



Homogeneity tests for one-way models with dependent errors under correlated groups

Yuichi Goto¹ · Koichi Arakaki² · Yan Liu^{3,4} · Masanobu Taniguchi³

Received: 16 December 2021 / Accepted: 2 August 2022 / Published online: 2 September 2022
© The Author(s) under exclusive licence to Sociedad de Estadística e Investigación Operativa 2022

Abstract

We consider the problem of testing for the existence of fixed effects and random effects in one-way models, where the groups are correlated and the disturbances are dependent. The classical F -statistic in the analysis of variance is not asymptotically distribution-free in this setting. To overcome this problem, we propose a new test statistic for this problem without any distributional assumptions, so that the test statistic is asymptotically distribution-free. The proposed test statistic takes the form of a natural extension of the classical F -statistic in the sense of distribution-freeness. The new tests are shown to be asymptotically size α and consistent. The nontrivial power under local alternatives is also elucidated. The theoretical results are justified by numerical simulations for the model with disturbances from linear time series with innovations of symmetric random variables, heavy-tailed variables, and skewed variables, and furthermore from GARCH models. The proposed test is applied to log-returns for stock prices and uncovers random effects in sectors.

✉ Yuichi Goto
yuichi.goto@math.kyushu-u.ac.jp

Koichi Arakaki
arakaki74@akane.waseda.jp

Yan Liu
liu@ims.sci.waseda.ac.jp

Masanobu Taniguchi
taniguchi@waseda.jp

- ¹ Department of Mathematical Sciences, Faculty of Mathematics, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan
- ² Department of Pure and Applied Mathematics, Graduate School of Fundamental Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
- ³ Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
- ⁴ Institute for Mathematical Science, Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan

Keywords Dynamic panel data · Fixed effect · Homogeneity test · Longitudinal data · One-way model · Random effect

Mathematics Subject Classification 62G10 · 62H15 · 62J10

1 Introduction

Longitudinal data and panel data are omnipresent in the real world. Statistical methods to analyze such data have been studied for several decades (Diggle et al. 2002). The methods have a wide range of applications, e.g., analysis of stress in mothers (Zeger et al. 1985), the weight of infants (Hoover et al. 1998), and COVID-19 data (Bernardes et al. 2020; Lucas et al. 2020).

The analysis of variance (ANOVA) is a common method to test for equality among groups. An F -statistic, defined as the ratio of variance between groups to variance within groups, is designed to test for the homogeneity of groups for independent and identically distributed (i.i.d.) data. Numerous papers are devoted to ANOVA and related topics for i.i.d. data (see, e.g., Searle et al. 1992; Rashid 1995; Clarke 2008; Liu and Xu 2016, and references there in). By contrast, the statistic does not work for dependent data. To resolve the issue, Nagahata and Taniguchi (2018) studied a test for the equality of means among groups based on the Whittle likelihood for multivariate one-way fixed effect models. Their statistic can be rephrased as the classical F -statistic rescaled by the spectral density of disturbances. They showed their statistic is asymptotically Chi-square distributed, although they did not derive the consistency of the test and assumed independence of groups.

One-way models for time series are closely related to the analysis of longitudinal data and dynamic panel data. For dynamic panel data, Baltagi and Li (1991) constructed the consistent estimator of variance of random effects for dynamic panel data models with errors from the autoregressive process of order 1 (AR(1)), provided that the number of groups and sample size tends to infinity. Galbraith and Zinde-Walsh (1995) dealt with error components models for panel data models with errors from the autoregressive moving-average process of orders p and q . You and Zhou (2013) advocated semiparametric panel data partially linear additive models with errors from the AR(1). The statistical methods for longitudinal data also have been intensively investigated. For example, Tang and Leng (2011) estimated regression coefficients by the empirical likelihood. Li (2011) constructed an efficient estimator for semiparametric regression models. A panel data model with common shocks is proposed by Bai and Li (2014), and Ergemen and Velasco (2017) extended the model to a fractionally integrated panel data model with common shocks. Under high-dimensional settings, Zhong et al. (2019) considered a test for homogeneity of covariance matrices and constructed a change test for covariance matrices. Fang et al. (2020) proposed a test for regression parameters. However, the principal objective in these fields is not fixed and random effects but is regression coefficients.

The importance of fixed effects and random effects has been recognized, whereas, to our best knowledge, there are few references of diagnostic tests for fixed effects and random effects. On a related topic, Akharif et al. (2020) and Fihri et al. (2020)

established optimal tests for the existence of random coefficients for i.i.d. data based on the locally asymptotic normality for random coefficient regression models. The optimal test based on multivariate ranks for the existence of fixed effects for i.i.d. data proposed by Hallin et al. (2021). Recently, González et al. (2021) have discussed tests for the existence of fixed effects and interactions for two-way models for spatial point processes. Ditzhaus et al. (2021) proposed robust tests based on quantiles for fixed effects and interaction for i.i.d. random variables.

We propose a test for the existence of fixed or random effects in one-way models for correlated groups and derive the asymptotic null distribution. In addition, the consistency of the proposed test and the nontrivial power under the local alternatives are elucidated. The numerical study illustrates the finite sample performance of the proposed test and comparison with the classical test. In particular, we also include the skewness and the heteroscedasticity in the disturbance process, which reveals its own importance in practical applications (Cook and Weisberg 1983). In this study, we also compare our statistic with the classical statistic. The classical statistic, defined in Sect. 2, assumes independence between groups, which is a major drawback in its application. The new statistic, defined in Sect. 3, elaborately relaxes the strong assumption of independence between groups. We emphasize that our setting allows us to deal with correlated groups, and thus, our proposed method has a wide range of applications.

A motivated real data example with correlated groups is the analysis of stock prices. Stock prices can be categorized by industry. Equity-focused investors believe that the stock prices are linked by factors related to earnings. For example, stock prices of automobile companies are linked to exchange rates. In other words, equity-focused investors believe that there are random effects related to industries. Our test which takes into account correlations between groups can be applied to verify this hypothesis.

This paper is organized as follows: We briefly review spectra and the classical settings and test statistic in Sect. 2. In Sect. 3, we introduce the fixed effects model and propose a new test for the existence of fixed effects. In Sect. 4, we deal with the random effects model and derive the asymptotic results for the proposed test. Section 5 presents the simulation study. In Sect. 6, we apply our test for the existence of effects to the log-returns in stock prices. The discussion is provided in Sect. 7. Supplementary material includes all proofs of theorems and additional simulation results.

2 Preliminary

2.1 Spectral density

In the frequency-domain approach, the L^2 -based spectral density is a pivotal index to describe time-dependent structures of data. To recall the definition, let X_t be a strictly stationary process with the autocovariance function $\gamma_X(h) = EX_t X_{t+h}$ satisfying $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$. Then, the spectral density function is defined, for $\lambda \in [-\pi, \pi]$, as

$$f_X(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_X(h)e^{-ih\lambda}. \tag{1}$$

Since $\gamma_X(h) = \int_{-\pi}^{\pi} f_X(\lambda)e^{ih\lambda}d\lambda$, the information of the spectrum $f_X(\lambda)$ is equivalent to that of autocovariance functions for all lags $\{\gamma_X(h)\}_{h \in \mathbb{Z}}$. A multivariate spectral density function can be defined by replacing $\gamma_X(h)$ in (1) with $\Gamma_X(h) = E X_t X_{t+h}^T$ for a p -dimensional strictly stationary process X_t . Typical examples of spectra are the spectrum for ARMA models of orders (p, q) and the exponential type of the spectrum proposed by Bloomfield (1973), taking the forms of

$$f_{\text{ARMA}}(\lambda) = \frac{\sigma^2 |1 + \theta_1 e^{-i\lambda} + \dots + \theta_q e^{-iq\lambda}|^2}{2\pi |1 - \phi_1 e^{-i\lambda} - \dots - \phi_p e^{-ip\lambda}|^2}$$

and $f_{\text{EXP}}(\lambda) = \frac{\sigma^2}{2\pi} \exp\left(2 \sum_{r=1}^d \varsigma_r \cos(r\lambda)\right)$,

where $\sigma, \theta_1, \dots, \theta_q, \phi_1, \dots, \phi_p, \varsigma_1, \dots, \varsigma_d$ are parameters, respectively. Other examples can be found by, e.g., Chiu (1988). We refer readers to von Sachs (2020) for review.

2.2 Classical setting and statistic

Nagahata and Taniguchi (2018) discussed one-way models with independent groups; for a fixed group size a , a growing sample size n_i of the i th group ($i = 1 \dots, a$), and a fixed dimension p of time series in each group,

$$y_{it} = \mu + \tau_i + e_{it}, \quad i = 1, \dots, a; t = 1, \dots, n_i, \tag{2}$$

where $y_{it} = (y_{it1}, \dots, y_{itp})^T$ is a t th p -dimensional observation of an i th group, $\mu = (\mu_1, \dots, \mu_p)^T$ is a general mean, $\tau_i = (\tau_{i1}, \dots, \tau_{ip})^T$ is a fixed effect such that $\sum_{i=1}^a \tau_i = \mathbf{0}$, and $e_{it} = (e_{it1}, \dots, e_{itp})^T$ is a centered strictly stationary sequence such that $\{e_{it}\}_{t \in \mathbb{Z}}$ is independent of $\{e_{jt}\}_{t \in \mathbb{Z}}$ for $j \neq i$ and e_{it} has a p -by- p spectral density matrix $f(\lambda) = (f_{j_1 j_2}(\lambda))_{j_1, j_2=1, \dots, p}$ which is independent of i . For a test for existence of fixed effects defined in (5), they proposed the following test statistic

$$S_n = n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^T \left(2\pi \tilde{f}_n(0)\right)^{-1} (\bar{y}_i - \bar{y}_{..}), \tag{3}$$

where $\bar{y}_i = \sum_{t=1}^{n_i} y_{it}/n_i, \bar{y}_{..} = \sum_{i=1}^a \sum_{t=1}^{n_i} y_{it}/(an_i), \tilde{f}_n(0)$ is defined as

$$\tilde{f}_n(0) = \frac{1}{a} \sum_{i=1}^a \hat{f}_{ii}(\lambda)/\rho_i,$$

where $\hat{f}_{ii}(\lambda)$ is given in (6), and $\rho_i = n_i/n$ with $n = \sum_{i=1}^a n_i$. This statistic is standardized within groups, and thus, the test based on S_n is asymptotically distribution-free in the case of independent groups (see Sect. 7). However, it does not hold when groups are correlated. This paper focuses on data with correlated groups such as stock prices are considered. In stock prices, sectors correspond to groups. We propose the test statistic standardized not only within groups but also between groups, defined in (7) so that our test statistic is asymptotically distribution-free. In this sense, our statistic takes the form of the natural extension of S_n .

3 Test for existence of fixed effects

In this section, we scrutinize one-way fixed effects model with dependent disturbance processes when the number of groups is fixed and the number of observations for each group diverges. Let us consider the model

$$y_{it} = \mu + \tau_i + e_{it}, \quad i = 1, \dots, a; \quad t = 1, \dots, n_i, \tag{4}$$

where $y_{it} = (y_{it1}, \dots, y_{itp})^T$ is a t th p -dimensional observation of an i th group, $\mu = (\mu_1, \dots, \mu_p)^T$ is a general mean, $\tau_i = (\tau_{i1}, \dots, \tau_{ip})^T$ is a fixed effect such that $\sum_{i=1}^a \tau_i = \mathbf{0}$, and $e_{it} = (e_{it1}, \dots, e_{itp})^T$ is a centered strictly stationary sequence. Suppose that an observed stretch $\{y_{it}; i = 1, \dots, a, t = 1, \dots, n_i\}$ is available, and $(e_{1t}^T, \dots, e_{at}^T)^T$ has an ap -by- ap spectral density matrix $f(\lambda) = (f_{ij}(\lambda))_{i,j=1,\dots,a}$ for $\lambda \in [-\pi, \pi]$. In addition, there exists $\rho_i \in (0, 1)$ such that $n_i = \rho_i n$ with $n = \sum_{i=1}^a n_i$. The number of groups, the length of time series from an i th group, and the dimension of time series from each group at each time are denoted as a, n_i , and p , respectively. The role of p is to include the multivariate analysis of variance (MANOVA) case. Obviously, $p = 1$ corresponds to the univariate ANOVA.

Remark 1 The above one-way model defined in (4) seems that only one time series for each group can be coped with, whereas we can handle the case that there are more than one time series for each group by reconfiguring the settings as follows: taking p as pq for $q \in \mathbb{N}$, $y_{it} = (y_{it11}, \dots, y_{it1q}, y_{it21}, \dots, y_{itp1}, \dots, y_{itpq})^T$, where $\mathbf{1}_q$ is a q -dimensional vector with all elements equal to one, $\mu = (\mu_1 \mathbf{1}_q^T, \dots, \mu_p \mathbf{1}_q^T)^T$, $\tau_i = (\tau_{i1} \mathbf{1}_q^T, \dots, \tau_{ip} \mathbf{1}_q^T)^T$, and $e_{it} = (e_{it11}, \dots, e_{it1q}, e_{it12}, \dots, e_{itp1}, \dots, e_{itpq})^T$. Moreover, p and q can depend on i . In this case, p and q represent the dimension of time series from each group at each time and the number of time series in each group, respectively.

Remark 2 The condition $\sum_{i=1}^a \tau_i = \mathbf{0}$ is not essential. When $\sum_{i=1}^a \tau_i \neq \mathbf{0}$, we can redefine μ as $\mu - \sum_{i=1}^a \tau_i$ and τ_i as $\tau_i - \sum_{i=1}^a \tau_i$.

Let the null hypothesis H_0 and the alternative K_0 be

$$H_0 : \tau_1 = \dots = \tau_a \quad \text{vs} \quad K_0 : \tau_i \neq \mathbf{0} \quad \text{for some } i. \tag{5}$$

Under the assumption $\sum_{i=1}^a \tau_i = \mathbf{0}$, the null hypothesis is equivalent to $\tau_i = \mathbf{0}$ for all $i \in \{1, \dots, a\}$.

Let $\hat{f}_n(\lambda) = (\hat{f}_{ij}(\lambda))_{i,j=1,\dots,a}$ be the nonparametric spectral density estimator defined as

$$\hat{f}_{ij}(\lambda) = \frac{1}{2\pi} \sum_{\{h \in \mathbb{Z}; |h| \leq \min\{n_i, n_j\} - 1\}} \omega\left(\frac{h}{M_n}\right) \hat{\Gamma}_{ij}(h) e^{-ih\lambda}, \quad \lambda \in [-\pi, \pi], \quad (6)$$

where $\omega(x) = \int_{-\infty}^{\infty} W(t)e^{ixt} dt$ and the function $W(\cdot)$ satisfy Assumption 3.2. Here, M_n is a positive sequence such that $M_n \rightarrow \infty$ and $M_n/\min_{i=1,\dots,a} n_i \rightarrow 0$ as $\min_{i=1,\dots,a} n_i \rightarrow \infty$, for $h \in \{0, \dots, \min\{n_i, n_j\} - 1\}$,

$$\hat{\Gamma}_{ij}(h) = \frac{1}{\min\{n_i, n_j\} - |h|} \sum_{t=1}^{\min\{n_i, n_j\} - |h|} (\mathbf{y}_{i(t+h)} - \bar{\mathbf{y}}_i)(\mathbf{y}_{jt} - \bar{\mathbf{y}}_j)^T,$$

for $h \in \{-\min\{n_i, n_j\} + 1, \dots, 0\}$, and

$$\hat{\Gamma}_{ij}(h) = \frac{1}{\min\{n_i, n_j\} - |h|} \sum_{t=-h+1}^{\min\{n_i, n_j\}} (\mathbf{y}_{i(t+h)} - \bar{\mathbf{y}}_i)(\mathbf{y}_{jt} - \bar{\mathbf{y}}_j)^T,$$

where $\bar{\mathbf{y}}_i = \sum_{t=1}^{n_i} \mathbf{y}_{it}/n_i$, and $\bar{\mathbf{y}}_{..} = \sum_{i=1}^a \sum_{t=1}^{n_i} \mathbf{y}_{it}/(an_i)$. Let $\hat{\mathbf{V}}_n = (\hat{\mathbf{V}}_{ij})_{i,j=1,\dots,a}$ be

$$\begin{aligned} \hat{\mathbf{V}}_{ij} &= \frac{2\pi \min\{\rho_i, \rho_j\}}{\rho_i \rho_j} \hat{f}_{ij}(0) - \frac{2\pi}{a} \sum_{s=1}^a \left\{ \frac{\min\{\rho_s, \rho_j\}}{\rho_s \rho_j} \hat{f}_{sj}(0) + \frac{\min\{\rho_i, \rho_s\}}{\rho_i \rho_s} \hat{f}_{is}(0) \right\} \\ &+ \frac{2\pi}{a^2} \sum_{s,k=1}^a \frac{\min\{\rho_s, \rho_k\}}{\rho_s \rho_k} \hat{f}_{sk}(0). \end{aligned}$$

The test statistic for H_0 is proposed as

$$T_n = n(\bar{\mathbf{y}}_{1.}^T - \bar{\mathbf{y}}_{..}^T, \dots, \bar{\mathbf{y}}_{a.}^T - \bar{\mathbf{y}}_{..}^T) \hat{\mathbf{V}}_n^- (\bar{\mathbf{y}}_{1.}^T - \bar{\mathbf{y}}_{..}^T, \dots, \bar{\mathbf{y}}_{a.}^T - \bar{\mathbf{y}}_{..}^T)^T, \quad (7)$$

where $\hat{\mathbf{V}}_n^-$ denotes the Moore–Penrose inverse of $\hat{\mathbf{V}}_n$. Using the Moore–Penrose inverse $\hat{\mathbf{V}}_n^-$ in T_n is essential since $\hat{\mathbf{V}}_n$ is a singular matrix. Actually, $\sum_{i=1}^a \hat{\mathbf{V}}_{ij} = \mathbf{O}_p$ for any j , where \mathbf{O}_p is an p -by- p zero matrix; thus, 0 is an eigenvalue of $\hat{\mathbf{V}}_n$. It is worth mentioning that our proposed test statistic T_n is scale-invariant. Since $(\bar{\mathbf{y}}_{1.}^T - \bar{\mathbf{y}}_{..}^T, \dots, \bar{\mathbf{y}}_{a.}^T - \bar{\mathbf{y}}_{..}^T)$ converges in distribution to a centered normal distribution with variance \mathbf{V} , defined in Theorem 1, and \mathbf{V} is the function of the spectral density matrix $\mathbf{f}(\lambda)$ (see Lemma 1 in Section A in the supplementary material), $\mathbf{f}(\lambda)$ appears.

To state the assumptions, we define, for a random variables $\{X_t\}$, the cumulant of order ℓ of (X_1, \dots, X_ℓ) as

$$\text{cum}(X_1, \dots, X_\ell) = \sum_{(v_1, \dots, v_p)} (-1)^{p-1} (p-1)! \left(E \prod_{j \in v_1} X_{v_j} \right) \dots \left(E \prod_{j \in v_p} X_{v_j} \right),$$

where the summation $\sum_{(v_1, \dots, v_p)}$ extends over all partitions (v_1, \dots, v_p) of $\{1, 2, \dots, \ell\}$ (see Brillinger 1981, p. 19). The following assumptions are made throughout the paper.

Assumption 3.1 For all $\ell \in \mathbb{N}$, $(k_1, \dots, k_\ell) \in \{1, \dots, a\}^\ell$, and $(r_1, \dots, r_\ell) \in \{1, \dots, p\}^\ell$,

$$\sum_{s_2, \dots, s_\ell = -\infty}^{\infty} \left(1 + \sum_{j=1}^{\ell} |s_j| \right) \left| \kappa_{r_1 \dots r_\ell}^{k_1 \dots k_\ell}(s_2, \dots, s_\ell) \right| < \infty,$$

where $\kappa_{r_1 \dots r_\ell}^{k_1 \dots k_\ell}(s_2, \dots, s_\ell) = \text{cum}\{e_{k_1 0 r_1}, e_{k_2 s_2 r_2}, \dots, e_{k_\ell s_\ell r_\ell}\}$.

Assumption 3.2 $W(\cdot)$ is a real, bounded, nonnegative, even function such that $\int_{-\infty}^{\infty} W(t) dt = 1$ and $\int_{-\infty}^{\infty} W^2(t) dt < \infty$ with a bounded derivative.

Assumption 3.3 $\text{rank}(\hat{V}_n)$ converges in probability to $\text{rank}(V)$, where V is defined in Theorem 1, as $\min_{i=1, \dots, a} n_i \rightarrow \infty$.

We briefly explain all assumptions. Assumption 3.1 is an assumption often imposed for dependent observations (see Brillinger 1981, p. 26). It implies the asymptotic normality of $(\bar{y}_1^T - \bar{y}_\cdot^T, \dots, \bar{y}_a^T - \bar{y}_\cdot^T)$. This assumption can be relaxed as Remark 1 in Section A in the supplementary material. Assumption 3.2 is a natural assumption for the nonparametric spectral density estimator. In conjunction with Assumption 3.1, $\hat{f}_n(\lambda)$ is a consistent estimator (see Brillinger 1981, Corollaries 5.6.1 and 5.6.2 and Theorem 5.9.1). Other conditions which ensure the consistency of the nonparametric spectral density estimator can be seen in Robinson (1991). Assumption 3.3 is a technical assumption to ensure \hat{V}_n^- converges in probability to V^- as $\min_{i=1, \dots, a} n_i \rightarrow \infty$ (see Rakocevic 1997; Stewart 1969).

Remark 3 When we assume independence of groups and $f_{11}(0) = \dots = f_{aa}(0)$, \hat{V}_n fulfills Assumption 3.3. As an illustration, we set $p = 1$ and $a = 3$. Then,

$$V = \begin{pmatrix} 1 - 1/a & -1/a & -1/a \\ -1/a & 1 - 1/a & -1/a \\ -1/a & -1/a & 1 - 1/a \end{pmatrix} 2\pi f(0),$$

and for matrices,

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 - 1/a & -1/a & -1/a \\ -1/a & 1 - 1/a & -1/a \end{pmatrix} 2\pi f(0),$$

it holds that

$$P\mathbf{V} = \begin{pmatrix} \mathbf{B} \\ 0 & 0 & 0 \end{pmatrix}.$$

Also, the matrix $P\hat{\mathbf{V}}_n$ takes the form of

$$\begin{pmatrix} \hat{\mathbf{B}}_n \\ 0 & 0 & 0 \end{pmatrix},$$

where $\hat{\mathbf{B}}_n$ is an appropriate $(a - 1)$ -by- a matrix. Since \mathbf{B} is a full rank matrix and the set of all full rank $(a - 1)$ -by- a matrices is open, $\hat{\mathbf{B}}_n$ is a full rank matrix for large n . Hence, the condition is confirmed.

Then, we obtain the following asymptotic null distribution based on Rao and Mitra (1971, Theorem 9.2.3, p. 173).

Theorem 1 *Suppose Assumptions 3.1–3.3 hold. Under H_0 , T_n converges in distribution to the Chi-square distribution with r degrees of freedom as $\min_{i=1,\dots,a} n_i \rightarrow \infty$, where $r = \text{rank}(\mathbf{V})$ and $\mathbf{V} = (\mathbf{V}_{ij})_{i,j=1,\dots,a}$ with*

$$\begin{aligned} \mathbf{V}_{ij} = & \frac{2\pi \min\{\rho_i, \rho_j\}}{\rho_i \rho_j} f_{ij}(0) - \frac{2\pi}{a} \sum_{s=1}^a \left\{ \frac{\min\{\rho_s, \rho_j\}}{\rho_s \rho_j} f_{sj}(0) + \frac{\min\{\rho_i, \rho_s\}}{\rho_i \rho_s} f_{is}(0) \right\} \\ & + \frac{2\pi}{a^2} \sum_{s,k=1}^a \frac{\min\{\rho_s, \rho_k\}}{\rho_s \rho_k} f_{sk}(0). \end{aligned}$$

From Theorem 1, we obtain an asymptotically size α test whether we reject H_0 when $T_n \geq \chi_{\hat{r}_n}^2 [1 - \alpha]$, where $\hat{r}_n = \text{rank}(\hat{\mathbf{V}}_n^-)$ and $\chi_{\hat{r}_n}^2 [1 - \alpha]$ denotes the upper α -percentiles of the Chi-square distribution with \hat{r}_n degrees of freedom.

We elucidate the theoretical power of the test in the next theorem.

Theorem 2 *Suppose Assumptions 3.1–3.3 hold. Under the alternative K_0 , the power of the above test based on T_n converges to 1, as $\min_{i=1,\dots,a} n_i \rightarrow \infty$. In other words, the test is consistent.*

To see the nontrivial power of the proposed test, let us consider local alternative hypotheses. Provided the perturbations $\mathbf{h}_1, \dots, \mathbf{h}_a$ satisfying $\sum_{i=1}^a \mathbf{h}_i = \mathbf{0}$, the local alternative is defined as

$$K_0^{(n)} : \boldsymbol{\tau}_i = \frac{\mathbf{h}_i}{\sqrt{n}} \quad (i = 1, \dots, a).$$

Theorem 3 *Suppose Assumptions 3.1–3.3 hold. Under the local alternatives $K_0^{(n)}$, T_n converges in distribution to the noncentral Chi-square distribution with r degrees of freedom and the noncentrality parameter $\delta = (\mathbf{h}_1^T, \dots, \mathbf{h}_a^T) \mathbf{V}^- (\mathbf{h}_1^T, \dots, \mathbf{h}_a^T)^T$, as $\min_{i=1,\dots,a} n_i \rightarrow \infty$.*

In view of this theorem, the nontrivial asymptotic power of the test under the local alternatives can be expressed as

$$1 - \Psi_{r,\delta}(\chi_r^2[1 - \alpha]),$$

where $\Psi_{r,\delta}$ is the cumulative distribution function of the noncentral Chi-square with r degrees of freedom and the noncentrality parameter δ .

Remark 4 In case that the number of time series in each group is greater than one ($q \geq 2$, see Remark 1), the multiple comparison problem occurs since our test provides different p-values for different orders of time series. For example, for $p = 1$, we obtain $(q!)^{a-1}$ different p-values in total. To avoid the multiple comparison problem, we propose that $\mathbf{y}_{it} = (y_{it1}, y_{it2}, \dots, y_{itp})^T$, where $y_{itpq} = \sum_{j=1}^q y_{itpj}/q$, is used instead of $(y_{it11}, \dots, y_{it1q}, y_{it21}, \dots, y_{itp1}, \dots, y_{itpq})^T$.

4 Test for existence of random effects

In this section, we consider the one-way random effects model with a series of strictly stationary residuals when the number of groups is fixed and the number of observations for each group diverges. The only difference from the fixed effects model (4) is that $\boldsymbol{\tau}_i$ is random effect of the i th group. To be simple, we assume $(\boldsymbol{\tau}_1^T, \dots, \boldsymbol{\tau}_a^T)^T$ follows the ap -dimensional centered normal distribution with variance $\boldsymbol{\Sigma}^\tau = (\boldsymbol{\Sigma}_{ij}^\tau)_{i,j=1,\dots,a}$. Here, $\{\boldsymbol{\tau}_j\}$ are supposed to be independent of any disturbance process $\{e_{it}; t = 1, \dots, n_i\}$. In this random effects model, the spectral density of \mathbf{y}_{it} does not exist due to the random effects.

Let the null hypothesis H_1 and the alternative K_1 for the existence of random effects be

$$H_1 : \boldsymbol{\Sigma}^\tau = \mathbf{O}_{ap} \quad \text{vs} \quad K_1 : \boldsymbol{\Sigma}^\tau \neq \mathbf{O}_{ap}, \quad (8)$$

where \mathbf{O}_{ap} is an ap -by- ap zero matrix. The test statistic T_n , defined in (7), is still available in this situation. The following theorem shows that the asymptotic null distribution is exactly the same as that for the fixed effects model.

Theorem 4 *Suppose Assumptions 3.1–3.3 hold. Under the null H_1 , T_n converges in distribution to the Chi-square distribution with r degrees of freedom as $\min_{i=1,\dots,a} n_i \rightarrow \infty$.*

In consequence, we reject H_1 in favor of K_1 if $T_n \geq \chi_{r_n}^2[1 - \alpha]$. The consistency of the test is shown as follows.

Theorem 5 *Suppose Assumptions 3.1–3.3 hold. Under the alternative K_1 , the proposed test is consistent. More precisely, under the alternative K_1 , $\text{pr}(T_n \geq \chi_{r_n}^2[1 - \alpha]) \rightarrow 1$, as $\min_{i=1,\dots,a} n_i \rightarrow \infty$.*

Now we consider the local alternative hypothesis to study the nontrivial power of the test based on T_n . Let $\mathbf{H} = (\mathbf{H}_{ij})_{i,j=1,\dots,a}$ be an ap -by- ap symmetric, positive definite matrix, and the local alternatives $K^{(n)}$ be defined as

$$K_1^{(n)} : \Sigma^\tau = \frac{\mathbf{H}}{n}.$$

The nontrivial power of the proposed test is elucidated in the next result.

Theorem 6 *Suppose Assumptions 3.1–3.3 hold. Under the alternatives $K_1^{(n)}$, we have*

$$\lim_{\min_{i=1,\dots,a} n_i \rightarrow \infty} \text{pr}(T_n \geq \chi_{r_n}^2 [1 - \alpha]) = \text{pr}(\mathbf{Z}^\top \mathbf{V}^{-1} \mathbf{Z} \geq \chi_r^2 [1 - \alpha]),$$

where \mathbf{Z} follows an ap -dimensional centered normal distribution with variance $\tilde{\mathbf{H}} + \mathbf{V}$; Here, $\tilde{\mathbf{H}} = (\tilde{\mathbf{H}}_{ij})_{i,j=1,\dots,a}$ is determined in terms of the matrix \mathbf{H} as

$$\tilde{\mathbf{H}}_{ij} = \mathbf{H}_{ij} - \frac{1}{a} \sum_{s=1}^a (\mathbf{H}_{sj} + \mathbf{H}_{is}) + \frac{1}{a^2} \sum_{s,k=1}^a \mathbf{H}_{sk}.$$

Remark 5 We can generalize the random effects $(\boldsymbol{\tau}_1^\top, \dots, \boldsymbol{\tau}_a^\top)^\top$ to an ap -dimensional random vector and show corresponding theorems to Theorems 4–6.

5 Numerical study

The finite sample performance of the proposed test based on T_n and comparison with the classical test based on S_n are illustrated in this section. To be specific, we let the dimension of time series from each group at each time p , the number of time series in each group q , and the number of groups a be $p = 1, q = 1$, and $a = 3, 9$. The sample sizes are set as (I) $n_1 = \dots = n_a = 1000$, (II) $n_{3k-1} = n_{3k-2} = 2000$ and $n_{3k} = 1000$ for $k \leq a/3$, (III) $n_1 = \dots = n_a = 2000$. (I) and (III) are cases of the sample size of each group being equal (balanced design). (II) is the case of the sample size of each group being unequal (unbalanced design). For each $1 \leq t \leq \max_{i=1,\dots,a} n_i$, denote $(e_{1t}, \dots, e_{at})^\top$ by $(\mathbf{e}_t) = (\mathbf{e}_{it})_{i=1,\dots,a}$.

We consider two scenarios, independent groups (Case 1) and correlated groups (Case 2). The disturbance process $\{\mathbf{e}_{it}\}$ is supposed to follow a multivariate moving-average model or a generalized autoregressive conditional heteroscedasticity model. Let $\{\boldsymbol{\varepsilon}_t\}$ be an i.i.d. sequence in the following.

As for Processes 1–3, we suppose $\mathbf{e}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Phi} \mathbf{e}_{t-1}$, with the coefficient matrix $\boldsymbol{\Phi} = (\Phi_{ij})$, where, in Case 1, $\boldsymbol{\Phi} = 0.5 \mathbf{I}_a$ and, in Case 2, $\Phi_{3k-2,3k-2} = 0.7, \Phi_{3k-1,3k-1} = -0.5, \Phi_{3k,3k} = 0.3, \Phi_{3k,3k-2} = 0.3, \Phi_{3k,3k-1} = -0.1$ for positive integer $k \leq a/3$; and otherwise $\Phi_{ij} = 0$.

Process 1: In Case 1, each component of $\boldsymbol{\varepsilon}_t$ follows a centered normal distribution with unit variance, which is of independent other components of $\boldsymbol{\varepsilon}_t$. In Case 2, $\boldsymbol{\varepsilon}_t$

is distributed as a zero mean multivariate normal distribution with covariance matrix $\Sigma = (\Sigma_{ij})$, where $\Sigma_{ii} = 1$ and $\Sigma_{j,j+1} = \Sigma_{j+1,j} = 0.5$ for $1 \leq i \leq a, 1 \leq j \leq a - 1$.

Process 2: In Case 1, each component of ϵ_t follows a centered t-distribution with 5 degrees of freedom, which is of independent other components of ϵ_t . In Case 2, ϵ_t is distributed as a zero mean multivariate t-distribution with 5 degrees of freedom, with the scale matrix Σ defined in Process 1.

Process 3: In Case 1, each component of ϵ_t follows a centered skew normal distribution with location parameter 0, scale 1 and shape parameter 50, which is of independent other components of ϵ_t . The noncentered skew normal distribution has a nonzero mean $50\sqrt{2}/\sqrt{\pi(1 + 50^2)}$. In Case 2, ϵ_t is distributed as a centered multivariate skew normal distribution with location parameter $\mathbf{0}_a$, correlation matrix Σ defined in Process 1, and shape parameter $\zeta = 50\mathbf{1}_a$, where $\mathbf{0}_a$ and $\mathbf{1}_a$ are a -dimensional vectors with every component being zero and one, respectively. The skewed process is found in Chan and Tong (1986); The joint density function of multivariate skew normal distribution is given, for $\mathbf{x} \in \mathbb{R}^a$, by

$$f_{SN}(\mathbf{x}; \Sigma, \zeta) = 2v_a(\mathbf{x}; \Sigma)\Upsilon(\zeta^T \mathbf{x}),$$

where $v_a(\cdot; \Sigma)$ is the probability density function of the a -dimensional centered multivariate normal distribution with a correlation matrix Σ and $\Upsilon(\cdot)$ is the cumulative distribution function of the standard normal distribution. Note that the noncentered process has a nonzero mean $\sqrt{2/(\pi(1 + \zeta^T \Sigma \zeta))} \Sigma \zeta$ unless $\zeta = \mathbf{0}$, so we need subtract the mean. The more details of multivariate skew normal distribution can be found in Azzalini and Valle (1996), Azzalini and Capitanio (1999).

As for Process 4, we suppose

Process 4: $\{e_t\}$ follows the generalized autoregressive conditional heteroscedasticity model

$$e_{it} = h_{it}^{1/2} \epsilon_{it}, \quad i = 1, \dots, a, \quad \begin{pmatrix} h_{1t} \\ \vdots \\ h_{at} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 0.1\Phi \begin{pmatrix} e_{1t}^2 \\ \vdots \\ e_{at}^2 \end{pmatrix} + \begin{pmatrix} 0.1h_{1,t-1} \\ \vdots \\ 0.1h_{a,t-1} \end{pmatrix},$$

where ϵ_t is distributed as a zero mean multivariate normal distribution with covariance I_a in case 1 and Σ in case 2.

R package `mvtnorm` (Genz et al. 2021) is available to produce innovation processes for Processes 1 and 2. Process 4 can be produced by R package `cogarch` (Nakatani 2014). The skew normal distribution can be generated by R package `sn` (Azzalini 2022).

Features of Processes 1–4 as follows: Process 1 is the most standard setting. Fifth and higher moments of Process 2 do not exist. Processes 3 and 4 have a nonzero skewness and conditional heteroskedasticity, respectively.

We report the rejection probabilities of our proposed test T_n and the classical tests S_n in Figs. 1, 2 and 3 over 1000 simulations for the following situations: (i) $\tau = \mathbf{0}_a$; (ii) $\tau = (\tau_1, \dots, \tau_a)^T$, where $\tau_{3k-2} = -0.03, \tau_{3k-1} = 0$, and $\tau_{3k} = 0.03$ for $k \leq a/3$; and (iii) τ is distributed as a zero mean multivariate normal with covariance matrix

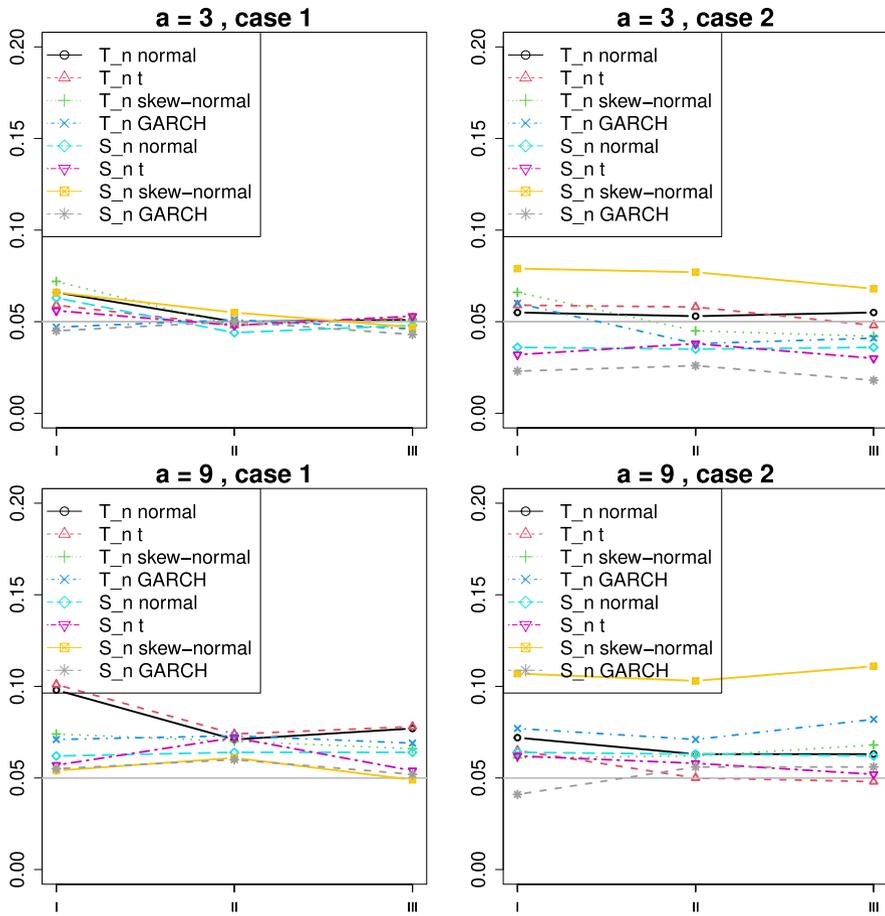


Fig. 1 Empirical size of tests for the existence of fixed and random effects based on T_n and S_n . The upper and lower plots correspond to $a = 3$ and $a = 9$, respectively. The left and right plots correspond to the cases 1 (independent groups) and 2 (correlated groups), respectively. The tick marks of the x-label (I), (II), and (III) correspond to the sample size $n_1 = \dots = n_a = 1000, n_{3k-1} = n_{3k-2} = 2000$ and $n_{3k} = 1000$ for $k \leq a/3$, and $n_1 = \dots = n_a = 2000$, respectively

Σ^τ . We let Σ^τ be a block diagonal matrix whose off-diagonal blocks are all 3×3 zero matrix and main-diagonal blocks are all the same 3×3 matrix $\tilde{\Sigma}^\tau = (\tilde{\Sigma}_{ij}^\tau)/5000$, where $\tilde{\Sigma}_{11}^\tau = 3, \tilde{\Sigma}_{22}^\tau = 2, \tilde{\Sigma}_{33}^\tau = \tilde{\Sigma}_{12}^\tau = \tilde{\Sigma}_{21}^\tau = 1, \tilde{\Sigma}_{23}^\tau = \tilde{\Sigma}_{32}^\tau = -0.5$, and $\tilde{\Sigma}_{13}^\tau = \tilde{\Sigma}_{31}^\tau = 0.008$. The significance level is set to be 0.05.

The situation (i) corresponds to both null hypotheses H_0 and H_1 defined in (5) and (8), respectively, and (ii) and (iii) correspond to the alternatives K_0 and K_1 , respectively. Note that fixed effects and random effects are chosen as tiny so that power become less than one to compare performances of tests against Processes 1–4. In the supplementary material, the consistency can be confirmed by results (see Tables 1–6 in Section B.1).

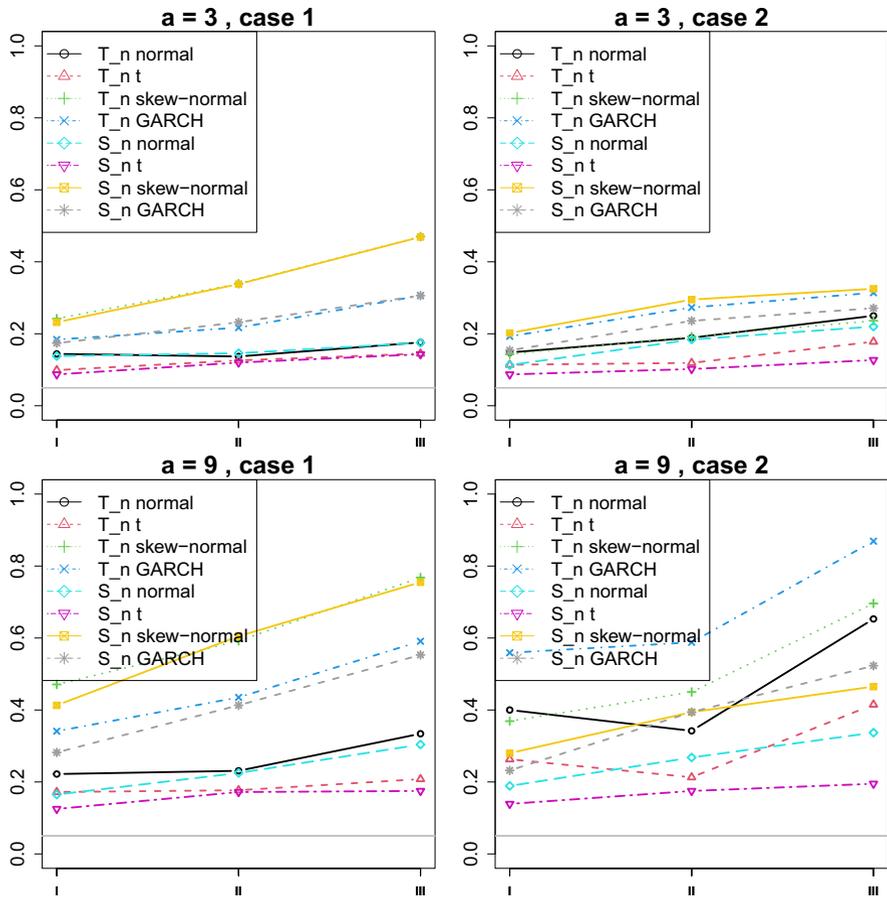


Fig. 2 Empirical power of tests for the existence of fixed effects based on T_n and S_n for fixed effects $\tau = (\tau_1, \dots, \tau_a)$, where $\tau_{3k-2} = -0.03$, $\tau_{3k-1} = 0$, and $\tau_{3k} = 0.03$ for $k \leq a/3$. The upper and lower plots correspond to $a = 3$ and $a = 9$, respectively. The left and right plots correspond to the cases 1 (independent groups) and 2 (correlated groups), respectively. The tick marks of the x-label (I), (II), and (III) correspond to the sample size $n_1 = \dots = n_a = 1000$, $n_{3k-1} = n_{3k-2} = 2000$ and $n_{3k} = 1000$ for $k \leq a/3$, and $n_1 = \dots = n_a = 2000$, respectively

Figure 1 shows the empirical size of the tests. Both tests work well for $a = 3$ and the case 1 (the top left plot) for all processes. Our proposed test based on T_n has good size for $a = 3$ and the case 2 (the top right plot). On the other hand, our test has small size distortion for $a = 9$ and the cases 1 and 2 (the lower plots). This distortion has occurred by the accumulation of estimating errors of the large matrix V^- (see Figures 1 and 2 in Section B.2 in the supplementary material). As expected, the classical test based on S_n has size distortion for both $a = 3, 9$ and the case 2 (the right plots) since the correlated groups are dealt with.

Figures 2 and 3 show the empirical power of the tests. Figures 2 and 3 for both $a = 3, 9$ and the case 1 display that empirical power of both tests are nearly equal for each model.

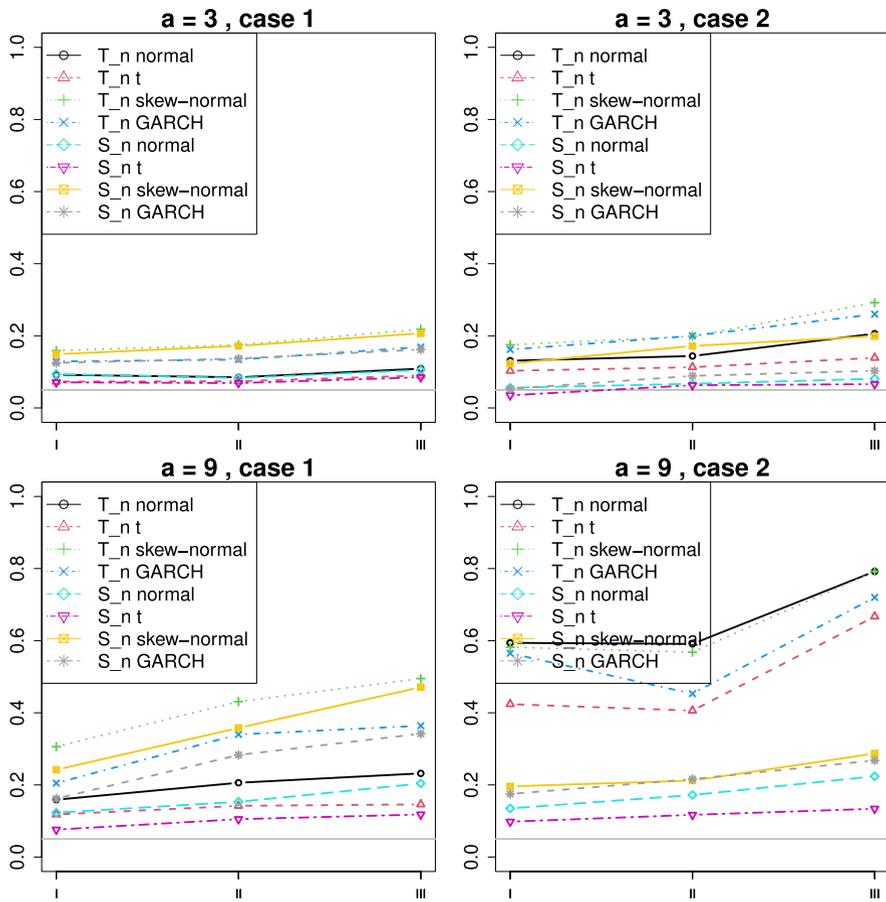


Fig. 3 Empirical power of tests for the existence of random effects based on T_n and S_n for random effects τ distributed as a zero mean multivariate normal with covariance matrix Σ^τ , where $\tilde{\Sigma}^\tau = (\tilde{\Sigma}_{ij}^\tau)/5000$ is a block diagonal matrix whose main-diagonal blocks are all the same 3×3 matrix such as $\tilde{\Sigma}_{11}^\tau = 3, \tilde{\Sigma}_{22}^\tau = 2, \tilde{\Sigma}_{33}^\tau = \tilde{\Sigma}_{12}^\tau = \tilde{\Sigma}_{21}^\tau = 1, \tilde{\Sigma}_{23}^\tau = \tilde{\Sigma}_{32}^\tau = -0.5, \tilde{\Sigma}_{13}^\tau = \tilde{\Sigma}_{31}^\tau = 0.008$, and $\tau_{3k} = 0.03$ for $k \leq a/3$. The upper and lower plots correspond to $a = 3$ and $a = 9$, respectively. The left and right plots correspond to the cases 1 (independent groups) and 2 (correlated groups), respectively. The tick marks of the x-label (I), (II), and (III) correspond to the sample size $n_1 = \dots = n_a = 1000, n_{3k-1} = n_{3k-2} = 2000$ and $n_{3k} = 1000$ for $k \leq a/3$, and $n_1 = \dots = n_a = 2000$, respectively

In most cases, size and power for the unbalanced design (II) $n_{3k-1} = n_{3k-2} = 2000$ and $n_{3k} = 1000$ for $k \leq a/3$ fall between results for the balanced design (I) $n_1 = \dots = n_a = 1000$ and (III) $n_1 = \dots = n_a = 2000$. There are the cases that the empirical power for unbalanced design (II) is worse than that for balanced design (I) regardless of the fact that the total sample size of (II) is larger than that of (I), e.g., the power of Processes 1 and 2 for $a = 9$ and case 2 in Fig. 3. Further, we implemented some additional experiments in the supplementary material and confirmed that the consistency of our test, i.e., the empirical power goes to one (see Tables 1–6 in Section

B.1). Overall, our proposed test works well to detect the existence of fixed or random effects. In summary, our test outperforms the classical test when groups are correlated and a is moderate.

6 Application to real data

Data analysis on stock prices often does not take random effects into account. However, for some portfolio of stocks, random effects cannot always be ignored. In fact, equity-focused investors take into account the sensitivities of currency, oil prices, market, etc. in determining their equity portfolios. In other words, equity-focused investors believe that the factors related to earnings and stock prices are linked. For example, stock prices of trading companies are linked to oil prices. It can be rephrased that equity-focused investors believe that random effects with respect to industries exist. In this empirical study, we pursue the question of whether random effects really exist for a portfolio that combines the automobile, telecom, and trading companies. We analyze the log-return in stock prices from January 4, 2016, to December 30, 2019. The companies we investigate are Itochu Corp., Mitsubishi Corp, Mitsui & Co., Ltd., and Marubeni Corp. from trading companies, Honda Motor Co. Ltd., Nissan Motor Co., Ltd., Suzuki Motor Corp., and Subaru Corp. from car companies, and KDDI Corp., Hikari Tsushin Inc. and NTT Data Corp. from telecom companies. The length of each time series is 978. These data can be downloaded from the website <https://www.investing.com>.

For this dataset, the number of groups a is three (trading, car, and telecom sectors), the dimension of time series p from each firm is one which corresponds to univariate ANOVA, the number of firms q is three for telecom sector and four for car and trading sectors, and the number of observations is $n_1 = n_2 = n_3 = 978$.

The plots of the log-returns are shown in Fig. 4. The dataset seems stationary and we cannot tell the difference between sectors. Table 1 gives that sample means and variances of the log-returns. The sample means of Suzuki and Hikari appear to be large compared to the sample means for other car and telecom companies, respectively. As for sample variances, the variances of Suzuki and Subaru are a little larger than other companies. Figure 5 shows the heatmap of sample correlations. These data have correlations between and within groups. This implies the classical F -test statistic should not be applied in this situation since it is designed for independent groups. Interesting observations for the data are as follows: Within-group correlations of telecom and trading companies are low and rather high, respectively. This may be because of a similar product mix for trading companies and a different product mix for telecom companies. Between-group correlations for car and trading companies are higher than those for telecom and car companies and those for telecom and trading companies. This may be ascribed to the facts that car and trading companies' stocks are cyclical, and by contrast, telecom companies' stocks are defensive.

We apply our test and the classical test as a comparison to this dataset (see Remark 4) and obtain the values 5.517 and 2.401 and the corresponding p-values 0.0634 and 0.301, respectively.

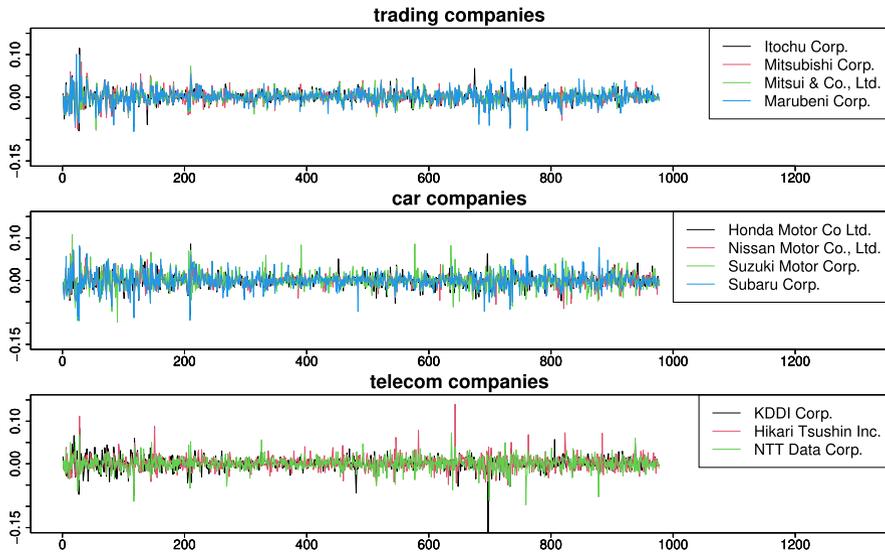


Fig. 4 Plots of log-return for stock prices

Table 1 Sample means and sample variances of log-returns

Sector	Company	Mean $\times 10^{-4}$	Variance $\times 10^{-4}$
Trading company	Itochu	5.89	2.15
	Mitsubishi	3.74	2.57
	Mitsui	3.15	2.16
	Marubeni	2.80	2.81
Car company	Honda	-1.90	2.75
	Nissan	-6.80	2.35
	Suzuki	2.51	3.91
	Subaru	-5.91	3.58
Telecom company	KDDI	0.71	2.45
	Hikari	12.61	2.86
	NTT	2.53	2.71

Therefore, the null hypothesis H_1 does not rejected under the significance level 0.05 for the existence of random effects for both tests. However, the p-values of our test is close to 0.05, and for the significance level 0.1, our tests rejects the hypothesis, but the classical test does not. From the observations that (i) our dataset has between-group correlations, and thus, the classical test is not appropriate, (ii) there exists the tendency of sample means: Car companies tend to have negative sample mean; in contrast, telecom and trading companies tend to positive sample mean, and (iii) the p-value of our test is close to 0.05, we conclude random effects should be taken into consideration for modeling log-return for stock prices. This result ensures equity-

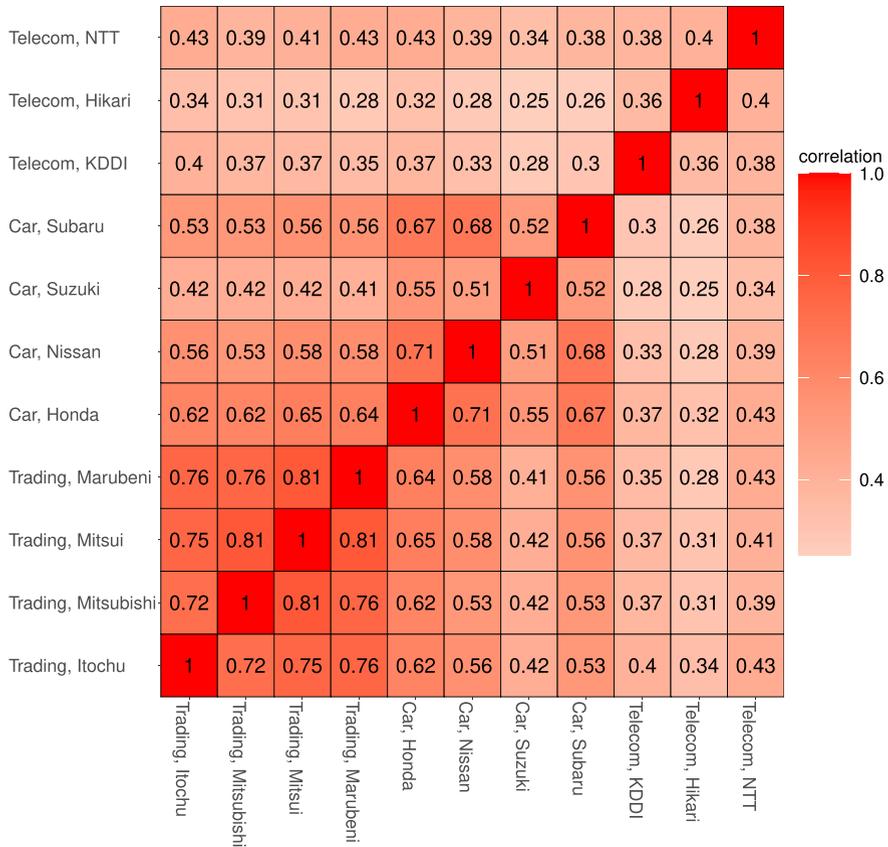


Fig. 5 Heatmap of sample correlations between companies

focused investors’ thoughts that different industries have different factors that affect corporate profits of companies and corporate profits influence stock prices such as profits of trading companies are linked to the price of crude oil.

Our result is convincing from portfolio theory. In that field, it is well known that portfolios of stocks have systematic risks related to the whole market and unsystematic risks related to sectors and companies. Many studies taking into account unsystematic risk have been conducted and emphasized the importance of unsystematic risks (see Aber 1976; Hsu and Jang 2008, and references therein). Industry effects corresponds to unsystematic risks in our case.

7 Additional thoughts/remarks

Nagahata and Taniguchi (2018) showed the asymptotic null distribution of S_n under the independence of groups. The following lines show that the independence of groups can be relaxed to uncorrelated groups. A simple algebra gives

$$\begin{aligned}
 S_n &= n \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^T \left(2\pi \tilde{f}_n(0) \right)^{-1} (\bar{y}_i - \bar{y}_{..}) \\
 &= n \begin{pmatrix} \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_1 \\ \vdots \\ \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_a \end{pmatrix}^T \{ (\mathbf{I}_a - \mathbf{J}_a/a) \otimes \mathbf{I}_p \} \begin{pmatrix} \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_1 \\ \vdots \\ \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_a \end{pmatrix}.
 \end{aligned}$$

Under Assumption 3.1 and the balanced design ($n_1 = \dots = n_a$), it holds that $\sqrt{n} \left(\left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_1, \dots, \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_a \right)$ converges in distribution to $N(\mathbf{0}, \mathbf{I}_{ap})$ as $n \rightarrow \infty$.

The idempotence of $(\mathbf{I}_a - \mathbf{J}_a/a) \otimes \mathbf{I}_p$, $\text{rank} \{ (\mathbf{I}_a - \mathbf{J}_a/a) \otimes \mathbf{I}_p \} = (a - 1)p$, the positive definiteness of the spectral density matrix, and the continuous mapping theorem yield that S_n converges in distribution to the Chi-square distribution with $(a - 1)p$ degrees of freedom under the independence of groups. The consistency of the test under the alternative and the power of the test under the local alternative can also be derived along the same line as our proof.

The independence or uncorrelatedness of groups is quite restrictive and impractical. In the case that groups are correlated, the asymptotic null limit distribution of S_n depends on the process since the nondiagonal elements of the asymptotic variance of the vector $\sqrt{n} \left(\left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_1, \dots, \left(2\pi \tilde{f}_n(0) \right)^{-1/2} \bar{e}_a \right)$ are not equal to zero. Thus, the p -value of the test based on S_n is not easy to compute. On the other hand, our proposed test statistic T_n is asymptotically distribution-free under the null. Based on the numerical studies, we realized the proposed test statistic has some size distortion under the null for large a . One direction to solve this problem is using S_n and applying a bootstrap method to obtain critical value. Homogeneity tests specialized for this type of models will be investigated in our future work.

8 Discussion

In this paper, the tests for the existence of fixed and random effects for one-way model with correlated groups were considered. The new test statistic was proposed and out tests are shown to be asymptotically size α under the null and consistent. The nontrivial power of tests is derived under the local alternative. In the numerical study, we confirmed our test performs well for several settings. In particular, our test is superior to the classical test when groups are correlated and a is moderate. The empirical study suggests the random effects are better to take into account in the analysis of stock prices .

Supplementary information

Supplementary material includes all proofs of theorems and additional simulation results.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11749-022-00828-9>.

Acknowledgements The authors are grateful to the editor and two referees for their instructive comments. The authors gratefully acknowledge Mr. Takeshi Tamaoka, the chief executive officer of Ananas Japan Co. Ltd, and Mr. Yuki Nakayasu, the chief executive officer of Minsetsu Inc., for their comments from practical points of view on the real data analysis. This work was supported by JSPS Grant-in-Aid for Research Activity Start-up under Grant Number JP21K20338 (Y.G.); JSPS Grant-in-Aid for Scientific Research (C) under Grant Number JP20K11719 (Y.L.); JSPS Grant-in-Aid for Scientific Research (S) under Grant Number JP18H05290 (M.T.); and the Research Institute for Science & Engineering of Waseda University (M.T.). This work was mainly carried out when the first author was affiliated with Waseda University.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Aber JW (1976) Industry effects and multivariate stock price behavior. *J Financ Quant Anal* 11(4):617–624
- Akharif A, Fihri M, Hallin M, Mellouk A (2020) Optimal pseudo-Gaussian and rank-based random coefficient detection in multiple regression. *Electron J Stat* 14(2):4207–4243
- Azzalini A (2022) The R package *sn*: the skew-normal and related distributions such as the skew-*t* and the SUN (version 2.0.2). Università degli Studi di Padova, Italia
- Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. *J R Stat Soc Ser B* 61(3):579–602
- Azzalini A, Valle AD (1996) The multivariate skew-normal distribution. *Biometrika* 83(4):715–726
- Bai J, Li K (2014) Theory and methods of panel data models with interactive effects. *Ann Stat* 42(1):142–170
- Baltagi BH, Li Q (1991) A transformation that will circumvent the problem of autocorrelation in an error-component model. *J Econom* 48(3):385–393
- Bernardes J, Mishra N, Tran F, Bahmer T, Best L, Blase J, Bordoni D, Franzenburg J, Geisen U, Josephs-Spaulling J, Köhler P, Künstner A, Rosati E, Aschenbrenner A, Bacher P, Baran N, Boysen T, Brandt B, Bruse N, Dörr J, Dräger A, Elke G, Ellinghaus D, Fischer J, Forster M, Franke A, Franzenburg S, Frey N, Friedrichs A, J. Fuß, Glück A, Hamm J, Hinrichsen F, Hoepfner M, Imm S, Junker R, Kaiser S, Kan Y, Knoll R, Lange C, Laue G, Lier C, Lindner M, Marinos G, Markewitz R, Nattermann J, Noth R, Pickkers P, Rabe K, Renz A, Röcken C, Rupp J, Schaffarzyk A, Scheffold A, Schulte-Schrepping J, Schunk D, Skowasch D, Ulas T, Wandinger K, Wittig M, Zimmermann J, Busch H, Hoyer B, Kaleta C, Heyckendorf J, Kox M, Rybniker J, Schreiber S, Schultze J, Rosenstiel P, DeCOI (2020) Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* 53(6):1296–1314
- Bloomfield P (1973) An exponential model for the spectrum of a scalar time series. *Biometrika* 60(2):217–226
- Brillinger DR (1981) *Time series: data analysis and theory*. Holden-Day, San Francisco
- Chan K, Tong H (1986) A note on certain integral equations associated with non-linear time series analysis. *Probab Theory Relat Fields* 73(1):153–158
- Chiu ST (1988) Weighted least squares estimators on the frequency domain for the parameters of a time series. *Ann Stat* 16(3):1315–1326
- Clarke BR (2008) *Linear models: the theory and application of analysis of variance*. Wiley

- Cook RD, Weisberg S (1983) Diagnostics for heteroscedasticity in regression. *Biometrika* 70(1):1–10
- Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of longitudinal data*. Oxford University Press, Oxford
- Ditzhaus M, Fried R, Pauly M (2021) QANOVA: quantile-based permutation methods for general factorial designs. *TEST* 30:960–979
- Ergemen YE, Velasco C (2017) Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *J Econom* 196(2):248–258
- Fang EX, Ning Y, Li R (2020) Test of significance for high-dimensional longitudinal data. *Ann Stat* 48(5):2622–2645
- Fihri M, Akharif A, Mellouk A, Hallin M (2020) Efficient pseudo-Gaussian and rank-based detection of random regression coefficients. *J Nonparam Stat* 32(2):367–402
- Galbraith JW, Zinde-Walsh V (1995) Transforming the error-components model for estimation with general ARMA disturbances. *J Econom* 66(1–2):349–355
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021) *mvtnorm: multivariate normal and t distributions*. R package version 1.1-3
- González JA, Lagos-Álvarez BM, Mateu J (2021) Two-way layout factorial experiments of spatial point pattern responses in mineral flotation. *TEST* 30:1046–1075
- Hallin M, Hlubinká D, Hudecova S (2021) Efficient fully distribution-free center-outward rank tests for multiple-output regression and MANOVA. *J Am Stat Assoc* 66:1–43
- Hoover DR, Rice JA, Wu CO, Yang LP (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85(4):809–822
- Hsu LT, Jang S (2008) The determinant of the hospitality industry's unsystematic risk: a comparison between hotel and restaurant firms. *Int J Hosp Tour Admin* 9(2):105–127
- Li Y (2011) Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* 98(2):355–370
- Liu X, Xu X (2016) Confidence distribution inferences in one-way random effects model. *TEST* 25(1):59–74
- Lucas, C., P. Wong, J. Klein, T.B. Castro, J. Silva, M. Sundaram, M.K. Ellingson, T. Mao, J.E. Oh, B. Israelow, T. Takahashi, M. Tokuyama, P. Lu, A. Venkataraman, A. Park, S. Mohanty, H. Wang, A.L. Wyllie, C.B.F. Vogels, R. Earnest, S. Lapidus, I.M. Ott, A.J. Moore, M.C. Muenker, J.B. Fournier, M. Campbell, C.D. Odio, A. Casanovas-Massana, Y.I. Team, R. Herbst, A.C. Shaw, R. Medzhitov, W.L. Schulz, N.D. Grubaugh, C.D. Cruz, S. Farhadian, A.I. Ko, S.B. Omer, and A. Iwasaki (2020) Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* 584(7821):463–469
- Nagahata H, Taniguchi M (2018) 4. Analysis of variance for multivariate time series. *Metron* 76:69–82
- Nakatani T (2014) *ccgarch: an R package for modelling multivariate GARCH models with conditional correlations*
- Rakocevic V (1997) On continuity of the Moore–Penrose and Drazin inverses. *Mater Vesn* 49(3–4):163–172
- Rao CR, Mitra SK (1971) *Generalized inverse of matrices and its applications*. Wiley, New York
- Rashid MM (1995) Robust analysis of two-way models with repeated measures on both factors. *TEST* 4(1):39–62
- Robinson PM (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59(5):1329–1363
- Searle SR, Casella G, McCulloch CE (1992) *Variance Components*. Wiley, New York
- Stewart G (1969) On the continuity of the generalized inverse. *SIAM J Appl Math* 17(1):33–45
- Tang CY, Leng C (2011) Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika* 98(4):1001–1006
- von Sachs R (2020) Nonparametric spectral analysis of multivariate time series. *Annu Rev Stat Appl* 7:361–386
- You J, Zhou X (2013) Efficient estimation in panel data partially additive linear model with serially correlated errors. *Stat Sin* 23:271–303
- Zeger SL, Liang KY, Self SG (1985) The analysis of binary longitudinal data with time independent covariates. *Biometrika* 72(1):31–38
- Zhong PS, Li R, Santo S (2019) Homogeneity tests of covariance matrices with high-dimensional longitudinal data. *Biometrika* 106(3):619–634

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.