.63
100	1.00	0.89	0.97	0.86	0.92

Table 9
Sensitivity of SPECO power to k parameter (Oscillatory DGP, $\theta = 0.5$, $N = 50$, $n_{\text{rep}} = 200$).

k	Power
3	0.125
5	0.110
7	0.130
10	0.150

Table 10
Comparison of combination methods (Oscillatory DGP, $\theta = 0.5$, $N = 50$, $n_{\text{rep}} = 200$).

Method	Power
Cauchy (CCT)	0.130
MCM	0.130
CMC	0.115

3.6. Heterogeneous panel

Table 8 reports power under unit-specific factor loadings drawn from $\beta_i \sim \mathcal{N}(\theta, 0.5\theta)$.

SPECO maintains 51%–97% power across panel sizes, demonstrating meaningful robustness to estimation-induced heterogeneity from unit-specific slopes. The CD test dominates at all panel sizes (93%–100%), as the heterogeneous loadings produce pervasive average correlation that CD is designed to detect. SPECO's intermediate performance is consistent with heterogeneous factor structures expressing strongly in mean correlation, even when the leading spectral signatures are less sharply separated. At moderate-to-strong signals ($\theta \geq 0.5$), all tests achieve 100% power.

3.7. Sensitivity to k parameter

Table 9 reports power for varying eigenvalue truncation levels under the Oscillatory DGP at moderate signal strength.

Power exhibits stability at approximately 11%–13% for $k \in \{3, 5, 7\}$ before increasing to 15% at $k = 10$. This pattern suggests that the default choice of $k = 5$ captures the primary spectral features while avoiding excessive noise contamination. The modest gain at $k = 10$ indicates that in some scenarios larger truncation may be beneficial by incorporating distributional information beyond the leading few eigenvalues. Nevertheless, the narrow range of power across $k \in \{3, 10\}$ (12.5%–15%) provides reassurance that the main conclusions remain qualitatively robust to this tuning parameter.

3.8. Combination method comparison

Table 10 compares aggregation schemes under the Oscillatory DGP at moderate signal strength.

Both the standard Cauchy and Max–Mean Combination achieve 13% power, while the Cauchy–Max Combination yields 11.5%. In this moderate-signal scenario, the three methods perform comparably, with differences well within simulation uncertainty ($\pm 2.4\%$ at the $n_{\text{rep}} = 200$ level). The CMC's slight underperformance may reflect the fact that its Bonferroni component is conservative when evidence is spread across multiple indicators, whereas the mean-based Cauchy statistic aggregates diffuse signals more efficiently. The standard Cauchy is retained as the default for its simplicity, theoretical tractability, and competitive performance across all scenarios examined.

Remark on minor numeric differences across auxiliary tables. The k-sensitivity (Table 9), combination-method comparison (Table 10), and ablation study (Table 11) are generated as separate Monte Carlo experiments; small differences in baseline power across these tables reflect simulation noise rather than a change in the underlying DGP.

3.9. Feature ablation study

Table 11 reports results from a leave-one-out analysis assessing the contribution of each spectral feature.

Removing trace or LSS produces the largest power drops (14% to 12%), confirming that these mass-based features contribute meaningfully to detecting distributed spectral patterns. The edge, gap, and ratio features show minimal

Table 11
Feature ablation study (Oscillatory DGP, $\theta = 0.5$, $N = 50$, $n_{\text{rep}} = 200$). "None" denotes the full six-feature SPECO.

Removed	Power
None (full)	0.140
Edge	0.145
Gap	0.145
Ratio	0.150
Trace	0.120
LSS	0.120
Entropy	0.150

individual impact in this oscillatory scenario, consistent with the DGP not generating a dominant single-factor spike. Entropy removal also has minimal effect here, though its contribution can be more pronounced under other dependence structures. The modest overall spread (12%–15%) reflects the moderate signal strength ($\theta = 0.5$) at which these differences manifest; at stronger signals all feature subsets converge to high power. This analysis supports retaining all six features as the default, since no single feature is redundant across the full range of alternatives considered.

4. Conclusions and future work

The Spectral Omnibus Test for Cross-Sectional Dependence (SPECO) provides a compact, finite-sample diagnostic for cross-sectional independence in panel models. By aggregating six complementary features of the residual correlation spectrum — edge, gaps, trace, LSS, and entropy — into a single p-value via Cauchy combination, the test achieves robust performance across global (linear and nonlinear), structured (sparse and block), and robustness (temporal and non-Gaussian) dependence structures.

The extended simulation study encompasses twelve data-generating processes spanning linear and nonlinear patterns, sparse and dense structures, temporal dependence, heavy-tailed errors, Gaussian mixtures, and heterogeneous panels. Direct comparisons with recent random matrix theory benchmarks — specifically the Max–Sum test of Feng et al. (2022) and the Unified RMT approach of Huang et al. (2025) — reveal that while all tests achieve comparable power under linear and monotonic dependence, SPECO exhibits clear advantages under oscillatory and non-monotonic structures. At $\theta = 0.8$, $N = 50$, SPECO reaches 87.5% power under oscillatory dependence where RMT achieves 76.5% and MaxSum 67.5%, with CD effectively powerless at 6.5%.

Sensitivity analyses establish that power remains stable across eigenvalue truncation levels $k \in \{3, 5, 7, 10\}$ (approximately 11%–15%), supporting the default recommendation of $k = 5$. Comparison of combination methods reveals that the standard Cauchy, MCM, and CMC achieve comparable power (11.5%–13%) under the oscillatory DGP at moderate signal strength, supporting the standard Cauchy as a stable and parsimonious default.

Practical recommendations emerge naturally from these findings. When the true dependence structure is unknown — the typical situation facing practitioners — SPECO provides a balanced primary diagnostic that performs competitively across a broad range of alternatives without requiring prior specification of the dependence form. In settings characterized by clear block or dense correlation patterns, combining SPECO with the CD test may leverage complementary strengths, while variance-driven dependence structures may warrant pairing with the LM test. High-dimensional panels with $N \geq 100$ units may benefit from larger eigenvalue truncation ($k \in \{7, 10\}$).

Several limitations suggest directions for future work. Theoretical foundations remain incomplete: formal derivations of finite-(N, T) null distributions for the six spectral features and characterization of power envelopes under specific alternatives would strengthen the methodological foundation. Software implementation represents a natural next step, with open-source `R` and `Python` packages featuring versioned null caches enabling widespread adoption. Real-data applications across macro-finance, regional growth, and trade networks would demonstrate practical value and quantify re-specification gains from detecting previously unrecognized residual patterns. Extensions incorporating spatial and network structures, panel VAR residuals, and dynamic factor models would broaden applicability, while adaptive aggregation schemes — such as BIC-based feature selection or data-driven combination weights — might further improve power while preserving finite-sample validity.

By providing a single, well-calibrated test that performs competitively across diverse dependence structures, SPECO reduces reliance on multiple sequential diagnostics and offers practitioners a reliable tool for residual validation in modern panel workflows.

CRediT authorship contribution statement

Marcell T. Kurbucz: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Betsabé Pérez Garrido:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Antal Jakovác:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Conceptualization.

Code availability

The `R` code used to generate the simulations, construct the null caches, and produce the figures is available from the corresponding author upon reasonable request. The implementation builds on the `RSpectra`, `Matrix`, and `ggplot2` packages.

Funding

This work was supported by the ÚNKP-23-4-II-CORVINUS-11 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgments

The authors are grateful to the editors and referees for their constructive comments and suggestions, which led to substantial improvements of the manuscript. M.T. Kurbucz acknowledges support from the ÚNKP-23-4-II-CORVINUS-11 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

Appendix. Detailed simulation results

A.1. Linear CSD (Common factor)

See Table A.1.

A.2. Monotonic (scale) CSD

See Table A.2.

A.3. Oscillatory (non-monotonic) CSD

See Table A.3.

A.4. Adversarial mix CSD

See Table A.4.

A.5. Time warp CSD

See Table A.5.

A.6. Checkerboard CSD

See Table A.6.

Table A.1
Empirical power under Linear CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\rho = 0.2$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.2$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.2$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.5$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.5$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.8$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.8$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.2
Empirical power under Monotonic (Scale) CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.185	1.00	0.075	0.96	1.00	1.00	1.00
$\theta = 0.2$, $N = 50$	0.225	1.00	0.070	1.00	1.00	1.00	1.00
$\theta = 0.2$, $N = 100$	0.170	1.00	0.125	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 30$	0.565	1.00	0.100	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	0.490	1.00	0.135	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	0.540	1.00	0.210	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	0.595	1.00	0.155	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	0.665	1.00	0.160	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	0.600	1.00	0.195	1.00	1.00	1.00	1.00

Table A.3
Empirical power under Oscillatory CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.060	0.060	0.055	0.060	0.035	0.040	0.070
$\theta = 0.2$, $N = 50$	0.080	0.035	0.045	0.020	0.045	0.045	0.025
$\theta = 0.2$, $N = 100$	0.065	0.060	0.070	0.060	0.040	0.055	0.055
$\theta = 0.5$, $N = 30$	0.060	0.115	0.035	0.190	0.160	0.085	0.200
$\theta = 0.5$, $N = 50$	0.065	0.085	0.040	0.100	0.235	0.105	0.110
$\theta = 0.5$, $N = 100$	0.040	0.065	0.075	0.095	0.085	0.105	0.100
$\theta = 0.8$, $N = 30$	0.095	0.835	0.070	0.840	0.965	0.825	0.890
$\theta = 0.8$, $N = 50$	0.065	0.695	0.070	0.635	0.875	0.675	0.765
$\theta = 0.8$, $N = 100$	0.070	0.410	0.065	0.405	0.625	0.360	0.475

Table A.4
Empirical power under Adversarial Mix CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.175	1.00	0.125	0.990	1.00	1.00	1.00
$\theta = 0.2$, $N = 50$	0.150	1.00	0.165	1.00	1.00	1.00	1.00
$\theta = 0.2$, $N = 100$	0.095	1.00	0.110	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 30$	0.470	1.00	0.395	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	0.410	1.00	0.395	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	0.385	1.00	0.420	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	0.560	1.00	0.520	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	0.565	1.00	0.490	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	0.605	1.00	0.605	1.00	1.00	1.00	1.00

Table A.5
Empirical power under Time Warp CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.055	0.045	0.055	0.085	0.055	0.030	0.075
$\theta = 0.2$, $N = 50$	0.035	0.105	0.040	0.090	0.110	0.115	0.100
$\theta = 0.2$, $N = 100$	0.060	0.070	0.060	0.065	0.040	0.055	0.070
$\theta = 0.5$, $N = 30$	0.120	1.00	0.085	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	0.080	0.995	0.085	0.990	1.00	0.995	1.00
$\theta = 0.5$, $N = 100$	0.050	0.935	0.075	0.905	0.990	0.905	0.950
$\theta = 0.8$, $N = 30$	0.230	1.00	0.205	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	0.125	1.00	0.135	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	0.075	1.00	0.100	1.00	1.00	1.00	1.00

Table A.6
Empirical power under Checkerboard CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.050	0.045	0.050	0.035	0.030	0.020	0.045
$\theta = 0.2$, $N = 50$	0.050	0.050	0.025	0.045	0.060	0.055	0.045
$\theta = 0.2$, $N = 100$	0.035	0.045	0.045	0.065	0.045	0.035	0.055
$\theta = 0.5$, $N = 30$	0.055	0.710	0.035	0.740	0.930	0.630	0.820
$\theta = 0.5$, $N = 50$	0.060	0.460	0.060	0.470	0.825	0.460	0.545
$\theta = 0.5$, $N = 100$	0.045	0.255	0.050	0.255	0.460	0.205	0.335
$\theta = 0.8$, $N = 30$	0.170	1.00	0.140	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	0.085	1.00	0.070	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	0.055	0.995	0.075	0.985	1.00	0.995	1.00

Table A.7
Empirical power under Sparse CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.140	0.075	0.140	0.065	0.150	0.090	0.080
$\theta = 0.2$, $N = 50$	0.220	0.165	0.230	0.155	0.275	0.195	0.200
$\theta = 0.2$, $N = 100$	0.645	0.400	0.610	0.375	0.725	0.395	0.470
$\theta = 0.5$, $N = 30$	0.640	0.995	0.620	0.845	0.995	0.990	0.995
$\theta = 0.5$, $N = 50$	0.910	1.00	0.915	0.995	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	0.870	1.00	0.905	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	0.995	1.00	0.995	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.8
Empirical power under Block (Dense) CSD ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.835	0.300	0.810	0.270	0.430	0.305	0.355
$\theta = 0.2$, $N = 50$	0.990	0.590	0.980	0.570	0.660	0.600	0.665
$\theta = 0.2$, $N = 100$	1.00	0.965	1.00	0.940	0.985	0.960	0.985
$\theta = 0.5$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	1.00	1.00	1.00	1.00	0.995	1.00	1.00
$\theta = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.9
Empirical power under Temporal Dependence (AR(1), $T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\rho_{\text{AR}} = 0.2$, $N = 30$	0.050	0.185	0.055	0.290	0.225	0.230	0.290
$\rho_{\text{AR}} = 0.2$, $N = 50$	0.055	0.460	0.050	0.490	0.380	0.380	0.585
$\rho_{\text{AR}} = 0.2$, $N = 100$	0.050	0.950	0.030	0.905	0.735	0.895	0.975
$\rho_{\text{AR}} = 0.5$, $N = 30$	0.125	1.00	0.110	1.00	0.995	1.00	1.00
$\rho_{\text{AR}} = 0.5$, $N = 50$	0.145	1.00	0.140	1.00	1.00	1.00	1.00
$\rho_{\text{AR}} = 0.5$, $N = 100$	0.095	1.00	0.095	1.00	1.00	1.00	1.00
$\rho_{\text{AR}} = 0.8$, $N = 30$	0.240	1.00	0.280	1.00	1.00	1.00	1.00
$\rho_{\text{AR}} = 0.8$, $N = 50$	0.280	1.00	0.255	1.00	1.00	1.00	1.00
$\rho_{\text{AR}} = 0.8$, $N = 100$	0.285	1.00	0.275	1.00	1.00	1.00	1.00

A.7. Sparse CSD (20% correlated units)

See Table A.7.

A.8. Block (Dense) CSD

See Table A.8.

A.9. Temporal dependence (AR(1))

See Table A.9.

A.10. Heavy-tailed errors (t_5 distribution)

See Table A.10.

A.11. Gaussian mixture errors (80%–20%)

See Table A.11.

A.12. Heterogeneous panel (unit-specific loadings)

See Table A.12.

Table A.10
Empirical power under Heavy-Tailed Errors (t_5, $T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	1.00	0.995	1.00	1.00	1.00	0.995	1.00
$\theta = 0.2$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.2$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.11
Empirical power under Gaussian Mixture Errors ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	1.00	0.690	1.00	0.985	0.900	0.745	0.735
$\theta = 0.2$, $N = 50$	1.00	0.885	1.00	1.00	0.985	0.890	0.920
$\theta = 0.2$, $N = 100$	1.00	0.990	1.00	1.00	1.00	0.995	0.995
$\theta = 0.5$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table A.12
Empirical power under Heterogeneous Panel ($T = 50$, $\alpha = 0.05$, $n_{\text{rep}} = 200$).

Signal, N	CD	LM	Rank	RankAbs	SPECO	MaxSum	RMT
$\theta = 0.2$, $N = 30$	0.925	0.240	0.925	0.320	0.505	0.275	0.315
$\theta = 0.2$, $N = 50$	1.00	0.555	1.00	0.565	0.810	0.505	0.625
$\theta = 0.2$, $N = 100$	1.00	0.890	1.00	0.860	0.970	0.860	0.920
$\theta = 0.5$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.5$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 30$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 50$	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\theta = 0.8$, $N = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00

A.13. Parametric bootstrap (per-dataset Monte Carlo calibration)

When panel dimensions do not match any precomputed (N, T) null cache, or when additional dataset-specific calibration is desired, a parametric bootstrap procedure under the Gaussian null is recommended. The algorithm begins by computing the observed six spectral statistics $\mathbf{s}_{\text{obs}} = (s_1, \ldots, s_6)$ from the residual matrix $\mathbf{E}$. For each bootstrap iteration $b = 1, \ldots, B_{\text{sim}}$ (with $B_{\text{sim}} = 200$ suggested), a null panel $\mathbf{E}^{(b)} \sim \mathcal{N}(0, \mathbf{I}_{N\times T})$ is generated, row-standardization is applied, and the corresponding statistics $\mathbf{s}^{(b)} = (s_1^{(b)}, \ldots, s_6^{(b)})$ are computed. Empirical p-values are obtained as

$$p_j = \frac{\#\{s_j^{(b)} \geq s_{j,\text{obs}}\} + 1}{B_{\text{sim}} + 1}, \tag{A.1}$$

for right-tailed indicators, with appropriate modification for left-tailed entropy. The six p-values are then aggregated using the Cauchy combination to obtain p_{SPECO}.

This procedure differs from the precomputed null cache (Section 2.3) only in that null panels are generated afresh for each dataset rather than drawn from a stored cache. While this approach maintains nominal size, it increases computational cost by approximately a factor of 20 relative to cache-based inference. It is therefore recommended only when precomputed null distributions are unavailable or when per-dataset calibration is required. For moderate departures from normality, as demonstrated in the heavy-tail and mixture experiments, the standard cache-based approach maintains adequate size and power.

Data availability

This study uses exclusively simulated data. The corresponding author will provide the simulation outputs, empirical null caches, figure sources, and configuration logs upon reasonable request.

References

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econ.: J. Econ. Soc. 1287–1294.
Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specification in econometrics. Rev. Econ. Stud. 47 (1), 239–253.
Chen, Z., 2022. Robust tests for combining p-values under arbitrary dependency structures. Sci. Rep. 12 (1), 3158.
Feng, L., Jiang, T., Liu, B., Xiong, W., 2022. Max-sum tests for cross-sectional independence of high-dimensional panel data. Ann. Statist. 50 (2), 1124–1143.
Feng, L., Zhao, P., Ding, Y., Liu, B., 2021. Rank-based tests of cross-sectional dependence in panel data models. Comput. Statist. Data Anal. 153, 107070.
Frees, E.W., 1995. Assessing cross-sectional correlation in panel data. J. Econometrics 69 (2), 393–414.
Garba, M.K., Oyejola, B.A., Yahya, W.B., 2013. Investigations of certain estimators for modelling panel data under violations of some basic assumptions. Math. Theory Model. 3 (10), 47–53.
Hadri, K., Rao, Y., 2008. Panel stationarity test with structural breaks. Oxf. Bull. Econ. Stat. 70 (2), 245–269.
Hansen, B.E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. J. Econometrics 93 (2), 345–368.
Huang, Z., Li, Z., Yao, J., 2025. Unified and robust tests for cross sectional independence in large panel data models. Electron. J. Stat. 19 (2), 4867–4913.
Im, K.S., Pesaran, M.H., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. J. Econometrics 115 (1), 53–74.
Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econom. Lett. 6 (3), 255–259.
Levin, A., Lin, C.-F., Chu, C.-S.J., 2002. Unit root tests in panel data: asymptotic and finite-sample properties. J. Econometrics 108 (1), 1–24.
Liu, Y., Xie, J., 2020. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J. Amer. Statist. Assoc. 115 (529), 393–402.
Meijer, E., Spierdijk, L., Wansbeek, T., 2017. Consistent estimation of linear panel data models with measurement error. J. Econometrics 200 (2), 169–180.
Pesaran, M.H., 2004. General diagnostic tests for cross section dependence in panels. Faculty of Economics, University of Cambridge, Cambridge Working Papers in Economics, No. 0435.
Pesaran, M.H., 2021. General diagnostic tests for cross-sectional dependence in panels. Empir. Econ. 60 (1), 13–50.
Pesaran, M.H., Ullah, A., Yamagata, T., 2008. A bias-adjusted LM test of error cross-section independence. Econom. J. 11 (1), 105–127.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econ.: J. Econ. Soc. 817–838.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT Press.
Xie, Y., Pesaran, M.H., 2022. A bias-corrected cd test for error cross-sectional dependence in panel data models with latent factors. Available At SSRN 4198155, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4198155.