Statistica Sinica Preprint No: SS-2023-0348	
Title	Fisher’s Combined Probability Test for Cross-Sectional Independence in Panel Data Models with Serial Correlation
Manuscript ID	SS-2023-0348
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0348
Complete List of Authors	Hongfei Wang, Binghui Liu, Long Feng and Yanyuan Ma
Corresponding Authors	Long Feng
E-mails	flnankai@nankai.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

Fisher's combined probability test for cross-sectional independence in panel data models with serial correlation

Hongfei Wang[1], Binghui Liu[2*],Long Feng[3*], Yanyuan Ma[4]

[1] *Nanjing Audit University*

[2] *Northeast Normal University*

[3] *Nankai University*

[4] *Pennsylvania State University*

Abstract:

Testing cross-sectional independence in panel data models is of fundamental importance in econometric analysis with high-dimensional panels. Recently, econometricians began to turn their attention to the problem in the presence of serial dependence. The existing procedure for testing cross-sectional independence with serial correlation is based on the sum of the sample cross-sectional correlations, which generally performs well when the alternative has dense cross-sectional correlations, but suffers from low power against sparse alternatives. To deal with sparse alternatives, we propose a test based on the maximum of the squared sample cross-sectional correlations. Furthermore, we propose a combined test

*co-corresponding authors.

to combine the p-values of the max based and sum based tests, which performs well under both dense and sparse alternatives. The combined test relies on the asymptotic independence of the max based and sum based test statistics, which we show rigorously. We show that the proposed max based and combined tests have attractive theoretical properties and demonstrate the superior performance via extensive simulation results. We demonstrate the practicality of the proposed tests through two empirical applications.

Key words and phrases: Asymptotic independence, Cross-sectional dependence, Heterogeneous panel data models, High dimensionality

1. Introduction

In this paper, we consider the problem of testing cross-sectional independence in heterogeneous panel data models. In statistics and econometrics, panel data occur frequently, which contain observations of various types obtained over multiple time periods for any single unit. In the study of panel data models, the cross-sectional dependency is an important concept, described as the interaction between cross-sectional units, which could arise from the behavioral interaction between units Breusch and Pagan (1980); Feng et al. (2022). To make theoretical study easier, experts often assume cross-sectional independence in the model setups. If data across units are dependent, inferences under the assumption of cross-sectional independence

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

would be inaccurate and misleading; see literature on spatial econometrics, such as Anselin and Bera (1998); Kelejian and Prucha (1999); Kapoor et al. (2007); Lee (2007); Lee and Yu (2010) for examples of cross-sectional dependence. Therefore, testing the existence of cross-sectional dependence is important and attracts increasing attention.

A large number of literatures on testing cross-sectional dependence are available, among which the most widely known is likely the Lagrange Multiplier (LM) test proposed by Breusch and Pagan (1980). The LM test is based on the sum of the squared cross-sectional correlations of residuals, which is applicable when the sample size T is large and the dimension N is finite, but is not a valid test when $N \to \infty$. To develop tests applicable when both N and T are large, two limit schemes have been considered in the literature. One is the sequential limit scheme: $T \to \infty$, followed by $N \to \infty$, and the other is the simultaneous limit scheme: N and T tend to infinity simultaneously, i.e. $(N, T) \to \infty$. Under the sequential limit scheme, Pesaran (2004) proposed a scaled version of the LM test, as well as a test based on the sum of the residual correlations, instead of the squared sum, to address the issue that the former test may suffer from substantial size distortions in the case of large N and small T; Pesaran et al. (2008) proposed a bias-adjusted LM test based on the scaled LM test. Later, under

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

the simultaneous limit scheme, Pesaran (2015) established the asymptotical properties of the sum based test by Pesaran (2004); Feng et al. (2022) established the asymptotical properties of the bias-adjusted LM test proposed by Pesaran et al. (2008). In addition to these tests, there are also works on testing cross-sectional independence under other related models, such as the fixed effect panel data models; see, for instance, Baltagi et al. (2012); Feng et al. (2020).

All the above tests are sum based tests, i.e., they are based on the sum of the correlations or squared correlations of the residuals. These tests generally perform well under dense alternatives, but may suffer from low power against sparse alternatives. To deal with sparse alternatives, Feng et al. (2022) proposed a max based test for testing cross-sectional dependence, and further developed a combined test that integrates the advantages of the max based and sum based tests, by establishing the asymptotical independence between the test statistics.

All the tests mentioned above make the common assumption that the errors in the panel data models are independent across time. However, the existence of serial dependence is likely to be the rule rather than the exception; see, for instance, Wei (2006); Hong (2010); Box et al. (2015). In many applications, serial dependence may have great impact on statistical inference, such as leading to deviation of the limiting spectral distribution of the sample covariance matrix (Gao et al., 2017). Hence, in analyzing panel data with serially correlated errors, using tests for cross-sectional independence that are based on the assumption that the errors are serially independent may lead to wrong conclusions.

To solve this problem, Baltagi et al. (2016) proposed a test for cross-sectional correlation under heterogeneous panel data models with serial correlations by adjusting the sum based test by Pesaran (2004). Similarly, Lan et al. (2017) proposed a test for cross-sectional independence under fixed effects panel data models with serial correlations. Both tests are sum based, hence the scope of their application is limited to dense alternatives. As far as we know, research on max based tests for sparse alternatives or combined tests regardless of whether alternatives are sparse or not is not yet available for data with serial dependence.

We aim to fill this gap in this work. To this end, for testing cross-sectional independence under heterogeneous panel data models with serial correlations, we propose a max based test based on the maximum of the squared cross-sectional residual correlations to deal with sparse alternatives. The method follows the framework of Chen and Liu (2018) for testing independence of correlated samples, while relaxing their Gaussian sample assumption. Furthermore, we propose a Fisher's combined probability test by combining the p-values of our proposed max based test and the sum based test by Baltagi et al. (2016). This combined test is applicable regardless the alternatives are sparse or dense. We derive the asymptotic null distribution of the proposed combined test, by first rigorously establishing the asymptotic independence of the max based and sum based test statistics. In summary, there are two main contributions of our work.

(1) We propose a max based test for testing cross-sectional independence in models with serial correlation and sub-Gaussian error. The new test is powerful in detecting sparse alternatives.

(2) We establish the asymptotic independence between the sum based and max based test statistics. We propose a combined test for testing cross-sectional independence in models with serial correlation and sub-Gaussian error. The new test is powerful overall.

The rest of the paper is organized as follows. We propose two new tests and establish their asymptotic properties in Section 2. Simulation results of the proposed tests and their comparison to some existing methods are demonstrated in Section 3, followed by an empirical application in Section 4. We conclusion the paper in Section 5 and relegate some additional simulation results, an additional empirical application and the technical proofs to the Supplementary Material.

Notataion. For any square matrix $\mathbf{A}$, $(\mathbf{A})_{ij}$ denotes the (i,j)-th entry of $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the maximum and the minimum eigenvalues of $\mathbf{A}$, respectively, $\|\mathbf{A}\|_{\mathrm{F}}$ denotes the Frobenius norm of $\mathbf{A}$, and $\mathbf{A}^{1/2}$ denotes the principal square root matrix of $\mathbf{A}$ if $\mathbf{A}$ is a positive definite matrix. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix for each positive integer n. For any two real numbers x and y, let $x \vee y = \max(x,y)$ and $x \wedge y = \min(x,y)$. For any vector v, $\|v\|$ denotes the Euclidean norm of v. Let $\mathcal{N}(a,b)$ denote the normal distribution with mean $a \in \mathbb{R}$ and variance $b \geq 0$, t_v denote the t-distribution with degree of freedom v, χ^2_v denote the chi-square distribution with degree of freedom v , and $U[a,b]$ denote the uniform distribution over the interval $[a,b]$, where both a and b are real numbers. Let $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal distribution. We use $(N,T) \to \infty$ to denote the joint convergence of N and T to infinity, i.e. $\min(N,T) \to \infty$.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

2. The proposed tests

2.1 Problem description

We consider the heterogeneous panel data model taking the form

$$y_{it} = x_{it}'\beta_i + \epsilon_{it},\ 1 \le i \le N,\ 1 \le t \le T, \tag{2.1}$$

where i indexes the cross-sectional units, and t indexes the time dimension. $y_{it} \in \mathbb{R}$ is the dependent variable, $x_{it} \in \mathbb{R}^p$ is the exogenous regressors with slope parameters $\beta_i \in \mathbb{R}^p$ that are allowed to vary across i and $\epsilon_{it} \in \mathbb{R}$ is the corresponding idiosyncratic error term. For each $1 \le i \le N$, let $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})' \in \mathbb{R}^{T\times p}$, $y_i = (y_{i1}, \ldots, y_{iT})' \in \mathbb{R}^T$ and $\epsilon_{i\cdot} = (\epsilon_{i1}, \ldots, \epsilon_{iT})' \in \mathbb{R}^T$. For each $1 \le t \le T$, let $\epsilon_{\cdot t} = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$. The null hypothesis of cross-sectional independence can be written as

$$H_0 : \epsilon_{1\cdot}, \epsilon_{2\cdot}, \ldots, \epsilon_{N\cdot} \text{ are independent random vectors.} \tag{2.2}$$

2.2 Related works

Early studies on testing cross-sectional independence in (2.2) are based on the assumption that there is no serial dependence in $\{\epsilon_{\cdot t}\}_{t=1}^T$, where $\epsilon_{\cdot t}$ is assumed to be iid over time t. The earliest work is the LM test (Breusch and Pagan, 1980), with test statistic

$$\mathrm{LM}_{\mathrm{BP}} = T \sum\sum_{1 \le i < j \le N} \hat{\rho}_{ij}^2,$$

where $\hat{\rho}_{ij}$ is the sample correlation constructed by the OLS residuals $\hat{\epsilon}_{it} = y_{it} - x_{it}'\hat{\beta}_i$, with

$$\hat{\beta}_i = \left(\mathbf{x}_i'\mathbf{x}_i\right)^{-1} \mathbf{x}_i' y_i \quad \text{and} \quad \hat{\rho}_{ij} = \frac{\sum_{t=1}^T \hat{\epsilon}_{it}\hat{\epsilon}_{jt}}{\sqrt{\sum_{t=1}^T \hat{\epsilon}_{it}^2 \sum_{t=1}^T \hat{\epsilon}_{jt}^2}}.$$

The asymptotic null distribution of $\mathrm{LM}_{\mathrm{BP}}$ is a chi-squared distribution with $N(N-1)/2$ degrees of freedom, which is established when N is fixed and T diverges to infinity. Hence, it is not applicable to the case of large N. To overcome the size distortions of the scaled version of the $\mathrm{LM}_{\mathrm{BP}}$ test proposed by Pesaran (2004) for large N and small T, Pesaran et al. (2008) proposed a bias-adjusted test, with test statistic

$$\mathrm{LM}_{\mathrm{PUY}} = \sqrt{\frac{2}{N(N-1)}} \sum\sum_{1 \le i < j \le N} \frac{(T-p)\hat{\rho}_{ij}^2 - \mu_{Tij}}{v_{Tij}},$$

where

$$\mu_{Tij} = \frac{\mathrm{tr}\left\{E\left(\mathbf{P}_i\mathbf{P}_j\right)\right\}}{T-p}, v_{Tij}^2 = \left[\mathrm{tr}^2\left\{E\left(\mathbf{P}_i\mathbf{P}_j\right)\right\}\right] a_{1T} + 2\mathrm{tr}\left[\left\{E\left(\mathbf{P}_i\mathbf{P}_j\right)\right\}^2\right] a_{2T},$$

$$a_{1T} = a_{2T} - \frac{1}{(T-p)^2} \text{ and } a_{2T} = 3\left\{\frac{(T-p-8)(T-p+2)+24}{(T-p+2)(T-p-2)(T-p-4)}\right\}^2.$$

Here, for each $1 \le i \le N$, $\mathbf{P}_i = \mathbf{I}_T - \mathbf{x}_i\left(\mathbf{x}_i'\mathbf{x}_i\right)^{-1}\mathbf{x}_i'$, hence $\hat{\epsilon}_{i\cdot} = (\hat{\epsilon}_{i1}, \ldots, \hat{\epsilon}_{iT})' = \mathbf{P}_i\epsilon_{i\cdot}$. The asymptotic null distribution of $\mathrm{LM}_{\mathrm{PUY}}$ is $\mathcal{N}(0,1)$, which is established when $T \to \infty$ first and then $N \to \infty$. Later, Feng et al. (2022) established that $\mathrm{LM}_{\mathrm{PUY}} \to \mathcal{N}(0,1)$ in distribution when $\min(N,T) \to \infty$, and also established that $\mathrm{LM}_{\mathrm{FJLX}} \to \mathcal{N}(0,1)$ in distribution when $\min(N,T) \to \infty$, where

$$\mathrm{LM}_{\mathrm{FJLX}} = \frac{1}{N} \sum\sum_{1 \le i < j \le N} T \hat{\rho}_{ij}^2 - \frac{\mu_N}{N} \text{ and } \mu_N = \frac{T}{(T-p)^2} \sum\sum_{1 \le i < j \le N} \mathrm{tr}\left(\mathbf{P}_i \mathbf{P}_j\right).$$

These tests are all based on the sum of squared sample correlations $\sum\sum_{1 \le i < j \le N} \hat{\rho}_{ij}^2$. In contrast, Pesaran (2004) proposed a test for cross-sectional independence directly based on the sum of sample correlations, with test statistic

$$\mathrm{CD}_{\mathrm{P}} = \sqrt{\frac{2T}{N(N-1)}} \sum\sum_{1 \le i < j \le N} \hat{\rho}_{ij}.$$

Pesaran (2015) established the asymptotic null distribution of CD_{P} to be $\mathcal{N}(0,1)$ when $\min(N,T) \to \infty$. Recently, to test cross-sectional correlation with serially correlated errors, Baltagi et al. (2016) proposed a new test with test statistic

$$S_N = \sqrt{\frac{2}{N(N-1)}} \sum\sum_{1 \le i < j \le N} \hat{\rho}_{ij}, \tag{2.3}$$

which is very similar to CD_{P}, but has different asymptotic variance. Under certain assumptions, they established that under the null hypothesis,

$$S_N / \hat{\sigma}_{S_N} \to \mathcal{N}(0,1) \text{ in distribution when } \min(N,T) \to \infty, \tag{2.4}$$

hence a level-α test can be performed by rejecting H_0 when $|S_N/\hat{\sigma}_{S_N}|$ is larger than the $(1-\alpha/2)$-quantile $z_\alpha = \Phi^{-1}(1-\alpha/2)$. Here,

$$\hat{\sigma}^2_{S_N} = \frac{2}{N(N-1)} \sum\sum_{1\le i<j\le N} v_j' \left(v_i - \bar{v}_{ij}\right) v_i' \left(v_j - \bar{v}_{ij}\right), \tag{2.5}$$

$\bar{v}_{ij} = \sum_{1<k\neq i,j<N} v_k/(N-2)$ and $v_k = \hat{\epsilon}_{k\cdot}/\|\hat{\epsilon}_{k\cdot}\|$ for all $1 \le k \le N$.

2.3 Max based test

All tests mentioned in Section 2.2 are sum based tests, which generally perform very well under dense alternatives, i.e. alternatives with dense cross-sectional correlation matrices, but may suffer from low power against sparse alternatives, i.e. alternatives with sparse cross-sectional correlation matrices. To deal with sparse alternatives, we now propose a max based test based on the maximum of the squares of the sample cross-sectional correlations. The test statistic we propose is

$$L_N = \max_{1\le i<j\le N} \hat{\rho}^2_{ij}. \tag{2.6}$$

Max based tests have been widely studied in testing independence among variables, e.g., Li and Xue (2015), Chen and Liu (2018) and Feng et al. (2022). Specifically, Feng et al. (2022) used it to test cross-sectional independence under the assumption of no serial correlation between the errors, and established that $TL_N - 4\log N + \log\log N \to G(y)$ in distribution, where $G(y) = \exp\left\{-\exp\left(-y/2\right)/\sqrt{8\pi}\right\}$. However, this test may perform poorly if blindly used in the situation where serial correlation exists.

To utilize the test statistic in (2.6) for data with serial correlations, we must reinvestigate the asymptotic properties of L_N. We impose the following assumptions.

Assumption 1. Assume that $\mathbf{E} = (\epsilon_{1\cdot}, \ldots, \epsilon_{N\cdot})' = \mathbf{Z}\mathbf{R}'$, where $\mathbf{R} \in \mathbb{R}^{T\times T}$ is an invertible matrix, and all elements of $\mathbf{Z} = (Z_{1\cdot}, \ldots, Z_{N\cdot})' \in \mathbb{R}^{N\times T}$ with $Z_{i\cdot} = (Z_{i1}, \ldots, Z_{iT})$ are iid variables with mean zero and variance one. The density function of $(\mathbf{Z})_{it}$ is symmetric and the sub-Gaussian norm of $(\mathbf{Z})_{it}$ is bounded by K, i.e., $E[\exp\{(\mathbf{Z})_{it}^2/K^2\}] \leq 2$, for each $1 \leq i \leq N$ and $1 \leq t \leq T$.

Assumption 2. (i) Assume that $p > 0$ is fixed, and the regressors x_{it} are strictly exogenous, such that

$$E\left(\epsilon_{it}|\mathbf{x}_i\right) = 0, \text{ for all } 1 \leq i \leq N \text{ and } 1 \leq t \leq T.$$

(ii)Assume that $\mathbf{x}_i'\mathbf{x}_i/T = \sum_{t=1}^T x_{it}x_{it}'/T$ is non-singular and $\mathbf{x}_i\left(\mathbf{x}_i'\mathbf{x}_i/T\right)^{-1}\mathbf{x}_i'$ is stochastically bounded for all $1 \leq i \leq N$.

Assumption 3. (i) Assume that for some constant $C > 0$, $C^{-1} \leq \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) \leq C$, where $\mathbf{\Sigma} = \mathbf{R}\mathbf{R}'$. (ii) Assume that $\max_{j=1,\ldots,T}\sum_{k=1}^T |(\mathbf{\Sigma})_{jk}|^\tau \leq C'$ for some $0 < \tau < 2$ and $C' > 0$.

Under Assumption 1, the error is $\epsilon_{i\cdot} = \mathbf{R}Z_{i\cdot}$, where $\mathbf{\Sigma} = \mathbf{R}\mathbf{R}'$ is the covariance matrix of $\epsilon_{i\cdot}$. The assumption of symmetry ensures that the expectations of both S_N and $\hat{\rho}_{ij}$ are precisely equal to zero. Pesaran et al. (2008) also assume the symmetry when the error are independent across time. If the error distribution is not symmetric, there would be a non-negligible bias term in S_N, see numerical experiments in Supplementary Material. How to calculate the bias term of S_N under the asymmetric assumption needs some further studies.

Assumption 2(i) is used in Pesaran et al. (2008), which is a common condition for panel data model in (2.1). Assumption 2(ii) is imposed to ensure that the difference between the distribution of the max based statistic based on the residuals and that based on the errors is negligible.

Assumption 3(i) is the same as Condition (C1) in Chen and Liu (2018), which is a common eigenvalue assumption in high-dimensional inference literature such as in Cai et al. (2016) and contains many important types of covariance matrices, including the bandable, Toeplitz and sparse covariance matrices. Assumption 3(ii) is the same as Condition (C2) in Chen and Liu (2018), which assumes the sparsity of $\mathbf{\Sigma}$. Note that the sparsity of $\mathbf{\Sigma}$ is imposed to ensure good estimation property. To eliminate the complexity caused by the serial correlation across t within each row, ideally we hope to work with $\mathbf{E}(\mathbf{R}')^{-1} = \mathbf{Z}$, and to test the cross-sectional independence of $\mathbf{Z}$. However, obtaining a sufficiently good estimate of $(\mathbf{R}')^{-1}$ may be very difficult. Thus, in our test procedure, we will need to estimate a $\boldsymbol{\Sigma}$ dependent quantity $\mathrm{tr}^2(\boldsymbol{\Sigma})/\|\boldsymbol{\Sigma}\|_{\mathrm{F}}^2$. Since $\boldsymbol{\Sigma}$ is a high dimensional matrix, sparsity is a common assumption to regularize the properties of the related estimation.

To better explain the assumptions imposed, below we provide a simple example that satisfies these assumptions. In fact, many time series models satisfy the above assumptions, such as the first-order autoregressive model: $\epsilon_{it} = a\epsilon_{it-1} + Z_{it}$ with $|a| < 1$. Clearly, this model satisfies Assumptions 1 with

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a^{T-1} & a^{T-2} & \cdots & 1 \end{bmatrix}, \tag{2.7}$$

and the covariance matrix $\boldsymbol{\Sigma} = \mathbf{R}\mathbf{R}'$ based on (2.7) satisfies Assumption 3. In addition, Assumption 2(i) is commonly used (Pesaran et al., 2008). Assumption 2(ii) holds if the two conditions, $\|x_{it}\|$ and $\sum_{t=1}^{T} x_{it}x_{it}'/T = \mathbf{x}_i'\mathbf{x}_i/T$ are randomly bounded for all $1 \leq i \leq N$ and $1 \leq t \leq T$, hold. These two conditions are also commonly used, such as in Assumption A of Bai (2009), Assumption 3 of Gao et al. (2023), Assumption 4 of Baltagi et al. (2016) and Assumption 2(iii) of Baltagi et al. (2012).

Throughout the text, we let γ be a positive finite constant.

Theorem 1. *Under Assumptions 1-3 and the null hypothesis in (2.2), when $\min(N,T)\to\infty$ and $N/T\to\gamma$, for any $y\in\mathbb{R}$, we have*

$$P\left(\frac{\mathrm{tr}^2(\mathbf{\Sigma})}{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2}L_N-4\log N+\log\log N\le y\right)\to G(y),$$

where $G(y)=\exp\big\{-\exp\left(-y/2\right)/\sqrt{8\pi}\big\}$ is a type-I Gumbel distribution function.

Remark 1. The main idea of the proof of Theorem 1 is summarized as follows. Let $\mathbf{\Lambda}=\mathbf{R}'\mathbf{R}$ and

$$\tilde{T}_{ij}\doteq\frac{\epsilon_{i\cdot}'\epsilon_{j\cdot}}{\|\mathbf{\Lambda}\|_{\mathrm{F}}}=\frac{Z_{i\cdot}'\mathbf{\Lambda}Z_{j\cdot}}{\|\mathbf{\Lambda}\|_{\mathrm{F}}},\qquad \rho_{ij}=\frac{\epsilon_{i\cdot}'\epsilon_{j\cdot}}{\|\epsilon_{i\cdot}\|\times\|\epsilon_{j\cdot}\|}.$$

To establish Theorem 1, we initially demonstrate that

$$P\biggl(\max_{1\le i<j\le N}\tilde{T}_{ij}^2-4\log N+\log\log N\le y\biggr)\to\exp\Big\{-\frac{1}{\sqrt{8\pi}}\exp\Big(-\frac{y}{2}\Big)\Big\}.$$

Subsequently, we prove $\max_{1\le i<j\le N}\tilde{T}_{ij}^2-\mathrm{tr}^2(\mathbf{\Lambda})/\|\mathbf{\Lambda}\|_{\mathrm{F}}^2\max_{1\le i<j\le N}\rho_{ij}^2\to 0$ in distribution, and $\mathrm{tr}^2(\mathbf{\Lambda})/\|\mathbf{\Lambda}\|_{\mathrm{F}}^2\max_{1\le i<j\le N}(\rho_{ij}^2-\hat{\rho}_{ij}^2)\to 0$ in distribution. By combining the aforementioned results, we derive that

$$P\biggl(\frac{\mathrm{tr}^2(\mathbf{\Lambda})}{\|\mathbf{\Lambda}\|_{\mathrm{F}}^2}\max_{1\le i<j\le N}\hat{\rho}_{ij}^2-4\log N+\log\log N\le y\biggr)\to\exp\Big\{-\frac{1}{\sqrt{8\pi}}\exp\Big(-\frac{y}{2}\Big)\Big\}.$$

Furthermore, due to $\mathrm{tr}(\mathbf{\Sigma})=\mathrm{tr}(\mathbf{\Lambda})$ and $\|\mathbf{\Sigma}\|_{\mathrm{F}}^2=\|\mathbf{\Lambda}\|_{\mathrm{F}}^2$, we eventually obtain the desired result.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

The column covariance matrix $\mathbf{\Sigma}$, which is needed in the construction of the max based test statistic, can be assessed via a procedure similar to that proposed in Chen and Liu (2018) for testing independence among variables using correlated samples. First, define the column sample covariance matrix $\hat{\mathbf{\Sigma}} = (\hat{\sigma}_{ij})_{1\le i,j\le T}$ with

$$\hat{\sigma}_{ij} = \frac{1}{N-1}\sum_{l=1}^{N}\left(\hat{\epsilon}_{li} - \bar{\hat{\epsilon}}_{\cdot i}\right)\left(\hat{\epsilon}_{lj} - \bar{\hat{\epsilon}}_{\cdot j}\right), \quad \bar{\hat{\epsilon}}_{\cdot j} = \frac{1}{N}\sum_{l=1}^{N}\hat{\epsilon}_{lj} \;\text{ and }\; \theta_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$$

for each $1 \le i,j \le T$. Then, define $\tilde{\mathbf{\Sigma}} = (\tilde{\sigma}_{ij})_{1\le i,j\le T}$ with

$$\tilde{\sigma}_{ij} = \begin{cases} \hat{\sigma}_{ij} I\left(\dfrac{|\theta_{ij}|}{1-\theta_{ij}^2} \ge \nu\sqrt{\dfrac{\hat{P}_N \log T}{N}}\right), & i \ne j, \\ \hat{\sigma}_{ii}, & i = j \end{cases} \tag{2.8}$$

for each $1 \le i,j \le T$. Here, $\nu > \sqrt{2}$ and $\hat{P}_N = \left[\|\hat{\mathbf{\Phi}}\|_{\mathrm{F}}^2 - \frac{1}{T}\{\mathrm{tr}(\hat{\mathbf{\Phi}})\}^2\right]/N$, where for each $1 \le i,j \le N$, $(\hat{\mathbf{\Phi}})_{ij} = \hat{\epsilon}_{i\cdot}'\hat{\epsilon}_{j\cdot}/\mathrm{tr}(\hat{\mathbf{\Sigma}})$. Based on $\tilde{\mathbf{\Sigma}}$, $\mathrm{tr}^2(\mathbf{\Sigma})$ and $\|\mathbf{\Sigma}\|_{\mathrm{F}}^2$ are estimated by $\mathrm{tr}^2(\tilde{\mathbf{\Sigma}})$ and $\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2$, respectively. Hence, $\mathrm{tr}^2(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|_{\mathrm{F}}^2$ is estimated by $\mathrm{tr}^2(\tilde{\mathbf{\Sigma}})/\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2$, as indicated in the following theorem, which is a ratio-consistent estimator of $\mathrm{tr}^2(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|_{\mathrm{F}}^2$.

Theorem 2. *Under Assumptions 1-3 and the null hypothesis in (2.2), for any $\nu > \sqrt{2}$,*

$$\frac{\mathrm{tr}^2(\tilde{\mathbf{\Sigma}})}{\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2}\frac{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2}{\mathrm{tr}^2(\mathbf{\Sigma})} = 1 + O_p\left\{\left(\frac{\log T}{N}\right)^{\frac{1}{2}\wedge(1-\frac{\tau}{2})}\right\}, \tag{2.9}$$

when $\min(N,T) \to \infty$ and $N/T \to \gamma$.

Remark 2. The main idea of the proof of Theorem 2 is summarized as follows. Initially, we establish that $\hat{\sigma}_{ij}$ converges to $(\boldsymbol{\Sigma})_{ij}$ in distribution uniformly for all $1 \leq i, j \leq T$, and $\hat{P}_N$ converges to 1 in distribution. Subsequently, we leverage these findings to demonstrate the desired result.

Combining Theorems 1 and 2, we have that for any $y \in \mathbb{R}$,

$$P\left(\frac{\mathrm{tr}^2(\tilde{\boldsymbol{\Sigma}})}{\|\tilde{\boldsymbol{\Sigma}}\|_{\mathrm{F}}^2} L_N - 4\log N + \log\log N \leq y\right) \to G(y). \tag{2.10}$$

Based on (2.10), for a given significance level α, the null hypothesis in (2.2) will be rejected by the established max based test when $L_N \mathrm{tr}^2(\tilde{\boldsymbol{\Sigma}})/\|\tilde{\boldsymbol{\Sigma}}\|_{\mathrm{F}}^2 \geq w_\alpha + 4\log N - \log\log N$, where w_α is the $1-\alpha$ quantile of the type-I Gumbel distribution with the cumulative distribution function $G(y)$ and has the specific form of $w_\alpha = \log(8\pi) - 2\log\log(1-\alpha)^{-1}$.

Next, we turn to the power analysis of the proposed max based test, in which the following two assumptions need to be imposed.

Assumption 4. Assume that $\mathbf{E} = (\epsilon_{1\cdot}, \ldots, \epsilon_{N\cdot})' = \mathbf{L}\mathbf{Z}\mathbf{R}'$, where $\mathbf{L} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{T\times T}$ are invertible matrices, all elements of $\mathbf{Z} \in \mathbb{R}^{N\times T}$ are iid variables with mean zero and variance one. The density function of $(\mathbf{Z})_{it}$ is symmetric, and the sub-Gaussian norm of $(\mathbf{Z})_{it}$ is bounded by K, that is, for each $1 \leq i \leq N$ and $1 \leq t \leq T$, $E[\exp\{(\mathbf{Z})_{it}^2/K^2\}] \leq 2$.

Assumption 5. (i) For a certain constant $C > 0$, assume $C^{-1} \leq \lambda_{\min}(\mathbf{\Phi}) \leq \lambda_{\max}(\mathbf{\Phi}) \leq C$, where $\mathbf{\Phi} = \mathbf{L}\mathbf{L}'$. (ii) For a certain $0 < \tau < 2$ and $C' > 0$, assume $\max_{1\leq j\leq N} \sum_{k=1}^{N} |(\mathbf{\Phi})_{jk}|^{\tau} \leq C'$. (iii) Assume that the diagonal elements of $\mathbf{\Phi}$ are all equal to 1.

Under Assumption 4, the error matrix is $\mathbf{E} = \mathbf{L}\mathbf{Z}\mathbf{R}'$. It is well known that when $(\mathbf{Z})_{it} \overset{iid}{\sim} N(0,1)$, the error matrix $\mathbf{E}$ has a matrix normal distribution $N(\mathbf{0}, \mathbf{\Phi} \otimes \mathbf{\Sigma})$, where $\otimes$ represents the Kronecker product. Chen and Liu (2018) studied the asymptotic properties of L_N under the matrix normal distribution. Assumption 5 (i) is the same as Condition (C1) in Chen and Liu (2018), which is a common eigenvalue assumption. Assumption 5 (ii) is the same as Condition (C2) in Chen and Liu (2018), which constrains the sparsity of $\mathbf{\Phi}$. In addition, Assumption 5 (iii) makes the structural model of the errors identifiable.

Theorem 3. *Under Assumptions 2-5, suppose that for some* $\delta > 2$, *some* $\nu > 0$, *and sufficiently large* N *and* T,

$$\psi_N = \max_{1\leq i<j\leq N} |(\mathbf{\Phi})_{ij}| \geq \delta\sqrt{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 \log N / \mathrm{tr}^2(\mathbf{\Sigma})},$$

then

$$P\left(\frac{\mathrm{tr}^2(\tilde{\mathbf{\Sigma}})}{\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2} L_N - 4\log N + \log\log N > w_\alpha\right) \to 1, \tag{2.11}$$

as $\min(N,T) \to \infty$ *and* $N/T \to \gamma$.

Remark 3. Here, we outline the primary steps in the proof of Theorem 3. Initially, we establish the following facts:

$$\max_{1\le i<j\le N} |\hat{\rho}_{ij} - \rho_{ij}| = O_p\Big(\frac{\sqrt{\mathrm{tr}\,(\mathbf{\Sigma}^2)}}{\sqrt{\log N}\mathrm{tr}\,(\mathbf{\Sigma})}\Big)$$

and

$$\frac{\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2}{\mathrm{tr}^2(\tilde{\mathbf{\Sigma}})}\frac{\mathrm{tr}^2(\mathbf{\Sigma})}{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2} = 1 + O_p\Big\{\Big(\sqrt{\frac{\log T}{N}}\Big)^{\min(1,2-\tau)}\Big\},$$

respectively. Subsequently, by leveraging these formulations and that fact of $\psi_N \geq \delta\sqrt{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 \log N/\mathrm{tr}^2(\mathbf{\Sigma})}$, we obtain the desired result.

Theorem 3 indicates that the proposed max based test is consistent under the sparse alternative in which the maximum non diagonal entries of $\mathbf{\Phi}$ is sufficiently large. More specifically, because $\sqrt{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2/\mathrm{tr}^2(\mathbf{\Sigma})} \asymp T^{-1/2}$ under Assumption 3, the test is able to detect the dependence as long as a single covariance is at the order of $(\log N/T)^{1/2}$, which leads to an overall detection rate of $(\log N/T)^{1/2}$. In contrast, the sum based test can detect the departure from null if each covariance reaches $T^{-1/2}N^{-1}$ (Theorem 4 of Baltagi et al. (2016)), hence the overall detection rate is $N/T^{1/2}$. This theoretical result is consistent with the simulation results in Section 3, where the max based test performs better than the sum based test in terms of empirical power under sparse alternatives. The presence of the possible serial correlation makes it more difficult to establish asymptotic properties of L_N. For example, we need to reestablish the asymptotic variance of L_N that depends on $\mathbf{\Sigma}$. This is the reason why we replace the test statistic $TL_N - 4\log N + \log\log N$ in Feng et al. (2022), where no serial correlation is considered, by $L_N \mathrm{tr}^2(\mathbf{\Sigma})/\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 - 4\log N + \log\log N$ and $L_N \mathrm{tr}^2(\tilde{\mathbf{\Sigma}})/\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2 - 4\log N + \log\log N$.

On the other hand, we note that the max based test for testing independence among normally distributed variables with correlated samples has been studied in Chen and Liu (2018), whereas in this paper we consider the max based test under panel data models with strictly exogenous regressors, and we relax the Gaussian assumption to sub-Gaussian. The proof under sub-Gaussianity is much more challenging because many properties associated with Gaussianity, such as rotation invariance, can no longer be used. To establish the theoretical results under sub-Gaussian distributions, more advanced technical tools, such as the Hanson-Wright inequality (Rudelson and Vershynin, 2013), need to be engaged in deriving the asymptotical distribution of the max based test statistic L_N. In addition, the expressions of the moments of the quadratic forms under sub-Gaussian distributions are also much more complex than that under Gaussian distributions, which further complicate the proof.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

2.4 Fisher's combined probability test

fisher

It is intuitively expected that the sum based test S_N performs well under dense alternatives, whereas the max based test L_N performs well under sparse alternatives. However, in practice, it is usually unknown whether the correlation matrix of the errors is sparse or not, hence it is difficult to decide which test to use. For this reason, in this subsection, we propose using Fisher's combined probability test by combining the sum based and max based tests, which is expected to take advantage of both tests. To construct the test, we first need to establish the asymptotic independence between the two statistics S_N and L_N under the null hypothesis.

Theorem 4. *Under Assumptions 1-3 and the null hypothesis in* (2.2), $L_N \mathrm{tr}^2(\tilde{\mathbf{\Sigma}})/\|\tilde{\mathbf{\Sigma}}\|_{\mathrm{F}}^2 - 4\log N + \log\log N$ *and* $S_N/\hat{\sigma}_{S_N}$ *are asymptotically independent, as* $\min(N,T) \to \infty$ *and* $N/T \to \gamma$.

Remark 4. Here, we outline the main idea for establishing the above asymptotical independence under the null hypothesis. Firstly, we establish the premise that if $\tilde{T}_{\max}$ and S_N are asymptotically independent, we can infer that L_N and S_N are asymptotically independent, where $\tilde{T}_{\max} = \max_{1\le i<j\le N}(\epsilon_{i\cdot}'\epsilon_{j\cdot})^2/\,\|\mathbf{\Sigma}\|_{\mathrm{F}}^2$. Hence, to demonstrate the asymptotic indepen-

dence, it suffices to show that for any x and y,

$$\begin{aligned}&\lim_{\min(N,T)\to\infty} P\Big(\frac{S_N}{\sigma_{S_N}} \le x, \tilde{T}_{\max} > a_N\Big)\\ =&\lim_{\min(N,T)\to\infty} P\Big(\frac{S_N}{\sigma_{S_N}} \le x\Big) \lim_{\min(N,T)\to\infty} P\big(\tilde{T}_{\max} > a_N\big),\end{aligned} \tag{2.12}$$

where $a_N = 4\log N - \log\log N + y$. Let $\Lambda_N = \{(i,j); 1 \le i < j \le N\}$, $A_N = \{S_N/\sigma_{S_N} \le x\}$, $B_I = \{|\epsilon_{i\cdot}'\epsilon_{j\cdot}| \ge l_N\}$ for any $I = (i,j) \in \Lambda_N$, and

$$l_N = \sqrt{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 \left(4\log N - \log\log N + y\right)} = \sqrt{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 \, a_N}.$$

Then, the left side of (2.12) can be expressed as

$$P\Big(\frac{S_N}{\sigma_{S_N}} \le x, \max_{1\le i<j\le N} \frac{(\epsilon_{i\cdot}'\epsilon_{j\cdot})^2}{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2} > a_N\Big) = P\Big(\bigcup_{I\in\Lambda_N} A_N B_I\Big).$$

Due to the principle of inclusion-exclusion, when m is sufficiently large, we can deduce that

$$\begin{aligned}P\left(\bigcup_{I\in\Lambda_N} A_N B_I\right) \approx& \sum_{I_1\in\Lambda_N} P\left(A_N B_{I_1}\right) - \sum_{I_1<I_2\in\Lambda_N} P\left(A_N B_{I_1} B_{I_2}\right) + \cdots\\ &+ (-1)^{m+1} \sum_{I_1<I_2<\cdots<I_m\in\Lambda_N} P\left(A_N B_{I_1} B_{I_2}\cdots B_{I_m}\right),\end{aligned}$$

where the symbol "$\approx$" indicates that the difference between the two sides tends to zero as $\min(N,T) \to \infty$ with $\lim_{\min(N,T)\to\infty} N/T = \gamma \in (0,+\infty)$. Next, we establish two key facts. The first is given by the equation:

$$P\left(A_N B_{I_1} B_{I_2}\cdots B_{I_m}\right) \approx P\left(A_N\right)\cdot P\left(B_{I_1} B_{I_2}\cdots B_{I_m}\right). \tag{2.13}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

The second fact is that for any $m \geq 1$, the following sum is negligible:

$$\sum_{I_1<I_2<\cdots<I_m\in\Lambda_N} \{P(A_N B_{I_1} B_{I_2}\cdots B_{I_m}) - P(A_N)\cdot P(B_{I_1}B_{I_2}\cdots B_{I_m})\} \to 0. \tag{2.14}$$

Finally, based on the above results, we obtain the desired result.

Note that in many related studies, such as Li and Xue (2015) and Feng et al. (2022), the asymptotic independence between the sum based and max based statistics is established under the assumption that all components of the random vectors concerned are independent and identically distributed. In contrast, we establish the asymptotic independence in the situation where the error vectors may have serial correlation, which makes the proof of the asymptotic independence much more challenging and requires more complex technical treatments and tools.

Let $P_{L_N} = 1 - G\Big(L_N \mathrm{tr}^2(\tilde{\Sigma})/\|\tilde{\Sigma}\|_{\mathrm{F}}^2 - 4\log N + \log\log N\Big)$ and $P_{S_N} = 2 - 2\Phi\left(|S_N/\hat{\sigma}_{S_N}|\right)$. We construct Fisher's combined probability test statistic as

$$T_C = -2\log(P_{L_N}) - 2\log(P_{S_N}). \tag{2.15}$$

By Theorem 4, (2.4) and (2.10), we see that P_{L_N} and P_{S_N} are asymptotically independent under the null hypothesis and each has limit distribution

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

$U[0, 1]$, the uniform distribution on $[0, 1]$. We thus obtain the following corollary.

Corollary 1. *Assume that the assumptions in Theorem 4 hold, then we have* $T_C \to \chi^2_4$ *in distribution when* $\min(N, T) \to \infty$ *and* $N/T \to \gamma$.

According to Corollary 1, we proposed Fisher's combined probability tes at level α by rejecting the null hypothesis in (2.2) if $T_C \geq q_\alpha$, where q_α is the $1 - \alpha$ quantile of χ^2_4.

3. Simulation studies

We now conduct simulations to investigate the finite sample performance of the two tests proposed in this paper, i.e. the test based on L_N and the Fisher's combined probability test based on T_C. For comparison, we also implement three other existing tests, i.e. the test based on S_N proposed by Baltagi et al. (2016), the $\mathrm{LM}_{\mathrm{PUY}}$ test proposed by Pesaran et al. (2008) and the CD_{P} test proposed by Pesaran (2004). Here, the max based test is implemented by setting $\nu = 1.42$ in (2.8) as in Chen and Liu (2018). For simplicity, we will abbreviate these five tests as L_N, T_C, S_N, $\mathrm{LM}_{\mathrm{PUY}}$ and CD_{P}, respectively.

We consider the data generating process

$$y_{it} = \alpha_i + \sum_{l=2}^{p} x_{li,t}\beta_{li} + \epsilon_{it},\ 1 \le i \le N,\ 1 \le t \le T, \tag{3.1}$$

where $x_{it} = (1, x_{2i,t}, \ldots, x_{pi,t})' \in \mathbb{R}^p$ and $\beta_i = (\alpha_i, \beta_{2i}, \ldots, \beta_{pi})' \in \mathbb{R}^p$. We generate $\alpha_i \overset{iid}{\sim} \mathcal{N}(0,1)$ and $\beta_{li} \overset{iid}{\sim} \mathcal{N}(1, 0.04)$ for $2 \le l \le p$. The strictly exogenous regressors are generated by $x_{li,t} = 0.6x_{li,t-1} + v_{li,t}$ for $1 \le i \le N$, $-50 \le t \le T$ and $2 \le l \le p$, where $x_{li,-51} = 0$ and $v_{li,t}$ are independently and identically distributed from $\mathcal{N}\left(0, \psi_{li}^2/\left(1-0.6^2\right)\right)$, where $\psi_{li}^2 \overset{iid}{\sim} \chi_6^2/6$.

Consider the following two settings of serial correlation of the errors ϵ_{it}.

(i) The errors follow an auto-regressive (AR) model of order one over time, i.e. AR(1): $\epsilon_{i1} = e_{i1}$ and $\epsilon_{it} = 0.6\epsilon_{it-1} + e_{it}$ for $2 \le t \le T$ and $1 \le i \le N$.

(ii) The errors follow an auto-regressive and moving average (ARMA) model of order (1,1) over time, i.e. ARMA(1,1): $\epsilon_{i1} = e_{i1}$ and $\epsilon_{it} = 0.6\epsilon_{it-1} + e_{it} + 0.2e_{it-1}$ for $2 \le t \le T$ and $1 \le i \le N$.

To produce data under the null hypothesis, e_{it} are independently generated from the following three distributions: (1) $\mathcal{N}(0,1)$; (2) $t_6/\sqrt{6/4}$; (3) $(\chi_5^2 - 5)/\sqrt{10}$.

Then, we turn to produce data under the alternative hypothesis. The data generating process is specified as

$$y_{it} = \alpha_i + \sum_{l=2}^{p} x_{li,t}\beta_{li} + \epsilon_{it}^*,\ 1 \le i \le N,\ 1 \le t \le T, \tag{3.2}$$

where ϵ_{it}^* are generated from the following two settings.

(I) Non-sparse case. Spatial moving average (SMA) model with order one, i.e. SMA(1): for all $1 \le t \le T$, $\epsilon_{1t}^* = 0.5\delta\epsilon_{2t} + \epsilon_{1t}$, $\epsilon_{Nt}^* = 0.5\delta\epsilon_{N-1t} + \epsilon_{Nt}$ and $\epsilon_{it}^* = \delta\left(0.5\epsilon_{i-1,t} + 0.5\epsilon_{i+1,t}\right) + \epsilon_{it}$, where $2 \le i \le N-1$, $\delta = 0.2$ and ϵ_{it} are generated from settings (i)-(ii) with distributions (1)-(3).

(II) Sparse case. Let $(\epsilon_{1\cdot}^*, \cdots, \epsilon_{N\cdot}^*)' = \mathbf{W}^{1/2}(\epsilon_{1\cdot}, \cdots, \epsilon_{N\cdot})'$, where $\epsilon_{i\cdot}^* = (\epsilon_{i1}^*, \cdots, \epsilon_{iT}^*)'$ and $\epsilon_{i\cdot} = (\epsilon_{i1}, \cdots, \epsilon_{iT})'$, for all $1 \le i \le N$. Here, $\epsilon_{i\cdot}$ are generated from settings (i)-(ii) with distributions (1)-(3). $\mathbf{W}$ is constructed as follows. Randomly select a subset $S \subset \{1, \cdots, N\}$ with cardinality $\lceil N^{0.3} \rceil$, the smallest integer grater than or equal to $N^{0.3}$. Let $(\mathbf{W})_{ij} = 1$ if $i = j$. For $i < j$, define $(\mathbf{W})_{ij} = 0$ if $i \notin S$ or $j \notin S$, and $(\mathbf{W})_{ij} \overset{iid}{\sim} U\left[\sqrt{4\frac{\log T}{N}}, \sqrt{6\frac{\log T}{N}}\right]$ if $i, j \in S$.

In addition, for the choices of p, N and T, we set $p \in \{3, 5\}$, $N \in \{100, 200\}$, $T \in \{200, 300, 400, 500\}$.

The results of the empirical size and power of the five tests in the non-sparse and sparse cases of error correlation matrices are summarized in

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

Tables 1-3. The power curves are plotted in Figure 1. All results are based on 1,000 replications. We next analyze them in detail.

Table 1 indicates that in most cases T_C, L_N and S_N have empirical sizes not much larger than 5%. Here, the max based test L_N and Fisher's combined probability test T_C tend to have smaller empirical sizes than the sum based test S_N, especially as T is relatively small. This is not surprising and is common for many max based tests, because the convergence rate of the type-I Gumbel distribution is typically slow (Liu et al., 2008). $\mathrm{CD_P}$ and $\mathrm{LM_{PUY}}$ fail to control the empirical size because $E(\hat{\rho}_{ij}^2)$ may be seriously affected by serial correlation. This is also observed by Baltagi et al. (2016).

Tables 2 and 3 show the empirical powers in both non-sparse and sparse cases of correlation matrices. Since $\mathrm{CD_P}$ and $\mathrm{LM_{PUY}}$ fail to control the empirical size, we exclude them from the empirical power results. Table 2 and 3 indicate that S_N and T_C generally perform better than L_N in non-sparse cases in terms of empirical powers, whereas L_N and T_C generally perform better than S_N in sparse cases. As expected, Fisher's combined probability test T_C has power advantages regardless the local alternative is sparse or not.

Figure 1 shows how the empirical powers of the three tests change as the degree of sparsity of the correlation matrix changes, where the x-axis is the level of density k and the y-axis represents the empirical power. To generate Figure 1, we designed the following simulation. Let $(\epsilon^*_{1\cdot}, \ldots, \epsilon^*_{N\cdot})' = \mathbf{W}^{1/2}(\epsilon_{1\cdot}, \ldots, \epsilon_{N\cdot})'$, where $\epsilon^*_{i\cdot} = (\epsilon^*_{i1}, \ldots, \epsilon^*_{iT})'$ and $\epsilon_{i\cdot} = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$, for all $1 \leq i \leq N$. Here, $\epsilon_{i\cdot}$ are generated from settings (i)-(ii) with distribution (1). We set $N = 100$, $T = 300$, $p = 3$, $k = 2, \ldots, 16$; a subset $S \subset \{1, \ldots, N\}$ is randomly selected with cardinality $|S| = k$; $(\mathbf{W})_{ij} = 1$ if $i = j$; for $i \neq j$, $(\mathbf{W})_{ij} = 0$ if $i \notin S$ or $j \notin S$, and $(\mathbf{W})_{ij} \overset{iid}{\sim} U\left[\sqrt{\frac{7}{k}\frac{\log T}{N}}, \sqrt{\frac{9}{k}\frac{\log T}{N}}\right]$ if $i \in S$ and $j \in S$. Hence, a larger k means a higher level of density.

Figure 1 indicates that the empirical power of Fisher's combined probability test T_C is always very close to the maximum power of both tests for all k, which suggests that it has robust empirical power performance regardless the alternative is sparse or not. In contrast, the empirical power curves of S_N, L_N are both monotone, with the sum based test S_N gains more power with the increase of the level of density, while the max based test L_N gains more power with the decrease of the level of density.

4. Empirical application

We now apply the proposed tests to the dataset analyzed in Serlenga and Shin (2007), which comprises bilateral trade flow data for 91 country pairs from 15 European countries, spanning a 42-year period from 1960 to 2001.

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

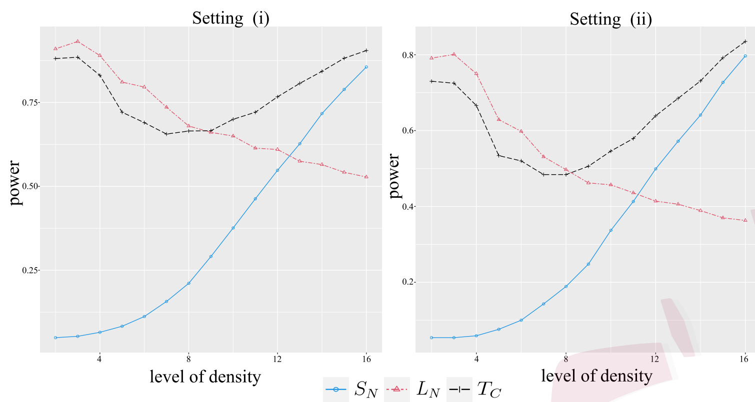

Figure 1: The empirical power curves of the three tests at 5% level under settings (i) and (ii).

Consider the following linear heterogeneous slope panel data model:

$$\begin{aligned} Y_{it} =& \alpha_i + \beta_{i1} GPD_{it} + \beta_{i2} RER_{it} + \beta_{i3} EMU_{it} + \gamma_i DIST_i \\ &+ \lambda_{i1} RERT_t + \lambda_{i2} FTRADE_t + \lambda_{i3} FGDP_t + \lambda_{i4} FRLF_t + \epsilon_{it}, \end{aligned} \tag{4.1}$$

where $1 \le i \le 91$ and $1 \le t \le 42$. For the i-th country pair in the t-th year, Y_{it} represents the bilateral trade flow, i.e. the sum of logged exports and imports, GPD_{it} represents the sum of the logged real GDPs, RER_{it} represents the logged bilateral real exchange rate, and EMU_{it} represents the dummy variable that is equal to 1 when both countries in this pair adopt the common currency. $DIST_i$ represents the geographical distance between capital cities, while $RERT_t$ denotes the logarithm of real exchange rates between the European currencies and the U.S. dollar. In addition,

$FTRADE_t$, $FGDP_t$ and $FRLF_t$ are the time specific common factors of the variables TRADE, GDP, RLF (a measure of relative factor endowments), respectively.

Before applying the proposed tests to this dataset, we conducted the Box-Pierce tests on the residual sequences to investigate whether there is serial correlation in the residual sequences. The histogram of the p-values of these tests, shown in Figure 2, suggests that for many residual sequences exhibit serial correlation. Therefore, it is reasonable to apply the proposed tests to this dataset. The p-values of applying the S_N, L_N and T_C tests to this dataset are 0.168, 0.015 and 0.018, respectively. To explore the reasons behind these results, we plotted the histogram of the cross-sectional correlations between all pairwise residual sequences as well as the heatmap of their absolute values in Figure 3, which indicate that many of these correlation values are non-zero, including some with very large absolute values, and the distribution of these correlations is nearly symmetric. The very large absolute values lead both the L_N and T_C tests to reject the null hypothesis. However, despite having many non-zero correlation values, the S_N test failed to reject the null hypothesis. This is because its test statistic is based on the sum of the correlation coefficients rather than the sum of the squared correlation coefficients. In fact, the symmetry of these correlations

Statistica Sinica: Preprint
doi:10.5705/ss.202023.0348

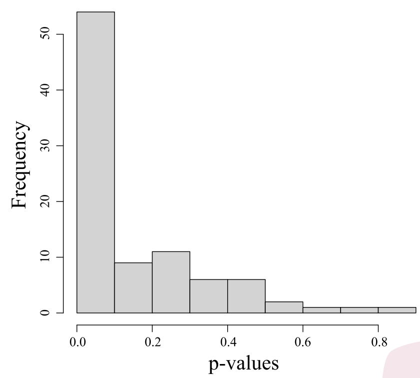

Figure 2: Histogram of the p-values of the Box-Pierce test for all residual sequences.

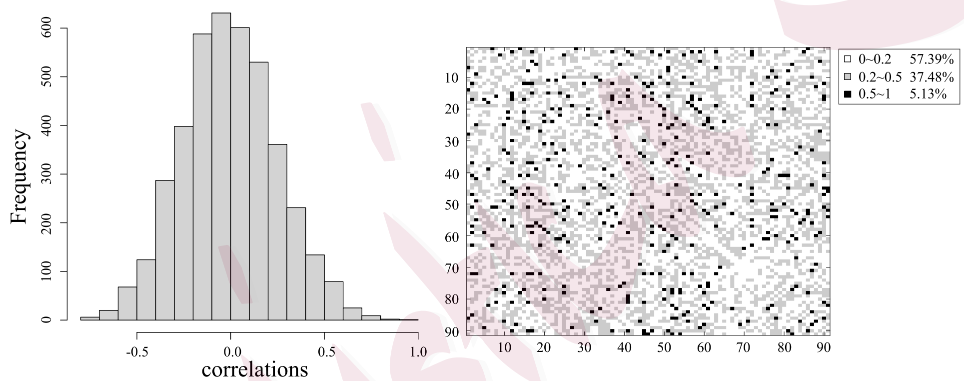

Figure 3: Heat map and histogram of the correlations between all pairwise residual sequences.

causes their sum to approach zero, thereby failing to detect deviations from the null hypothesis.

In the aforementioned application, the performance of L_N is better than S_N. In the Supplementary Material, we will provide another empirical application where the superiority of L_N or S_N depends on the different ways of handling the residuals.

5. Conclusion

In this paper, for testing cross-sectional independence under heterogeneous panel data models with serial correlation, we proposed a max based test based on the maximum of the squared cross-sectional correlations of residuals to deal with sparse alternatives. Furthermore, we proposed using Fisher's combined probability test by combining the p-values of the proposed max based test and the sum based test, which is applicable regardless alternatives are sparse or not. In constructing the combined test, the asymptotic null distribution is established strictly based on the asymptotic independence of the max based and sum based test statistics, which is an important contribution of this paper. In addition, we relaxed the distribution assumption made in the existing studies for testing independence with serial correlation from Gaussian to sub-Gaussian. Finally, simulation studies and empirical applications demonstrate the superiority of the proposed tests in comparison with some of their competitors and their practicality.

Supplementary Material

The Supplementary Material contains some additional numerical results, an additional empirical application and the technical proofs.

Acknowledgments

We express our sincere gratitude to the editor, associate editor, and anonymous reviewers for their invaluable guidance, support, and constructive feedback. Wang and Liu's research was partially supported by National Natural Science Foundation of China grant 12171079 and the China National Key R&D Program grant 2020YFA0714100. Long Feng is partially supported by Shenzhen Wukong Investment Company, the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education, Tianjin Science Fund for Outstanding Young Scholar (23JCJQJC00150), the Fundamental Research Funds for the Central Universities under Grant No. ZB22000105 and 63233075, the China National Key R&D Program (Grant Nos. 2019YFC1908502, 2022YFA1003703, 2022YFA1003802, 2022YFA1003803) and the National Natural Science Foundation of China Grants (Nos. 12271271, 11925106, 12231011, 11931001 and 11971247). Ma's research was partially supported by grants from national institute of health. Binghui liu and Long

Feng are co-corresponding authors and contributed equally to this paper.

References

Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics textbooks and monographs 155*, 237–290.

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

Baltagi, B. H., Q. Feng, and C. Kao (2012). A lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics 170*(1), 164–177.

Baltagi, B. H., Q. Feng, and C. Kao (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics 191*(1), 176–195.

Baltagi, B. H., C. Kao, and B. Peng (2016). Testing cross-sectional correlation in large panel data models with serial correlation. *Econometrics 4*(4), 1–24.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Breusch, T. S. and A. R. Pagan (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies 47*(1), 239–253.

Cai, T. T., Z. Ren, and H. H. Zhou (2016). Estimating structured high-dimensional covariance and precision matrices: optimal rates and adaptive estimation. *Electronic Journal of Statistics 10*(1), 1–59.

Chen, X. and W. Liu (2018). Testing independence with high-dimensional correlated samples.

The Annals of Statistics 46(2), 866–894.

Feng, L., Y. Ding, and B. Liu (2020). Rank-based tests for cross-sectional dependence in large (n, t) fixed effects panel data models. *Oxford Bulletin of Economics and Statistics 82*(5), 1198–1216.

Feng, L., T. Jiang, B. Liu, and W. Xiong (2022). Max-sum tests for cross-sectional dependence of high-demensional panel data. *The Annals of Statistics 50*(2), 1124–1143.

Gao, J., X. Han, G. Pan, and Y. Yang (2017). High dimensional correlation matrices: The central limit theorem and its applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(3), 677–693.

Gao, J., F. Liu, B. Peng, and Y. Yan (2023). Binary response models for heterogeneous panel data with interactive fixed effects. *Journal of Econometrics 235*(2), 1654–1679.

Hong, Y. (2010). Serial correlation and serial dependence. In *Macroeconometrics and Time Series Analysis*, pp. 227–244. Springer.

Kapoor, M., H. H. Kelejian, and I. R. Prucha (2007). Panel data models with spatially correlated error components. *Journal of econometrics 140*(1), 97–130.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review 40*(2), 509–533.

Lan, W., R. Pan, R. Luo, and Y. Cheng (2017). High dimensional cross-sectional dependence test under arbitrary serial correlation. *Science China Mathematics 60*(2), 345–360.

Lee, L.-f. (2007). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*(2), 489–514.

Lee, L.-f. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of econometrics 154*(2), 165–185.

Li, D. and L. Xue (2015). Joint limiting laws for high-dimensional independence tests. *arXiv:1512.08819*.

Liu, W.-D., Z. Lin, and Q.-M. Shao (2008). The asymptotic distribution and berry–esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability 18*(6), 2337–2366.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels (iza discussion paper no. 1240). *Institute for the Study of Labor (IZA)*.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric reviews 34*(6-10), 1089–1117.

Pesaran, M. H., A. Ullah, and T. Yamagata (2008). A bias-adjusted lm test of error cross-section independence. *The Econometrics Journal 11*(1), 105–127.

Rudelson, M. and R. Vershynin (2013). Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability 18*, 1–9.

Serlenga, L. and Y. Shin (2007). Gravity models of intra-eu trade: application of the ccep-ht estimation in heterogeneous panels with unobserved common time-specific factors. *Journal* *of applied econometrics 22*(2), 361–381.

Wei, W. (2006). *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley.

Hongfei Wang

Department of Statistics, Nanjing Audit University

E-mail: whfei561@163.com

Binghui Liu

School of Mathematics and Statistics & KLAS, Northeast Normal University

E-mail: liubh100@nenu.edu.cn

Long Feng

School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University

E-mail: flnankai@nankai.edu.cn

Yanyuan Ma

Department of Statistics, Pennsylvania State University

E-mail: (yzm63@psu.edu)

Table 1: The empirical size of the five tests under settings (i) and (ii) with distributions (1)-(3) at 5% level.

		Setting (i)								Setting (ii)							
p		3				5				3				5			
N	T	200	300	400	500	200	300	400	500	200	300	400	500	200	300	400	500
Normal distribution																	
100	S_N	3.9	4.8	4.1	4.2	3.8	4.9	4.2	3.3	3.7	4.9	3.9	4.1	4.0	5.1	4.4	3.2
	L_N	3.4	3.2	3.9	4.9	3.6	4.4	3.9	3.9	2.2	2.5	3.7	4.4	3.4	3.9	3.4	4.0
	T_C	3.8	4.5	4.0	4.0	3.3	4.6	4.1	4.0	2.9	3.8	3.8	3.5	2.8	4.2	3.8	3.6
	$\mathrm{CD_P}$	15.9	16.2	16.1	17.0	16.6	18.4	16.2	16.0	19.0	19.7	20.6	21.7	20.4	23.1	20.1	20.6
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
200	S_N	4.4	4.1	4.1	5.9	4.4	4.8	4.8	5.0	4.5	4.2	4.2	5.4	4.3	4.8	4.9	5.1
	L_N	2.1	2.9	3.6	3.6	1.8	2.4	2.2	3.1	1.5	2.7	2.6	3.4	0.9	1.9	1.7	2.1
	T_C	2.4	2.8	4.8	4.3	3.1	4.4	3.6	3.3	2.0	2.6	4.3	3.5	2.6	3.4	3.6	2.5
	$\mathrm{CD_P}$	15.7	15.9	15.8	18.6	14.1	16.8	16.8	16.8	20.2	20.1	20.2	22.8	18.1	20.9	21.2	19.2
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
t_6-distribution																	
100	S_N	3.7	4.9	4.5	4.6	4.2	5.1	5.0	4.6	3.9	5.2	4.4	4.1	4.6	5.2	4.8	4.9
	L_N	3.3	4.5	3.4	5.2	2.0	2.9	3.8	4.3	2.3	3.7	3.1	5.1	1.5	2.6	3.7	3.9
	T_C	3.0	4.7	4.1	4.2	2.8	4.0	4.3	4.1	2.8	4.1	3.5	4.4	2.5	3.8	3.7	3.9
	$\mathrm{CD_P}$	15.3	15.2	15.7	16.4	16.0	17.1	16.7	16.7	17.8	19.4	19.9	19.3	19.3	22.5	20.7	20.5
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
200	S_N	4.0	4.3	4.7	5.1	5.0	5.9	5.3	4.8	4.2	4.1	4.6	4.7	4.8	5.5	5.5	4.2
	L_N	2.4	3.5	4.2	4.2	3.0	2.5	3.2	3.2	1.4	2.1	2.9	3.9	2.0	2.4	3.0	2.8
	T_C	3.0	4.3	4.1	4.2	4.1	4.7	3.8	4.3	2.7	3.5	3.8	3.6	3.5	4.2	3.7	3.8
	$\mathrm{CD_P}$	14.4	15.1	15.2	17.9	13.9	17.3	17.9	16.1	19.2	19.8	19.2	23.2	18.2	20.6	21.9	19.5
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
χ^2_5-distribution																	
100	S_N	4.2	3.8	5.1	5.7	5.2	5.1	4.4	4.2	4.5	3.5	5.1	5.6	4.9	5.2	4.3	4.4
	L_N	3.4	3.4	3.4	3.1	3.4	2.8	4.6	3.7	2.5	2.3	2.9	3.5	2.1	2.6	4.1	3.3
	T_C	3.6	3.9	4.4	5.2	3.6	3.3	3.6	4.0	3.4	3.4	4.0	5.0	3.3	3.2	3.5	3.9
	$\mathrm{CD_P}$	15.8	15.9	16.2	17.6	15.0	16.5	16.4	15.9	20.5	20.9	19.8	21.2	18.6	19.7	20.3	20.3
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
200	S_N	5.2	4.5	3.6	4.0	4.8	4.5	4.5	5.0	5.6	4.5	3.9	4.2	4.6	4.9	4.6	4.9
	L_N	1.9	3.3	3.9	4.9	3.1	3.9	4.1	3.7	1.2	2.7	3.2	3.8	1.7	3.4	3.6	3.2
	T_C	4.3	3.7	4.4	4.3	3.3	4.0	4.0	4.2	4.2	3.2	4.0	3.8	2.7	3.1	3.9	3.7
	$\mathrm{CD_P}$	16.5	16.7	16.2	17.5	14.4	15.8	16.1	17.2	20.6	21.2	20.8	21.9	18.8	19.6	19.0	21.5
	$\mathrm{LM_{PUY}}$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Table 2: The empirical power of the three tests at 5% level under case (I).

		Setting (i)								Setting (ii)							
p		3				5				3				5			
N	T	200	300	400	500	200	300	400	500	200	300	400	500	200	300	400	500
Normal distribution																	
100	S_N	71.3	83.6	91.5	96.6	69.0	82.9	91.6	96.0	65.4	77.6	85.5	93.3	61.8	75.6	87.0	92.8
	L_N	40.9	83.3	98.7	100	35.6	82.9	98.9	100	23.5	59.2	90.4	99.7	22.1	63.7	89.4	99.0
	T_C	81.6	96.2	99.7	100	77.8	96.9	99.8	100	68.1	89.6	98.0	99.9	67.8	91.1	98.5	100
200	S_N	69.7	84.9	91.5	96.4	69.4	84.4	92.5	96.2	61.8	78.3	86.2	93.1	63.2	78.2	87.1	93.4
	L_N	27.9	77.7	98.2	100	25.2	77.0	97.3	100	12.7	53.0	88.7	98.8	12.6	52.9	86.5	99.2
	T_C	75.8	97.6	99.8	100	74.7	97.0	99.9	100	61.6	89.5	98.4	99.7	62.6	90.8	98.2	99.9
t_6-distribution																	
100	S_N	71.3	84.2	92.3	96.1	70.5	83.1	90.7	96.1	63.9	78.9	87.7	93.0	63.7	77.8	86.2	92.5
	L_N	35.9	79.3	98.4	100	34.9	81.5	97.9	100	21.9	57.8	90.3	99.4	20.4	60.9	88.6	99.0
	T_C	79.1	95.7	99.7	99.9	78.5	96.4	99.8	100	67.2	89.8	98.5	99.9	65.1	90.5	98.4	99.9
200	S_N	69.9	85.9	92.1	96.4	70.8	84.3	92.0	96.9	62.4	77.4	87.1	93.0	64.2	78.7	87.6	93.5
	L_N	27.3	78.8	98.6	100	24.3	75.4	98.4	100	13.0	54.7	87.3	98.1	12.9	51.2	86.5	99.2
	T_C	77.0	96.7	99.7	100	75.7	96.0	99.7	100	63.4	89.1	97.7	99.9	63.7	88.9	97.5	99.8
χ_5^2-distribution																	
100	S_N	70.8	85.4	92.5	96.5	71.5	84.6	91.3	97.0	65.0	79.2	88.3	93.1	65.9	77.1	87.7	93.9
	L_N	41.4	82.5	98.2	100	36.8	81.7	98.0	100	24.7	62.9	90.3	98.8	21.3	60.9	88.4	99.1
	T_C	81.3	97.1	99.9	100	80.3	95.9	99.5	100	69.6	91.5	98.6	99.9	70.2	89.7	97.8	100
200	S_N	69.6	84.7	94.0	98.3	72.8	84.4	94.3	96.5	63.9	78.8	89.4	95.1	65.9	77.9	90.4	93.1
	L_N	31.4	78.6	98.5	100	33.8	79.1	98.4	100	16.5	54.1	90.2	99.3	17.4	54.4	88.6	99.0
	T_C	78.0	96.7	99.9	100	80.8	97.1	99.9	100	66.0	90.0	99.1	100	66.8	90.5	99.2	99.9

Table 3: The empirical power of the three tests at 5% level under case (II).

		Setting (i)								Setting (ii)							
p		3				5				3				5			
N	T	200	300	400	500	200	300	400	500	200	300	400	500	200	300	400	500
Normal distribution																	
100	S_N	7.5	8.7	8.4	11.7	8.9	10.5	10.6	10.9	7.6	8.2	8.4	10.8	7.9	9.8	9.2	9.3
	L_N	99.1	100	100	100	100	100	100	100	97.0	100	100	100	97.1	100	100	100
	T_C	98.4	100	100	100	99.6	100	100	100	94.7	100	100	100	93.7	100	100	100
200	S_N	5.9	6.4	5.9	8.7	5.5	7.0	8.1	7.6	5.6	5.8	5.3	8.6	5.7	6.9	7.3	6.6
	L_N	58.3	97.2	99.6	100	48.5	97.1	100	100	35.6	88.7	99.2	99.9	32.3	87.5	98.4	100
	T_C	46.2	93.3	99.1	100	39.4	94.2	99.8	100	27.1	80.2	98.3	99.9	24.6	80.2	97.3	100
t_6-distribution																	
100	S_N	6.9	9.0	9.4	10.3	8.4	9.3	8.9	10.3	6.5	9.4	8.3	9.7	7.8	8.8	8.7	9.5
	L_N	99.8	100	100	100	99.6	100	100	100	96.0	100	100	100	97.2	100	100	100
	T_C	99.7	100	100	100	98.4	100	100	100	92.7	100	100	100	93.0	100	100	100
200	S_N	5.7	5.9	6.3	7.3	5.6	8.6	8.2	7.2	5.5	5.9	5.4	7.4	5.6	7.7	8.1	6.3
	L_N	65.3	98.2	100	100	60.9	96.0	100	100	35.8	89.1	99.3	100	33.5	88.8	99.0	100
	T_C	53.9	95.7	99.9	100	48.8	92.7	99.8	100	27.8	79.6	98.4	100	25.1	81.5	98.3	99.8
χ^2_5-distribution																	
100	S_N	8.6	9.5	9.7	12.1	8.8	8.3	8.9	11.7	7.6	7.8	9.3	11.4	8.0	8.8	8.6	11.6
	L_N	99.9	100	100	100	100	100	100	100	97.0	100	100	100	97.7	100	100	100
	T_C	99.8	100	100	100	99.8	100	100	100	93.2	100	100	100	93.9	100	100	100
200	S_N	7.1	7.0	7.9	8.9	5.5	7.2	7.8	7.3	7.2	6.4	7.1	8.5	6.7	6.8	8.1	6.8
	L_N	56.3	98.1	99.9	100	59.1	97.9	100	100	33.9	87.4	99.4	100	35.4	85.5	99.1	100
	T_C	47.2	96.2	99.8	100	48.4	95.0	100	100	25.2	81.0	98.5	100	27.7	78.5	97.7	100