Marginal homogeneity tests with panel data[*]

Federico A. Bugni
Department of Economics
Northwestern University
federico.bugni@northwestern.edu

Jackson Bunting
Department of Economics
Texas A&M University
jbunting@tamu.edu

Muyang Ren
Department of Economics
Duke University
muyang.ren@duke.edu

August 28, 2024

Abstract

A panel dataset satisfies marginal homogeneity if the time-specific marginal distributions are homogeneous or time-invariant. Marginal homogeneity is relevant in economic settings such as dynamic discrete games. In this paper, we propose several tests for the hypothesis of marginal homogeneity and investigate their properties. We consider an asymptotic framework in which the number of individuals n in the panel diverges, and the number of periods T is fixed. We implement our tests by comparing a studentized or non-studentized T-sample version of the Cramér-von Mises statistic with a suitable critical value. We propose three methods to construct the critical value: asymptotic approximations, the bootstrap, and time permutations. We show that the first two methods result in asymptotically exact hypothesis tests. The permutation test based on a non-studentized statistic is asymptotically exact when $T = 2$, but is asymptotically invalid when $T > 2$. In contrast, the permutation test based on a studentized statistic is always asymptotically exact. Finally, under a time-exchangeability assumption, the permutation test is exact in finite samples, both with and without studentization.

Keywords and phrases: Marginal homogeneity test, panel data, Cramér-von Mises statistic, Asymptotic approximation, Bootstrap, Permutation tests.

JEL classification: C12, C57, C63.

---

1   Introduction

This paper considers a hypothesis testing problem for panel data, $\{\{X_{i,t}\}_{t=1}^T\}_{i=1}^n$. We assume that the data are independent and identically distributed (i.i.d.) across units $i = 1, \ldots, n$, but allow for arbitrary dependence across time. For each period $t = 1, \ldots, T$, let F_t denote the common marginal cumulative distribution function (CDF) of $X_{i,t}$. We say that the data generating process satisfies *marginal homogeneity* if these marginal distributions are homogeneous or time-invariant, i.e.,

$$F_1 \ = \ F_2 \ = \ \ldots \ = \ F_T. \tag{1.1}$$

In this paper, we propose several tests for the hypothesis of marginal homogeneity and investigate their properties.

Marginal homogeneity is relevant in economic models such as dynamic discrete games. In this case, researchers often observe $X_{i,t} = (A_{i,t}, S_{i,t})$, representing action and state variables for units $i = 1, \ldots, n$ (individuals, firms, etc.) over $t = 1, \ldots, T$ periods, respectively. In this context, a standard approach is to assume that the conditional choice probabilities (i.e., $P(A_{i,t} = a \mid S_{i,t} = s)$) and state transition probabilities (i.e., $P(S_{i,t+1} = s' \mid A_{i,t} = a, S_{i,t} = s)$) are homogeneous, and posit a structural model for them. Under standard assumptions, these objects yield a homogeneous structural model for $P(X_{i,t+1} = x' \mid X_{i,t} = x) = f_\theta(x', x)$ with parameter θ. This posited structure forms the basis of inference on θ in dynamic discrete choice games. Notably, this inference does not invoke the marginal homogeneity hypothesis in (1.1). However, if this condition holds, it provides valuable efficiency gains in the estimation of θ. To see this, note that marginal homogeneity in this context implies that $X_{i,t}$ is in a steady state with a marginal CDF F, which yields the following structural equation:

$$dF(x') \ = \ \int_{x \in \mathcal{X}} f_\theta(x', x) \times dF(x) \quad \text{for all } x' \in \mathcal{X} \tag{1.2}$$

Imposing (1.2) in the estimation of θ can deliver substantial efficiency gains relative to the standard method that does not impose it. In this sense, the marginal homogeneity hypothesis in (1.1) is a source of efficiency gains in the structural estimation of dynamic discrete choice games.

Beyond dynamic discrete games, marginal homogeneity tests have been applied to financial data. For example, Ditzhaus and Gaigall (2022) (see also references therein) tests for possible dependence between two stock market indices. In terms of our notation, we can express their data as $\{\{X_{i,t}\}_{t=1}^2\}_{i=1}^n$, where $\{X_{i,1}\}_{i=1}^n$ represents the monthly returns of the first stock market (e.g., Nikkei 225 Stock Average) and $\{X_{i,2}\}_{i=1}^n$ represents the monthly returns of the second market (e.g., Dow Jones Industrial Average). Invoking classical models for stock prices, Ditzhaus and Gaigall (2022) posit monthly returns to be i.i.d. across $i = 1, \ldots, n$ (i.e., months), while the interconnectedness of global financial markets implies that there may be dependence across $t = 1, 2$ (i.e., stock markets).

In this context, the marginal homogeneity hypothesis in (1.1) implies that the various stock markets have equally distributed returns. Finally, we note that their paper focuses on pairs of stock markets (i.e., $T = 2$), whereas we consider applications with $T > 2$.

An inherent feature of the preceding examples is that the data are likely to exhibit dependence across time periods. In dynamic discrete games, actions and states depend on their past values. In fact, this is precisely what gives the discrete game its dynamic nature. In the application presented in Ditzhaus and Gaigall (2022), stock returns across the globe are likely to be interrelated. Beyond these two examples, dependence over time is common when dealing with panel data. For this reason, the classical goodness-of-fit testing literature focused on independent samples (e.g., Lehmann and Romano, 2022, Section 17.2.1), and its generalization to T-sample problems, does not apply.

This paper studies the T-sample testing problem with possibly dependent data. Namely, we implement our tests by comparing a studentized or non-studentized T-sample version of the Cramér-von Mises statistic with a suitable critical value. We consider three possible methods to construct the critical value: asymptotic approximations, the bootstrap, and time permutations. We show that the first two methods lead to asymptotically exact hypothesis tests, with or without studentization. Results for the permutation test are more nuanced: the permutation test based on a non-studentized statistic is asymptotically exact when $T = 2$, but is asymptotically invalid when $T > 2$. Once studentized, the permutation test is always asymptotically exact. Finally, under a time-exchangeability assumption, the permutation test is exact in finite samples, both with or without studentization. On the other hand, relative to the non-studentized case, the asymptotic analysis of the studentized statistics requires an additional assumption: the variance-covariate matrix used in the studentization must be non-singular, an assumption that can fail in practice. See the related discussion in Section 2.2 and our empirical application in Section 6.

For independent cross-sectional data, the marginal homogeneity hypothesis in (1.1) becomes the standard equality-of-distribution hypothesis for T-sample data, which has been thoroughly studied in the literature. In such case, Lehmann and Romano (2022, Theorem 17.2.1) shows that permutation tests of homogeneity are finite-sample exact. Chung and Romano (2013) explores the behavior of permutation tests with studentized test statistics. Relatedly, Bugni and Horowitz (2021) studies the application of permutation tests to functional cross-sectional data. Relatively speaking, the test for marginal homogeneity hypothesis in (1.1) with panel data (i.e., allowing for time dependence) has received less attention. Gaigall (2020) is among the first papers to test for marginal homogeneity in panel data with $T = 2$. In subsequent work, Ditzhaus and Gaigall (2022) broadened the analysis with $T = 2$ to paired functional data. These papers show that permutation tests are asymptotically valid with two periods of panel data. Neither of these papers considers the case with $T > 2$, which is common in economic applications. To our knowledge, our paper is the first one to consider testing for marginal homogeneity in panel data models, allowing for $T > 2$. In this respect, one remarkable finding of our paper reveals that the asymptotic validity of the

non-studentized permutation test breaks down when we consider $T > 2$.

A related testing problem in dynamic discrete choice games is concerned with evaluating the homogeneity of the state transition probabilities, i.e., $P(X_{i,t+1} = x' \mid X_{i,t} = x) = P(X_{i,t'+1} = x' \mid X_{i,t'} = x)$ for all $t, t' < T-1$. See Otsu et al. (2016) and Bugni et al. (2024) for recent contributions on this topic. The homogeneity of state transition probabilities and the marginal homogeneity in (1.1) are non-nested hypotheses, and so our contribution is complementary to but distinct from these references.

In other related work, Pauly et al. (2015) and Friedrich et al. (2017) investigate the validity of permutation tests to evaluate the presence of treatment effects in experiments under factorial designs. There are important differences between these papers and ours. The first key difference is in the class of data permutations used to implement the tests. Pauly et al. (2015) and Friedrich et al. (2017) generate their test by permuting observations in both units and time indices. In contrast, we generate our test by permuting the time index of our observations. We show that the classes of distributions under which the two types of permutation tests are finite-sample valid are non-nested; see Lemma A.2. Another difference is that Pauly et al. (2015) and Friedrich et al. (2017) focus on studentized statistics, while we consider both studentized and non-studentized statistics. In this respect, it is relevant to point out that analyzing studentized statistics requires additional assumptions compared to non-studentized ones; see discussion in Section 2.2. Finally, our Monte Carlo simulations suggest that our permutation test appears more powerful than theirs in finite samples. This is related to the fact that our permutation test only considers time index permutations, which provide a better contrast to detect departures from the marginal homogeneity hypothesis in (1.1).

The remainder of the paper is organized as follows. Section 2 introduces the hypothesis test problem in greater detail. Section 3 contains our main theoretical results. Section 4 discusses the power of the proposed inference methods. In Section 5, we evaluate the finite-sample performance of these tests via Monte Carlo simulations. Section 6 considers an empirical application based on Igami and Yang (2016). Section 7 concludes. The paper's appendix collects all of the proofs and several auxiliary results.

2 The hypothesis testing problem

As the introduction explains, this paper considers a hypothesis-testing problem for panel data with n units and T time periods. Inspired by the typical application in economics, we consider an asymptotic framework in which n grows and T remains fixed. We denote the data by $\mathbf{X}_n = \{\{X_{i,t}\}_{t=1}^T\}_{i=1}^n$. As explained earlier, we allow the data to be arbitrarily dependent across time $t = 1, \ldots, T$, and assume i.i.d. across units $i = 1, \ldots, n$. We formalize this assumption next.

Assumption 1. For all $t = 1, \ldots, T$, $\{X_{i,t}\}_{i=1}^n$ are i.i.d. with marginal CDF F_t.

Our goal is to test whether the marginal homogeneity hypothesis in (1.1) holds in the data, i.e.,

$$H_0 \; : \; (1.1) \text{ holds} \quad \text{vs.} \quad H_1 \; : \; (1.1) \text{ does not hold.} \tag{2.1}$$

We propose implementing this hypothesis test by rejecting H_0 in (2.1) whenever a test statistic exceeds a suitable critical value. That is, for any significant level of $\alpha \in (0, 1)$, we propose

$$\phi_n\left(\alpha\right) \; = \; 1\left\{S_n > c_n(1 - \alpha)\right\}, \tag{2.2}$$

where $\phi_n\left(\alpha\right)$ indicates the test function, S_n denotes the test statistic, and $c_n(1 - \alpha)$ indicates the critical value. In the remainder of this section, we describe the test statistic (Section 2.1) and establish its asymptotic distribution under the null hypothesis of marginal homogeneity (Section 2.2). With these results in place, Section 3 provides three inference methods, each based on a different type of critical values.

2.1 Test statistics

We propose implementing our test using the Cramér-von Mises (CvM) statistic, given by the sample-weighted sum of squared differences of the empirical CDFs for all consecutive periods. For simplicity, we evaluate these differences on a finite number of user-defined points on the real line $\mathcal{U}_K = \{u_1, u_2, \ldots, u_K\}$ with $u_0 := -\infty < u_1 < u_2 < \cdots < u_K$. For reasons that will be explained soon, we refer to this as the *non-studentized* CvM statistic, given by

$$S_n \; = \; n \sum_{k=1}^K \sum_{t=1}^{T-1} [\hat{F}_t(u_k) - \hat{F}_{t+1}(u_k)]^2 \hat{P}(u_k), \tag{2.3}$$

where $\hat{F}_t$ is the empirical CDF in period $t = 1, \ldots, T$, and $\hat{P}(u_k)$ is the empirical analog of the aggregate probability in the interval $(u_{k-1}, u_k]$, i.e., $k = 1, \ldots K$,

$$\hat{P}(u_k) \; = \; \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n 1(u_{k-1} < X_{i,t} \le u_k). \tag{2.4}$$

It is easy to see that (2.3) can be reexpressed as follows

$$S_n \; = \; \hat{Z}' \hat{Z},$$

where $\hat{Z} \in \mathbb{R}^{(T-1)K}$ and, for all $(t,k) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$,

$$\hat{Z}_{(t-1)K+k} = \sqrt{n\hat{P}(u_k)}[\hat{F}_t(u_k) - \hat{F}_{t+1}(u_k)]. \tag{2.5}$$

Under H_0 in (2.1), our formal arguments (see the proof of Theorem 2.1) reveal that $\hat{Z}$ is asymptotically distributed according to $N(\mathbf{0}, \Sigma_Z)$, where for each $(t,k), (\tilde{t}, \tilde{k}) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$,

$$\begin{aligned}
\Sigma_Z[(t-1)K + k, (\tilde{t}-1)K + \tilde{k}] &= \\
\sqrt{P(u_k)P(u_{\tilde{k}})} &\times cov[1(X_{i,t} \le u_k) - 1(X_{i,t+1} \le u_k), 1(X_{i,\tilde{t}} \le u_{\tilde{k}}) - 1(X_{i,\tilde{t}+1} \le u_{\tilde{k}})],
\end{aligned} \tag{2.6}$$

and $P(u_k)$ is the aggregate probability in the interval $(u_{k-1}, u_k]$,

$$P(u_k) = \frac{1}{T} \sum_{t=1}^{T} [F_t(u_k) - F_t(u_{k-1})]. \tag{2.7}$$

As a corollary, S_n has a generalized chi-squared asymptotic distribution under H_0 in (2.1), with weights determined by the eigenvalues of Σ_Z. The dependence of the limiting distribution on Σ_Z explains why we refer to S_n as the *non-studentized* CvM statistic.

If Σ_Z is a non-singular matrix, it is natural to also consider a studentized version of the CvM statistic. To this end, we consider the *studentized* CvM statistic, given by

$$\bar{S}_n = \hat{Z}' \hat{\Sigma}_Z^- \hat{Z}, \tag{2.8}$$

where $\hat{\Sigma}_Z^-$ is the generalized inverse of the empirical analog of Σ_Z, denoted by $\hat{\Sigma}_Z$. For each $(t,k), (\tilde{t}, \tilde{k}) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$, we define $\hat{\Sigma}_Z$ as follows:

$$\begin{aligned}
&\hat{\Sigma}_Z[(t-1)K + k, (\tilde{t}-1)K + \tilde{k}] \equiv \\
&\sqrt{\hat{P}(u_k)\hat{P}(u_{\tilde{k}})} \times \frac{1}{n} \sum_{i=1}^{n} \left[\begin{matrix} \left(1(X_{i,t} \le u_k) - \hat{F}_t(u_k) - 1(X_{i,t+1} \le u_k) + \hat{F}_{t+1}(u_k)\right) \times \\ \left(1(X_{i,\tilde{t}} \le u_{\tilde{k}}) - \hat{F}_{\tilde{t}}(u_{\tilde{k}}) - 1(X_{i,\tilde{t}+1} \le u_{\tilde{k}}) + \hat{F}_{\tilde{t}+1}(u_{\tilde{k}})\right) \end{matrix} \right]
\end{aligned} \tag{2.9}$$

for each $(t,k), (\tilde{t}, \tilde{k}) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$. If H_0 in (2.1) holds and Σ_Z is a non-singular, $\bar{S}_n$ has a chi-squared asymptotic distribution with $(T-1)K$ degrees of freedom. The lack of dependence of this limiting distribution on Σ_Z justifies referring to $\bar{S}_n$ as the *studentized* CvM statistic.

Remark 2.1 (On the choice of test-statistics). The test statistics in (2.3) and (2.8) are CvM-type statistics evaluated over a finite set of points $\mathcal{U}_K$. In principle, our analysis extends to other types of test statistics evaluated over these points, such as the Kolmogorov-Smirnov statistic. It is also worth reiterating that $\mathcal{U}_K$ can be arbitrarily chosen by the researcher. In this respect, the most important simplifying aspect in (2.3) and (2.8) is that we use a finite number of points rather than an infinite number of points, such as a continuum.

There are several reasons to prefer a finite set $\mathcal{U}_K$ over a continuum. First and foremost, it leads to simpler asymptotic analysis regarding the studentization of test statistics. Second, for applications in which the data are discrete and with finite support S_X, one can set $\mathcal{U}_K = S_X \backslash \{\max S_X\}$ without any loss of information (note that equality of marginal distributions for all points in $S_X \backslash \{\max S_X\}$ is equivalent to H_0). Many empirical applications, including the one in Section 6, feature discrete data with finite support. Third, the extension to the continuum case could be implemented along the lines of the Khmaladze (2016) transformation. See also the related work by Chung and Olivares (2021). Having said this, this extension requires new and considerably more complicated arguments, which we consider out of the scope of our paper.

∎

2.2 Asymptotic distribution under the null hypothesis

In this section, we derive the asymptotic distribution of the non-studentized CvM statistic S_n in (2.3) and the studentized version $\bar{S}_n$ in (2.8). Our characterization of the asymptotic distribution of the studentized CvM statistic relies on the following assumption.

Assumption 2. Σ_Z is positive definite.

We note that Assumption 2 is required for our asymptotic analysis of the studentized CvM statistic and is not necessary in the case of the non-studentized version. For a suitable choice of $\mathcal{U}_K$, Assumption 2 is widely applicable to many data-generating processes. We now describe scenarios in which this assumption does not hold. First, note that Assumption 2 would fail if $\mathcal{U}_K$ includes any point that is "irrelevant" with respect to the support of $\mathbf{X}_n$, i.e., $\hat{P}(u_k) = 0$ for some $k = 1, \ldots K$. An example of this occurs for any $u_k \in \mathcal{U}_K$ that lies below the support of $\mathbf{X}_n$. In this case, one can always restore the validity of Assumption 2 by removing all of these irrelevant points. Second, one should never include $u_k \in \mathcal{U}_K$ that equals or exceeds the support of $\mathbf{X}_n$. Doing this would result in $1(X_{i,t} \leq u_k) = 1(X_{i,t+1} \leq u_k) = 1$, leading to $1(X_{i,t} \leq u_k) - 1(X_{i,t+1} \leq u_k) = 0$, rendering Σ_Z singular. Finally, Σ_Z would be singular if there is no full communication between some states. For example, consider a Markov chain for $X_{i,t} \in \{1, 2, 3, 4\}$ with $P(X_{i,t+1} = x' | X_{i,t} = x) = \Pi[x, x']$ for all $x, x' \in \{1, 2, 3, 4\}$, where for any $\rho_1, \rho_2, \rho_3, \rho_4 \in (0, 1)$,

$$
\Pi = \begin{bmatrix} \rho_1 & 1 - \rho_1 & 0 & 0 \\ 1 - \rho_2 & \rho_2 & 0 & 0 \\ 0 & 0 & \rho_3 & 1 - \rho_3 \\ 0 & 0 & 1 - \rho_4 & \rho_4 \end{bmatrix}
$$

This transition matrix implies no communication between the first and last two states. Then, $1(X_{i,t} \leq u_k) - 1(X_{i,t+1} \leq u_k) = 0$ for any $u_k \in [2, 3)$, producing a singular Σ_Z.

The next result establishes the asymptotic distribution of the non-studentized and studentized CvM statistics under the marginal homogeneity hypothesis in (2.1).

Theorem 2.1. Let Assumption 1 hold.

(a) Under H_0 in (2.1),

$$S_n \overset{d}{\to} S \equiv \sum_{j=1}^{(T-1)K} \lambda_j \zeta_j^2, \tag{2.10}$$

where $\{\lambda_j\}_{j=1}^{(T-1)K}$ are the eigenvalues of Σ_Z in (2.6) and $\{\zeta_j\}_{j=1}^{(T-1)K}$ are i.i.d. $N(0,1)$.

(b) Under Assumption 2 and H_0 in (2.1),

$$\bar{S}_n \overset{d}{\to} \chi^2_{(T-1)K}, \tag{2.11}$$

where $\chi^2_{(T-1)K}$ denotes the chi-squared distribution with $(T-1)K$ degrees of freedom.

Theorem 2.1 shows that under the marginal homogeneity hypothesis, the non-studentized CvM statistic in (2.3) converges to a generalized chi-square distribution with the weights determined by the eigenvalues of Σ_Z. Since Σ_Z is not necessarily positive definite, some eigenvalues may be zero, leading to reduced degrees of freedom. When Σ_Z is positive definite, then the limiting distribution of studentized CvM statistic in (2.8) is chi-square distributed. These results are the basis of the critical values proposed in the following section.

3 Critical values and validity of inference

This section describes three critical values for the CvM test statistics proposed in Section 2.1. Each critical value gives rise to a different hypothesis test according to equation (2.2). We formally study the validity of each one of these methods.

3.1 Asymptotic approximation

In this section, we propose a hypothesis test for (2.1) by approximating the quantiles of the asymptotic distribution in Theorem 2.1. To this end, we now introduce some notations. For any $\ell \in \mathbb{R}^{(T-1)K}$, let $S(\ell) \geq 0$ denote a random variable with the generalized chi-square distribution of weights equal to ℓ, unit vector of degrees of freedom, zero vector of non-centrality parameters, and no constant or normal terms. Also, for any $(x, \ell) \in \mathbb{R} \times \mathbb{R}^{(T-1)K}$, let $G(x, \ell)$ denote the CDF of $S(\ell)$ evaluated at $x \in \mathbb{R}$. This function can be numerically computed with arbitrary accuracy by simulating its empirical distribution.

For the non-studentized CvM statistic, we propose

$$c_n^A(1-\alpha) \;=\; \inf\left\{x \in \mathbb{R} : G(x, \hat{\lambda}_n) \geq 1 - \alpha\right\}, \tag{3.1}$$

where $\hat{\lambda}_n$ denotes the eigenvalues of $\hat{\Sigma}_Z$, and the following hypothesis test:

$$\phi_n^A(\alpha) \;=\; 1\left\{S_n > c_n^A(1-\alpha)\right\}. \tag{3.2}$$

For the studentized CvM statistic, we propose the following hypothesis test:

$$\bar{\phi}_n^A(\alpha) \;=\; 1\left\{\bar{S}_n > \bar{c}_n^A(1-\alpha)\right\}, \tag{3.3}$$

where $\bar{c}_n^A(1-\alpha)$ equals the $(1-\alpha)$-quantile of the (standard) chi-squared distribution with $(T-1)K$ degrees of freedom.

The next result shows that the hypothesis tests in (3.2) and (3.3) are asymptotically valid.

Theorem 3.1. Let Assumption 1 and H_0 in (2.1) hold, and let $\alpha \in (0,1)$.

    (a) $\lim_{n\to\infty} E_P[\phi_n^A(\alpha)] \leq \alpha$. Furthermore, the inequality becomes an equality if $\Sigma_Z \neq 0_{(T-1)K \times (T-1)K}$.

    (b) Under Assumption 2, $\lim_{n\to\infty} E_P[\bar{\phi}_n^A(\alpha)] = \alpha$.

3.2 Bootstrap

This section proposes a hypothesis test for (2.1) via the bootstrap. To this end, we repeatedly resample the data with replacement across units $i = 1, \ldots, n$ to construct a bootstrap sample, denoted by $\{X_i^*\}_{i=1}^n$. For each bootstrap sample, the bootstrap analog of the non-studentized and studentized CvM statistics are given by

$$S_n^* \;=\; (\hat{Z}^*)'(\hat{Z}^*) \quad \text{and} \quad \bar{S}_n^* \;=\; (\hat{Z}^*)'\hat{\Sigma}_Z^-(\hat{Z}^*). \tag{3.4}$$

where, for all $(t,k) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$,

$$\hat{Z}_{(t-1)K+k}^* \;\equiv\; \sqrt{\hat{P}(u_k)}\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\left[1(X_{i,t}^* \leq u_k) - \hat{F}_t(u_k)\right] - \left[1(X_{i,t+1}^* \leq u_k) - \hat{F}_{t+1}(u_k)\right]\right). \tag{3.5}$$

Remark 3.1. One could also define $\bar{S}_n^*$ in (3.4) with $\hat{\Sigma}_Z$ replaced by its bootstrap analog. Our main text omits this option for brevity, but we include it in our Monte Carlo simulations. ∎

By repeating the bootstrap sampling sufficiently many times, we can approximate the conditional distributions $P\left(S_n^* \leq x | \mathbf{X}_n\right)$ and $P\left(\bar{S}_n^* \leq x | \mathbf{X}_n\right)$ with arbitrary accuracy.

For the non-studentized CvM statistic, we propose

$$c_n^B (1 - \alpha) \; = \; \inf \{x \in \mathbb{R} : P\left(S_n^* \le x | \mathbf{X}_n\right) \ge 1 - \alpha\}, \tag{3.6}$$

and the following hypothesis test:

$$\phi_n^B (\alpha) \; = \; 1\left\{S_n > c_n^B (1 - \alpha)\right\}. \tag{3.7}$$

For the studentized CvM statistic, we propose

$$\bar{c}_n^B (1 - \alpha) \; = \; \inf \{x \in \mathbb{R} : P\left(\bar{S}_n^* \le x | \mathbf{X}_n\right) \ge 1 - \alpha\}, \tag{3.8}$$

and the following hypothesis test:

$$\bar{\phi}_n^B (\alpha) \; = \; 1\left\{\bar{S}_n > \bar{c}_n^B (1 - \alpha)\right\}. \tag{3.9}$$

The next result shows that the hypothesis tests in (3.6) and (3.8) are asymptotically valid.

Theorem 3.2. Let Assumption 1 and H_0 in (2.1) hold, and let $\alpha \in (0, 1)$.

(a) $\lim_{n\to\infty} E_P[\phi_n^B(\alpha)] \le \alpha$. Furthermore, the inequality becomes an equality if $\Sigma_Z \ne 0_{(T-1)K \times (T-1)K}$.

(b) Under Assumption 2, $\lim_{n\to\infty} E_P[\bar{\phi}_n^B(\alpha)] = \alpha$.

3.3 Permutations

In this section, we propose a hypothesis test for (2.1) by random permutations of the data. Our permutations are motivated by the marginal homogeneity hypothesis in (1.1), and they consist of randomly permuting the time index $t = 1, \ldots, T$ for each unit $i = 1, \ldots, n$.

These tests require the following notation. Let $\mathcal{M}$ denote the set of all permutations of the indices $\{1, \ldots, T\}$, and $\mathcal{M}^n$ is defined as the set of all possible permutations of the time indices over n observations. A typical element of $\mathcal{M}^n$ is given by $\pi^n \equiv \{\{\pi_i(1), \ldots, \pi_i(T)\}\}_{i=1}^n$, where $\pi_i(t)$ denotes the permuted time index of the observation $X_{i,t}$, and $\{\pi_i(1), \ldots, \pi_i(T)\}$ is an arbitrary time permutation that belongs to the set $\mathcal{M}$. In other words, the permuted version of the data $\mathbf{X}_n = \{\{X_{i,t}\}_{t=1}^T\}_{i=1}^n$ can be written as $\mathbf{X}_n^\pi \equiv \{\{X_{i,\pi_i(t)}\}_{t=1}^T\}_{i=1}^n$.

For each permutation $\pi^n \in \mathcal{M}^n$, the permutation analog of the non-studentized and studentized CvM statistics are given by

$$S_n^\pi = (\hat{Z}^\pi)'(\hat{Z}^\pi) \quad \text{and} \quad \bar{S}_n^\pi = (\hat{Z}^\pi)'\hat{\Sigma}_{Z^\pi}^-(\hat{Z}^\pi), \tag{3.10}$$

where, for all $(t, k) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$,

$$\hat{Z}^\pi_{(t-1)K+k} \equiv \sqrt{n\hat{P}^\pi(x)}\big(\hat{F}^\pi_t(u_k) - \hat{F}^\pi_{t+1}(u_k)\big) : (t, k) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}, \qquad (3.11)$$

and, for all $(t, k), (\tilde{t}, \tilde{k}) \in \{1, \ldots, T-1\} \times \{1, \ldots, K\}$,

$$\hat{\Sigma}_{Z^\pi}[(t-1)K + k, (\tilde{t}-1)K + \tilde{k}] \equiv$$
$$\sqrt{\hat{P}^\pi(u_k)\hat{P}^\pi(u_{\tilde{k}})} \times \frac{1}{n} \sum_{i=1}^{n} \left[\begin{array}{c} \big(1(X_{i,\pi_i(t)} \leq u_k) - \hat{F}^\pi_t(u_k) - 1(X_{i,\pi_i(t+1)} \leq u_k) + \hat{F}^\pi_{t+1}(u_k)\big) \times \\ \big(1(X_{i,\pi_i(\tilde{t})} \leq u_{\tilde{k}}) - \hat{F}^\pi_{\tilde{t}}(u_{\tilde{k}}) - 1(X_{i,\pi_i(\tilde{t}+1)} \leq u_{\tilde{k}}) + \hat{F}^\pi_{\tilde{t}+1}(u_{\tilde{k}})\big) \end{array} \right],$$
$$(3.12)$$

where $\hat{P}^\pi(u_k) \equiv \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} 1\big(u_{k-1} < X_{i,\pi_i(t)} \leq u_k\big)$ and $\hat{F}^\pi_t(u_k) \equiv \frac{1}{n} \sum_{i=1}^{n} 1(X_{i,\pi_i(t)} \leq u_k)$.

These permutation test statistics can be used to construct permutation-based tests along the lines of Lehmann and Romano (2022, Section 17.2.1). For the non-studentized CvM statistic, we propose a critical value $c^\pi_n(1-\alpha)$, which is the $(1-\alpha)$-quantile over all possible permutations of the non-studentized CvM statistics (denoted by $\{S^\pi_n : \pi^n \in \mathcal{M}^n\}$). The corresponding hypothesis test is defined as:

$$\phi^\pi_n(\alpha) = 1\{S_n > c^\pi_n(1-\alpha)\}. \qquad (3.13)$$

For the studentized CvM statistic, we propose the analogous object but for the studentized CvM statistic. That is, $\bar{c}^\pi_n(1-\alpha)$ equals to the $(1-\alpha)$-quantile over all the possible permutations for the studentized CvM statistics (given by $\{\bar{S}^\pi_n : \pi^n \in \mathcal{M}^n\}$). The corresponding hypothesis test is given by:

$$\bar{\phi}^\pi_n(\alpha) = 1\{\bar{S}_n > \bar{c}^\pi_n(1-\alpha)\}. \qquad (3.14)$$

Remark 3.2. The tests in (3.13) and (3.14) are the non-random versions of the standard permutation tests described in Lehmann and Romano (2022, Section 17.2.1). The key difference between the non-random and the random versions lies in the handling of ties between the test statistic and the critical value: the non-random version does not reject, while the random version rejects with a specific probability. While being more conservative, the non-random version is preferred because it involves a simpler decision rule similar to the one used in previous tests. ∎

The next result studies the asymptotic validity of the hypothesis tests in (3.13) and (3.14).

Theorem 3.3. Let Assumption 1 and H_0 in (2.1) hold, and let $\alpha \in (0, 1)$.

(a) For $T = 2$, $\lim_{n\to\infty} E_P[\phi^\pi_n(\alpha)] \leq \alpha$. Furthermore, the inequality becomes an equality if $\Sigma_Z \neq 0_{(T-1)K \times (T-1)K}$.

(b) When $T > 2$, the non-studentized permutation test in (3.13) is invalid. That is, it is possible to have $\liminf_{n\to\infty} E_P[\phi^\pi_n(\alpha)] > \alpha$ for some distribution P.

(c) Under Assumption 2, $\lim_{n\to\infty} E_P[\bar{\phi}_n^\pi(\alpha)] = \alpha$.

We first describe the result for the non-studentized test in (3.13). Part (a) shows that the non-studentized test is asymptotically valid for $T = 2$. This result aligns with the analysis in Gaigall (2020) and Ditzhaus and Gaigall (2022). Part (b) reveals that the previous result fails for $T > 2$. This finding is apparently new in the literature and empirically relevant, as many economic applications involve panel data with more than two periods. To gain intuition about (a) and (b), it is useful to compare the variance-covariance of the ith observation, i.e., $\{X_{i,t}\}_{t=1}^T$, before and after permutations. If the variance-covariance remains the same, then the permutation test can be shown to be asymptotically valid. See the proof of Theorem 3.3 for a formal justification. When $T = 2$, the ith observation is $(X_{i,1}, X_{i,2})$ before the permutation and a mixture of $(X_{i,1}, X_{i,2})$ and $(X_{i,2}, X_{i,1})$ after the permutation. Under H_0 in (2.1), these two random vectors share the same variance-covariance matrix. When $T > 2$, this equivalence breaks down. For example, when $T = 3$, the ith observation is $(X_{i,1}, X_{i,2}, X_{i,3})$ before the permutation and a mixture of $(X_{i,1}, X_{i,2}, X_{i,3})$, $(X_{i,1}, X_{i,3}, X_{i,2})$, $(X_{i,2}, X_{i,1}, X_{i,3})$, $(X_{i,2}, X_{i,3}, X_{i,1})$, $(X_{i,3}, X_{i,1}, X_{i,2})$, and $(X_{i,3}, X_{i,2}, X_{i,1})$ after the permutation. The variance-covariance matrix of the former and the latter differ even under H_0 in (2.1).

Finally, part (c) of Theorem 3.3 shows that the studentized test in (3.14) is asymptotically valid for any $T \geq 2$. This result is in line with several studies in the literature that prove the asymptotic validity of permutation tests for suitably normalized test statistics, e.g., Janssen (1997); Chung and Romano (2013); DiCiccio and Romano (2017); Chung and Olivares (2021). This result is also related to the earlier discussion regarding variance-covariance matrices of the ith observation before and after permutations. Once the data is studentized, the variance-covariance matrix of the ith observation is asymptotically invariant to permutations. By the previous paragraph's logic, the permutation test is asymptotically valid for the studentized test statistic.

An important advantage of the permutation tests over the ones described in previous subsections lies in their finite-sample validity under an important class of distributions that satisfy H_0 in (2.1). To explain this clearly, let Ω_{TE} denote the set of distributions that satisfy time exchangeability, i.e., for any $i = 1, \ldots, n$, $\{X_{i,t}\}_{t=1}^T$ has the same distribution as $\{X_{i,\pi(t)}\}_{t=1}^T$ for each $\pi \in \mathcal{M}$. The next result describes the finite-sample validity of our test under suitable conditions.

Theorem 3.4. Let Assumption 1 hold. Then,

(a) $P \in \Omega_{\mathrm{TE}}$ implies that P satisfies H_0 in (2.1). However, the converse does not hold.

(b) For each $P \in \Omega_{\mathrm{TE}}$, our permutation tests are finite-sample valid, i.e.,

$$E_P[\phi_n^\pi(\alpha)] \leq \alpha \quad \text{and} \quad E_P[\bar{\phi}_n^\pi(\alpha)] \leq \alpha. \tag{3.15}$$

12

Theorem 3.4 implies that our permutation tests are finite-sample valid under suitable conditions. Part (a) says that a time-exchangeable distribution satisfies the null hypothesis H_0 in (2.1) under our maintained Assumption 1. Under such distribution, the permutation tests provide finite-sample size control. We stress that size control is not exact (i.e., the inequalities in (3.15) might be strict) only because we are using a non-randomized permutation test; see Remark 3.2. That is, if we replaced our permutation test with its random version, these would enjoy exact size control (i.e., both inequalities in (3.15) would hold with equality).

The combination of Theorems 3.3 and 3.4 justifies the use of studentized permutation test in (3.14) (and also the non-studentized one in (3.13) when $T = 2$). Theorem 3.3 indicates that this test is asymptotically exact, and Theorem 3.4 shows that it is finite-sample valid for an important class of distributions in Ω_{TE}. As already mentioned, the finite-sample validity makes the permutation test an exceptionally attractive inference method compared to those discussed in previous subsections.

Remark 3.3. It is worth noting that our class of time permutations differs from the class of "all permutations" that uniformly permute both time periods and units. The latter has been used in the previous literature, such as Friedrich et al. (2017). Naturally, our time permutations form a strict subset of the class of "all permutations". On the flip side, under our maintained Assumption 1, the class of distributions that satisfy exchangeability over "all permutations" is *more restrictive* than the set of distributions that satisfy time exchangeability only; see Lemma A.2. In other words, if we used a permutation test based on "all permutations", we could only demonstrate Theorem 3.4 for a substantially smaller class of distributions than Ω_{TE}. ∎

4 Power analysis

This section briefly describes the power properties of the various hypothesis-testing procedures considered in this paper. Given the results in Section 3, we restrict attention to the hypothesis tests that are asymptotically valid under Assumptions 1–2:

- The asymptotic approximation-based test, both non-studentized and studentized.

- The bootstrap-based test, both non-studentized and studentized.

- The studentized permutation-based test.

- The non-studentized permutation-based test for $T = 2$.

As explained in Section 2.1, our CvM test statistic is defined to detect differences in marginal CDFs at any point in $\mathcal{U}_K$. We thus focus our power analysis on the following subset of H_1:

$$\bar{H}_1 : F_t(u) \neq F_r(u) \quad \text{for some} \ \ t, r = 1, \ldots, T \ \ \text{and} \ \ u \in \mathcal{U}_K. \tag{4.1}$$

For all fixed hypotheses in $\bar{H}_1$, it is not hard to see that the CvM test statistic in (2.3) and (2.8) diverges. At the same time, one can establish that the critical values described throughout this paper remain bounded in probability. For this reason, all asymptotically valid tests are consistent against any fixed hypothesis in $\bar{H}_1$.

We now compare the local power properties of these tests. We consider local alternative hypotheses under $\bar{H}_1$, which are sequences of DGPs whose marginal CDFs $\{F_{n,t} : t = 1, \ldots, T\}$ satisfy $\{\sqrt{n}(F_{n,t}(u_k) - F_{n,r}(u_k)) : t, r = 1, \ldots, T, u \in \mathcal{U}_K\} \to c \in \mathbb{R}^{T^2 K}$ with $c'c > 0$. Under these sequences of distributions, we can repeat the arguments used to prove Theorem 2.1 to establish the asymptotic distribution of the CvM test statistic in (2.3) and (2.8). It is not hard to see that these become the non-central versions of the asymptotic distributions in (2.10) and (2.11) under H_0. Moreover, under these local alternatives, the critical values considered throughout this paper can be shown not to change their asymptotic behavior. As a corollary, the asymptotically valid tests based on the non-studentized statistics share the same local power properties, and the same holds for the asymptotically valid tests based on the studentized statistics.

5   Monte Carlo simulations

This section investigates the finite-sample performance of the various tests for marginal homogeneity tests proposed in this paper. To this end, we repeatedly simulate independent panel datasets $\mathbf{X}_n = \{\{X_{i,t}\}_{t=1}^T\}_{i=1}^n$ where, for each $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$X_{i,t} = (\varepsilon_{i,t} + \xi_i)\alpha_t$$

with

$$\begin{pmatrix} \{\varepsilon_{i,t}\}_{t=1}^T \\ \xi_i \end{pmatrix} \sim N\left(0_{(T+1)\times 1}, \begin{bmatrix} \Xi & 0_{T\times 1} \\ 0_{1\times T} & 1 \end{bmatrix}\right),$$

where $\{\alpha_t\}_{t=1}^T$ is a sequence of constants and Ξ is a constant positive-definite matrix. By definition, ξ_i is a random effect and $\{\varepsilon_{i,t}\}_{t=1}^T$ are transient shocks with variance-covariate matrix Ξ. We consider two specifications for Ξ:

- WN: $\Xi = I_{T\times T}$, i.e., $\{\varepsilon_{i,t}\}_{t=1}^T$ is a white noise process with zero mean and unit variance. The distribution of $\mathbf{X}_n$ is time-exchangeable under this specification.

- AR(1): for all $t_1, t_2 = 1, \ldots, T$ with $t_1 \neq t_2$, $\Xi[t_1, t_1] = 2$ and $\Xi[t_1, t_2] = 2(-0.9)^{|t_1-t_2|}$, i.e., $\{\varepsilon_{i,t}\}_{t=1}^T$ is an AR(1) process zero mean, variance two, and correlation coefficient -0.9. In this case, the distribution of $\mathbf{X}_n$ is not time-exchangeable.

We consider two options for $\{\alpha_t\}_{t=1}^T$, which determines whether the data satisfies marginal homogeneity or not. For simulations under H_0 in (2.1), we use $\alpha_t = 1$ for $t = 1, \ldots, T$. For simulations

under H_1 in (2.1), we set $\alpha_t = \sqrt{1 + (t-1)/2}$ for $t = 1, \ldots, T$.

To compute the CvM statistics we set $\mathcal{U}_K$ equal to the 1/6, 2/6, 3/6, 4/6, and 5/6 empirical quantiles of $\{\{X_{i,t}\}_{t=1}^T\}_{i=1}^n$ (thus, $K = 5$). For each dataset, we implemented the following tests with a significance level of $\alpha = 5\%$:

- AA: Asymptotic approximation tests in Section 3.1, non-studentized as in (3.2) and studentized as in (3.3).

- BS: Bootstrap tests in Section 3.2, non-studentized as in (3.7) and studentized as in (3.9).

- BS2: Studentized bootstrap tests based on a bootstrapped covariance estimator, as described in Remark 3.1.

- PT: Permutation tests in Section 3.3, non-studentized as in (3.13) and studentized as in (3.14).

- PT2: Permutation tests based on the class of "all permutations" (i.e., time periods and units), as described in Remark 3.3. These can be non-studentized and studentized.

We consider simulations with $n \in \{30, 60, 120, 240, 480\}$ units and $T \in \{2, 3\}$ periods. The results shown in the tables are obtained from $S = 5,000$ independent panel data draws based on the design described.

Table 1 describes the rejection rates of the various tests under marginal homogeneity, i.e., H_0 in (2.1). When $T = 2$, all our proposed hypothesis tests (AA, BS, PT, studentized or not) provide exact size control when n is sufficiently large. These results are consistent with our asymptotic analysis. This conclusion also seems to apply to the BS2 and the studentized PT2 tests. On the other hand, the non-studentized PT2 test does not control size. This is consistent with our discussion in Lemma A.2, where we argue that the permutation class used to implement our PT tests is qualitatively different from the one used for the PT2 tests. Our simulations also allow us to examine how our asymptotically valid methods perform when n is relatively small. In this respect, one interesting finding is that the non-studentized versions of the AA and BS tests outperform their corresponding studentized ones when n is small. On the other hand, the PT test performs equally well with and without studentization for small n. The results for $T = 3$ are qualitatively similar to those for $T = 2$, except for the PT test. For $T > 2$, our formal results show that the studentized PT test is asymptotically valid, but the non-studentized PT test is not. This is clearly evidenced in our simulations with AR(1) shocks, where the non-studentized PT test exhibits rejection rates close to 13%. However, if data is time-exchangeable as with the WN shocks, then we observe that PT test is finite-sample valid regardless of studentization even if $T = 3$.

Table 2 explores the performance of the same test for data configurations that do not satisfy marginal homogeneity. To make the comparison fair, we focus on inference methods that are

T	Ξ	Test type	Critical value	$n = 30$	$n = 60$	$n = 120$	$n = 240$	$n = 480$
$T = 2$	WN	Non-stud.	AA	5.62	5.40	5.22	5.44	4.58
			BS	5.64	5.30	5.26	5.42	4.74
			PT	4.64	4.98	5.00	5.28	4.70
			PT2	1.00	0.96	0.70	0.88	0.84
		Studentized	AA	14.20	8.92	7.00	6.42	5.66
			BS	14.20	8.92	6.92	6.36	5.68
			BS2	3.18	4.58	4.66	5.38	5.14
			PT	5.18	5.18	4.92	5.52	5.24
			PT2	4.88	5.04	4.76	5.44	5.24
	AR(1)	Non-stud.	AA	6.06	5.74	5.34	5.26	4.96
			BS	6.08	5.80	5.44	5.12	4.82
			PT	4.86	5.14	4.98	5.18	4.74
			PT2	7.62	7.58	7.70	7.80	7.50
		Studentized	AA	14.22	9.02	6.74	5.86	5.20
			BS	14.38	9.02	6.78	6.04	5.06
			BS2	3.42	4.70	4.80	4.86	4.68
			PT	4.94	4.92	4.86	4.96	4.64
			PT2	4.98	5.00	5.00	4.88	4.72
$T = 3$	WN	Non-Studentized	AA	5.58	5.38	5.86	5.20	5.42
			BS	5.56	5.40	5.84	5.14	5.44
			PT	5.10	5.14	5.66	4.94	5.40
			PT2	0.90	0.84	0.66	0.52	0.76
		Studentized	AA	35.94	17.56	11.06	7.16	6.20
			BS	35.94	17.74	11.04	7.18	6.12
			BS2	0.58	3.66	4.92	5.30	4.94
			PT	5.50	5.42	5.26	5.28	4.92
			PT2	4.20	4.98	5.00	5.22	4.96
	AR(1)	Non-stud.	AA	5.80	5.62	4.96	5.36	5.28
			BS	5.80	5.68	5.14	5.36	5.26
			PT	13.06	13.74	13.14	13.74	13.08
			PT2	9.70	10.30	10.18	10.78	10.16
		Studentized	AA	35.66	17.98	9.26	8.02	6.00
			BS	36.06	17.98	9.36	7.94	6.02
			BS2	1.66	2.78	3.62	5.46	4.96
			PT	5.26	5.02	4.40	5.74	4.92
			PT2	4.34	4.58	4.12	5.50	4.82

Table 1: Empirical rejection rates (in %) under H_0 in (2.1) for $\alpha = 5\%$ based on $S = 5,000$ i.i.d. panel datasets for the various designs.

asymptotically valid. For $T = 2$, this includes studentized and non-studentized versions of the AA, BS, and PT tests, and studentized BS2. Our results indicate that studentized tests are considerably more powerful than their corresponding non-studentized versions. The main difference in the case of $T = 3$, is that we now have to eliminate the non-studentized PT test. With this exception, the results with $T = 3$ are qualitatively similar to those for $T = 2$.

Tables 1 and 2 offer interesting conclusions regarding the finite sample properties of the asymptotically valid tests. When the sample size is relatively small, the non-studentized tests appear to provide better size control than their studentized counterparts, though at the expense of lower power. As the sample size increases, the studentized tests improve their size control. As a result, the studentized tests seem to be a better option for larger sample sizes: they offer adequate size control and relatively higher power compared to their non-studentized versions.

6   Empirical application

This section applies our marginal homogeneity tests to the state variable in the dynamic discrete choice game in Igami and Yang (2016). In this paper, the authors develop and estimate a dynamic entry model oligopoly game among Canada's five main hamburger chains: A&W, Burger King, Harvey's, McDonald's, and Wendy's. They use yearly data from 400 geographical markets[1] located in seven major Canadian cities between 1970 and 2004, i.e., $n = 400$ and $T = 35$. For each market-year pair (i, t), they observe the number of stores for each chain, population, and income.

The state variable used in Igami and Yang (2016) is a discrete categorical variable whose value represents the number of stores of each chain, population, and income of a given market-year pair. We now briefly explain its construction, and defer to their Igami and Yang (2016, Section 4.1) for details. The paper restricts the number of stores per chain to three, and divides population and income into quartiles.

For each period $t = 1, \ldots, T = 35$ and market $i = 1, \ldots, n = 400$, $X_{i,t}$ is uniquely determined by the number of stores (up to three) for each chain $j = 1, 2, 3, 4, 5$, $N_{i,t,j} \in \{0, 1, 2, 3\}$, the population quartile $P_{i,t} \in \{1, 2, 3, 4\}$, and the income quartile $I_{i,t} \in \{1, 2, 3, 4\}$. So, $X_{i,t} = 1$ indicates that $N_{i,t,j} = 0$ for all $j = 1, 2, 3, 4, 5$, $P_{i,t} = 1$, and $I_{i,t} = 1$, $X_{i,t} = 2$ indicates that $N_{i,t,j} = 0$ for all $j = 1, 2, 3, 4, 5$, $P_{i,t} = 1$, and $I_{i,t} = 2$, and so on, until $X_{i,t} = 16,384$ indicates that $N_{i,t,j} = 3$ for all $j = 1, 2, 3, 4, 5$, $P_{i,t} = 4$, and $I_{i,t} = 4$. While $X_{i,t}$ could take up to $4^7 = 16,384$ possible values, it only takes 467 distinct values in the entire dataset.

The data spans an extensive period in which the Canadian fast-food industry grew considerably. As Igami and Yang (2016, Section 3.2) reports, the average number of shops per market was less

---

[1]The paper defines a market as a cluster of stores located within a 0.5-mile radius at any point of their sample period. Markets in downtown areas are omitted as these experience a different nature of competition.

T	Ξ	Test type	Critical value	$n=30$	$n=60$	$n=120$	$n=240$	$n=480$
$T=2$	WN	Non-stud.	AA	8.56	10.40	16.14	32.68	65.32
			BS	8.60	10.48	16.06	32.66	65.36
			PT	7.34	9.90	15.66	32.60	65.16
			PT2	1.38	1.64	2.76	7.74	28.80
		Studentized	AA	19.76	17.66	26.94	47.28	78.70
			BS	19.80	17.62	26.90	47.30	78.86
			BS2	5.04	10.68	22.22	44.74	77.78
			PT	8.22	11.64	22.64	44.90	77.64
			PT2	7.72	11.58	22.52	44.94	77.84
	AR(1)	Non-stud.	AA	6.96	7.04	8.32	13.20	27.66
			BS	7.00	7.10	8.28	13.16	27.72
			PT	5.34	6.28	7.98	12.96	27.42
			PT2	8.34	9.36	12.16	20.36	40.56
		Studentized	AA	19.28	16.84	24.40	42.68	72.80
			BS	19.58	16.90	24.62	42.88	73.00
			BS2	5.56	10.28	19.84	39.76	71.56
			PT	7.30	10.90	20.18	39.64	71.74
			PT2	7.54	10.76	20.20	39.86	71.78
$T=3$	WN	Non-stud.	AA	6.90	8.34	13.42	27.50	73.36
			BS	6.80	8.28	13.46	27.46	73.52
			PT	6.28	7.78	12.62	26.90	73.58
			PT2	1.12	1.04	2.10	3.50	19.32
		Studentized	AA	50.84	44.38	60.98	88.94	99.78
			BS	50.96	44.36	60.90	88.96	99.74
			BS2	1.98	16.12	45.92	85.26	99.72
			PT	11.00	21.54	47.26	85.38	99.72
			PT2	9.00	20.00	46.58	85.34	99.72
	AR(1)	Non-stud.	AA	6.20	6.78	7.02	8.60	15.18
			BS	6.30	6.70	7.06	8.52	14.96
			PT	13.62	16.10	17.14	23.48	46.72
			PT2	10.40	12.54	13.50	17.86	34.88
		Studentized	AA	60.96	66.26	90.00	99.78	100
			BS	61.48	66.50	89.98	99.78	100
			BS2	4.70	28.58	80.14	99.54	100
			PT	14.44	38.26	82.02	99.58	100
			PT2	12.52	36.56	81.38	99.56	100

Table 2: Empirical rejection rates (in %) under H_1 in (2.1) for $\alpha = 5\%$ based on $S = 5,000$ i.i.d. panel datasets for the various designs.
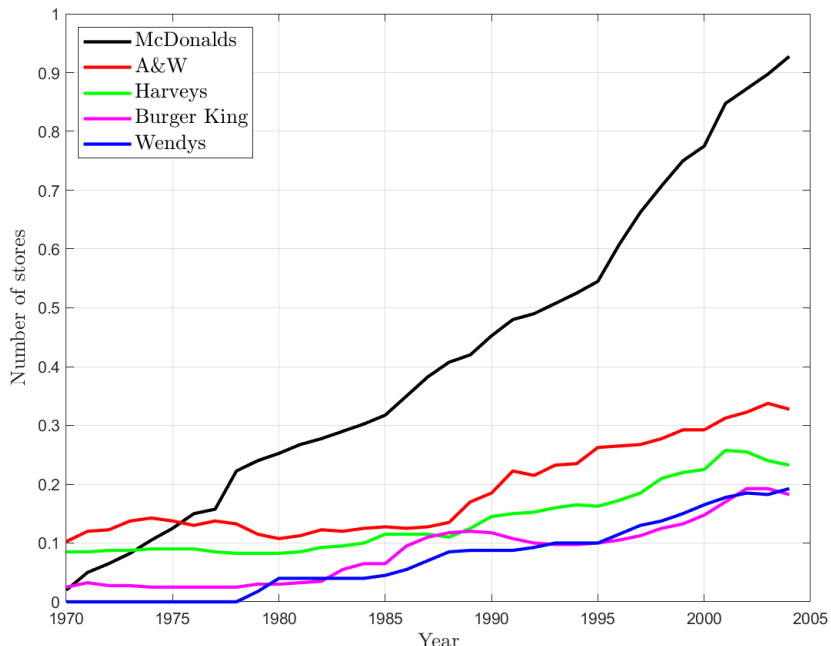
Figure 1: Average number of stores per market over time for each chain.

than 0.5 during 1970's to approximately 1.8 in the early 2000's. We now provide further evidence about the evolution of this industry. Figure 1 shows the average number of stores per market over time desegregated by chain. This figure shows that the average number of stores per market has increased across all chains, with the most significant increase observed for McDonald's compared to the other chains. Figure 2 shows the average number of competitors per market over time. This figure reveals that the frequency of empty markets decreased steadily between 1970 and 2000, while the frequencies of monopoly, duopoly, and triopoly or more steadily increased over the same period. In contrast, during 2000-2004, the frequency of each market type has remained relatively stable. This evidence suggests that the Canadian fast-food industry has been evolving between 1970 and the early 2000's, and may have reached a steady state during the last years of the sample. The hypothesis tests developed in this paper can be used to evaluate whether the state distribution is homogeneous over any of the periods in the sample.

As explained in Section 1, the marginal homogeneity of the state variable can be a source for efficiency gains in the estimation of dynamic discrete games. With this motivation in mind, we now apply our marginal homogeneity tests to our panel data of the state variable. Given the large number of values that the variable takes, Assumption 2 does not hold over the sample period. For this reason, we only consider non-studentized tests. We implement our tests for two subsets of periods. Since we consider panel data with $T > 2$, the non-studentized asymptotic approximation and bootstrap tests are valid, but the permutation test is not. Table 3 presents the results of
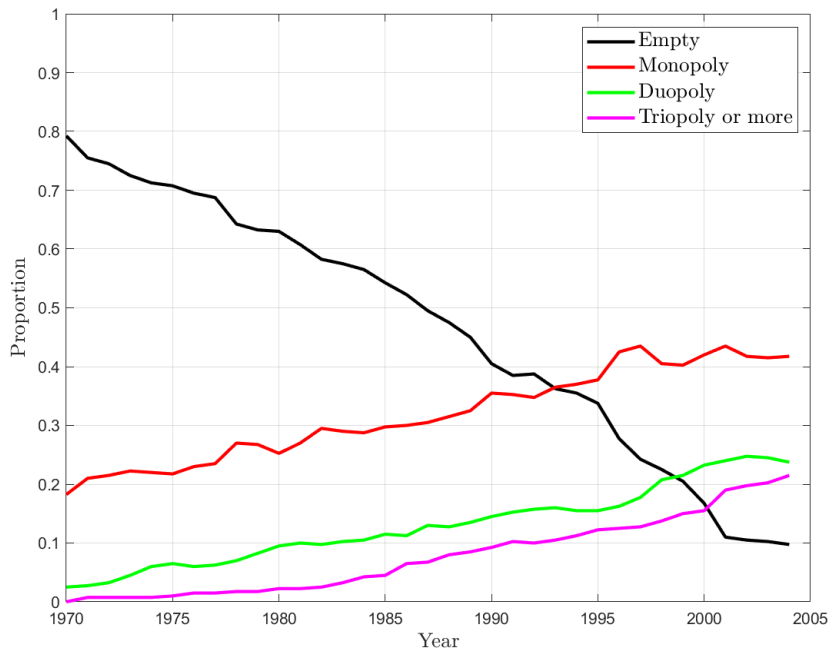
Figure 2: Average type of market over time.

our hypothesis tests for two subsets of sample periods. First, we consider a subset of our data every five years, i.e., 1970, 1975, 1980, 1985, 1990, 1995, and 2000. In this case, our tests strongly reject the hypothesis of marginal homogeneity. This result is expected, as it is consistent with the informal discussion in the previous paragraph regarding the growth of the Canadian fast-food industry between 1970 and 2000. Second, we repeat the analysis for the last four years in the sample period, i.e., 2001, 2002, 2003, and 2004. In this case, our tests do not reject the hypothesis of marginal homogeneity. These results suggest that the Canadian fast-food industry may be in a steady state in the latter part of the sample period.

Sample	Test type	Test statistic	Critical Value with $\alpha = 5\%$		
			Asy. approx.	Bootstrap	Permutation
Pre-2000	Non-stud.	10.76	0.88	0.82	2.35
Post-2000	Non-stud.	0.08	0.13	0.13	0.23

Table 3: Marginal homogeneity tests applied to panel data of the state variable in Igami and Yang (2016) with $\alpha = 5\%$. The Pre-2000 sample refers to the subsample with $t \in \{1970, 75, 80, 90, 95, 00\}$. The Post-2000 sample refers to the subsample with $t \in \{2001, 02, 03, 04\}$.

7   Conclusions

This paper proposes hypothesis tests to evaluate whether panel data satisfies the hypothesis of marginal homogeneity. As we argue in the paper, marginal homogeneity is a relevant property in economic settings such as dynamic discrete games.

Our asymptotic framework for panel data considers a diverging number of units n and a fixed number of periods T. We implement our tests by comparing a studentized or non-studentized T-sample version of the Cramér-von Mises statistic with a suitable critical value. Relative to the non-studentized case, the asymptotic analysis of the studentized statistics requires an additional assumption: the variance-covariate matrix used in the studentization must be non-singular. It is relevant to note that this condition can fail in practice. In fact, it failed in our empirical application.

We investigate three methods to construct the critical value: asymptotic approximations, the bootstrap, and time permutations. We prove that the asymptotic approximation and bootstrap tests are asymptotically valid, regardless of whether we use studentized or non-studentized test statistics. The permutation test based on a non-studentized statistic is asymptotically exact when $T = 2$, but is asymptotically invalid when $T > 2$. In contrast, the permutation test based on a studentized statistic is always asymptotically exact. Finally, under a time-exchangeability assumption, the permutation test is valid in finite samples, both with and without studentization.

We also study the power of the various methods. The asymptotically valid tests we consider are consistent and have non-trivial asymptotic power under suitable local alternatives. Moreover, the asymptotically valid tests based on the non-studentized statistics share the same local power properties, and the same holds for the asymptotically valid tests based on the studentized statistics.

Our Monte Carlo simulations investigate the finite sample behavior of our tests. The non-studentized tests exhibit better finite-sample size control than their studentized counterparts, though this comes at the cost of lower power. Finally, we apply our test to the state variable of the dynamic oligopoly model of the Canadian fast-food industry in Igami and Yang (2016). Our findings suggest that the industry evolved between 1970 and 2000, and appears to have reached a steady state since then.

A   Proofs

Throughout this appendix, we use RHS, LLN, CMT, CLT, PSD, PD, and to abbreviate "right-hand side", "law of large numbers", "continuous mapping theorem", "central limit theorem", "positive semi-definite", and "positive definite", respectively. We also define the sequence of random vectors $\{Z_i\}_{i=1}^n$ with

$$Z_i \equiv \left\{ \sqrt{P(u_k)}[1(X_{i,t} \leq u_k) - 1(X_{i,t+1} \leq u_k)] : \quad t = 1, \ldots, T-1, \quad k = 1, \ldots, K \right\} \in \mathbb{R}^{(T-1)K}, \quad \text{(A.1)}$$

where $P(u_k)$ is the aggregated probability given in (2.7).

Let $\Omega \in \mathbb{R}^{(T-1)K \times (T-1)K}$ denote a PSD matrix with a vector of eigenvalues $\ell \in \mathbb{R}^{(T-1)K}$. By the Principal-Axis Theorem (Scheffe, 1959, pp. 397, 418), $\epsilon'\Omega\epsilon$ with $\epsilon \sim N(0_{(T-1)K \times 1}, \Omega)$, has the same distribution as $S(\ell)$ defined in Section 3.1. That is, $\epsilon'\Omega\epsilon$ has a generalized chi-square distribution of weights equal to ℓ, unit vector of degrees of freedom, zero vector of non-centrality parameters, and no constant or normal terms. Throughout this appendix, we use $c^\chi(1 - \alpha; \Omega)$ to denote the $(1 - \alpha)$-quantile of $\epsilon'\Omega\epsilon$.

Proof of Theorem 2.1. <u>Part (a)</u>. Under Assumption 1 and H_0, $\{Z_i\}_{i=1}^n$ are i.i.d. with $E[Z_i] = 0_{1 \times (T-1)K}$ and $V[Z_i] = \Sigma_Z$ as defined in (2.6). By the CLT,

$$\tilde{Z} \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i' \xrightarrow{d} \xi \sim N(0_{(T-1)K \times 1}, \Sigma_Z). \tag{A.2}$$

For any $t = 1, \ldots, T-1$ and $k = 1, \ldots, K$, the $((t-1)K + k)$-component of $\hat{\delta} \equiv \hat{Z} - \tilde{Z}$ satisfies

$$\sqrt{n\hat{P}(u_k)}[\hat{F}_t(u_k) - \hat{F}_{t+1}(u_k)] - \sqrt{nP(u_k)}[\hat{F}_t(u_k) - \hat{F}_{t+1}(u_k)]$$
$$\overset{(1)}{=} \sqrt{n}[(\hat{F}_t(u_k) - F_t(u_k)) - (\hat{F}_{t+1}(u_k) - F_{t+1}(u_k))] \left(\sqrt{\hat{P}(u_k)} - \sqrt{P(u_k)}\right)$$
$$\overset{(2)}{=} O_p(1)o_p(1) = o_p(1), \tag{A.3}$$

where (1) holds by $F_t(u_k) = F_{t+1}(u_k)$, which is implied by H_0, and (2) by $\sqrt{n}[(\hat{F}_t(u_k) - F_t(u_k)) - (\hat{F}_{t+1}(u_k) - F_{t+1}(u_k))] = O_p(1)$ and $\hat{P}(u_k) - P(u_k) = o_p(1)$, which are implied by the CLT and LLN, respectively.

Then, consider the following derivation.

$$S_n \overset{(1)}{=} \tilde{Z}'\tilde{Z} + 2\hat{\delta}'\tilde{Z} + \hat{\delta}'\hat{\delta} \overset{(2)}{\underset{}{\xrightarrow{d}}} \xi'\xi. \tag{A.4}$$

where (1) holds by $\hat{\delta} = \hat{Z} - \tilde{Z}$, and (2) by (A.2) and (A.3).

To complete the proof, it then suffices to show that RHS of (A.4) can be expressed as (2.10). To this end, consider an orthogonal decomposition of $\Sigma_Z = H'\Lambda H$, where $H \in \mathbb{R}^{(T-1)K \times (T-1)K}$ is an orthogonal matrix (i.e., $HH' = I_{(T-1)K \times (T-1)K}$) and $\Lambda = diag\{\{\lambda_j\}_{j=1}^{(T-1)K}\}$ is the diagonal matrix of eigenvalues of Σ_Z. The desired result then holds by the following derivation:

$$\xi'\xi \overset{(1)}{=} \epsilon'\epsilon \overset{(2)}{=} \sum_{j=1}^{(T-1)K} \lambda_j \zeta_j^2, \tag{A.5}$$

where (1) holds for $\epsilon \equiv H\xi$ which implies $\epsilon'\epsilon = \xi'H'H\xi = \xi'\xi$, and (2) by $\epsilon \sim N(0_{(T-1)K \times 1}, \Lambda)$, and so $\epsilon = \{\lambda_j^{1/2}\zeta_j\}_{j=1}^{(T-1)K}$ for $\zeta \sim N(0_{(T-1)K}, I_{(T-1)K \times (T-1)K})$.

<u>Part (b)</u>. Under Assumption 1 and the LLN, $\hat{\Sigma}_Z = \Sigma_Z + o_p(1)$. Then, Assumption 2 and the CMT imply that $\hat{\Sigma}_Z^- = \Sigma_Z^{-1} + o_p(1)$. We can then repeat the arguments in part (a) to get

$$\hat{\delta} = o_p(1) \quad \text{and} \quad \tilde{Z}'\hat{\Sigma}_Z^-\tilde{Z} \xrightarrow{d} \xi'\Sigma_Z^{-1}\xi \sim \chi_{(T-1)K}^2. \tag{A.6}$$

From here, the desired result follows from the next derivation:

$$\bar{S}_n \overset{(1)}{=} \tilde{Z}'\hat{\Sigma}_Z^-\tilde{Z} + 2\hat{\delta}'\hat{\Sigma}_Z^-\tilde{Z} + \hat{\delta}'\hat{\Sigma}_Z^-\hat{\delta} \overset{d}{\to} \xi'\Sigma_Z^{-1}\xi \overset{(3)}{\sim} \chi^2_{(T-1)K},$$

where (1) holds by $\hat{\delta} = \hat{Z} - \tilde{Z}$, (2) by (A.6), and (3) by $\xi \sim N(0_{(T-1)K}, \Sigma_Z)$. ∎

Proof of Theorem 3.1. <u>Part (a).</u> We divide the argument into two cases.

<u>Case 1: $\Sigma_Z \neq 0_{(T-1)K \times (T-1)K}$.</u> Let $\lambda \in \mathbb{R}^{(T-1)K}$ denote the vector of eigenvalues of Σ_Z. Since λ is a continuous function of Σ_Z and $\hat{\Sigma}_Z \overset{p}{\to} \Sigma_Z$ holds, the CMT implies that $\hat{\lambda}_n \overset{p}{\to} \lambda$. Since Σ_Z is positive semi-definite, $\Sigma_Z \neq 0_{(T-1)K \times (T-1)K}$ implies that $\lambda \neq 0_{(T-1)K \times 1}$.

Note that $G(x, \ell)$ is continuous in ℓ for any $x \in \mathbb{R}$ and $\ell \in \mathbb{R}^{(T-1)K} \backslash 0_{(T-1)K \times 1}$. To see why, consider an arbitrary $x \in \mathbb{R}$ and sequence $\ell^{(n)} \to \ell \in \mathbb{R}^{(T-1)K} \backslash 0_{(T-1)K \times 1}$. The characteristic function of $S(\ell)$ is

$$\varphi(t, \ell) = \prod_{j=1}^{(T-1)K} (1 - it\ell_j)^{-1/2}$$

and satisfies $\varphi(t, \ell^{(n)}) \to \varphi(t, \ell)$ for all $t \in \mathbb{R}$. From this and Levy's Continuity Theorem (e.g., see Davidson (1994, Theorem 22.17)) we deduce that $S(\ell^{(n)}) \overset{d}{\to} S(\ell)$. This is equivalent to $G(x, \ell^{(n)}) \to G(x, \ell)$ since $S(\ell)$ is continuously distributed (as $\ell \neq 0_{(T-1)K \times 1}$). Since the choices of $x \in \mathbb{R}$ and $\ell \in \mathbb{R}^{(T-1)K} \backslash 0_{(T-1)K \times 1}$ were arbitrary, the desired result follows.

Given the continuity of $G(x, \cdot)$ for all $x \in \mathbb{R}$, the CMT gives $G(x, \hat{\lambda}_n) \overset{p}{\to} G(x, \lambda) = P(S \leq x)$. In turn, since $G(x, \hat{\lambda}_n)$ and $P(S \leq x)$ are weakly increasing and bounded, and $P(S \leq x)$ is continuous (as $\lambda \neq 0_{(T-1)K \times 1}$), an argument along the lines of van der Vaart (1998, Lemma 2.11) implies that

$$\sup_{x \in \mathbb{R}} |G(x, \hat{\lambda}_n) - P(S \leq x)| \overset{p}{\to} 0. \tag{A.7}$$

We now show that

$$c_n^A(1 - \alpha) \overset{p}{\to} c^X(1 - \alpha; \Sigma_Z). \tag{A.8}$$

Fix $\varepsilon > 0$ arbitrarily. Since the CDF of S is continuous and strictly increasing at $c^X(1 - \alpha; \Sigma_Z) > 0$, $\exists \delta = \delta(\varepsilon) > 0$ such that

$$c^X(1 - \alpha; \Sigma_Z) - c^X(1 - \alpha - \delta; \Sigma_Z) \leq \varepsilon \quad \text{and} \quad c^X(1 - \alpha + \delta; \Sigma_Z) + \delta - c^X(1 - \alpha; \Sigma_Z) \leq \varepsilon. \tag{A.9}$$

Then, let E_n be defined by

$$E_n = \left\{ \sup_{x \in \mathbb{R}} |G(x, \hat{\lambda}_n) - P(S \leq x)| < \delta \right\}.$$

Under E_n, we have

$$\delta \overset{(1)}{\geq} G(c_n^A(1 - \alpha), \hat{\lambda}_n) - P(S \leq c_n^A(1 - \alpha)) \overset{(2)}{\geq} (1 - \alpha) - P(S \leq c_n^A(1 - \alpha)),$$

23

where (1) holds by E_n, and (2) by (3.1), as it implies $G(c_n^A(1-\alpha), \hat{\lambda}_n) \geq 1 - \alpha$. This yields

$$P(S \leq c_n^A(1-\alpha)) \ \geq \ (1-\alpha) - \delta. \tag{A.10}$$

From here, we can get

$$c_n^A(1-\alpha) \ \overset{(1)}{\geq} \ c^\chi(1-\alpha-\delta; \Sigma_Z) \ \overset{(2)}{\geq} \ c^\chi(1-\alpha; \Sigma_Z) - \varepsilon, \tag{A.11}$$

where (1) holds by (A.10) and (2) by the first condition in (A.9). Also under E_n, we have

$$-\delta \ \overset{(1)}{\leq} \ G(c_n^A(1-\alpha) - \delta, \hat{\lambda}_n) - P(S \leq c_n^A(1-\alpha) - \delta) \ \overset{(2)}{<} \ (1-\alpha) - P(S \leq c_n^A(1-\alpha) - \delta),$$

where (1) holds by E_n, and (2) by $G(c_n^A(1-\alpha) - \delta, \hat{\lambda}_n) < 1 - \alpha$. This implies that

$$P(S \leq c_n^A(1-\alpha) - \delta) \ < \ (1-\alpha) + \delta. \tag{A.12}$$

From here, we can get

$$c_n^A(1-\alpha) \ \overset{(1)}{\leq} \ c^\chi((1-\alpha) + \delta; \Sigma_Z) + \delta \ \overset{(2)}{\leq} \ c^\chi(1-\alpha; \Sigma_Z) + \varepsilon, \tag{A.13}$$

where (1) holds by (A.12) and (2) by the second condition in (A.9). By combining (A.11) and (A.13), we conclude that $|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon$. From this argument, we deduce that

$$P(E_n) \ \leq \ P(|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon). \tag{A.14}$$

Since $P(E_n) \to 1$ by (A.7), we conclude from (A.14) that $P(|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon) \to 1$. Since the choice of $\varepsilon > 0$ was arbitrary, (A.8) follows.

For any $\varepsilon \in (0, c^\chi(1-\alpha; \Sigma_Z))$, consider the following argument.

$$
\begin{aligned}
&P(\{S_n < c^\chi(1-\alpha; \Sigma_Z) - \varepsilon\}) + P(\{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon\}) - 1 \\
&\leq \ P(\{S_n < c^\chi(1-\alpha; \Sigma_Z) - \varepsilon\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon\}) \\
&\leq \ \left\{ \begin{array}{c} P(\{S_n < c^\chi(1-\alpha; \Sigma_Z) - \varepsilon\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon\}) \\ +P(\{S_n < c_n^A(1-\alpha)\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| > \varepsilon\}) \end{array} \right\} \\
&\leq \ \left\{ \begin{array}{c} P(\{S_n < c_n^A(1-\alpha)\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon\})+ \\ P(\{S_n < c_n^A(1-\alpha)\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| > \varepsilon\}) \end{array} \right\} \\
&= \ P(S_n < c_n^A(1-\alpha)) \\
&\overset{(1)}{\leq} \ 1 - E[\phi_n^A(\alpha)] \\
&\overset{(2)}{\leq} \ P(S_n \leq c_n^A(1-\alpha)) \\
&= \ \left\{ \begin{array}{c} P(\{S_n \leq c_n^A(1-\alpha)\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| \leq \varepsilon\})+ \\ P(\{S_n \leq c_n^A(1-\alpha)\} \cap \{|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| > \varepsilon\}) \end{array} \right\} \\
&\leq \ P(S_n \leq c^\chi(1-\alpha; \Sigma_Z) + \varepsilon) + P(|c_n^A(1-\alpha) - c^\chi(1-\alpha; \Sigma_Z)| > \varepsilon),
\end{aligned}
\tag{A.15}
$$

where (1) and (2) hold by $\{S_n < c_n^A(1-\alpha)\} \subseteq \{\phi_n^A(\alpha) = 0\} \subseteq \{S_n \leq c_n^A(1-\alpha)\}$. By taking sequential limits on (A.15) as $n \to \infty$ and $\varepsilon \to 0$, and combined with (A.8), Theorem 2.1(a), and the fact that S is continuously distributed, we conclude that $E[\phi_n^A(\alpha)] \to P(S > c^\chi(1-\alpha; \Sigma_Z)) = \alpha$, as desired.

Case 2: $\Sigma_Z = 0_{(T-1)K \times (T-1)K}$. Under H_0 in (2.1), we also have $E[Z_i] = 0_{(T-1)K \times 1}$, and so $Z_i = 0_{(T-1)K \times 1}$ a.s. By (A.1), this implies that $1(X_{i,t} \leq u_k) = 1(X_{i,t+1} \leq u_k)$ a.s. for all $t = 1, \ldots, T-1$ and $k = 1, \ldots K$. In turn, this gives that $\hat{Z} = 0_{(T-1)K \times 1}$ a.s., with $\hat{Z}$ defined as in (2.5). Then, $S_n = 0$ holds a.s. By this and $c_n^A(1-\alpha) \geq 0$, we get $E[\phi_n^A(\alpha)] = P(S_n > c_n^A(1-\alpha)) = 0 < \alpha$. From here, the desired result holds by taking limits as $n \to \infty$.

Part (b) The conclusion follows directly from combining (3.3), Theorem 2.1(b), and that the CDF of $\chi^2_{(T-1)K}$ is continuous at $c^\chi(1-\alpha; I_{(T-1)K \times (T-1)K}) > 0$. ∎

Proof of Theorem 3.2. Part (a). We divide the argument into two cases.

Case 1: $\Sigma_Z \neq 0_{(T-1)K \times (T-1)K}$. For each $i = 1, \ldots, n$, $t = 1, \ldots, T-1$, and $k = 1, \ldots, K$, let

$$Z_{i,(t-1)K+k}^* \equiv \sqrt{P(u_k)}[1(X_{i,t}^* \leq u_k) - \hat{F}_t(u_k) - (1(X_{i,t+1}^* \leq u_k) - \hat{F}_{t+1}(u_k))],$$

and let $Z_i^* = \{Z_{i,(t-1)K+k}^* : k = 1, \ldots, K, t = 1, \ldots, T-1\} \in \mathbb{R}^{(T-1)K}$ and $\tilde{Z}^* \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i^{*\prime}$. By van der Vaart and Wellner (1996, Theorem 3.6.2),

$$\{\tilde{Z}^* | \mathbf{X}_n\} \xrightarrow{d} \xi \sim N(0_{(T-1)K \times 1}, \Sigma_Z) \quad \text{a.s.} \tag{A.16}$$

Let $\hat{\delta}^* \equiv \hat{Z}^* - \tilde{Z}^*$ with $\hat{Z}^*$ as in (3.5). For any $t = 1, \ldots, T-1$ and $k = 1, \ldots, K$, then conditional on $\mathbf{X}_n$, the $((t-1)K+k)$-component of $\hat{\delta}^*$ satisfies

$$\sqrt{n\hat{P}(u_k)}[\hat{F}_t^*(u_k) - \hat{F}_t(u_k) - (\hat{F}_{t+1}^*(u_k) - \hat{F}_{t+1}(u_k))] - \sqrt{nP(u_k)}[\hat{F}_t^*(u_k) - \hat{F}_t(u_k) - (\hat{F}_{t+1}^*(u_k) - \hat{F}_{t+1}(u_k))]$$
$$= \sqrt{n}[\hat{F}_t^*(u_k) - \hat{F}_t(u_k) - (\hat{F}_{t+1}^*(u_k) - \hat{F}_{t+1}(u_k))](\sqrt{\hat{P}(u_k)} - \sqrt{P(u_k)})$$
$$\overset{(1)}{=} O_p(1)o(1) = o_p(1), \quad \text{w.p.a.1}, \tag{A.17}$$

where (1) holds by van der Vaart and Wellner (1996, Theorem 3.6.2) (which implies that $\{\sqrt{n}[\hat{F}_t^*(u_k) - \hat{F}_t(u_k) - (\hat{F}_{t+1}^*(u_k) - \hat{F}_{t+1}(u_k))] | \mathbf{X}_n\} = O_p(1)$ a.s.) and that $\hat{P}(u_k) - P(u_k) = o_p(1)$ by LLN.

Then, consider the following derivation. Conditional on $\mathbf{X}_n$,

$$S_n^* \overset{(1)}{=} (\tilde{Z}^*)'(\tilde{Z}^*) + 2(\hat{\delta}^*)'(\tilde{Z}^*) + (\hat{\delta}^*)'(\hat{\delta}^*) \overset{(2)}{\xrightarrow{d}} S, \quad \text{w.p.a.1}, \tag{A.18}$$

where (1) holds by $\hat{\delta}^* = \hat{Z}^* - \tilde{Z}^*$ and (2) by (A.16) and (A.17). As a corollary of (A.18) and also by the condition that Σ_Z is a nonzero matrix, we deduce that $P(S_n^* \leq x | \mathbf{X}_n) \xrightarrow{P} P(S \leq x)$ for all points $x \in \mathbb{R}$. From this point onward, the rest of the proof is identical to that of part (a) in Theorem 3.1.

Case 2: $\Sigma_Z = 0_{(T-1)K \times (T-1)K}$. This result holds by the same argument as in Theorem 3.1, except that $\phi_n^A(\alpha)$ and $c_n^A(1-\alpha)$ are replaced by $\phi_n^B(\alpha)$ and $c_n^B(1-\alpha)$, respectively.

25

<u>Part (b).</u> By the proof of part (b) in Theorem 3.1, $\hat{\Sigma}_Z^- = \Sigma_Z^{-1} + o_p(1)$. Then, conditional on $\mathbf{X}_n$,

$$\hat{\Sigma}_Z^- \to \Sigma_Z^{-1} \quad \text{w.p.a.1.} \tag{A.19}$$

We can then repeat the arguments in part (a) to get that, conditional on $\mathbf{X}_n$,

$$\hat{\delta}^* = o_p(1) \quad \text{and} \quad (\tilde{Z}^*)'\hat{\Sigma}_Z^-(\tilde{Z}^*) \overset{d}{\to} \xi'\Sigma_Z^{-1}\xi \sim \chi^2_{(T-1)K} \quad \text{w.p.a.1.} \tag{A.20}$$

From here, the desired result follows from the next derivation. Conditional on $\mathbf{X}_n$,

$$\bar{S}_n^* \overset{(1)}{=} (\tilde{Z}^*)'\hat{\Sigma}_Z^-(\tilde{Z}^*) + 2(\hat{\delta}^*)'\hat{\Sigma}_Z^-(\tilde{Z}^*) + (\hat{\delta}^*)'\hat{\Sigma}_Z^-(\hat{\delta}^*) \overset{(2)}{\underset{d}{\to}} \chi^2_{(T-1)K}, \quad \text{w.p.a.1}, \tag{A.21}$$

where (1) holds by $\hat{\delta}^* = \hat{Z}^* - \tilde{Z}^*$, and (2) by (A.19) and (A.20). As a corollary of (A.21), we have that $P(\bar{S}_n^* \leq x|\mathbf{X}_n) \overset{p}{\to} P(\chi^2_{(T-1)K} \leq x)$ for all $x \in \mathbb{R}$. From this point onward, the rest of the proof follows from arguments in part (a) in Theorem 3.1. ∎

Theorem A.1. Let Assumption 1 hold, and let $\hat{Z}^{\boldsymbol{\pi}}$ be as in (3.11), where $\boldsymbol{\pi}$ is a uniformly chosen *random permutation* in $\mathcal{M}^n$. Then,

$$P(\hat{Z}^{\boldsymbol{\pi}} \leq x|\mathbf{X}_n) \overset{p}{\to} P(\xi \leq x) \tag{A.22}$$

for all x such that $P(\xi \leq x)$ is continuous, where $\xi \sim N(0_{(T-1)K \times 1}, \Omega_Z^{\pi})$ with

$$\Omega_Z^{\pi} \equiv \frac{1}{T!} \sum_{\pi \in \mathcal{M}} B(\pi)\Omega_Z B(\pi)', \tag{A.23}$$

$\Omega_Z = E[Z_i Z_i']$ with Z_i as in (A.1), and $\{B(\pi) \in \{-1,0,1\}^{(T-1)K \times (T-1)K} : \pi \in \mathcal{M}\}$ are known matrices defined within in the proof.

Proof. We divide the proof into several steps.

    <u>Step 1.</u> Introduce suitable notation.

    For each $i = 1, \ldots, n$ and $k = 1, \ldots, K$, let

$$V_{i,k} \equiv \begin{bmatrix} 1(X_{i,1} \leq u_k) - 1(X_{i,2} \leq u_k) \\ \vdots \\ 1(X_{i,T-1} \leq u_k) - 1(X_{i,T} \leq u_k) \end{bmatrix} \quad \text{and} \quad V_i \equiv \begin{bmatrix} V_{i,1} \\ \vdots \\ V_{i,K} \end{bmatrix} \in \{-1,0,1\}^{(T-1)K \times 1}. \tag{A.24}$$

Also, let

$$M \equiv diag\{\sqrt{P(u_1)}, \ldots, \sqrt{P(u_K)}\} \otimes I_{T-1} \in [0,1]^{(T-1)K \times (T-1)K}$$
$$\hat{M} \equiv diag\{\sqrt{\hat{P}(u_1)}, \ldots, \sqrt{\hat{P}(u_K)}\} \otimes I_{T-1} \in [0,1]^{(T-1)K \times (T-1)K}, \tag{A.25}$$

where $\otimes$ denotes the Kronecker product. Note that $Z_i = M \times V_i$ for all $i = 1, \ldots, n$.

    Let $\pi^n = \{\pi_i : i = 1, \ldots, n\} \in \mathcal{M}^n$ denote a fixed permutation. For any $i = 1, \ldots, n$ and $k = 1, \ldots, K$,

the π-permutation analogs of $V_{i,k}$ and V_i

$$V_{i,k}^\pi \equiv \begin{bmatrix} 1(X_{i,\pi_i(1)} \le u_k) - 1(X_{i,\pi_i(2)} \le u_k) \\ \vdots \\ 1(X_{i,\pi_i(T-1)} \le u_k) - 1(X_{i,\pi_i(T)} \le u_k) \end{bmatrix} \quad \text{and} \quad V_i^\pi \equiv \begin{bmatrix} V_{i,1}^\pi \\ \vdots \\ V_{i,K}^\pi \end{bmatrix} \in \{-1,0,1\}^{K(T-1)\times 1}. \quad \text{(A.26)}$$

We note that $\hat{M}$ is invariant to π^n. To see why, note that for each $k = 1, \ldots, K$, and $t = 1, \ldots, T$,

$$\hat{P}^\pi(u_k) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T 1(u_{k-1} < X_{i,\pi_i(t)} \le u_k) \stackrel{(1)}{=} \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T 1(u_{k-1} < X_{i,t} \le u_k) = \hat{P}(u_k), \quad \text{(A.27)}$$

where (1) holds because the sum over $t = 1, \ldots, T$ is invariant across the permutation.

<u>Step 2.</u> For any $i = 1, \ldots, n$ and $\pi_i \in \mathcal{M}$, we define the matrix $B(\pi_i)$ described in the statement and establish the following representation:

$$V_i^\pi = B(\pi_i) \times V_i. \quad \text{(A.28)}$$

This result follows from expressing $V_{i,k}^\pi$ as a particular linear combination of $V_{i,k}$. To see why, fix $t = 1, \ldots, T$ arbitrarily. If $\pi_i(t) < \pi_i(t+1)$, then we have

$$1(X_{i,\pi_i(t)} \le u_k) - 1(X_{i,\pi_i(t+1)} \le u_k) = \begin{bmatrix} 1(X_{i,\pi_i(t)} \le u_k) - 1(X_{i,\pi(t)+1} \le u_k) + \ldots \\ +1(X_{i,\pi_i(t+1)-1} \le u_k) - 1(X_{i,\pi(t+1)} \le u_k) \end{bmatrix}$$

$$= (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0) \times V_{i,k},$$

where the ones are located at time periods corresponding to $\pi_i(t), \ldots, \pi_i(t+1) - 1$. Conversely, if $\pi_i(t) > \pi_i(t+1)$, then we have

$$1(X_{i,\pi_i(t)} \le u_k) - 1(X_{i,\pi_i(t+1)} \le u_k) = \begin{bmatrix} -(1(X_{i,\pi_i(t)} \le u_k) - 1(X_{i,\pi(t)+1} \le u_k)) - \ldots \\ -(1(X_{i,\pi_i(t+1)-1} \le u_k) - 1(X_{i,\pi(t+1)} \le u_k)) \end{bmatrix}$$

$$= (0, \ldots, 0, -1, \ldots, -1, 0, \ldots, 0) \times V_{i,k},$$

where the minus ones are located at time periods corresponding to $\pi_i(t), \ldots, \pi_i(t+1) - 1$. Since $\pi_i(t)$ and $\pi_i(t+1)$ were arbitrarily chosen, we can define a matrix $\tilde{B}(\pi_i) \in \{-1,0,1\}^{(T-1)\times(T-1)}$ such that $V_{i,k}^\pi = \tilde{B}(\pi_i) \times V_{i,k}$. By collecting results for $k = 1, \ldots K$, and setting

$$B(\pi_i) \equiv diag\{\tilde{B}(\pi_i), \ldots, \tilde{B}(\pi_i)\}, \quad \text{(A.29)}$$

(A.28) follows. Finally, by repeating this operation for all $\pi = \pi_i \in \mathcal{M}$, we define the collection of matrices $\{B(\pi) \in \{-1,0,1\}^{(T-1)K\times(T-1)K} : \pi \in \mathcal{M}\}$.

<u>Step 3.</u> Establish the Hoeffding's condition for $\frac{1}{\sqrt{n}} \sum_{i=1}^n B(\pi_i)Z_i$, where $\pi^n = \{\pi_i : i = 1, \ldots, n\}$ denotes a randomly chosen permutation in $\mathcal{M}^n$. That is,

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n B(\pi_i)Z_i, \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\pi}_i)Z_i \right) \stackrel{d}{\to} (\xi, \tilde{\xi}), \quad \text{(A.30)}$$

where $\boldsymbol{\pi}^n = \{\boldsymbol{\pi}_i : i = 1, \ldots, n\}$ and $\tilde{\boldsymbol{\pi}}^n = \{\tilde{\boldsymbol{\pi}}_i : i = 1, \ldots, n\}$ denote two mutually independent random permutations chosen uniformly from $\mathcal{M}^n$ and independent of the data, and ξ and $\tilde{\xi}$ are i.i.d. $N(0_{(T-1)K \times 1}, \Omega_Z^\pi)$.

We establish (A.30) using the Cramér-Wold device. That is, for arbitrary $\lambda, \nu \in \mathbb{R}^{(T-1)K \times 1}$, (A.30) follows from showing that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i \overset{d}{\to} N(0_{(T-1)K \times 1}, \lambda' \Omega_Z^\pi \lambda + \nu' \Omega_Z^\pi \nu). \tag{A.31}$$

We begin the argument by showing that $\{(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i\}_{i=1}^n$ is an i.i.d. sequence. To see why, note that $\{Z_i\}_{i=1}^n$ is i.i.d. by Assumption 1. Also, since $\boldsymbol{\pi}^n = \{\boldsymbol{\pi}_i : i = 1, \ldots, n\}$ and $\tilde{\boldsymbol{\pi}}^n = \{\tilde{\boldsymbol{\pi}}_i : i = 1, \ldots, n\}$ are defined as i.i.d. sequences, we conclude that $\{B(\boldsymbol{\pi}_i)\}_{i=1}^n$ and $\{B(\tilde{\boldsymbol{\pi}}_i)\}_{i=1}^n$ are also i.i.d. By combining these facts, we get that $\{(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i\}_{i=1}^n$ is an i.i.d. sequence.

As a next step, we now show that

$$\sum_{\pi \in \mathcal{M}} B(\pi) = 0_{(T-1)K \times (T-1)K}. \tag{A.32}$$

Since $B(\pi) \equiv diag\{\tilde{B}(\pi), \ldots, \tilde{B}(\pi)\}$, it suffices to show that $\sum_{\pi \in \mathcal{M}} \tilde{B}(\pi) = 0_{(T-1) \times (T-1)}$. For each $t = 1, \ldots, T-1$, denote the t'th row of $\tilde{B}(\pi)$, denoted $\tilde{B}_t(\pi)$, can be expressed as follows:

$$\tilde{B}_t(\pi) = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0) \times (1[\pi(t) < \pi(t+1)] - 1[\pi(t) > \pi(t+1)]),$$

where the sequence of ones appears in the positions $\min\{\pi(t), \pi(t+1)\}$ through $\max\{\pi(t), \pi(t+1)\} - 1$. From here, we get that $\sum_{\pi \in \mathcal{M}} \tilde{B}_t(\pi) = 0_{1 \times (T-1)}$, as the occurrence of $(0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)$ in the sum over $\mathcal{M}$ cancels with the corresponding $(0, \ldots, 0, -1, \ldots, -1, 0, \ldots, 0)$ when $\pi(t)$ and $\pi(t+1)$ are reversed.

Next, we show that $E[B(\boldsymbol{\pi}_i)] = E[B(\tilde{\boldsymbol{\pi}}_i)] = 0_{(T-1)K \times (T-1)K}$ for all $i = 1, \ldots, n$. To see why, fix $i = 1, \ldots, n$ arbitrarily and note that

$$E[B(\boldsymbol{\pi}_i)] \overset{(1)}{=} E[B(\tilde{\boldsymbol{\pi}}_i)] \overset{(2)}{=} \frac{1}{T!} \sum_{\pi \in \mathcal{M}} B(\pi) \overset{(3)}{=} 0_{(T-1)K \times (T-1)K}, \tag{A.33}$$

where (1) holds because $B(\boldsymbol{\pi}_i)$ and $B(\tilde{\boldsymbol{\pi}}_i)$ are equally distributed, (2) because there are $|\mathcal{M}| = (T!)$ possible permutations of $\{1, 2, \ldots, T\}$, all equally likely, and (3) by (A.32). Then, for all $i = 1, \ldots, n$,

$$E[(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i] \overset{(1)}{=} E[(\lambda' E[B(\boldsymbol{\pi}_i)] + \nu' E[B(\tilde{\boldsymbol{\pi}}_i)]) Z_i] \overset{(2)}{=} 0, \tag{A.34}$$

where (1) holds by $Z_i \perp (B(\boldsymbol{\pi}_i), B(\tilde{\boldsymbol{\pi}}_i))$ and (2) by (A.33).

From here, note that for all $i = 1, \ldots, n$,

$$\begin{aligned} V[(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i] &= E[(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i Z_i' (B(\boldsymbol{\pi}_i)' \lambda + B(\tilde{\boldsymbol{\pi}}_i)' \nu)] \\ &\overset{(1)}{=} E[(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) \Omega_Z (B(\boldsymbol{\pi}_i)' \lambda + B(\tilde{\boldsymbol{\pi}}_i)' \nu)] \\ &\overset{(2)}{=} \lambda' E[B(\boldsymbol{\pi}_i) \Omega_Z B(\boldsymbol{\pi}_i)'] \lambda + \nu' E[B(\tilde{\boldsymbol{\pi}}_i) \Omega_Z B(\tilde{\boldsymbol{\pi}}_i)'] \nu, \\ &\overset{(3)}{=} \lambda' \Omega_Z^\pi \lambda + \nu' \Omega_Z^\pi \nu, \end{aligned} \tag{A.35}$$

where (1) holds by $Z_i \perp (B(\boldsymbol{\pi}_i), B(\tilde{\boldsymbol{\pi}}_i))$ and $E[Z_i Z_i'] = \Omega_Z$, (2) by (A.33) and that $B(\boldsymbol{\pi}_i) \perp B(\tilde{\boldsymbol{\pi}}_i)$, and (3) by (A.23) and that there are $|\mathcal{M}| = (T!)$ possible permutations of $\{1, 2, \ldots, T\}$, and all are equally likely.

To conclude the step, note that (A.31) follows from the CLT, as $\{(\lambda' B(\boldsymbol{\pi}_i) + \nu' B(\tilde{\boldsymbol{\pi}}_i)) Z_i\}_{i=1}^n$ was shown to be an i.i.d. sequence that satisfies (A.34) and (A.35).

<u>Step 4.</u> Use the previous steps to conclude the proof.

By Chung and Romano (2016, Lemma A.1), (A.22) is equivalent to showing that $\hat{Z}^\pi$ satisfies the following Hoeffding condition:

$$(\hat{Z}^{\boldsymbol{\pi}}, \hat{Z}^{\tilde{\boldsymbol{\pi}}}) \xrightarrow{d} (\xi, \tilde{\xi}), \tag{A.36}$$

where $\hat{Z}^{\boldsymbol{\pi}}$ and $\hat{Z}^{\tilde{\boldsymbol{\pi}}}$ are permuted according to $\boldsymbol{\pi}^n = \{\boldsymbol{\pi}_i : i = 1, \ldots, n\}$ and $\tilde{\boldsymbol{\pi}}^n = \{\tilde{\boldsymbol{\pi}}_i : i = 1, \ldots, n\}$, respectively, which are two mutually independent random permutations chosen uniformly from $\mathcal{M}^n$ and independent of the data, and ξ and $\tilde{\xi}$ are i.i.d. according to $N(0_{(T-1)K \times 1}, \Omega_Z^\pi)$.

Before proving the desired result, we establish three preliminary results. First, by repeating the arguments in step 3 but with M replaced by $I \in \mathbb{R}^{K(T-1) \times K(T-1)}$, we have that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n B(\boldsymbol{\pi}_i) V_i = O_p(1) \quad \text{and} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\boldsymbol{\pi}}_i) V_i = O_p(1). \tag{A.37}$$

Second, note that Assumption 1, the LLN, and the CMT imply that

$$\hat{\delta} \equiv \hat{M} - M = o_p(1). \tag{A.38}$$

Third, note that for any permutation $\pi \in \mathcal{M}$, we have

$$\begin{aligned} M \times B(\pi) &= diag\{\sqrt{P(u_1)} \tilde{B}(\pi), \ldots, \sqrt{P(u_K)} \tilde{B}(\pi)\} \\ &= \{diag\{\tilde{B}(\pi), \ldots, \tilde{B}(\pi)\}\} \times (diag\{\sqrt{P(u_1)}, \ldots \sqrt{P(u_K)}\} \otimes I_{T-1}) \\ &= B(\pi) \times M. \end{aligned} \tag{A.39}$$

The desired result follows from the next derivation.

$$\begin{aligned} (\hat{Z}^{\boldsymbol{\pi}}, \hat{Z}^{\tilde{\boldsymbol{\pi}}}) &\overset{(1)}{=} \left(\frac{\hat{M}}{\sqrt{n}} \sum_{i=1}^n V_i^{\boldsymbol{\pi}}, \frac{\hat{M}}{\sqrt{n}} \sum_{i=1}^n V_i^{\tilde{\boldsymbol{\pi}}} \right) \\ &\overset{(2)}{=} \left(\frac{\hat{M}}{\sqrt{n}} \sum_{i=1}^n B(\boldsymbol{\pi}_i) V_i, \frac{\hat{M}}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\boldsymbol{\pi}}_i) V_i \right) \\ &\overset{(3)}{=} \left(\begin{array}{c} \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\boldsymbol{\pi}_i) Z_i + \hat{\delta} \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\boldsymbol{\pi}_i) V_i, \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\boldsymbol{\pi}}_i) Z_i + \hat{\delta} \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\boldsymbol{\pi}}_i) V_i \end{array} \right) \\ &\overset{(4)}{=} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n B(\boldsymbol{\pi}_i) Z_i, \frac{1}{\sqrt{n}} \sum_{i=1}^n B(\tilde{\boldsymbol{\pi}}_i) Z_i \right) + o_p(1) \\ &\overset{(5)}{\xrightarrow{d}} (\xi, \tilde{\xi}), \end{aligned}$$

as desired, where (1) holds by (3.11), (A.26), and (A.27); (2) holds by (A.28); (3) by (A.39) and $Z_i = M V_i$

29

for all $i = 1, \ldots, n$; (4) by (A.37) and (A.38); and (5) by (A.30). ∎

Proof of Theorem 3.3. Throughout the proof, we continuously invoke the results and notations from Theorem A.1. Recall that this result derives the asymptotic distribution of $\hat{Z}^{\boldsymbol{\pi}}$ as in (3.11), where $\boldsymbol{\pi}$ is a uniformly chosen random permutation in $\mathcal{M}^n$. Recall from this theorem that $\Omega_Z \equiv E[Z_i Z_i']$, which equals $\Sigma_Z \equiv var(Z_i)$ under H_0 in (2.1); Ω_Z^π, defined in (A.23), represents the asymptotic variance of randomization distribution; $B(\pi)$, defined below (A.29) for $\pi \in \mathcal{M}$, denotes a known matrix taking values in $\{-1, 0, 1\}$. With these notations in mind, we present the proof below.

Part (a). We divide the argument into two cases.

Case 1: $\Sigma_Z \neq 0_{T(K-1) \times T(K-1)}$. By $T = 2$, there are $T! = 2$ permutations of $(1, 2)$, hence $\mathcal{M} = \{(1, 2), (2, 1)\}$. Following the construction in step 2 of Theorem A.1, $B((1, 2)) = I_{K \times K}$ and $B((2, 1)) = -I_{K \times K}$. Therefore,

$$\Omega_Z^\pi \overset{(1)}{=} \tfrac{1}{2} B(1, 2) \Omega_Z B(1, 2)' + \tfrac{1}{2} B((2, 1)) \Omega_Z B((2, 1))' \overset{(2)}{=} \Omega_Z \overset{(3)}{=} \Sigma_Z,$$

where (1) holds by the definition of Ω_Z^π, (2) holds by the definition of $B(\pi)$, and (3) follows by $\Sigma_Z \equiv var(Z_i) = E[Z_i Z_i'] \equiv \Omega_Z$ under H_0 in (2.1).

Theorem A.1 then implies that $P(\hat{Z}^{\boldsymbol{\pi}} \leq x | \mathbf{X}_n) \overset{p}{\to} P(\xi \leq x)$ for all x such that $P(\xi \leq x)$ is continuous, where $\xi \sim N(0_{(T-1)K \times 1}, \Sigma_Z)$. Then, the continuous mapping theorem from Chung and Romano (2016, Lemma A.6) implies that for non-studentized statistic:

$$P(S_n^{\boldsymbol{\pi}} \leq x | \mathbf{X}_n) \overset{p}{\to} P(S \leq x), \tag{A.40}$$

for all $x \in \mathbb{R}$, where S is as in (2.10). This convergence relies on the fact that $\Sigma_Z \neq 0_{T(K-1) \times T(K-1)}$, which implies $P(S \leq x)$ is continuous for all $x \in \mathbb{R}$. From this point onward, the rest of the proof follows from arguments in part (a) of Theorem 3.1.

Case 2: $\Sigma_Z = 0_{T(K-1) \times T(K-1)}$. By the same arguments as in Theorem 3.1, we have that $1(X_{i,t} \leq u_k) = 1(X_{i,t+1} \leq u_k)$ a.s. for all $t = 1, \ldots, T-1$ and $k = 1, \ldots K$, and so $S_n = 0$ a.s. Furthermore, for all $\pi \in \mathcal{M}$, we have that $1(X_{i,\pi(t)} \leq u_k) = 1(X_{i,\pi(t+1)} \leq u_k)$ a.s. for all $t = 1, \ldots, T-1$, $k = 1, \ldots K$. This implies that $S_n^{\boldsymbol{\pi}} = 0$ a.s., and so $S_n = c_n^\pi(1 - \alpha) = 0$. The desired result follows from this and the construction of the test in (3.13).

Part (b). We construct an example with $T = 3$. For $\tau_1, \tau_2 \in (0, 1)$, we focus on a Markov chain with two states, $s_1 = 0$ and $s_2 = 1$, and a transition matrix given by

$$\begin{pmatrix} P(X_{i,t+1} = 0 | X_{i,t} = 0) & P(X_{i,t+1} = 1 | X_{i,t} = 0) \\ P(X_{i,t+1} = 0 | X_{i,t} = 1) & P(X_{i,t+1} = 1 | X_{i,t} = 1) \end{pmatrix} = \begin{pmatrix} \tau_1 & 1 - \tau_1 \\ 1 - \tau_2 & \tau_2 \end{pmatrix}.$$

In the steady state, the marginal distribution is such that for all $t = 1, 2, 3$,

$$P(X_{i,t} = 0) = \frac{1 - \tau_2}{2 - \tau_1 - \tau_2} \quad \text{and} \quad P(X_{i,t} = 1) = \frac{1 - \tau_1}{2 - \tau_1 - \tau_2}. \tag{A.41}$$

To assess the marginal homogeneity of this Markov chain on only two support points, it suffices to test the hypothesis at one of the two points (as the other is just its complement). For this reason, we construct

30

our test statistic with $K = 1$ and $u_1 = 0$. It follows then

$$\Omega_Z = \frac{(1-\tau_1)(1-\tau_2)^2}{(2-\tau_1-\tau_2)^2} \begin{bmatrix} 2 & -(2-(\tau_1+\tau_2)) \\ -(2-(\tau_1+\tau_2)) & 2 \end{bmatrix}.$$

By $T = 3$, the $T! = 6$ permutations of $(1,2,3)$ are $\mathcal{M} = \{(1,2,3),(1,3,2),(2,1,3),(2,3,1),(3,1,2),(3,2,1)\}$. Following the construction in step 2 of Theorem A.1, we have

$$B((1,2,3)) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B((1,3,2)) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}, \quad B((2,1,3)) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$B((2,3,1)) = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \quad B((3,1,2)) = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}, \quad B((3,2,1)) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix},$$

and

$$\Omega_Z^{\pi} = \tfrac{1}{6} \sum_{\pi \in \mathcal{M}} B(\pi) \Omega_Z B(\pi)'.$$

It is not hard to verify that Ω_Z^{π} is PD and $\Omega_Z^{\pi} \neq \Omega_Z$. By the same arguments as in part (a), we conclude that

$$P(S_n^{\pi} \leq x | \mathbf{X}_n) \overset{p}{\to} P(S^{\pi} \leq x), \tag{A.42}$$

for all $x \in \mathbb{R}$, where $S^{\pi} = \sum_{j=1}^{2} \lambda_j^{\pi} \zeta_j^2$ with $\{\zeta_j\}_{j=1}^{2}$ being i.i.d. $N(0,1)$, and $\{\lambda_j^{\pi}\}_{j=1}^{2}$ are the eigenvalues of Ω_Z^{π}. From this point onward, we can repeat arguments in part (a) of Theorem 3.1 to show that

$$c_n^{\pi}(1-\alpha) \overset{p}{\to} c^{\mathcal{X}}(1-\alpha; \Omega_Z^{\pi}) \quad \text{and} \quad E_P[\phi_n^{\pi}(\alpha)] \to P[S \leq c^{\mathcal{X}}(1-\alpha; \Omega_Z^{\pi})], \tag{A.43}$$

where $S = \sum_{j=1}^{2} \lambda_j \zeta_j^2$ with $\{\lambda_j\}_{j=1}^{2}$ equal to the eigenvalues of $\Sigma_Z = \Omega_Z$. Since $\Omega_Z^{\pi} \neq \Omega_Z$, we have that $c^{\mathcal{X}}(1-\alpha; \Omega_Z^{\pi}) \neq c^{\mathcal{X}}(1-\alpha; \Omega_Z)$ and therefore $P[S \leq c^{\mathcal{X}}(1-\alpha; \Omega_Z^{\pi})] \neq \alpha$. To show the asymptotic overrejection, it suffices to find examples of parameters in which $P[S \leq c^{\mathcal{X}}(1-\alpha; \Omega_Z^{\pi})] > \alpha$. For instance, by choosing $\tau_1 = \tau_2 = 0.1$, we obtain

$$c^{\mathcal{X}}(0.9; \Omega_Z^{\pi}) = 1.56 < c^{\mathcal{X}}(0.9; \Sigma_Z) = 2.36,$$
$$c^{\mathcal{X}}(0.95; \Omega_Z^{\pi}) = 2.12 < c^{\mathcal{X}}(0.95; \Sigma_Z) = 3.33,$$
$$c^{\mathcal{X}}(0.99; \Omega_Z^{\pi}) = 3.49 < c^{\mathcal{X}}(0.99; \Sigma_Z) = 5.72,$$

with the following asymptotic overrejection:

$$E_P[\phi_n^{\pi}(0.1)] \to 0.1828 > 0.1, \quad E_P[\phi_n^{\pi}(0.05)] \to 0.1194 > 0.05, \quad E_P[\phi_n^{\pi}(0.01)] \to 0.0446 > 0.01.$$

Part (c). Under H_0, we have $\Omega_Z = \Sigma_Z$. Then, Assumption 2 implies Ω_Z is PD. Thus, $B(\pi)\Omega_Z B(\pi)'$ is PSD for all $\pi \in \mathcal{M}$, where $B(\pi)$ is defined in the proof of Theorem A.1. Furthermore, by choosing $\pi = (1,\ldots,T)$, step 2 of Theorem A.1 implies that $B((1,\ldots,T)) = I_{(T-1)K \times (T-1)K}$, and so $B((1,\ldots,T))\Omega_Z B((1,\ldots,T))' = \Omega_Z$. Then, $\Omega_Z^{\pi} = \frac{1}{T!} \sum_{\pi \in \mathcal{M}} B(\pi)\Omega_Z B(\pi)'$ is PD, as it is the sum of PSD matrices with at least one PD matrix.

By Theorem A.1, Lemma A.1, and the fact that Ω_Z^{π} is PD, the Slutsky's theorem from Chung and Romano (2016, Lemma A.5) implies that $P(\bar{S}_n^{\pi} \leq x | \mathbf{X}_n) \overset{p}{\to} P(\chi^2_{(T-1)K} \leq x)$ for all $x \in \mathbb{R}$. From this point

31

onward, the rest of the proof follows from arguments in part (a) in Theorem 3.1. ∎

Proof of Theorem 3.4. Part (a). Fix $(i,t,s) \in \{1, \ldots, n\} \times \{1, \ldots, T\} \times \{1, \ldots, T\}$ arbitrarily, and let $\pi \in \mathcal{M}$ be any permutation that interchanges t and s. For any $x \in \mathbb{R}$,

$$
\begin{aligned}
F_t(x) &= P(X_{i,t} \leq x) \\
&= \lim_{\{u_j\}_{j \neq t} \to \infty} P(\{X_{i,1}, \ldots, X_{i,t}, \ldots, X_{i,s}, \ldots, X_{i,T}\} \leq (u_1, \ldots, x, \ldots, u_s, \ldots, u_T)) \\
&\stackrel{(1)}{=} \lim_{\{u_j\}_{j \neq t} \to \infty} P(\{X_{i,\pi(1)}, \ldots, X_{i,\pi(t)}, \ldots, X_{i,\pi(s)}, \ldots, X_{i,\pi(T)}\} \leq (u_1, \ldots, x, \ldots, u_s, \ldots, u_T)) \\
&\stackrel{(2)}{=} \lim_{\{u_j\}_{j \neq t} \to \infty} P(\{X_{i,\pi(1)}, \ldots, X_{i,s}, \ldots, X_{i,t}, \ldots, X_{i,\pi(T)}\} \leq (u_1, \ldots, x, \ldots, u_s, \ldots, u_T)) \\
&= P(X_{i,s} \leq x) = F_s(x),
\end{aligned}
$$

where (1) holds by $P \in \Omega_{\mathrm{TE}}$ and (2) by the specification of π. Since $x \in \mathbb{R}$ and $(t,s) \in \{1, \ldots, T\} \times \{1, \ldots, T\}$ were arbitrary, H_0 in (2.1) holds.

To see that the reverse implication fails, consider the following example: $\mathbf{X}_n = \{(X_{i,1}, X_{i,2}, X_{i,3})\}_{i=1}^n$ i.i.d. with $X_{i,1} = X_{i,2} \perp X_{i,3}$ and $X_{i,t} \sim N(0,1)$. It is not hard to verify that this distribution satisfies Assumption 1 but does not belong to Ω_{TE}.

Part (b). Let $\mathbf{X}_n^\pi \equiv \{\{X_{i,\pi_i(t)}\}_{t=1}^T\}_{i=1}^n$ denote the sample permuted according to an arbitrary permutation $\pi^n = \{\pi_i\}_{i=1}^n \in \mathcal{M}^n$. Then,

$$
F_{\mathbf{X}_n} \stackrel{(1)}{=} \prod_{i=1}^n F_{X_{i,1}, \ldots, X_{i,T}} \stackrel{(2)}{=} \prod_{i=1}^n F_{X_{i,\pi_i(1)}, \ldots, X_{i,\pi_i(T)}} \stackrel{(3)}{=} F_{\mathbf{X}_n^\pi}, \tag{A.44}
$$

where (1) and (3) hold by Assumption 1, and (2) by $P \in \Omega_{\mathrm{TE}}$. We note that (A.44) implies that the randomization hypothesis (i.e., Lehmann and Romano (2022, Definition 17.2.1)) holds. From here, Lehmann and Romano (2022, Theorem 17.2.1) implies that the permutation test described in Lehmann and Romano (2022, Section 17.2.1) satisfies (3.15) with equality. In turn, this implies that our permutation test (i.e., the non-random version of the test in Lehmann and Romano (2022, Section 17.2.1)) satisfies (3.15). ∎

Lemma A.1. Under Assumption 1,

$$
\hat{\Sigma}_{Z^\pi} \xrightarrow{p} \Omega_Z^\pi.
$$

Proof. This proof relies on notation and arguments in the proof of Theorem A.1. First, we show that

$$
\frac{1}{n} \sum_{i=1}^n B(\boldsymbol{\pi}_i) Z_i Z_i' B(\boldsymbol{\pi}_i)' \xrightarrow{p} \Omega_Z^\pi. \tag{A.45}
$$

By similar arguments as in step 3 of Theorem A.1, we note that $\{B(\boldsymbol{\pi}_i) Z_i Z_i' B(\boldsymbol{\pi}_i)'\}_{i=1}^n$ is i.i.d. with

$$
E[B(\boldsymbol{\pi}_i) Z_i Z_i' B(\boldsymbol{\pi}_i)'] \stackrel{(1)}{=} E[B(\boldsymbol{\pi}_i) \Omega_Z B(\boldsymbol{\pi}_i)'] \stackrel{(2)}{=} \Omega_Z^\pi.
$$

where (1) holds by $\{B(\boldsymbol{\pi}_i)\}_{i=1}^n \perp \{Z_i\}_{i=1}^n$ and $E[Z_i Z_i'] = \Omega_Z$, and (2) by (A.23) and the fact that $\boldsymbol{\pi}_i$ is uniformly distributed in $\mathcal{M}$. From these observations and the LLN, (A.45) follows.

Second, we note that

$$\frac{1}{n}\sum_{i=1}^{n} B(\boldsymbol{\pi}_i)Z_i V_i' B(\boldsymbol{\pi}_i)' = O_p(1), \quad \frac{1}{n}\sum_{i=1}^{n} B(\boldsymbol{\pi}_i)V_i Z_i' B(\boldsymbol{\pi}_i)' = O_p(1), \text{ and } \quad \frac{1}{n}\sum_{i=1}^{n} B(\boldsymbol{\pi}_i)V_i V_i' B(\boldsymbol{\pi}_i)' = O_p(1).$$

$$(A.46)$$

These can be shown by the arguments that yield (A.45), except that $Z_i Z_i'$ is replaced by $Z_i V_i'$, $V_i Z_i'$, and $V_i V_i'$, respectively.

To conclude the proof, consider the following derivation.

$$\hat{\Sigma}_{Z^{\boldsymbol{\pi}}} \overset{(1)}{=} \frac{1}{n}\sum_{i=1}^{n}\hat{M}V_i^{\boldsymbol{\pi}}(\hat{M}V_i^{\boldsymbol{\pi}})' - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{M}V_i^{\boldsymbol{\pi}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{M}V_i^{\boldsymbol{\pi}}\right)'$$

$$\overset{(2)}{=} \hat{M}\left(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i V_i' B(\boldsymbol{\pi}_i)'\right)\hat{M}' - \hat{M}\left(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i\right)\left(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i\right)'\hat{M}'$$

$$\overset{(3)}{=} (M + o_p(1))\left(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i V_i' B(\boldsymbol{\pi}_i)'\right)(M + o_p(1))' + o_p(1)$$

$$\overset{(4)}{=} \left[\begin{array}{c} \frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)Z_i Z_i' B(\boldsymbol{\pi}_i)' + (\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)Z_i V_i' B(\boldsymbol{\pi}_i)')o_p(1) + \\ o_p(1)(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i V_i' B(\boldsymbol{\pi}_i)') + o_p(1)(\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i Z_i' B(\boldsymbol{\pi}_i)')o_p(1) \end{array}\right] + o_p(1)$$

$$\overset{(5)}{=} \Omega_Z^{\boldsymbol{\pi}} + o_p(1),$$

as desired, where (1) holds by (A.25), (A.26), and (A.27), (2) by (A.28), (3) by (A.38) and $\frac{1}{n}\sum_{i=1}^{n}B(\boldsymbol{\pi}_i)V_i = o_p(1)$ (implied by (A.37)), (4) by $Z_i = MV_i$ for all $i = 1, \ldots, n$ and (A.39), and (5) by (A.45) and (A.46). ∎

Lemma A.2. Let Ω_{FE} denote the class of distributions that are "fully" exchangeable over units and time periods, i.e., $\mathbf{X}_n = \{\{X_{i,t}\}_{i=1}^{n}\}_{t=1}^{T}$ has the same distribution as $\{\{X_{\lambda(i,t)}\}_{i=1}^{n}\}_{t=1}^{T}$ where λ denotes an arbitrary permutation of units and time periods. Under Assumption 1, the following statements hold:

(a) If $n = 1$, $\Omega_{\text{FE}} = \Omega_{\text{TE}}$.

(b) If $n > 1$, $\Omega_{\text{FE}} \subsetneq \Omega_{\text{TE}}$.

Proof. Part (a) is straightforward, so we prove part (b). We begin by showing a useful intermediate result: $\mathbf{X}_n \sim P \in \Omega_{\text{FE}}$ implies that its CDF can be written as

$$P(\mathbf{X}_n \le y) = \prod_{i=1}^{n}\prod_{t=1}^{T} F(y_{i,t}) \quad \text{for all } y \in \mathbb{R}^{nT}, \tag{A.47}$$

where F is the CDF of $X_{1,1}$. Since Assumption 1 already implies independence across units:

$$P(\mathbf{X}_n \le y) = \prod_{i=1}^{n} \tilde{F}(\{y_{i,t}\}_{t=1}^{T}), \tag{A.48}$$

where $\tilde{F}$ is the CDF of the vector $\{X_{1,t}\}_{t=1}^{T}$. Then the desired result (A.47) follows immediately from (A.48) provided that $\{X_{1,t}\}_{t=1}^{T}$ are i.i.d. with marginal CDF F. We now establish this result in two steps.

First, we show $\{X_{1,t}\}_{t=1}^T$ is an independent sequence. To this end, fix $\{x_t\}_{t=1}^T \in \mathbb{R}^T$ arbitrarily. For any $s = 1, \ldots, T-1$, consider the following permutation: $\lambda(1,t) = (1,t)$ for $t \le s$ and $\lambda(1,t) = (2,t)$ for $t > s$, $\lambda(2,t) = (2,t)$ for $t \le s$ and $\lambda(2,t) = (1,t)$ for $t > s$, and $\lambda(j,t) = (j,t)$ for all $j > 2$ and $t = 1, \ldots, T$. Then,

$$
\begin{aligned}
&P(X_{1,1} \le x_1, \ldots, X_{1,s} \le x_s, X_{1,s+1} \le x_{s+1}, \ldots, X_{1,T} \le x_T) \\
&\stackrel{}{=} \lim_{u \to \infty} P\left(\begin{array}{c} X_{1,1} \le x_1, \ldots, X_{1,s} \le x_s, X_{1,s+1} \le x_{s+1}, \ldots, X_{1,T} \le x_T \\ X_{j,t} \le u \text{ for all } j \ge 2 \text{ and } t = 1, \ldots, T \end{array} \right) \\
&\stackrel{(1)}{=} \lim_{u \to \infty} P\left(\begin{array}{c} X_{\lambda(1,1)} \le x_1, \ldots, X_{\lambda(1,s)} \le x_s, X_{\lambda(1,s+1)} \le x_{s+1}, \ldots, X_{\lambda(1,T)} \le x_T \\ X_{\lambda(j,t)} \le u \text{ for all } j \ge 2 \text{ and } t = 1, \ldots, T \end{array} \right) \\
&\stackrel{(2)}{=} \lim_{u \to \infty} P\left(\begin{array}{c} X_{1,1} \le x_1, \ldots, X_{1,s} \le x_s, X_{2,s+1} \le x_{s+1}, \ldots, X_{2,T} \le x_T \\ X_{\lambda(j,t)} \le u \text{ for all } j \ge 2 \text{ and } t = 1, \ldots, T \end{array} \right) \\
&= P(X_{1,1} \le x_1, \ldots, X_{1,s} \le x_s, X_{2,s+1} \le x_{s+1}, \ldots, X_{2,T} \le x_T) \\
&\stackrel{(3)}{=} P(X_{1,1} \le x_1, \ldots, X_{1,s} \le x_s) P(X_{2,s+1} \le x_{s+1}, \ldots, X_{2,T} \le x_T), \quad\quad\quad\quad\quad (A.49)
\end{aligned}
$$

where (1) holds by $P \in \Omega_{\mathrm{FE}}$, (2) by the specification of λ, and (3) by Assumption 1. By taking limits of (A.49) as $x_1, \ldots, x_s \to \infty$, we get

$$
P(X_{1,s+1} \le x_{s+1}, \ldots, X_{1,T} \le x_T) = P(X_{2,s+1} \le x_{s+1}, \ldots, X_{2,T} \le x_T). \quad\quad (A.50)
$$

Since (A.50) holds for all $\{x_t\}_{t=1}^T \in \mathbb{R}^T$, we can combine it with (A.49) to get

$$
(X_{1,1}, \ldots, X_{1,s}) \perp (X_{1,s+1}, \ldots, X_{1,T}) \text{ for any } s = 1, \ldots, T-1. \quad\quad (A.51)
$$

The desired result follows by considering (A.51) sequentially for $s = 1$, $s = 2$, and so on.

Second, we show that $\{X_{1,t}\}_{t=1}^T$ is an identically distributed sequence. To this end, fix $x \in \mathbb{R}$ arbitrarily. For any $s \ne 1$, consider the following permutation: $\lambda(1,1) = (1,s)$, $\lambda(1,s) = \lambda(1,1)$, and $\lambda(i,t) = (i,t)$ otherwise. Then,

$$
\begin{aligned}
P(X_{1,1} \le x) &= \lim_{u \to \infty} P(X_{1,1} \le x, X_{i,t} \le u \text{ for all } (i,t) \ne (1,1)) \\
&\stackrel{(1)}{=} \lim_{u \to \infty} P(X_{\lambda(1,1)} \le x, X_{\lambda(i,t)} \le u \text{ for all } (i,t) \ne (1,1)) \\
&\stackrel{(2)}{=} \lim_{u \to \infty} P(X_{1,s} \le x, X_{\lambda(i,t)} \le u \text{ for all } (i,t) \ne (1,1)) \\
&= P(X_{1,s} \le x), \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (A.52)
\end{aligned}
$$

where (1) holds by $P \in \Omega_{\mathrm{FE}}$ and (2) by the specification of λ. Since (A.52) holds for all $x \in \mathbb{R}$, $X_{1,1}$ and $X_{1,s}$ have the same distribution. Since the choice of $s = 2, \ldots, T$ was arbitrary, the desired result follows.

Finally, we conclude the proof by finding a distribution $P \in \Omega_{\mathrm{TE}}$ but $P \notin \Omega_{\mathrm{FE}}$ so the inclusion is strict. Consider the following example: $\mathbf{X}_n = \{\{X_{i,t}\}_{i=1}^n\}_{t=1}^T$ with $n = 2$, $T = 2$, where $X_{1,1} = X_{1,2} = Z_1$, $X_{2,1} = X_{2,2} = Z_2$, and $\{Z_1, Z_2\}$ are i.i.d. $N(0,1)$. It is trivial to see that this distribution satisfies Assumption 1 and $P \in \Omega_{\mathrm{TE}}$, however $P \notin \Omega_{\mathrm{FE}}$. $\blacksquare$

Bibliography

BUGNI, F. A., J. BUNTING, AND T. URA (2024): "Testing homogeneity in dynamic discrete games in finite samples," *arXiv preprint arXiv:2010.02297*.

BUGNI, F. A. AND J. L. HOROWITZ (2021): "Permutation tests for equality of distributions of functional data," *Journal of Applied Economics*, 36, 861–877.

CHUNG, E. AND M. OLIVARES (2021): "Permutation test for heterogeneous treatment effects with a nuisance parameter," *Journal of Econometrics*, 225, 148–174.

CHUNG, E. AND J. ROMANO (2013): "Exact and Asymptotically Robust Permutation Tests," *The Annals of Statistics*, 41, 484–507.

——— (2016): "Multivariate and multiple permutation tests," *Journal of Econometrics*, 193, 76–91.

DAVIDSON, J. (1994): *Stochastic Limit Theory*, Oxford University Press.

DICICCIO, C. J. AND J. P. ROMANO (2017): "Robust permutation tests for correlation and regression coefficients," *Journal of the American Statistical Association*, 112, 1211–1220.

DITZHAUS, M. AND D. GAIGALL (2022): "Testing marginal homogeneity in Hilbert spaces with applications to stock market returns," *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 31, 749–770.

FRIEDRICH, S., E. BRUNNER, AND M. PAULY (2017): "Permuting longitudinal data in spite of the dependencies," *Journal of Multivariate Analysis*, 153, 255–265.

GAIGALL, D. (2020): "Testing marginal homogeneity of a continuous bivariate distribution with possibly incomplete paired data," *Metrika*, 83, 437–465.

IGAMI, M. AND N. YANG (2016): "Unobserved heterogeneity in dynamic games: Cannibalization and preemptive entry of hamburger chains in Canada," *Quantitative Economics*, 7, 483–521.

JANSSEN, A. (1997): "Studentized permutation tests for non-iid hypotheses and the generalized Behrens-Fisher problem," *Statistics & probability letters*, 36, 9–21.

KHMALADZE, E. (2016): "Unitary transformations, empirical processes and distribution free testing," *Bernoulli*, 22, 563–588.

LEHMANN, E. L. AND J. P. ROMANO (2022): *Testing Statistical Hypothesis: Fourth edition*, Springer.

OTSU, T., M. PESENDORFER, AND Y. TAKAHASHI (2016): "Pooling data across markets in dynamic Markov Games," *Quantitative Economics*, 7, 523–559.

PAULY, M., E. BRUNNER, AND F. KONIETSCHKE (2015): "Asymptotic permutation tests in general factorial designs," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 77, 461–473.

SCHEFFE, H. (1959): *The Analysis of Variance*, John Wiley & Sons.

VAN DER VAART, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

VAN DER VAART, A. AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*, Springer.