Testing homogeneity in dynamic discrete games in finite samples*

Federico A. Bugni
Department of Economics
Duke University
federico.bugni@duke.edu

Jackson Bunting
Department of Economics
Duke University
jackson.bunting@duke.edu

Takuya Ura
Department of Economics
University of California, Davis
takura@ucdavis.edu

April 5, 2021

Abstract

The literature on dynamic discrete games often assumes that the conditional choice probabilities and the state transition probabilities are homogeneous across markets and over time. We refer to this as the "homogeneity assumption" in dynamic discrete games. This homogeneity assumption enables empirical studies to estimate the game's structural parameters by pooling data from multiple markets and from many time periods. In this paper, we propose a hypothesis test to evaluate whether the homogeneity assumption holds in the data. Our hypothesis test is the result of an approximate randomization test, implemented via a Markov chain Monte Carlo (MCMC) algorithm. We show that our hypothesis test becomes valid as the (user-defined) number of MCMC draws diverges, for any fixed number of markets, time periods, and players. We apply our test to the empirical study of the U.S. Portland cement industry in Ryan (2012).

Keywords and phrases: Dynamic discrete choice problems, dynamic games, Markov decision problem, randomization tests, Markov chain Monte Carlo (MCMC).

JEL classification: C12, C57, C63

*We thank Ivan Canay and seminar participants at Duke University and Toulouse School of Economics for helpful comments and discussion.

1 Introduction

In applications of dynamic discrete games, practitioners often assume that the conditional choice probabilities and the state transition probabilities are invariant across time and markets.[1] We refer to this as the "homogeneity assumption" in dynamic discrete games. This is a convenient assumption, as it allows the estimation of the model's structural parameters by pooling data from multiple markets and from many time periods.

Despite the widespread use of the homogeneity assumption in dynamic discrete games, it is plausible for this condition to fail in applications. We now provide a few examples. First, a game could suffer from a structural change or equilibrium changes due to "sunspots", which would invalidate the homogeneity assumption across time. Second, markets could be affected by permanent time-invariant heterogeneity that is observed to the players but not to the econometrician (e.g., Arcidiacono and Miller, 2011). This would invalidate the homogeneity assumption across markets. Third and relatedly, there may be multiplicity of equilibria, and different markets could be playing different equilibria. The literature has considered hypothesis testing for the multiplicity of equilibria in games. In particular, de Paula and Tang (2012) propose a test for the multiplicity of equilibria across markets in static games, while Otsu, Pesendorfer, and Takahashi (2016) consider a test for the multiplicity of equilibria across markets in dynamic games.

In this paper, we propose a hypothesis test for the homogeneity assumption. That is, our test is designed to capture the various possible violations to the homogeneity assumption described in the previous paragraph (both across markets and over time). Our test is implemented via Markov chain Monte Carlo (MCMC) methods, and it is justified by the theory of randomization tests (cf. Lehmann and Romano, 2005, Section 15.2). While our test is not exactly a randomization test, we establish its validity by coupling it with an underlying randomization test. The latter is exactly valid yet computationally infeasible. In this sense, we can interpret our proposed MCMC algorithm as a computationally feasible way to implement the infeasible randomization test. We formally show that the approximation error vanishes as the (user-defined) number of MCMC draws diverges. It is worth mentioning that our results hold for any fixed and finite number of players, markets, and time periods. This is an important aspect of our contribution, as the datasets used in empirical applications often have a small number of time periods and markets. For example, our empirical application is based on Ryan (2012), and has only $n = 23$ markets and either $T = 9$ or $T = 10$ time periods.

The econometric framework considered in this paper is arguably very general. It includes the single-agent dynamic discrete choice model (e.g., Rust (1987); Hotz and Miller (1993); Hotz, Miller, Sanders, and Smith (1994); Aguirregabiria and Mira (2002)) and the Markov equilibrium dynamic game model (e.g., Pakes, Ostrovsky, and Berry (2007); Aguirregabiria and Mira (2007);

[1] In this paper, we use "market" to denote a cross-sectional unit.

Bajari, Benkard, and Levin (2007); Pesendorfer and Schmidt-Dengler (2008, 2010)). Furthermore, it includes the Markov dynamic game model of Aguirregabiria and Magesan (2020), which allows some players to have biased beliefs. Importantly, our econometric framework does not impose any parametric structure on the model. In this sense, our test is non-parametric. We consider this feature to be an key advantage of our methodology. Our test's rejection of the homogeneity assumption cannot be influenced by the possible misspecification of functional forms imposed by the researcher. In addition, our test can be implemented without estimating any structural parameters.

In a recent paper, Otsu et al. (2016) propose several hypothesis tests for dynamic discrete games. Two of their proposals are directly related to the problem considered in our paper.[2] Specifically, they consider a method to test the homogeneity across markets of the conditional choice probabilities and the state transition probabilities, under the maintained assumption that these functions are time-homogeneous. Their inference method is based on the bootstrap, and its validity is shown in an asymptotic framework in which the number of time periods T diverges to infinity. Unfortunately, the number of time periods in applications is often small. Besides the aforementioned application of Ryan (2012) with $T = 9$ or $T = 10$, we can mention Sweeting (2013) with $T = 4$, Collard-Wexler (2013) with $T = 24$, and Dunne, Klimek, Roberts, and Xu (2013) with $T = 5$.

The most critical step of our MCMC algorithm is based on the so-called Euler algorithm (Kandel, Yossi, Unger, and Winkler, 1996). In related work, Besag and Mondal (2013) use this algorithm to test whether a time series of data has a time-homogeneous Markov structure. In terms of our setup, this corresponds to testing whether the data from a single market has a time-homogeneous state transition probability. Relative to this work, our paper incorporates several essential features of dynamic Markov discrete games. First, we recognize that the dataset in a typical dynamic game has information about actions and states. Second, our construction exploits the typical economic structure imposed in dynamic games, such as the conditional independence assumption (i.e., conditional on the current state variable, the current action variable is independent from the past information). This can be clearly evidenced in our MCMC algorithm, where we first transform the state variable data and then we transform the action variable data. Finally, while Besag and Mondal (2013) focus on data from a single market, our MCMC algorithm exploits the possibility that the data includes observations from multiple markets. This is an important aspect of our contribution, as the datasets used in empirical applications usually include data multiple markets, e.g., Ryan (2012) with $n = 23$, Sweeting (2013) with $n = 102$, Collard-Wexler (2013) with $n = 1,600$, and Dunne et al. (2013) with $n = 639$. In Section 5 of Besag and Mondal (2013), the authors briefly mention that their methods could be extended to the multiple market case. However, they do not explain how this can be implemented nor do they justify its validity. In contrast, our test uses a different procedure than theirs, incorporates the essential features of dynamic Markov discrete

[2] The other two testing methodologies are less related to our paper. One test assumes that the state distribution is in its steady state. This condition is not commonly imposed in the literature, and our test does not require it. The other test they propose is based on the frequencies of states conditional on the state distribution in the first period.

games, and we formally prove its validity by connecting it with the theory of randomization tests.

We explore the performance of our hypothesis test in Monte Carlo simulations. Our results show that our method provides excellent size control even in small samples, and can successfully detect relatively small deviations from the homogeneity hypothesis. In these two accounts, our test appears to work favorably in comparison with the bootstrap-based test in Otsu et al. (2016). As an empirical application, we investigate the homogeneity of the decisions in the U.S. Portland cement industry data used in Ryan (2012). This is a key assumption in Ryan (2012), as it allows him to pool data from multiple markets to estimate the model's parameters. Unlike Otsu et al. (2016)'s test, our test finds no evidence against the homogeneity hypothesis in the data.

The rest of the paper is organized as follows. Section 2 describes the dynamic discrete choice model and the hypothesis test. Section 3 specifies our hypothesis test and its implementation via the MCMC algorithm. In Section 4, we show that our test is an approximate implementation of a computationally infeasible randomization test. In Section 5, we evaluate the performance of our test in finite samples via Monte Carlo simulations. Section 6 considers an empirical application based on Ryan (2012). Section 7 concludes. The paper's appendix collects all of the proofs, several auxiliary results, and computational details related to the proposed MCMC algorithm.

2 The econometric model and the testing problem

2.1 The econometric model

We begin by describing the dynamic discrete game under consideration. We observe the outcome of n markets in which J players choose actions over T time periods. Our setup allows for $J = 1$, i.e., single-agent problems, or $J > 1$, i.e., multiple-agent games. This paper's inference results are valid for all finite n, T, and J.

We consider a setup in which the observed actions and state variables are discretely distributed, which is common in the dynamic discrete choice literature. For every market $i = 1, \ldots, n$ and period $t = 1, \ldots, T$, let $A_{i,t}$ be the random variable that specifies the actions chosen by the players in market i and period t, and let $S_{i,t}$ be the random variable that specifies the state variable of market i and period t. We define the following $n \times T$ matrices:

$$S \equiv (S_{i,t} : i = 1, \ldots, n, t = 1, \ldots, T),$$
$$A \equiv (A_{i,t} : i = 1, \ldots, n, t = 1, \ldots, T).$$

In this notation, the data are then given by

$$X \equiv (S, A).$$

We assume the common support of $S_{i,t}$ is a finite set $\mathcal{S}$, and the common support of $A_{i,t}$ is a finite set $\mathcal{A}$. Then, the support of X is represented by $\mathcal{X} \equiv \mathcal{S}^{nT} \times \mathcal{A}^{nT}$.

Remark 2.1. We have thus far assumed that we observe a balanced panel, i.e., all n markets are fully observed over T time periods. This is only for the simplicity of notation and exposition. All of the arguments in our paper extend immediately to the case in which each market $i = 1, \ldots, n$ is observed over a market-specific time period T_i. ■

Throughout this paper, we maintain the following assumption.

Assumption 2.1. The following conditions hold:

(a) $((S_{i,t}, A_{i,t}) : t = 1, \ldots, T)$ are independent across $i = 1, \ldots n$.

(b) $(S_{i,t}, A_{i,t})$ and $(S_{i,1}, A_{i,1}, \ldots, S_{i,t-2}, A_{i,t-2})$ are conditionally independent given $(S_{i,t-1}, A_{i,t-1})$ for every $i = 1, \ldots, n$ and $t = 3, \ldots T$.

(c) $A_{i,t}$ and $(S_{i,t-1}, A_{i,t-1})$ are conditionally independent given $S_{i,t}$ for every $i = 1, \ldots, n$ and $t = 2, \ldots T$.

Assumption 2.1 is standard in much of the literature on dynamic discrete games. Assumption 2.1(a) imposes that markets are independently distributed. Assumption 2.1(b) indicates that the observations of state and actions are a Markov process. Assumption 2.1(c) imposes that the current actions are independent of past information once we condition on the current state. Assumptions 2.1(b)-(c) are high-level restrictions that are typically imposed on the equilibrium strategies used by the players. In particular, they follow from the assumption that players use Markov strategies Maskin and Tirole (2001), as assumed in Pakes et al. (2007); Aguirregabiria and Mira (2007); Bajari et al. (2007); Pesendorfer and Schmidt-Dengler (2008). These conditions are implied even in models in which the players' beliefs are allowed to be out of equilibrium, i.e., do not coincide with the true equilibrium probabilities (e.g., Aguirregabiria and Magesan (2020)).

Assumption 2.1 is the only condition imposed in this paper. The assumption does not impose any parametric structure on the model. In this sense, our test is nonparametric. We view this as an important advantage of our methodology. Our test cannot be affected by the possible misspecification of functional forms imposed by the researcher.

We now introduce necessary notation to express our hypothesis of interest. We use $\sigma_{i,t}$ to denote the conditional choice probability for market i and period t, i.e., for every $(s, a) \in \mathcal{S} \times \mathcal{A}$,

$$\sigma_{i,t}(a|s) \equiv P(A_{i,t} = a|S_{i,t} = s).$$

We use $g_{i,t+1}$ to denote the state transition probability from period t to $t+1$ for market i, i.e., for every $(s, a, s') \in \mathcal{S} \times \mathcal{A} \times \mathcal{S}$,

$$g_{i,t+1}(s'|a, s) \equiv P(S_{i,t+1} = s'|(S_{i,t}, A_{i,t}) = (s, a)).$$

We use $m_i(s)$ to denote the marginal state distribution for market i in period 1, i.e., for every $s \in \mathcal{S}$,

$$m_i(s) \ \equiv \ P(S_{i,1} = s).$$

With this notation in place, we specify our hypothesis testing problem in the next section.

2.2 The hypothesis testing problem

Our goal is to test whether the "homogeneity assumption" holds in the data, i.e., whether the conditional choice probabilities and state transition probabilities are homogeneous across time and markets. That is,

$$H_0 : \sigma_{i,t}(a|s) = \sigma(a|s) \ \text{ and } \ g_{i,t+1}(s'|a,s) = g(s'|a,s) \quad \text{vs.} \quad H_1 : H_0 \text{ is false.} \tag{1}$$

Note that H_0 in (1) represents two types of homogeneity: time and market homogeneity, and involves two functions: conditional choice probabilities and state transition probabilities. In this sense, our hypothesis test evaluates four homogeneity conditions: time homogeneity of the conditional choice probabilities, market homogeneity of the conditional choice probabilities, time homogeneity of the state transition probabilities, and market homogeneity of the state transition probabilities. A rejection of H_0 in (1) would be indicative that one or more of these homogeneity conditions is violated. In certain applications, however, one may feel comfortable that some of the conditions are satisfied and should be part of our maintained assumptions. For example, in a given application, one may be confident that the conditional choice probability and state transition probability are time-homogeneous. Then, one could reinterpret H_0 in (1) as testing the market homogeneity of the conditional choice probabilities and state transition probabilities.

Under Assumption 2.1 and H_0 in (1), Lemma B.1 in the appendix shows that the likelihood of the data $X = (S, A)$ evaluated at any realization $\tilde{X} = (\tilde{S}, \tilde{A}) \in \mathcal{X}$ is as follows:

$$P(X = \tilde{X}) \ = \ \prod_{i=1}^{n} \left(m_i(\tilde{S}_{i,1}) \ \left(\prod_{t=1}^{T} \sigma(\tilde{A}_{i,t}|\tilde{S}_{i,t}) \right) \ \left(\prod_{t=1}^{T-1} g(\tilde{S}_{i,t+1}|\tilde{S}_{i,t}, \tilde{A}_{i,t}) \right) \right). \tag{2}$$

This expression reveals that the markets are independently distributed (Assumption 2.1(a)), but they are not necessarily identically distributed because $m_i(\cdot)$ depends on i. Even though the conditional choice probabilities and state transition probabilities are homogeneous under H_0, markets can still be heterogeneous due to differences in their initial state values. This is a desired feature in our testing problem, as the dynamic discrete choice literature usually allows the initial state distribution to be market-specific.

3 Our hypothesis test

In this paper, we propose to reject H_0 in (1) whenever the significance level α is larger than or equal to our p-value, which we denote by $\hat{p}_K$. That is,

$$\phi_K(X) \;\equiv\; 1\{\hat{p}_K \leq \alpha\}. \tag{3}$$

In turn, our p-value $\hat{p}_K$ is the result of constructing K transformations of the data via our MCMC algorithm, which we describe in Section 3.1 (see Algorithm 3.1). Our MCMC algorithm produces K sequential transformations of the data X, denoted by $(X^{(1)}, \ldots, X^{(K)})$. Our p-value is then computed as follows

$$\hat{p}_K \;\equiv\; \frac{1}{K}\sum_{k=1}^{K} 1\{\tau(X^{(k)}) \geq \tau(X)\}, \tag{4}$$

where $\tau : \mathcal{X} \to \mathbb{R}$ denotes the test statistic designed to detect departures from H_0 in the data. One notable feature of our hypothesis test is that its validity will not depend on the choice of the test statistic. This explains why it is not necessary to specify the test statistic or discuss it any further at this point.[3] The following is the main result of this paper.

Theorem 3.1. Under H_0 in (1), the test in (3) satisfies

$$\limsup_{K\to\infty} E[\phi_K(X)] \;\;\leq\;\; \alpha, \tag{5}$$

where the expectation is taken with respect to the randomness in $(X, X^{(1)}, \ldots, X^{(K)})$, i.e., both in the data X and in the random draws $(X^{(1)}, \ldots, X^{(K)})$ generated by our MCMC algorithm.

Theorem 3.1 establishes that the proposed test in (3) controls size as the length of the MCMC draws diverges. While this is an approximate result for a finite K, we note that the researcher controls the number of MCMC draws and that the approximation error becomes negligible by choosing a large value of K. Remarkably, Theorem 3.1 holds regardless of the number of markets n, time periods T, and players J, which can remain constant in our analysis. In addition and as promised earlier, this result also holds irrespective of the specific choice of test statistic τ used in the construction of the p-value in (4).

Our proposed test can be related to randomization tests, e.g., Lehmann and Romano (2005, Chapter 15.2). In particular, Theorem 3.1 follows from showing that the p-value in (4) approximates the p-value of a (computationally infeasible) randomization test for H_0 in (1). This observation provides intuition as to why Theorem 3.1 does not require the number of markets n, time periods T, or players J to grow. We provide a detailed explanation and additional results in Section 4. In the rest of this section, we introduce the MCMC algorithm used to construct $(X^{(1)}, \ldots, X^{(K)})$.

[3]For concreteness, (20) Section 5 proposes two possible test statistics designed to detect departures from market homogeneity in the conditional choice probability function.

3.1 The MCMC algorithm

Our MCMC algorithm requires some notation. Let $I = \{I_1, I_2\}$ denote an arbitrary pair of markets I_1 and I_2 in the data, i.e., $I_1, I_2 = 1, 2, \ldots, n$. We allow for $I_1 = I_2$. We use $\mathcal{I}$ to denote the collection of all such pairs of markets, i.e., $|\mathcal{I}| = n^2$. We also define several sets.

Definition 3.1. For any $I \in \mathcal{I}$ and $\breve{S} \in \mathcal{S}^{nT}$, $R_S(I, \breve{S})$ is the set of all $\tilde{S} \in \mathcal{S}^{nT}$ satisfying the following conditions:

(a) $\tilde{S}_{i,1} = \breve{S}_{i,1}$ for all $i = 1, \ldots, n$,

(b) $\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t} = s, \tilde{S}_{i,t+1} = s'\} = \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\}$ for all $s, s' \in \mathcal{S}$ and $i \in I^c$,

(c) $\sum_{i \in I} \sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t} = s, \tilde{S}_{i,t+1} = s'\} = \sum_{i \in I} \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\}$ for all $s, s' \in \mathcal{S}$.

In words, $R_S(I, \breve{S})$ is the set of all state configurations that result from "mixing" the hypothetical state data $\breve{S}$, subject to certain restrictions (given by conditions (a)-(c)). Under H_0, these restrictions imply that each state configuration in $R_S(I, \breve{S})$ has the same value of the likelihood function, provided that it is paired with a suitable action configuration. The corresponding suitable action configurations are precisely those in next definition.

Definition 3.2. For any $\tilde{S}, \breve{S} \in \mathcal{S}^{nT}$ and $\breve{A} \in \mathcal{A}^{nT}$, $R_A(\tilde{S}, (\breve{S}, \breve{A}))$ is the set of all $\tilde{A} \in \mathcal{A}^{nT}$ satisfying the following conditions:

(a) $\sum_{i=1}^{n} \sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t} = s, \tilde{A}_{i,t} = a, \tilde{S}_{i,t+1} = s'\} = \sum_{i=1}^{n} \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{A}_{i,t} = a, \breve{S}_{i,t+1} = s'\}$ for all $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}$,

(b) $\sum_{i=1}^{n} 1\{\tilde{S}_{i,T} = s, \tilde{A}_{i,T} = a\} = \sum_{i=1}^{n} 1\{\breve{S}_{i,T} = s, \breve{A}_{i,T} = a\}$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$.

By definition, $R_A(\tilde{S}, (\breve{S}, \breve{A}))$ is the set of action configurations that result from "mixing" the hypothetical action data $\breve{A}$, subject to certain restrictions (given by conditions (a)-(b)). Under H_0, these restrictions imply that the hypothetical data $(\breve{S}, \breve{A})$ has the same likelihood as the state configuration $\tilde{S}$ paired with any action configuration in $R_A(\tilde{S}, (\breve{S}, \breve{A}))$. With these definitions in place, we can now specify our MCMC algorithm.

Algorithm 3.1 (MCMC algorithm)**.** Let $(X^{(1)}, \ldots, X^{(K)})$ denote the following Markov chain.

Initiation. Initiate the chain with $X^{(1)} = X$.

Iteration. The rest of the chain is iteratively generated as follows. For any $k = 2, \ldots, K$ and given $(X^{(1)}, \ldots, X^{(k-1)})$, $X^{(k)} = (S^{(k)}, A^{(k)})$ is randomly generated as follows:

Step 1: Draw $I^{(k)}$ uniformly from $\mathcal{I}$, independently of $(X^{(1)}, \ldots, X^{(k-1)})$.

Step 2: Given $(X^{(1)}, \dots, X^{(k-1)}, I^{(k)})$, draw $S^{(k)}$ uniformly from $R_S(I^{(k)}, S^{(k-1)})$.

Step 3: Given $(X^{(1)}, \dots, X^{(k-1)}, I^{(k)}, S^{(k)})$, draw $A^{(k)}$ uniformly from $R_A(S^{(k)}, X^{(k-1)})$. ■

Several comments are in order. Steps 2 and 3 require randomly drawing state and action configurations uniformly over the sets $R_S(I^{(k)}, S^{(k-1)})$ and $R_A(S^{(k)}, X^{(k-1)})$, respectively. On the one hand, Step 3 is relatively easy to implement by permuting the action data in $A^{(k-1)}$ subject to the restrictions in $R_A(S^{(k)}, X^{(k-1)})$. Algorithm A.3 in Section A.2 explains how to implement this in practice and provides a justification (Lemma A.5). On the other hand, Step 2 is more involved. We propose to implement it using a modified version of the Euler algorithm (Kandel et al., 1996; Besag and Mondal, 2013). Section A.1 describes the original Euler algorithm (Algorithm A.1), our modification (Algorithm A.2), and formally shows that the latter satisfy Step 2 in Lemma A.2.

For any $k = 2, \dots, K$, $X^{(1)}, \dots, X^{(k-1)} \in \mathcal{X}$, $I \in \mathcal{I}$, and $\tilde{X} = (\tilde{S}, \tilde{A}) \in \mathcal{X}$, our MCMC algorithm implies the following transition probabilities:

$$P(I^{(k)} = I \mid X^{(1)}, \dots, X^{(k-1)}) = \frac{1}{|\mathcal{I}|}, \tag{6}$$

$$P(S^{(k)} = \tilde{S} \mid I^{(k)}, X^{(1)}, \dots, X^{(k-1)}) = \frac{1\{\tilde{S} \in R_S(I^{(k)}, S^{(k-1)})\}}{|R_S(I^{(k)}, S^{(k-1)})|}, \tag{7}$$

$$P(A^{(k)} = \tilde{A} \mid S^{(k)}, I^{(k)}, X^{(1)}, \dots, X^{(k-1)}) = \frac{1\{\tilde{A} \in R_A(S^{(k)}, X^{(k-1)})\}}{|R_A(S^{(k)}, X^{(k-1)})|}. \tag{8}$$

Note that (7) and (8) are well defined, as both denominators can be shown to be positive. In turn, these transition probabilities imply that our MCMC algorithm has the following transition probability:

$$P(X^{(k)} = \tilde{X} \mid X^{(1)}, \dots, X^{(k-1)}) = \begin{cases} \sum_{I \in \mathcal{I}} \frac{1\{\tilde{S} \in R_S(I, S^{(k-1)}), \tilde{A} \in R_A(\tilde{S}, X^{(k-1)})\}}{|\mathcal{I}| \times |R_S(I, S^{(k-1)})| \times |R_A(\tilde{S}, X^{(k-1)})|} & \text{if } |R_S(I, S^{(k-1)})| \times |R_A(\tilde{S}, X^{(k-1)})| > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

4 Our test as an approximate randomization test

This section provides the formal arguments that are necessary to prove Theorem 3.1 and, thus, justify our hypothesis test in (3). In particular, we show that our MCMC-based p-value in (4) is an approximation of the p-value of a specific randomization test. We argue that this randomization test is valid in finite samples but computationally infeasible, which explains why we propose the MCMC algorithm to approximate its p-value.

This section is organized as follows. Section 4.1 provides an alternative representation of the likelihood of the data under H_0 in (1). This result allows us to define a sufficient statistic of the data under H_0, denoted by $U(X)$. Section 4.2 introduces a transformation group of the data, which does not change the value of the sufficient statistic $U(X)$.[4] Section 4.3 defines a specific randomization test for (1), and argues that is both finite-sample valid and computationally infeasible. Section 4.4 shows that our MCMC-based test in (3) can successfully approximate the infeasible randomization test as the number of MCMC draws diverges.

4.1 An alternative representation of the likelihood

The following result provides an alternative representation of the likelihood under H_0 in (1).

Lemma 4.1. Under Assumption 2.1 and H_0 in (1), the likelihood of the data $X = (S, A)$ evaluated at $\tilde{X} = (\tilde{S}, \tilde{A}) \in \mathcal{X}$ with $\tilde{S} = (\tilde{S}_{i,t} : i = 1, \ldots, n, t = 1, \ldots, T) \in \mathcal{S}^{nT}$ and $\tilde{A} = (\tilde{A}_{i,t} : i = 1, \ldots, n, t = 1, \ldots, T) \in \mathcal{A}^{nT}$ is

$$P(X = \tilde{X}) \;=\; P(A = \tilde{A}|S = \tilde{S}) \;\times\; P(S = \tilde{S}), \tag{10}$$

where

$$\begin{aligned} P(A = \tilde{A}|S = \tilde{S}) = &\prod_{(s,a,s')\in\mathcal{S}\times\mathcal{A}\times\mathcal{S}} \left(\frac{\sigma(a|s)g(s'|s,a)}{\sum_{\tilde{a}\in\mathcal{A}} g(s'|\tilde{a},s)\sigma(\tilde{a}|s)}\right)^{\sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=s,\tilde{A}_{i,t}=a,\tilde{S}_{i,t+1}=s'\}} \\ &\times \prod_{(s,a)\in\mathcal{S}\times\mathcal{A}} \sigma(a|s)^{\sum_{i=1}^{n} 1\{\tilde{S}_{i,T}=s,\tilde{A}_{i,T}=a\}} \end{aligned} \tag{11}$$

and

$$P(S = \tilde{S}) = \left(\prod_{i=1}^{n} m_i(\tilde{S}_{i,1})\right) \times \prod_{(s,s')\in\mathcal{S}\times\mathcal{S}} \left(\sum_{a\in\mathcal{A}} g(s'|a,s)\sigma(\tilde{a}|s)\right)^{\sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=s,\tilde{S}_{i,t+1}=s'\}}. \tag{12}$$

From this result, we can deduce the following corollary.

Corollary 4.1. Under Assumption 2.1 and H_0 in (1), the sufficient statistic for $X = (S, A)$ is

$$U(X) = \left(\begin{array}{l} (S_{i,1} : i = 1\ldots,n)\,, \\ \left(\sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{S_{i,t} = s, A_{i,t} = a, S_{i,t+1} = s'\} : (s,a,s') \in \mathcal{S}\times\mathcal{A}\times\mathcal{S}\right), \\ (\sum_{i=1}^{n} 1\{S_{i,T} = s, A_{i,T} = a\} : (s,a) \in \mathcal{S}\times\mathcal{A}) \end{array} \right). \tag{13}$$

Corollary 4.1 implies that, under H_0, any transformation of the data that does not modify the value of $U(X)$ will not change the value of the likelihood. This observation provides the basis of the randomization test that we consider in the remaining sections.

[4]Hereafter, we use "transformation group" to denote the set defined in Lehmann and Romano (2005, pages 693-4).

4.2 A transformation group related to the proposed MCMC algorithm

Our proposed MCMC algorithm can be understood as an iteration of transformations to the data X. In particular, $X^{(1)} = X$ is the identity transformation, $X^{(2)}$ follows from applying Steps 1-3 to X, $X^{(3)}$ follows from applying Steps 1-3 twice to X, and so forth. In this section, we define a transformation group that is related to the transformations in our MCMC algorithm. To define this properly, we first require the following definition.

Definition 4.1. For any pair of markets $I = \{I_1, I_2\} \in \mathcal{I}$, let $\mathbf{G}(I)$ denote the set of all transformations of $\mathcal{X}$ onto itself such that, for any $g \in \mathbf{G}(I)$ and $(\breve{S}, \breve{A}) \in \mathcal{X}$, $(\tilde{S}, \tilde{A}) = g(\breve{S}, \breve{A})$ satisfies $\tilde{S} \in R_S(I, \breve{S})$ and $\tilde{A} \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$.

Lemma B.3 in the appendix shows that $\mathbf{G}(I)$ is a transformation group. By Definition 4.1, $\mathbf{G}(I)$ is the group representation of Steps 2-3 of our MCMC algorithm. Given a randomly chosen pair of markets $I^{(k)}$ in Step 1, Steps 2-3 obtain the next element of the Markov chain $X^{(k)} = (S^{(k)}, A^{(k)})$ by applying a randomly chosen transformation in $\mathbf{G}(I^{(k)})$ to the preceding element of the Markov chain, $X^{(k-1)}$. In this sense, Steps 2-3 of our MCMC algorithm are a specific way of choosing a particular transformation in $\mathbf{G}(I^{(k)})$.

By the description in the previous paragraph, our MCMC algorithm randomly chooses transformations in $\mathbf{G}(I)$ for random pairs of markets I, and iteratively applies them to the data. These iterative transformations are related to the set that we define next.

Definition 4.2. Let $\mathbf{G}$ be the set of all finitely many compositions of the elements in $\bigcup_{I \in \mathcal{I}} \mathbf{G}(I)$.

The next result states that $\mathbf{G}$ is a transformation group with desirable properties.

Lemma 4.2. $\mathbf{G} : \mathcal{X} \to \mathcal{X}$ is a transformation group of $\mathcal{X}$ such that, for any $g \in \mathbf{G}$ and $\tilde{X} \in \mathcal{X}$, $\tilde{X}$ and $g\tilde{X}$ have the same sufficient statistic in (13), i.e., $U(\tilde{X}) = U(g\tilde{X})$.

The properties shown in Lemma 4.2 imply that we can use $\mathbf{G}$ to define a valid randomization test. We do this in Section 4.3.

4.3 A randomization test

Following Lehmann and Romano (2005, Chapter 15.2), we can use the transformation group $\mathbf{G}$ to define a randomization test. This test rejects H_0 in (1) whenever the significance level α is larger than or equal to the randomization p-value, which we denote by $\hat{p}$. That is,

$$\phi(X) \;\equiv\; 1\{\hat{p} \leq \alpha\}, \tag{14}$$

where

$$\hat{p} \;\equiv\; \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} 1\{\tau(gX) \geq \tau(X)\}. \tag{15}$$

By the arguments in Lehmann and Romano (2005, Page 636), the randomization test in (14) is finite-sample valid. We record this in the next result.

Lemma 4.3. Under H_0 in (1) and for any $\alpha \in (0,1)$, the test in (14) satisfies

$$E[\phi(X)] \;\leq\; \alpha. \tag{16}$$

Unlike our proposed test in (3), the hypothesis test in (14) is computationally infeasible in typical applications of dynamic discrete choice games. The basic reason is that the transformation group $\mathbf{G}$ is usually impossible to enumerate. To see why, note that $\mathbf{G}$ is a set of transformations restricted by the condition on the sufficient statistics in (13). This condition is hard to impose without explicitly verifying that it holds. In turn, an explicit verification of this condition is not practically feasible, as it would require exploring all possible transformations that map $\mathcal{X}$ to $\mathcal{X}$. Even in applications in which n, T, and $|\mathcal{A}|$ and $|\mathcal{S}|$ are relatively small, the resulting state space of the data $\mathcal{X} = \mathcal{S}^{nT} \times \mathcal{A}^{nT}$ can be overwhelming.

In the randomization testing literature, it is not uncommon for the transformation set $\mathbf{G}$ to be huge. As Lehmann and Romano (2005, page 636) explains, one can still implement a random version of the test in (14) by drawing randomly from $\mathbf{G}$ *in a uniform fashion*. This point is routinely exploited in standard settings to construct tests based on permutations or sign changes. To the best of our knowledge, however, there is no feasible way of obtaining such random draws in the current setup, as the condition on the sufficient statistics in (13) is hard to impose without explicitly checking whether it holds. This explains why we cannot directly exploit the finite-sample result in Lemma 4.3. In any case, Section 4.4 reveals that our MCMC-based p-value in (4) approximates the infeasible p-value in (15) as the length of the MCMC diverges.

4.4 An MCMC approximation to the randomization test

Our main theoretical result is Theorem 3.1, which shows that the test in (3) controls size as the number of MCMC draws K diverges to infinity. The following lemma provides a fundamental stepping stone to prove this result.

Lemma 4.4. Conditional on X,

$$\sup_{t\in\mathbb{R}} \left| \frac{1}{K}\sum_{k=1}^{K} 1\{\tau(X^{(k)}) \geq t\} \;-\; \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} 1\{\tau(gX) \geq t\} \right| \;\overset{a.s.}{\to} 0 \quad \text{as } K \to \infty.$$

Lemma 4.4 shows that, as the number of MCMC draws diverges, the conditional distribution based on the MCMC algorithm converge to the conditional distribution of the computationally infeasible randomization test described in Section 4.3. By applying Lemma 4.4 with $t = \tau(X)$, we can deduce that the p-value in (4) approximates the p-value in (15) as the number of MCMC draws K diverges. That is, conditional on X,

$$\hat{p}_K \overset{a.s.}{\to} \hat{p} \quad \text{as } K \to \infty.$$

By combining this observation with the finite-sample validity of the infeasible randomization test in (14) (Lemma 4.3), it follows that our proposed MCMC-test becomes valid as the number of MCMC draws K diverges. This argument provides the intuition behind Theorem 3.1 (see Section A.3 of the appendix for the proof), and why it holds regardless of the number of markets n, time periods T, and players J.

Our analysis in this paper focuses on the properties of our test under the null hypothesis. While it would also be desirable to analyze the power properties of our test, we consider that this is out of the scope of this paper. As explained, our validity result holds in finite samples, i.e., for any number of markets n and time period T. As far as we understand, analyzing the power properties of our test under these conditions would be a formidable task in finite samples. A more manageable way of conducting such analysis would involve a diverging number of markets n or time periods T. Given the finite-sample nature of our analysis, we believe that this would be a significant departure in our framework. Having said this, the simulation evidence in the next section suggests that our test has desirable power properties (e.g., Table 2).

5 Monte Carlo simulations

In this section, we explore the performance of our proposed test in Monte Carlo simulations. We consider the Monte Carlo design used by Otsu et al. (2016, Section 4), which follows from the dynamic oligopoly discrete game in Pesendorfer and Schmidt-Dengler (2008, Section 7.1). We refer to these papers for the details on the setup. The simulated data are generated by two oligopolic firms deciding whether to enter or not into n markets, and over T time periods. This dynamic game has multiple equilibria, which we exploit to generate departures from the homogeneity assumption.

In each period $t = 1, \ldots, T$ and market $i = 1, \ldots, n$, there are four possible actions in this game: $A_{i,t} = 1$ denotes that neither firm entered the market, $A_{i,t} = 2$ denotes that only firm 2 enters, $A_{i,t} = 3$ denotes that only firm 1 enters, and $A_{i,t} = 4$ denotes that both firms enter. This implies that $\mathcal{A} = \{1, 2, 3, 4\}$. In addition, the state is equal to the last period's action, i.e.,

$$S_{i,t} \;=\; A_{i,t-1}, \tag{17}$$

and so $\mathcal{S} = \{1, 2, 3, 4\}$. Note that (17) implies that the state transition probabilities are homogeneous, and given by

$$g_{i,t+1}(s'|a, s) \;=\; 1\{s' = a\}. \tag{18}$$

We presume that (18) is known to the researcher, who replaces H_0 in (1) with the homogeneity of the conditional choice probabilities. In other words, the relevant hypothesis testing problem is

$$H_0 : \sigma_{i,t}(a|s) = \sigma(a|s) \quad \text{vs.} \quad H_1 : H_0 \text{ is false.} \tag{19}$$

Following Otsu et al. (2016, Eq. (4), (7)), we consider these test statistics

$$\begin{aligned} \tau_1(X) &\equiv \sum_{i=1}^{n} \sum_{(a,s)\in\mathcal{A}\times\mathcal{S}} (\hat{\sigma}_i(a|s) - \hat{\sigma}(a|s))^2 \left(\frac{\sum_{t=1}^{T} 1\{S_{i,t} = s\}}{\hat{\sigma}(a|s)} \right) \\ \tau_2(X) &\equiv 2\sum_{i=1}^{n} \sum_{(a,s)\in\mathcal{A}\times\mathcal{S}} \hat{\sigma}_i(a|s) \log \left(\frac{\hat{\sigma}_i(a|s)}{\hat{\sigma}(a|s)} \right) \sum_{t=1}^{T} 1\{S_{i,t} = s\}, \end{aligned} \tag{20}$$

where we interpret $0/0 = 0$ and $0 \times \log(0) = 0$, and we define

$$\begin{aligned} \hat{\sigma}_i(a|s) &\equiv \frac{\sum_{t=1}^{T} 1\{A_{i,t} = a, S_{i,t} = s\}}{\sum_{t=1}^{T} 1\{S_{i,t} = s\}} \\ \hat{\sigma}(a|s) &\equiv \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} 1\{A_{i,t} = a, S_{i,t} = s\}}{\sum_{i=1}^{n} \sum_{t=1}^{T} 1\{S_{i,t} = s\}}. \end{aligned}$$

The statistics in (20) compute weighted differences between market-specific empirical conditional choice probabilities and the overall counterpart.

The data produced by this game is a matrix $X = (S, A) \in \mathcal{X}$ constructed exactly as in Otsu et al. (2016, Section 4). We simulate data from a mixture of two data generating processes: DGP 1 and DGP 2. They represents Markov perfect equilibria of the dynamic game, which differ in the conditional choice probabilities $\sigma(a|s)$. The matrices of conditional choice probabilities in DGP 1 and DGP 2 are

$$\begin{pmatrix} 0.19 & 0.30 & 0.12 & 0.18 \\ 0.08 & 0.09 & 0.08 & 0.07 \\ 0.53 & 0.48 & 0.46 & 0.53 \\ 0.20 & 0.13 & 0.34 & 0.22 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0.18 & 0.48 & 0.03 & 0.16 \\ 0.20 & 0.21 & 0.14 & 0.23 \\ 0.29 & 0.22 & 0.13 & 0.26 \\ 0.33 & 0.09 & 0.70 & 0.35 \end{pmatrix},$$

respectively, where the columns index the value of the state $s \in \mathcal{S} = \{1, 2, 3, 4\}$, and the rows index the value of the action $a \in \mathcal{A} = \{1, 2, 3, 4\}$. Each market is sampled independently. Market $i = 1, \ldots, n$ behaves according to DGP 1 if $i/n \leq \lambda$ and to DGP 2 if $i/n > \lambda$. Therefore, $\lambda \in [0, 1]$ represents the proportion of markets that are in DGP 1. Each market is initialized with state equal to 1, and we simulate the corresponding action according to the corresponding conditional choice probabilities. This, in turn, determines the next period's state according to (17). We then proceed iteratively until we have simulated $T + 100$ periods for each market. Then, the first 100 periods are discarded, producing a sample of T periods for n markets, which are the data observed by the researcher.

For each simulated data, we implement our proposed test in (3) with $K = 10,000$. We consider simulations with $n \in \{20, 40, 80, 160\}$, $T \in \{5, 10, 20, 40, 80\}$, and $\lambda \in \{1, 0, 0.5, 0.9\}$. As explained earlier, λ represents the proportion of markets that are in DGP 1. If $\lambda = 1$ or $\lambda = 0$, all markets are sampled from the same distribution, and so the conditional choice probabilities are homogeneous across markets. This means that H_0 in (19) holds. In turn, if $\lambda = 0.5$ or $\lambda = 0.9$, each data is composed of markets from both distributions, and so the conditional choice probabilities are not homogeneous across markets. This means that H_0 in (19) fails to hold. Note that $\lambda = 0.5$ generates data in which both distributions are equally represented, and so the heterogeneity in the conditional choice probabilities should be more salient. On the other hand, the case with $\lambda = 0.9$ produces data with a vast majority of markets in DGP 1, and so the heterogeneity in the conditional choice probabilities should be harder to detect. For each simulation design, we compute rejection rates based on $1,000$ independently simulated datasets.

The results from the Monte Carlo simulation are shown in Table 1 for $\lambda \in \{0, 1\}$ and Table 2 for $\lambda \in \{0.5, 0.9\}$, respectively. For the sake of comparison, we also include the results from the test proposed by Otsu et al. (2016). Their test compares the same test statistics in (20) with critical values based on the bootstrap. As mentioned earlier, they show the validity of their test in an asymptotic framework with $T \to \infty$ and n fixed. In contrast, our main result in Theorem 3.1 is valid for any finite n and T.

Table 1 reveals that our test achieves relatively good size control for all values of time periods and market sizes under consideration. The table shows the result of running 80 hypothesis tests for different data configurations that satisfy H_0 in (19) (four market sizes, five time periods, two test statistics, and two distributions). Across these 80 numbers, our proposed test has an average rejection rate of 5.3, with a standard deviation of 0.7, and a range of 3.9 to 8. We note that Theorem 3.1 implies that our test should not produce over-rejection as K becomes large, but it is silent about the possibility of under-rejection. Table 1 reveals that our test does not seem to suffer from under-rejection in these simulations. For Otsu et al. (2016)'s test, the average rejection rate is 5.1, with a standard deviation is 2.2, and a range of 0.6 to 13.5. We note that these extreme rejection rates occur in simulations with $T = 5$, which is reasonable for a test whose validity is proven in an asymptotic framework in which T diverges.

Table 2 explores the performance of these tests for data configurations that do not satisfy H_0 in (19) due to the multiplicity of equilibria. We begin by explaining the results of the table that are common to both hypothesis tests. First, the value of λ denotes the proportion of the n markets in the data that are in DGP 1. As λ becomes closer to either zero or one, the data are increasingly coming from a single distribution, making the departure from the H_0 harder to detect. Second, as the number of markets n grows, the inference methods gain more evidence of the presence of multiplicity, resulting in higher rejection rates. The same phenomenon occurs as the number of time periods T increases. Third, $\tau_2(X)$ is designed to be a more efficient test statistic than $\tau_1(X)$, which explains why it produces higher rejection rates across the various simulation designs. We now turn to compare rejection rates between the two tests. In most simulation designs, our test appears to have a higher or equal rejection rate than Otsu et al. (2016)'s test. The few exceptions occur in designs with $n = 20$ and $T \in \{5, 10\}$, which correspond to designs in which Otsu et al. (2016)'s test over-rejects under the null hypothesis. In other words, their power advantage relative to our test may vanish if we consider a size-corrected version of their test.

		DGP 1 ($\lambda = 1$)				DGP 2 ($\lambda = 0$)			
n	T	Our test		OPT's test		Our test		OPT's test	
		$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$
20	5	5.6	5.2	13.2	5.9	4.8	4.6	13.5	12.7
20	10	5.5	4.9	7.0	4.5	6.4	5.4	7.9	8.4
20	20	4.1	4.9	4.4	5.0	4.9	5.2	5.8	7.1
20	40	6.0	5.9	5.1	6.2	5.4	4.9	4.8	5.4
20	80	5.6	5.0	5.7	6.6	5.7	5.5	5.1	5.2
40	5	5.5	4.8	6.5	2.3	3.9	4.3	8.0	6.4
40	10	5.5	4.9	3.8	2.7	5.8	6.2	5.2	4.9
40	20	4.6	4.3	4.3	3.4	4.8	5.0	6.2	6.9
40	40	5.7	5.8	4.5	5.3	6.2	4.9	3.9	5.3
40	80	5.4	5.2	5.3	5.3	4.7	4.9	5.6	4.5
80	5	6.1	6.0	5.3	1.5	4.6	4.5	4.6	3.7
80	10	5.4	4.6	3.2	1.2	5.5	5.4	5.6	5.1
80	20	4.9	4.3	5.2	3.5	5.8	4.8	4.9	5.7
80	40	6.5	6.1	3.9	3.9	4.7	4.6	4.7	5.1
80	80	5.2	5.6	4.7	4.6	6.0	6.3	5.3	5.0
160	5	7.0	8.0	4.9	0.6	6.1	6.3	4.0	1.4
160	10	5.5	5.2	3.4	0.9	4.3	5.2	4.7	3.9
160	20	5.2	5.0	3.3	2.4	5.9	5.8	4.5	4.5
160	40	4.7	5.4	4.8	4.8	6.0	5.6	6.3	5.5
160	80	4.7	4.8	4.5	4.6	5.6	6.0	5.5	5.2

Table 1: Simulation results under H_0 for $\alpha = 5\%$ based on 1,000 i.i.d. simulation draws. The results for $\lambda = 1$ corresponds to data sampled from DGP 1 and $\lambda = 0$ corresponds to data sampled from DGP 2. The test statistics $\tau_1(X)$ and $\tau_2(X)$ are defined in (20). Our test is computed according to (3) with $K = 10,000$. OPT refers to Otsu et al. (2016), whose results were copied from Tables 1 and 2 in their paper.

n	T	Mixture with $\lambda = 0.5$				Mixture with $\lambda = 0.9$			
		Our test		OPT's test		Our test		OPT's test	
		$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$
20	5	4.5	5.9	10.3	8.3	4.1	5.0	10.7	6.0
20	10	8.6	13.5	6.5	7.4	5.4	5.8	6.5	4.8
20	20	38.2	51.3	27.8	27.4	12.0	15.0	11.7	12.8
20	40	96.3	98.2	79.7	76.1	38.5	40.1	32.7	35.3
20	80	100	100	99.9	99.8	85.9	86.9	75.8	76.5
40	5	4.8	8.1	4.7	4.1	5.1	5.5	4.5	2.5
40	10	9.9	17.8	7.4	5.5	6.3	6.4	5.4	4.2
40	20	63.3	76.3	44.6	36.2	20.2	24.0	16.0	14.8
40	40	100	100	97.4	94.3	59.6	63.7	49.0	50.1
40	80	100	100	100	100	98.5	99.1	93.5	92.5
80	5	4.3	9.3	3.3	2.3	4.8	6.8	3.4	1.7
80	10	13.3	25.2	10.8	5.8	6.3	9.0	5.9	3.2
80	20	87.3	95.4	68.5	55.5	28.5	34.0	23.3	19.7
80	40	100	100	100	99.9	85.1	88.2	72.8	73.2
80	80	100	100	100	100	100	100	99.7	99.6
160	5	4.1	11.5	2.9	0.9	4.9	7.7	4.0	0.9
160	10	21.4	44.5	12.4	5.8	9.1	12.6	6.0	2.1
160	20	99.2	100	92.3	78.6	44.4	53.0	38.2	30.6
160	40	100	100	100	100	97.7	98.3	93.4	92.4
160	80	100	100	100	100	100	100	100	100

Table 2: Simulation results under H_1 for $\alpha = 5\%$ based on 1,000 i.i.d. simulation draws. The results for $\lambda = 0.5$ corresponds to data sampled from DGP 1 and DGP 2 in equal proportions, and the results for $\lambda = 0.9$ corresponds to data sampled from DGP 1 and DGP 2 with proportions 0.9 and 0.1, respectively. The test statistics $\tau_1(X)$ and $\tau_2(X)$ are defined in (20). Our test is computed according to (3) with $K = 10,000$. OPT refers to Otsu et al. (2016), whose results were copied from Tables 3 and 4 in their paper.

Sample	Average	Std. dev.	Minimum	Maximum
1980-1990	4,226.8	2,284.4	1,321.3	12,578.0
1991-1998	3,857.2	2,107.9	1,084.0	9,564.8

Table 3: Summary statistics for market capacity per year, measured in thousand of tons.

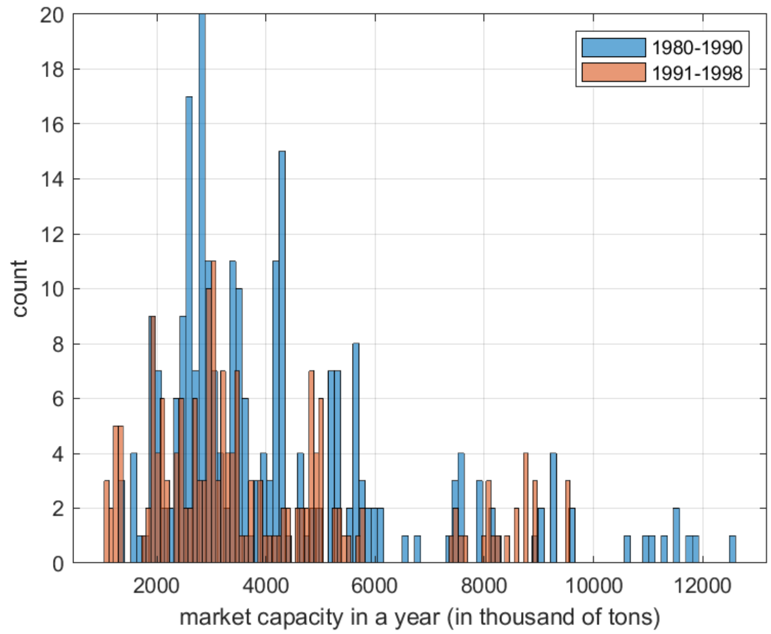

Figure 1: Histogram of market capacity per year, measured in thousand of tons.

6 Empirical application

In this section, we revisit the application in Ryan (2012), as studied in Otsu et al. (2016, Section 5). Ryan (2012) considers a dynamic discrete game to study the welfare costs of the 1990 Amendments to the Clean Air Act on the U.S. Portland cement industry. He develops a dynamic oligopoly game based on Ericson and Pakes (1995), and estimates it using the two-stage method developed by Bajari et al. (2007). This method's first stage is to estimate optimal entry, exit, and investment decisions as a function of production capacity, and it relies on the assumptions that markets are homogeneous. Our hypothesis test can be used to investigate the validity of this assumption.

We use the same data as in Otsu et al. (2016, Section 5). For each year in 1980-1998 and 23 geographically separated U.S. markets, we observe the sum of the production capacities for all the firms in that market. Table 1 provides summary statistics of this aggregate production capacity before and after the 1990 Amendments, and Figure 3 provides the corresponding histogram.

These data represent the result of the firms' optimal entry, exit, and investment decisions in the dynamic game estimated by Ryan (2012). We follow Otsu et al. (2016) and discretize the production capacity into 50 bins with equal intervals of 250 thousand tons each (0-250 thousand tons, 250-500 thousand tons, and so on). For each $i = 1, \dots, n = 23$ and year $t = 1, \dots, 19$, we use $A_{i,t} \in \mathcal{A} = \{1, \dots, 50\}$ to denote the production capacity bin. The state variable in any market is the previous period's action, i.e.,

$$S_{i,t} \;=\; A_{i,t-1}, \tag{21}$$

and so $S_{i,t} \in \mathcal{S} = \{1, \dots, 50\}$. We note that (21) implies that the state transition probabilities are homogeneous, and so H_0 in (1) is equivalent to the homogeneity of the conditional choice probabilities.

Following Ryan (2012) and Otsu et al. (2016), we allow the 1990 Amendments to affect the decision of the firms. We then test the homogeneity of the conditional choice probabilities for two subsets of data: before and after 1990. That is, we implement the following hypothesis tests:

$$H_0^{\text{before}} : \sigma_{i,t}(a|s) = \sigma(a|s) \text{ for } i = 1, \dots, 23,\ t = 1, \dots, 10 \quad \text{vs.} \quad H_1^{\text{before}} : H_0^{\text{before}} \text{ is false} \tag{22}$$

$$H_0^{\text{after}} : \sigma_{i,t}(a|s) = \sigma(a|s) \text{ for } i = 1, \dots, 23,\ t = 11, \dots, 19 \quad \text{vs.} \quad H_1^{\text{after}} : H_0^{\text{after}} \text{ is false} \tag{23}$$

We note that the two samples used to test the hypotheses in (22) and (23) have a relatively small number of time periods ($T = 10$ and $T = 9$ for (22) and (23), respectively) and markets (in both cases, $n = 23$). In this sense, this represents an ideal scenario for our proposed test, as its validity does not rely on either one of these dimensions diverging.

Table 4 shows the results of applying our procedure to test the hypotheses in (22) and (23). We consider both test statistics in (20), and we use $K = 10,000$.[5] At a significance level of $\alpha = 5\%$, we do not reject the homogeneity of the conditional choice probabilities. Table 4 also shows the results of the bootstrap-based tests proposed by Otsu et al. (2016), using the same test statistics. As opposed to our results, their methods reject the hypothesis of homogeneity of the conditional choice probabilities in the sample prior to 1990. Together with the simulation evidence in Section 5, these results confirm that our test and the one proposed by Otsu et al. (2016) can produce different conclusions. Since both tests rely on the same test statistic, these differences are entirely driven by the differences in the p-values.

7 Conclusions

This paper proposes a hypothesis test for the "homogeneity assumption" in dynamic discrete games. Our test is implemented by an MCMC algorithm and it is non-parametric, i.e., it does not rely on functional forms imposed by the researcher. We show that our test is valid as the (user-defined) number of MCMC draws diverges, regardless of the number of markets and time periods in the data.

[5] These tests were implemented in MATLAB using a laptop computer and took 8.7 minutes for the sample prior to 1990 and 3.6 minutes for the sample after 1990.

	Before 1990		After 1990	
	$\tau_1(X)$	$\tau_2(X)$	$\tau_1(X)$	$\tau_2(X)$
Test statistic	199.48	159.43	89.44	90.58
Our p-value	0.17	0.07	0.62	0.56
OPT's p-value	0.009	0.01	0.09	0.055

Table 4: Results of testing (1) separately before and after the passing of the 1990 Amendments. The test statistics $\tau_1(X)$ and $\tau_2(X)$ are defined in (20). Our test is computed according to (3) with $K = 10,000$. OPT refers to Otsu et al. (2016), whose results were copied from Table 6 in their paper.

This result contrasts with that of available methods in the literature, which require the number of time periods to diverge. We establish our validity result by showing that our proposed test is an MCMC approximation to a computationally infeasible randomization test, which happens to be finite-sample valid. Our Monte Carlo simulations confirm that our test has an excellent performance in finite samples, both in terms of size control and power.

A Appendix to Section 3

A.1 Implementation of Step 2 in Algorithm 3.1

For any $k = 2, \dots, K$, $S^{(k-1)} \in \mathcal{S}^{nT}$, and $I^{(k)}$ selected in Step 1 of Algorithm 3.1, Step 2 of Algorithm 3.1 draws $S^{(k)}$ uniformly within $R_S(I^{(k)}, S^{(k-1)})$. To implement this step, we propose a modification of the Euler Algorithm. For a description of the Euler Algorithm, see Kandel et al. (1996) and Besag and Mondal (2013). We first describe the original Euler Algorithm in Algorithm A.1 and then introduce our modification in Algorithm A.2. Throughout this section, we use 0 to represent an auxiliary value for the state variable that does not belong to the observed values of the state variable, as $0 \notin \mathcal{S} = \{1, 2, \dots, |\mathcal{S}|\}$.

Algorithm A.1 (Euler Algorithm)**.** Given any integer $V \geq 2$ and any $\breve{\xi} \in (\mathcal{S} \cup \{0\})^V$, $\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_V)$ is randomly generated as follows:

Step 1: For every $s, s' \in \mathcal{S} \cup \{0\}$, define

$$N^{(0)}(s, s') \;=\; 1\{(\breve{\xi}_V, \breve{\xi}_1) = (s, s')\} + \sum_{v=1}^{V} 1\{(\breve{\xi}_v, \breve{\xi}_{v+1}) = (s, s')\}.$$

Step 2: Define $\zeta_1 = \breve{\xi}_V$. Set $v = 1$ and do the following.

(a) Generate ζ_{v+1} according to the following distribution.

$$P(\zeta_{v+1} = s \mid \zeta_v = s') \;=\; \frac{N^{(0)}(s, s')}{\sum_{s'' \in \mathcal{S} \cup \{0\}} N^{(0)}(s'', s')},$$

(b) If $(\mathcal{S} \cup \{0\}) \not\subset \{\zeta_1, \dots, \zeta_{v+1}\}$, then increase v by one and go back to (a). If $(\mathcal{S} \cup \{0\}) \subset \{\zeta_1, \dots, \zeta_{v+1}\}$, then set $\bar{v} = v + 1$ and go to Step 3.

Step 3: Define $\tilde{\xi}_1 = \breve{\xi}_1$. Also, for every $s, s' \in \mathcal{S} \cup \{0\}$, set

$$N^{(1)}(s, s') \;=\; \sum_{v=1}^{V} 1\{(\breve{\xi}_v, \breve{\xi}_{v+1}) = (s, s')\} - 1\{s' = \zeta_{(\min\{v=1,\ldots,\bar{v}:\zeta_v = s\}-1)}\}.$$

Step 4: For every $v = 2, \ldots, V$, generate $\tilde{\xi}_v$ iteratively according to

$$P(\tilde{\xi}_v = s' \mid \tilde{\xi}_{v-1} = s) \;=\; \begin{cases} \dfrac{N^{(v-1)}(s, s')}{\sum_{s'' \in \mathcal{S}\cup\{0\}} N^{(v-1)}(s, s'')} & \text{if } \sum_{s''\in\mathcal{S}\cup\{0\}} N^{(v-1)}(s, s'') \geq 1, \\ 1\{s' = \zeta_{(\min\{v=1,\ldots,\bar{v}:\zeta_v = s\}-1)}\} & \text{otherwise,} \end{cases}$$

where, for every $s, s' \in \mathcal{S} \cup \{0\}$, $N^{(v)}(s, s') \;=\; N^{(v-1)}(s, s') - 1\{(\tilde{\xi}_{v-1}, \tilde{\xi}_v) = (s, s')\}$. ■

Before we describe the central property of the Euler algorithm, we first introduce the following definition.

Definition A.1. For any $\breve{\xi} \in (\mathcal{S} \cup \{0\})^V$, let $R_{S0}(\breve{\xi})$ denote the set of all $\tilde{\xi} \in (\mathcal{S} \cup \{0\})^V$ that satisfy the following conditions:

(a) $\tilde{\xi}_1 = \breve{\xi}_1$,

(b) $\sum_{v=1}^{V} 1\{\tilde{\xi}_v = s, \tilde{\xi}_{v+1} = s'\} = \sum_{v=1}^{V} 1\{\breve{\xi}_v = s, \breve{\xi}_{v+1} = s'\}$ for all $s, s' \in \mathcal{S} \cup \{0\}$.

Note that $\breve{\xi} \in R_{S0}(\breve{\xi})$, and so $R_{S0}(\breve{\xi}) \neq \emptyset$. Next, we give the main property of the Euler algorithm.

Lemma A.1. For any $\breve{\xi} \in (\mathcal{S} \cup \{0\})^V$, the outcome of the Euler algorithm given $\breve{\xi}$ (i.e., Algorithm A.1) is uniformly distributed over $R_{S0}(\breve{\xi})$ conditional on $\breve{\xi}$.

Proof. See Kandel et al. (1996, Theorem 2). ■

We now introduce our modification of the Euler algorithm to construct $S^{(k)}$ for any $k = 2, \ldots, K$.

Algorithm A.2 (Generation of $S^{(k)}$)**.** For any $k = 2, \ldots, K$ and given $(X^{(1)}, \ldots, X^{(k-1)}, I^{(k)})$, $S^{(k)}$ is randomly generated as follows:

Case 1: $I_1^{(k)} \neq I_2^{(k)}$.

Step 1: Set $\xi^{(k-1)} = (S^{(k-1)}_{I_1^{(k)},1}, \ldots, S^{(k-1)}_{I_1^{(k)},T}, 0, S^{(k-1)}_{I_2^{(k)},1}, \ldots, S^{(k-1)}_{I_2^{(k)},T}, 0)$.

Step 2: Generate $\xi^{(k)}$ as follows:

(a) Generate a random draw of ξ using the Euler algorithm given $\xi^{(k-1)}$.

(b) If $\xi_{T+1} = 0$, set $\xi^{(k)} = \xi$ and go to Step 3. Otherwise, return to (a).

Step 3: Given $\xi^{(k)}$, generate $S^{(k)}$ as follows:

(a) For every $i \notin I^{(k)}$, generate $(S^{(k)}_{i,1}, \ldots, S^{(k)}_{i,T})$ using the Euler algorithm given $(S^{(k-1)}_{i,1}, \ldots, S^{(k-1)}_{i,T})$.

(b) $(S^{(k)}_{I_1^{(k)},1}, \ldots, S^{(k)}_{I_1^{(k)},T}) = (\xi^{(k)}_1, \ldots, \xi^{(k)}_T)$.

(c) $(S^{(k)}_{I_2^{(k)},1}, \ldots, S^{(k)}_{I_2^{(k)},T}) = (\xi^{(k)}_{T+1}, \ldots, \xi^{(k)}_{2T+1})$.

Case 2: $I_1^{(k)} = I_2^{(k)}$. For every $i = 1, \ldots, n$, generate $(S^{(k)}_{i,1}, \ldots, S^{(k)}_{i,T})$ using the Euler algorithm given $(S^{(k-1)}_{i,1}, \ldots, S^{(k-1)}_{i,T})$. ■

Lemma A.2 shows that $S^{(k)}$ generated by Algorithm A.2 has the desired properties.

Lemma A.2. For any $k = 2, \dots, K$, $S^{(k)}$ generated by Algorithm A.2 satisfies the requirements of Step 2 of Algorithm 3.1, i.e., (7) holds.

Proof. We fix $k = 2, \dots, K$, $(X^{(1)}, \dots, X^{(k-1)})$, and a generic $\breve{S} \in \mathcal{S}^{nT}$ arbitrarily throughout this proof. We divide the proof in two cases.

Case 1: $I_1^{(k)} \neq I_2^{(k)}$. For $S^{(k-1)}$ and $S^{(k)}$ determined by $X^{(k-1)} = (S^{(k-1)}, A^{(k-1)})$ and $X^{(k)} = (S^{(k)}, A^{(k)})$, and for a generic $\breve{S} \in \mathcal{S}^{nT}$, we set

$$\begin{aligned}
\xi^{(k-1)} &= (S_{I_1^{(k-1)},1}, \dots, S_{I_1^{(k-1)},T}, 0, S_{I_2^{(k-1)},1}, \dots, S_{I_2^{(k-1)},T}, 0),\\
\xi^{(k)} &= (S_{I_1^{(k)},1}, \dots, S_{I_1^{(k)},T}, 0, S_{I_2^{(k)},1}, \dots, S_{I_2^{(k)},T}, 0),\\
\breve{\xi} &= (\breve{S}_{I_1^{(k)},1}, \dots, \breve{S}_{I_1^{(k)},T}, 0, \breve{S}_{I_2^{(k)},1}, \dots, \breve{S}_{I_2^{(k)},T}, 0).
\end{aligned}$$

Step 3 of Algorithm A.2 implies

$$P(S^{(k)} = \breve{S} \mid I^{(k)}, X^{(1)}, \dots, X^{(k-1)}) = \left\{ \begin{array}{c} P(\xi^{(k)} = \breve{\xi} \mid \xi^{(k-1)}) \times \\ \prod_{i \in (I^{(k)})^c} P((S_{i,1}^{(k)}, \dots, S_{i,T}^{(k)}) = (\breve{S}_{i,1}, \dots, \breve{S}_{i,T}) \mid S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)}) \end{array} \right\}, \tag{A-1}$$

Lemma A.1 implies that

$$P((S_{i,1}^{(k)}, \dots, S_{i,T}^{(k)}) = (\breve{S}_{i,1}, \dots, \breve{S}_{i,T}) \mid S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)}) = \frac{1\{(\breve{S}_{i,1}, \dots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})\}}{|R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})|} \tag{A-2}$$

for every $i \in (I^{(k)})^c$. In turn, Lemma A.3 implies that

$$P(\xi^{(k)} = \breve{\xi} \mid \xi^{(k-1)}) = \frac{1\{\breve{\xi} \in R_{S0}(\xi^{(k-1)}) : \breve{\xi}_{T+1} = 0\}}{|\{\tilde{\xi} \in R_{S0}(\xi^{(k-1)}) : \tilde{\xi}_{T+1} = 0\}|}. \tag{A-3}$$

By combining (A-1), (A-2), and (A-3),

$$P(S^{(k)} = \breve{S} \mid I^{(k)}, X^{(1)}, \dots, X^{(k-1)}) = \frac{1\{\breve{\xi} \in R_{S0}(\xi^{(k-1)}) : \breve{\xi}_{T+1} = 0\} \times \prod_{i \in (I^{(k)})^c} 1\{(\breve{S}_{i,1}, \dots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})\}}{|\{\tilde{\xi} \in R_{S0}(\xi^{(k-1)}) : \tilde{\xi}_{T+1} = 0\}| \times \prod_{i \in (I^{(k)})^c} |R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})|}. \tag{A-4}$$

To complete the proof, it suffices to show that the right-hand side of (A-4) is equal to the right-hand side of (7). To this end, it suffices to show that

$$\begin{aligned}
&1\{\breve{\xi} \in R_{S0}(\xi^{(k-1)}) : \breve{\xi}_{T+1} = 0\} \times \prod_{i \in (I^{(k)})^c} 1\{(\breve{S}_{i,1}, \dots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})\}\\
&= 1\{\breve{S} \in R_S(I^{(k)}, S^{(k-1)})\}
\end{aligned} \tag{A-5}$$

and

$$|\{\tilde{\xi} \in R_{S0}(\xi^{(k-1)}) : \tilde{\xi}_{T+1} = 0\}| \times \prod_{i \in (I^{(k)})^c} |R_{S0}(S_{i,1}^{(k-1)}, \dots, S_{i,T}^{(k-1)})| = |R_S(I^{(k)}, S^{(k-1)})|. \tag{A-6}$$

To show (A-5), consider the following derivation.

$$
\begin{aligned}
&1\{\breve{\xi} \in R_{S0}(\xi^{(k-1)}) : \breve{\xi}_{T+1} = 0\} \times \prod_{i \in (I^{(k)})^c} 1\{(\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})\} \\
&= \left\{ \begin{array}{c} 1\{(\breve{S}_{I_1^{(k)},1}, \ldots, \breve{S}_{I_1^{(k)},T}, 0, \breve{S}_{I_2^{(k)},1}, \ldots, \breve{S}_{I_2^{(k)},T}, 0) \in R_{S0}(\xi^{(k-1)})\} \\ \times \prod_{i \in (I^{(k)})^c} 1\{(\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})\} \end{array} \right\} \\
&\overset{(1)}{=} 1 \left\{ \begin{array}{c} \breve{S}_{i,1} = S_{i,1}^{(k-1)} \text{ for all } i = 1, \ldots, n, \\ \sum_{i \in I^{(k)}} \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\} = \sum_{i \in I^{(k)}} \sum_{t=1}^{T-1} 1\{S_{i,t}^{(k-1)} = s, S_{i,t+1}^{(k-1)} = s'\} \ \forall s, s' \in \mathcal{S}, \\ \sum_{i \in I^{(k)}} 1\{\breve{S}_{i,T} = s\} = \sum_{i \in I^{(k)}} 1\{S_{i,T}^{(k-1)} = s\} \text{ for all } s \in \mathcal{S}, \\ \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\} = \sum_{t=1}^{T-1} 1\{S_{i,t}^{(k-1)} = s, S_{i,t+1}^{(k-1)} = s'\} \text{ for all } s, s' \in \mathcal{S}, \ i \in (I^{(k)})^c \end{array} \right\} \\
&\overset{(2)}{=} 1 \left\{ \begin{array}{c} \breve{S}_{i,1} = S_{i,1}^{(k-1)} \text{ for all } i = 1, \ldots, n, \\ \sum_{i \in I^{(k)}} \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\} = \sum_{i \in I^{(k)}} \sum_{t=1}^{T-1} 1\{S_{i,t}^{(k-1)} = s, S_{i,t+1}^{(k-1)} = s'\} \ \forall s, s' \in \mathcal{S}, \\ \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{S}_{i,t+1} = s'\} = \sum_{t=1}^{T-1} 1\{S_{i,t}^{(k-1)} = s, S_{i,t+1}^{(k-1)} = s'\} \text{ for all } s, s' \in \mathcal{S}, \ i \in (I^{(k)})^c \end{array} \right\} \\
&\overset{(3)}{=} 1\{\breve{S} \in R_S(I^{(k)}, S^{(k-1)})\},
\end{aligned}
$$

as desired, where (1) follows from $I_1^{(k)} \neq I_2^{(k)}$ and applying Definition A.1, (2) follows from Lemma A.4, and (3) follows from Definition 3.1. To show (A-6), consider the following argument.

$$
\begin{aligned}
1 &= \sum_{\breve{S} \in \mathcal{S}} P(S^{(k)} = \breve{S} \mid I^{(k)}, X^{(1)}, \ldots, X^{(k-1)}) \\
&\overset{(1)}{=} \frac{\sum_{\breve{S} \in \mathcal{S}} 1\{\breve{S} \in R_S(I^{(k)}, S^{(k-1)})\}}{|\{\tilde{\xi} \in R_{S0}(\xi^{(k-1)}) : \tilde{\xi}_{T+1} = 0\}| \times \prod_{i \in (I^{(k)})^c} |R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})|} \\
&= \frac{|R_S(I^{(k)}, S^{(k-1)})|}{|\{\tilde{\xi} \in R_{S0}(\xi^{(k-1)}) : \tilde{\xi}_{T+1} = 0\}| \times \prod_{i \in (I^{(k)})^c} |R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})|},
\end{aligned}
$$

where (1) follows from combining (A-4) and (A-5). From here, (A-6) follows.

Case 2: $I_1^{(k)} = I_2^{(k)}$. Algorithm A.2 implies

$$
P(S^{(k)} = \breve{S} \mid I^{(k)}, X^{(1)}, \ldots, X^{(k-1)}) = \prod_{i=1}^{n} P((S_{i,1}^{(k)}, \ldots, S_{i,T}^{(k)}) = (\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \mid S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)}). \quad \text{(A-7)}
$$

Lemma A.1 implies that for every $i = 1, \ldots, n$,

$$
P((S_{i,1}^{(k)}, \ldots, S_{i,T}^{(k)}) = (\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \mid S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)}) = \frac{1\{(\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})\}}{|R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})|}. \quad \text{(A-8)}
$$

By combining (A-7) and (A-8)

$$
\begin{aligned}
P(S^{(k)} = \breve{S} \mid I^{(k)}, X^{(1)}, \ldots, X^{(k-1)}) &= \prod_{i=1}^{n} \frac{1\{(\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})\}}{|R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})|} \\
&= \frac{\prod_{i=1}^{n} 1\{(\breve{S}_{i,1}, \ldots, \breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})\}}{\prod_{i=1}^{n} |R_{S0}(S_{i,1}^{(k-1)}, \ldots, S_{i,T}^{(k-1)})|}. \quad \text{(A-9)}
\end{aligned}
$$

To complete the proof, it suffices to show that the right-hand side of (A-9) is equal to the right-hand side of

(7). To this end, it suffices to show that

$$\prod_{i=1}^{n} 1\{(\breve{S}_{i,1},\ldots,\breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)},\ldots,S_{i,T}^{(k-1)})\} = 1\{\breve{S} \in R_S(I^{(k)},S^{(k-1)})\} \tag{A-10}$$

$$\prod_{i=1}^{n} |R_{S0}(S_{i,1}^{(k-1)},\ldots,S_{i,T}^{(k-1)})| = |R_S(I^{(k)},S^{(k-1)})|. \tag{A-11}$$

To show (A-10), consider the following derivation.

$$\begin{aligned}
&\prod_{i=1}^{n} 1\{(\breve{S}_{i,1},\ldots,\breve{S}_{i,T}) \in R_{S0}(S_{i,1}^{(k-1)},\ldots,S_{i,T}^{(k-1)})\}\\
&\overset{(1)}{=} 1\left\{\begin{array}{c}\breve{S}_{i,1} = S_{i,1}^{(k-1)} \text{ for all } i=1,\ldots,n,\\ \sum_{t=1}^{T} 1\{\breve{S}_{i,t}=s,\breve{S}_{i,t+1}=s'\} = \sum_{t=1}^{T} 1\{S_{i,t}^{(k-1)}=s,S_{i,t+1}^{(k-1)}=s'\} \text{ for all } s,s'\in\mathcal{S} \text{ and } i=1,\ldots,n\end{array}\right\}\\
&\overset{(2)}{=} 1\{\breve{S} \in R_S(I^{(k)},S^{(k-1)})\},
\end{aligned}$$

as desired, where (1) follows from $I_1^{(k)} = I_2^{(k)}$ and applying Definition A.1 for each $i = 1,\ldots,n$, and (2) follows from Definition 3.1. Finally, (A-11) can be shown by using an argument that is analogous to the one used to prove (A-6). We omit this for the sake of brevity. ■

Lemma A.3. For any $k = 2,\ldots,K$, if $\xi^{(k)}$ is generated by Algorithm A.2, then $\xi^{(k)}$ is uniformly distributed over the set $\{\xi \in R_{S0}(\xi^{(k-1)}) : \xi_{T+1} = 0)\}$ conditional on $(I^{(k)},X^{(1)},\ldots,X^{(k-1)})$.

Proof. By Lemma A.1, $\tilde{\xi}$ in Step 2(a) of Algorithm A.2 follows the uniform distribution on $R_{S0}(\xi^{(k-1)})$, conditional on $(I^{(k)},X^{(1)},\ldots,X^{(k-1)})$. Steps 2(b) of Algorithm A.2 truncates the variable to the set $\{\xi \in R_{S0}(\xi^{(k-1)}) : \xi_{T+1} = 0\}$. The desired result then follows from the fact that a truncated version of a discrete uniform distribution is uniformly distributed on the truncated set. ■

Lemma A.4. For any $I \in \mathcal{I}$, if $\breve{S},\tilde{S} \in \mathcal{S}^{nT}$ satisfy the following conditions:

(a) $\tilde{S}_{i,1} = \breve{S}_{i,1}$ for all $i \in I$,

(b) $\sum_{i\in I}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=s,\tilde{S}_{i,t+1}=s'\} = \sum_{i\in I}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=s,\breve{S}_{i,t+1}=s'\}$ for all $s,s'\in\mathcal{S}$,

then, $\sum_{i\in I} 1\{\tilde{S}_{i,T}=s\} = \sum_{i\in I} 1\{\breve{S}_{i,T}=s\}$ for all $s\in\mathcal{S}$.

Proof. For every $i \in I$ and $s \in \mathcal{S}$, note that

$$\begin{aligned}
1\{\breve{S}_{i,T}=s\} &= \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t+1}=s\} - \sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=s\} + 1\{\breve{S}_{i,1}=s\}\\
&= \sum_{\bar{s}\in\mathcal{S}}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=\bar{s},\breve{S}_{i,t+1}=s\} - \sum_{\bar{s}\in\mathcal{S}}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=s,\breve{S}_{i,t+1}=\bar{s}\} + 1\{\breve{S}_{i,1}=s\}. \qquad \text{(A-12)}
\end{aligned}$$

By the same argument applied to $\tilde{S} \in R_S(I,\breve{S})$, we have that for every $i \in I$ and $s \in \mathcal{S}$,

$$1\{\tilde{S}_{i,T}=s\} = \sum_{\bar{s}\in\mathcal{S}}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=\bar{s},\tilde{S}_{i,t+1}=s\} - \sum_{\bar{s}\in\mathcal{S}}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=s,\tilde{S}_{i,t+1}=\bar{s}'\} + 1\{\tilde{S}_{i,1}=s\}. \tag{A-13}$$

To show this lemma, fix $s \in \mathcal{S}$ arbitrarily and consider the following argument.

$$\begin{aligned}\sum_{i\in I} 1\{\breve{S}_{i,T}=s\} &\overset{(1)}{=} \sum_{\bar{s}\in\mathcal{S}}\sum_{i\in I}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=\bar{s},\breve{S}_{i,t+1}=s\} - \sum_{\bar{s}\in\mathcal{S}}\sum_{i\in I}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t}=s,\breve{S}_{i,t+1}=\bar{s}\} + \sum_{i\in I} 1\{\breve{S}_{i,1}=s\}\\ &\overset{(2)}{=} \sum_{\bar{s}\in\mathcal{S}}\sum_{i\in I}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=\bar{s},\tilde{S}_{i,t+1}=s\} - \sum_{\bar{s}\in\mathcal{S}}\sum_{i\in I}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t}=s,\tilde{S}_{i,t+1}=\bar{s}\} + \sum_{i\in I} 1\{\tilde{S}_{i,1}=s\}\\ &\overset{(3)}{=} \sum_{i\in I} 1\{\tilde{S}_{i,T}=s\},\end{aligned}$$

where (1) holds by (A-12), (2) holds by conditions (a)-(b), and (3) holds by (A-13). ■

A.2 Implementation of Step 3 in Algorithm 3.1

For any $k = 2, \ldots, K$, $X^{(k-1)} \in \mathcal{X}$, and $S^{(k)} \in \mathcal{S}^{nT}$, Step 3 of Algorithm 3.1 draws $A^{(k)}$ uniformly within $R_A(S^{(k)}, X^{(k-1)})$. This can be implemented by the following algorithm.

Algorithm A.3 (Generation of $A^{(k)}$)**.** For any $k = 2, \ldots, K$ and given $(X^{(1)}, \ldots, X^{(k-1)}, I^k, S^{(k)})$, $A^{(k)}$ is randomly generated as follows

Step 1: For every $(s, s') \in \mathcal{S} \times \mathcal{S}$, define

$$\begin{aligned}&\text{Index}^{(k-1)}(s,s') = \{(i,t) \in \{1,\ldots,n\}\times\{1,\ldots,T-1\} : (S^{(k-1)}_{i,t}, S^{(k-1)}_{i,t+1}) = (s,s')\}\\ &\text{Index}^{(k)}(s,s') = \{(i,t) \in \{1,\ldots,n\}\times\{1,\ldots,T-1\} : (S^{(k)}_{i,t}, S^{(k)}_{i,t+1}) = (s,s')\}\\ &\text{Index}^{(k-1)}(s) = \{(i,T) : i \in \{1,\ldots,n\}, S^{(k-1)}_{i,T} = s\}\\ &\text{Index}^{(k)}(s) = \{(i,T) : i \in \{1,\ldots,n\}, S^{(k)}_{i,T} = s\}.\end{aligned}$$

Step 2: For every $(s, s') \in \mathcal{S} \times \mathcal{S}$, we generate $(A^{(k)}_{i,t} : (i,t) \in \text{Index}^{(k)}(s,s'))$ by uniformly sampling from $(A^{(k-1)}_{i,t} : (i,t) \in \text{Index}^{(k-1)}(s,s'))$ without replacement, i.e., a uniformly chosen permutation.

Step 3: For every $s \in \mathcal{S}$, we construct $(A^{(k)}_{i,T} : (i,T) \in \text{Index}^{(k)}(s))$ by uniformly sampling from the discrete set $(A^{(k-1)}_{i,T} : (i,T) \in \text{Index}^{(k-1)}(s))$ without replacement, i.e., a uniformly chosen permutation. ■

Lemma A.5 shows that $A^{(k)}$ generated by Algorithm A.3 has the desired properties.

Lemma A.5. For any $k = 2, \ldots, K$, the outcome, $A^{(k)}$, of Algorithm A.3 satisfies the requirements of Step 3 of Algorithm 3.1, i.e., (8) holds.

Proof. This follows from noting that any element of $R_A(S^{(k)}, X^{(k-1)})$ corresponds to a restricted set of permutations of the action data, and Algorithm A.3 chooses an element uniformly within this set. ■

A.3 Proof of Theorem 3.1

By (3), (5) is equivalent to $\liminf_{K\to\infty}(\alpha - P(\hat{p}_K \le \alpha)) \ge 0$. In this proof, we are going to show a stronger statement (cf. Lehmann and Romano, 2005, Eq. (15.6)):

$$\liminf_{K\to\infty} \inf_{u\in[0,1]} (u - P(\hat{p}_K \le u)) \quad\ge\quad 0.$$

Fix $\varepsilon > 0$ and $u \in [0,1]$ arbitrarily. The rest of the proof is going to show

$$u - P(\hat{p}_K \leq u) \geq -2\varepsilon$$

for sufficiently large K. For any positive integer K, let

$$\mathcal{E}_K \equiv \left\{ \sup_{t\in\mathbb{R}} \left| \frac{1}{K}\sum_{k=1}^{K} 1\{\tau(X^{(k)}) \geq t\} - \frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} 1\{\tau(g(X)) \geq t\} \right| > \varepsilon \right\}.$$

By Lemma 4.4, for sufficiently large K,

$$P(\mathcal{E}_K) \leq \varepsilon. \tag{A-14}$$

For any positive integer K, consider the following derivation:

$$\begin{aligned} P(\hat{p}_K \leq u) &= P\left(\frac{1}{K}\sum_{k=1}^{K} 1\{\tau(X^{(k)}) \geq \tau(X)\} \leq u\right) \\ &\leq P\left(\frac{1}{|\mathbf{G}|}\sum_{g\in\mathbf{G}} 1\{\tau(g(X)) \geq \tau(X)\} \leq u + \varepsilon\right) + P(\mathcal{E}_K) \\ &\overset{(1)}{\leq} u + \varepsilon + P(\mathcal{E}_K), \end{aligned} \tag{A-15}$$

where (1) holds by Lemma 4.3. By (A-15) and (A-14), we conclude that, for sufficiently large K, $P(\hat{p}_K \leq u) \leq u + 2\varepsilon$ or, equivalently, $u - P(\hat{p}_K \leq u) \geq -2\varepsilon$, as desired. ■

B Appendix to Section 4

B.1 Proof of lemmas

Proof of Lemma 4.1. Note that

$$P(S = \tilde{S}) = \prod_{i=1}^{n}\left(m_i(\tilde{S}_{i,1})\prod_{t=1}^{T-1}\left(\sum_{a\in\mathcal{A}} g(\tilde{S}_{i,t+1}|a,\tilde{S}_{i,t})\sigma(a|\tilde{S}_{i,t})\right)\right). \tag{B-16}$$

This equation follows from the following derivation

$$\begin{aligned} P(S = \tilde{S}) &\overset{(1)}{=} \prod_{i=1}^{n}\left(P(S_{i,1} = \tilde{S}_{i,1})\prod_{t=1}^{T-1} P(S_{i,t} = \tilde{S}_{i,t}|(S_{i,1},\ldots,S_{i,t-1}) = (\tilde{S}_{i,1},\ldots,\tilde{S}_{i,t-1}))\right) \\ &\overset{(2)}{=} \prod_{i=1}^{n}\left(P(S_{i,1} = \tilde{S}_{i,1})\prod_{t=1}^{T-1} P(S_{i,t} = \tilde{S}_{i,t}|S_{i,t-1} = \tilde{S}_{i,t-1})\right) \\ &= \prod_{i=1}^{n}\left(P(S_{i,1} = \tilde{S}_{i,1})\prod_{t=1}^{T-1}\left(\sum_{a\in\mathcal{A}}\left(\begin{array}{c} P(S_{i,t} = \tilde{S}_{i,t}|A_{i,t-1} = a, S_{i,t-1} = \tilde{S}_{i,t-1}) \\ \times P(A_{i,t-1} = a|S_{i,t-1} = \tilde{S}_{i,t-1}) \end{array}\right)\right)\right) \\ &\overset{(3)}{=} \prod_{i=1}^{n}\left(m_i(\tilde{S}_{i,1})\prod_{t=1}^{T-1}\left(\sum_{a\in\mathcal{A}} g(\tilde{S}_{i,t+1}|a,\tilde{S}_{i,t})\sigma(a|\tilde{S}_{i,t})\right)\right), \end{aligned}$$

where (1) holds by Assumption 2.1(a), (2) holds by Lemma B.2, and (3) holds holds under H_0 in (1).

To conclude the proof, it suffices to show (11) and (12). To this end, consider the following derivation.

$$\begin{aligned} P(X=\tilde{X}) &\overset{(1)}{=} \prod_{i=1}^{n}\left(m_i(\tilde{S}_{i,1})\sigma(\tilde{A}_{i,T}|\tilde{S}_{i,T})\prod_{t=1}^{T-1}\left(\sigma(\tilde{A}_{i,t}|\tilde{S}_{i,t})g(\tilde{S}_{i,t+1}|\tilde{S}_{i,t},\tilde{A}_{i,t})\right)\right)\\ &\overset{(2)}{=} P(S=\tilde{S})\left(\sigma(\tilde{A}_{i,T}|\tilde{S}_{i,T})\left(\prod_{t=1}^{T-1}\frac{\sigma(\tilde{A}_{i,t}|\tilde{S}_{i,t})g(\tilde{S}_{i,t+1}|\tilde{S}_{i,t},\tilde{A}_{i,t})}{\sum_{a\in\mathcal{A}}g(\tilde{S}_{i,t+1}|a,\tilde{S}_{i,t})\sigma(a|\tilde{S}_{i,t})}\right)\right), \end{aligned} \tag{B-17}$$

where (1) holds by (2), which is shown in Lemma B.1, and (2) holds by (B-16). By combining (10) and (B-17), we conclude that

$$P(A=\tilde{A}|S=\tilde{S}) = \prod_{i=1}^{n}\left(\sigma(\tilde{A}_{i,T}|\tilde{S}_{i,T})\left(\prod_{t=1}^{T-1}\frac{\sigma(\tilde{A}_{i,t}|\tilde{S}_{i,t})g(\tilde{S}_{i,t+1}|\tilde{S}_{i,t},\tilde{A}_{i,t})}{\sum_{a\in\mathcal{A}}g(\tilde{S}_{i,t+1}|a,\tilde{S}_{i,t})\sigma(a|\tilde{S}_{i,t})}\right)\right).$$

By re-expressing this equation in terms of counts of $(s,a,s')\in\mathcal{S}\times\mathcal{A}\times\mathcal{S}$, (11) follows. Moreover, (12) follows from re-expressing (B-16) in terms of individual counts of each $(s,s')\in\mathcal{S}\times\mathcal{S}$. ■

Proof of Lemma 4.2. We first show that $\mathbf{G}$ is a collection of transformations from $\mathcal{X}$ onto itself. Consider any $g\in\mathbf{G}$. By definition, g is the composition of a finite number of transformations in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$, i.e., $g=g^{(K)}\circ\cdots\circ g^{(1)}$ with $(g^{(1)},\ldots,g^{(K)})\in\mathbf{G}(I^{(1)})\times\cdots\times\mathbf{G}(I^{(K)})$ with $I^{(j)}\in\mathcal{I}$ for $j=1,\ldots,K$. By Lemma B.3, $g^{(j)}\in\mathbf{G}(I^{(j)})$ are onto transformations from $\mathcal{X}$ to itself. From this, we can conclude that $g=g^{(K)}\circ\cdots\circ g^{(1)}$ is an onto transformation from $\mathcal{X}$ to itself, as desired.

Second, we show that $\mathbf{G}$ is a group. To this end, it suffices to verify conditions (i)-(iv) in Lehmann and Romano (2005, Section A.1). To verify condition (i), consider arbitrary $g_1,g_2\in\mathbf{G}$. By definition, this implies g_1 and g_2 are compositions of a finite number of transformations in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$. Then, $g_2\circ g_1$ is a composition of a finite number of elements in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$, and so $g_2\circ g_1\in\mathbf{G}$. Condition (ii) follows from the argument in Lehmann and Romano (2005, page 693). Condition (iii) follows from the fact that $\mathbf{G}(I)$ is a group for any for any $I\in\mathcal{I}$ (shown in Lemma B.3), and so it includes the identity transformation. To verify condition (iv), consider the following argument for any arbitrary $g\in\mathbf{G}$. By definition, g is the composition of a finite number of transformations in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$, i.e., $g=g^{(K)}\circ\cdots\circ g^{(1)}$ with $(g^{(1)},\ldots,g^{(K)})\in\mathbf{G}(I^{(1)})\times\cdots\times\mathbf{G}(I^{(K)})$ with $I^{(j)}\in\mathcal{I}$ for $j=1,\ldots,K$. By Lemma B.3, $\mathbf{G}(I^{(j)})$ is a group for each $j=1,\ldots,K$. From this, we can conclude that $\exists(g^{(j)})^{-1}\in\mathbf{G}(I^{(j)})$ for each $j=1,\ldots,K$. Since $g\circ\tilde{g}$ and $\tilde{g}\circ g$ are equal to the identity transformation, $\tilde{g}=g^{-1}$. Finally, note that $g^{-1}=(g^{(1)})^{-1}\circ\cdots\circ(g^{(K)})^{-1}$ is the compositions of a finite number of transformations in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$ and so $g^{-1}\in\mathbf{G}$, as desired.

To complete the proof, it suffices to show that, for any $\tilde{X}\in\mathcal{X}$ and $g\in\mathbf{G}$, $\tilde{X}$ and $g\tilde{X}$ have the same sufficient statistics in (13). g is the composition of a finite number of transformations in $\bigcup_{I\in\mathcal{I}}\mathbf{G}(I)$, i.e., $g=g^{(K)}\circ\cdots\circ g^{(1)}$ with $(g^{(1)},\ldots,g^{(K)})\in\mathbf{G}(I^{(1)})\times\cdots\times\mathbf{G}(I^{(K)})$ with $I^{(j)}\in\mathcal{I}$ for $j=1,\ldots,K$. Therefore, $g\tilde{X}=g^{(K)}\circ\cdots\circ g^{(1)}\tilde{X}$. For each $j=1,\ldots,K$, Lemma B.4 implies that, for any $\check{X}\in\mathcal{X}$ and $g^{(j)}\in\mathbf{G}(I^{(j)})$, $g^{(j)}\check{X}$ and $\check{X}$ have the same sufficient statistic in (13). From these observations and by finite induction, it follows that $\tilde{X}$ and $g\tilde{X}$ have the same sufficient statistics in (13), as desired. ■

Proof of Lemma 4.3. By Lemma 4.2, we know (i) $\mathbf{G}$ is a finite group of transformations of $\mathcal{X}$ onto itself, and (ii) if X satisfies H_0 in (19), then X and gX have the same sufficient statistics in (13) for any $g\in\mathbf{G}$. The second statement, together with Lemma 4.1, implies that the randomization hypothesis holds (Lehmann and Romano 2005, Definition 15.2.1), i.e., if X satisfies H_0 in (1), its distribution is invariant under the transformations in $\mathbf{G}$. Under these conditions, the result follows from Lehmann and Romano (2005, Eq. (15.6) and Problem 15.2). ■

Proof of Lemma 4.4. We condition on $X \in \mathcal{X}$ throughout this proof. Let $(G^{(1)}, \ldots, G^{(K)})$ be as in Definition B.1. By Lemma B.5, it suffices to show that

$$\sup_{t \in \mathbb{R}} \left| \frac{1}{K} \sum_{k=1}^{K} 1\{\tau(G^{(k)}X) \geq t\} - \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} 1\{\tau(gX) \geq t\} \right| \overset{a.s.}{\to} 0 \quad \text{as } K \to \infty. \tag{B-18}$$

For any $k = 1, \ldots, K$, Definition B.1 implies that $G^{(k)}X \in \mathcal{X}$. Thus, $\tau(G^{(k)}X)$ takes values in the finite set $\{\tau(\tilde{X}) : \tilde{X} \in \mathcal{X}\}$. It then suffices to show the pointwise version of (B-18), i.e.,

$$\frac{1}{K} \sum_{k=1}^{K} 1\{\tau(G^{(k)}X) \geq t\} \quad \overset{a.s.}{\to} \quad \frac{1}{|\mathbf{G}|} \sum_{g \in \mathbf{G}} 1\{\tau(gX) \geq t\} \quad \text{as } K \to \infty.$$

By Definition B.1, $(G^{(1)}, \ldots, G^{(K)})$ is the result of a Markov chain with transition probability given in (B-22). By Robert and Casella (2004, Algorithm A-24 and pages 270-1), we can equivalently interpret $(G^{(1)}, \ldots, G^{(K)})$ as the outcome of a Metropolis-Hastings algorithm. For any $g, \breve{g} \in \mathbf{G}$, this Metropolis-Hastings algorithm has a conditional density $q(\breve{g} \mid g) \equiv P(G^{(k+1)} = \breve{g} | G^{(k)} = g)$, a target probability f defined by

$$f(g) \;\equiv\; \frac{1\{g \in \mathbf{G}\}}{|\mathbf{G}|}, \tag{B-19}$$

and Metropolis-Hastings acceptance probability equal to one. To show the latter, note that, for every $g, \breve{g} \in \mathbf{G}$,

$$\rho(g, \breve{g}) \;=\; \min\left\{ \frac{f(\breve{g})}{f(g)} \times \frac{q(g \mid \breve{g})}{q(\breve{g} \mid g)},\ 1 \right\} \overset{(1)}{=} 1,$$

where (1) uses that $f(\breve{g}) = f(g) = 1/|\mathbf{G}|$ and $q(g \mid \breve{g}) = q(\breve{g} \mid g)$ by (B-19) and Lemma B.9, respectively. By this and Robert and Casella (2004, Theorem 7.4), it suffices to show that the conditional density $q(\breve{g} \mid g)$ is f-irreducible. By Robert and Casella (2004, Theorem 6.15, part (i)), this follows from showing that, for any $g, \breve{g} \in \mathbf{G}$ (and so $f(g) > 0$ and $f(\breve{g}) > 0$), the Markov chain has a positive probability of transitioning from g to $\breve{g}$ after a sufficient number of steps. We devote the rest of the proof to show this.

Consider any arbitrary choice of $g, \breve{g} \in \mathbf{G}$. Since $\mathbf{G}$ is the group generated by finitely many compositions of elements in $\bigcup_{I \in \mathcal{I}} \mathbf{G}(I)$, there are $(g^{(1)}, \ldots, g^{(K_1+K_2)}) \in \mathbf{G}(I^{(1)}) \times \cdots \times \mathbf{G}(I^{(K_1+K_2)})$ with $I^{(j)} \in \mathcal{I}$ for $j = 1, \ldots, K$ such that $g = g^{(K_1)} \circ \cdots \circ g^{(1)}$ and $\breve{g} = g^{(K_1+K_2)} \circ \cdots \circ g^{(K_1+1)}$. By Lemma B.3, $\mathbf{G}(I^{(j)})$ is a group for all $j = 1, \ldots, K_1 + K_2$, and so $(g^{(j)})^{-1} \in \mathbf{G}(I^{(j)})$ for every $j = 1, \ldots, K_1 + K_2$. Then, note that

$$\begin{aligned} \breve{g} \;&=\; \breve{g} \circ g^{-1} \circ g \\ &\overset{(1)}{=} g^{(K_1+K_2)} \circ \cdots \circ g^{(K_1+1)} \circ (g^{(1)})^{-1} \circ \cdots \circ (g^{(K_1)})^{-1} \circ g \\ &\overset{(2)}{=} \breve{g}^{(K_1+K_2)} \circ \cdots \circ \breve{g}^{(K_1+1)} \circ \breve{g}^{(K_1)} \circ \cdots \circ \breve{g}^{(1)} \circ g, \end{aligned} \tag{B-20}$$

where (1) holds by setting $\breve{g} = g^{(K_1+K_2)} \circ \cdots \circ g^{(K_1+1)}$ and $g^{-1} = (g^{(1)})^{-1} \circ \cdots \circ (g^{(K_1)})^{-1}$, and (2) holds by defining $(\breve{g}^{(1)}, \ldots, \breve{g}^{(K_1+K_2)}) = ((g^{(K_1)})^{-1}, \ldots, (g^{(1)})^{-1}, g^{(K_1+1)}, \ldots, g^{(K_1+K_2)})$. Note that (B-20) provides a specific path for transitioning from g to $\breve{g}$ after $K_1 + K_2$ steps. We complete the proof by showing that

$P(G^{(K_1+K_2+k)} = \breve{g}|G^{(k)} = g) > 0$ for any posive integer k. To this end, we define $(\breve{I}^{(1)}, \ldots, \breve{I}^{(K_1+K_2)}) = (I^{(K_1)}, \ldots, I^{(1)}, I^{(K_1+1)}, \ldots, I^{(K_1+K_2)})$ and we consider the following argument:

$$
\begin{aligned}
P(G^{(K_1+K_2+k)} = \breve{g}|G^{(k)} = g) &\overset{(1)}{\geq} q(\breve{g}^{(1)} \circ g|g) \prod_{k=2}^{K} q(\breve{g}^{(k)} \circ \cdots \circ \breve{g}^{(1)} \circ g \mid \breve{g}^{(k-1)} \circ \cdots \circ \breve{g}^{(1)} \circ g) \\
&\overset{(2)}{\geq} \prod_{j=1}^{K_1+K_2} \frac{1}{|\mathcal{I}||\mathbf{G}(\breve{I}^j)|} \overset{(3)}{>} 0,
\end{aligned}
$$

where (1) uses the fact that the conditional distribution of $G^{(j+1)}$ given $G^{(j)}$ is q for all $j = 1, \ldots, K_1 + K_2$, (2) holds by (B-22) and $\breve{g}^{(j)} \in \breve{I}^{(j)}$ for all $j = 1, \ldots, K_1 + K_2$, and (3) holds because $\breve{I}^{(j)} \in \mathcal{I}$ for all $j = 1, \ldots, K_1 + K_2$. ■

B.2 Auxiliary Lemmas

Lemma B.1. Under Assumptions 2.1 and H_0 in (1), (2) holds.

Proof. Consider the following derivation.

$$
\begin{aligned}
P(X = \tilde{X}) &\overset{(1)}{=} \prod_{i=1}^{n} P((S_{i,t}, A_{i,t}) = (\tilde{S}_{i,t}, \tilde{A}_{i,t}) : t = 1, \ldots, T) \\
&\overset{(2)}{=} \prod_{i=1}^{n} \left[\begin{array}{c} P(S_{i,1} = \tilde{S}_{i,1}, A_{i,1} = \tilde{A}_{i,1}) \times \\ \prod_{t=2}^{T} P((S_{i,t}, A_{i,t}) = (\tilde{S}_{i,t}, \tilde{A}_{i,t}) | (S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,t-1}, \tilde{A}_{i,t-1})) \end{array} \right] \\
&\overset{(3)}{=} \prod_{i=1}^{n} \left[\begin{array}{c} P(S_{i,1} = \tilde{S}_{i,1}) \left(\prod_{t=1}^{T} P(A_{i,t} = \tilde{A}_{i,t} | S_{i,t} = \tilde{S}_{i,t}) \right) \\ \times \left(\prod_{t=1}^{T-1} P(S_{i,t+1} = \tilde{S}_{i,t+1} | S_{i,t} = \tilde{S}_{i,t}, A_{i,t} = \tilde{A}_{i,t}) \right) \end{array} \right] \\
&\overset{(4)}{=} \prod_{i=1}^{n} \left[m_i(\tilde{S}_{i,1}) \left(\prod_{t=1}^{T} \sigma(\tilde{A}_{i,t} | \tilde{S}_{i,t}) \right) \left(\prod_{t=1}^{T-1} g(\tilde{S}_{i,t+1} | \tilde{S}_{i,t}, \tilde{A}_{i,t}) \right) \right],
\end{aligned}
$$

where (1) holds by Assumption 2.1(a), (2) holds by Assumption 2.1(b), (3) holds by Assumption 2.1(c), and (4) holds under H_0 in (1). ■

Lemma B.2. Under Assumptions 2.1(b)-(c), the state variable is Markovian, i.e., for every $i = 1, \ldots, n$ and $t = 2, \ldots T$ and every $\tilde{S} \in \mathcal{S}^{nT}$,

$$
P(S_{i,t} = \tilde{S}_{i,t} | S_{i,t-1} = \tilde{S}_{i,t-1}) = P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})). \tag{B-21}
$$

Proof. Fix $i = 1, \ldots, n$, $t = 2, \ldots T$, and $\tilde{S} \in \mathcal{S}^{nT}$ arbitrarily. Consider the following argument.

$$
\begin{aligned}
& P((S_{i,1}, \ldots, S_{i,t}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t})) \\
& = \sum_{(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}} \begin{pmatrix} P((S_{i,1}, A_{i,1}, \ldots, S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,1}, a_1, \ldots, \tilde{S}_{i,t-1}, a_{t-1})) \times \\ P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,1}, A_{i,1}, \ldots, S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,1}, a_1, \ldots, \tilde{S}_{i,t-1}, a_{t-1})) \end{pmatrix} \\
& \overset{(1)}{=} \sum_{(a_1,\ldots,a_{t-1}) \in \mathcal{A}^{t-1}} \begin{pmatrix} P((S_{i,1}, A_{i,1}, \ldots, S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,1}, a_1, \ldots, \tilde{S}_{i,t-1}, a_{t-1})) \\ \times P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,t-1}, a_{t-1})) \end{pmatrix} \\
& = \sum_{a_{t-1} \in \mathcal{A}} \begin{pmatrix} P((S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1}), A_{i,t-1} = a_{t-1}) \\ \times P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,t-1}, a_{t-1})) \end{pmatrix} \\
& = \sum_{a_{t-1} \in \mathcal{A}} \begin{pmatrix} P(A_{i,t-1} = a_{t-1} | (S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})) \\ \times P((S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})) \\ \times P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,t-1}, a_{t-1})) \end{pmatrix} \\
& \overset{(2)}{=} \sum_{a_{t-1} \in \mathcal{A}} \begin{pmatrix} P(A_{i,t-1} = a_{t-1} | S_{i,t-1} = \tilde{S}_{i,1}) P((S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})) \\ \times P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,t-1}, A_{i,t-1}) = (\tilde{S}_{i,t-1}, a_{t-1})) \end{pmatrix} \\
& = P(S_{i,t} = \tilde{S}_{i,t} | S_{i,t-1} = \tilde{S}_{i,t-1}) P((S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})),
\end{aligned}
$$

where (1) holds by Assumption 2.1(b) and (2) holds by Assumption 2.1(c). Therefore,

$$
P(S_{i,t} = \tilde{S}_{i,t} | S_{i,t-1} = \tilde{S}_{i,t-1}) = P(S_{i,t} = \tilde{S}_{i,t} | (S_{i,1}, \ldots, S_{i,t-1}) = (\tilde{S}_{i,1}, \ldots, \tilde{S}_{i,t-1})),
$$

as desired. ∎

Lemma B.3. For any $I \in \mathcal{I}$, $\mathbf{G}(I)$ is a group.

Proof. We fix $I \in \mathcal{I}$ arbitrarily. It suffices to verify conditions (i)-(iv) in Lehmann and Romano (2005, Section A.1). Note that we can verify condition (ii) using the same argument as in Lehmann and Romano (2005, page 693).

We begin with condition (i). First, for any arbitrary $g_1, g_2 \in \mathbf{G}(I)$, we now verify that $g_2 \circ g_1 \in \mathbf{G}(I)$. Since $g_1, g_2 \in \mathbf{G}(I)$, g_1 and g_2 are both onto transformations of $\mathcal{X}$ onto itself, then $g_2 \circ g_1$ is an onto transformation of $\mathcal{X}$ onto itself. Now we will show that, for any $(\breve{S}, \breve{A}) \in \mathcal{X}$, the data configuration $(\tilde{S}, \tilde{A}) = (g_2 \circ g_1)(\breve{S}, \breve{A})$ satisfies $\tilde{S} \in R_S(I, \breve{S})$ and $\tilde{A} \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$. Define $(\dot{S}, \dot{A}) = g_1(\breve{S}, \breve{A})$. Now $(\dot{S}, \dot{A}) = g_1(\breve{S}, \breve{A})$ and $(\tilde{S}, \tilde{A}) = g_2(\dot{S}, \dot{A})$. Since $g_1, g_2 \in \mathbf{G}(I)$, all the conditions in Definitions 3.1 and 3.2 satisfy the transitive property as the equality condition, so that $\tilde{S} \in R_S(I, \breve{S})$ and $\tilde{A} \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$, as desired. By combining these results, we conclude that $g_2 \circ g_1 \in \mathbf{G}(I)$, as desired.

To verify condition (iii), we now show that the identity transformation belongs to $\mathbf{G}(I)$. To this end, we note that the identity transformation is an onto transformation of $\mathcal{X}$ onto itself, and $\breve{S} \in R_S(I, \breve{S})$ and $\breve{A} \in R_A(\breve{S}, (\breve{S}, \breve{A}))$.

To verify condition (iv), we now show that for any $g \in \mathbf{G}(I)$, $g^{-1} \in \mathbf{G}(I)$ holds. By definition $\mathbf{G}(I)$ is a collection of onto transformations that map a finite set $\mathcal{X}$ onto itself. By the pigeonhole principle, the transformations in $\mathbf{G}(I)$ are one to one, i.e., bijective, implying that g^{-1} is well defined. First, note that g^{-1} is a bijective transformation (hence, an onto transformation) of $\mathcal{X}$ onto itself. For the rest of the verification of Condition (iv), pick $\tilde{X} \in \mathcal{X}$ arbitrarily. Second, we would like to show that, for any $(\breve{S}, \breve{A}) \in \mathcal{X}$, the data configuration $(\tilde{S},\tilde{A}) = g^{-1}(\breve{S},\breve{A})$ satisfies $\tilde{S} \in R_S(I,\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},(\breve{S},\breve{A}))$. Since $g \in \mathbf{G}(I)$ and $g(\tilde{S},\tilde{A}) = g(g^{-1}(\breve{S},\breve{A})) = (\breve{S},\breve{A})$, we have $\breve{S} \in R_S(I,\tilde{S})$ and $\breve{A} \in R_A(\breve{S},(\tilde{S},\tilde{A}))$. Note that all the conditions in Definitions 3.1 and 3.2 treat $(\tilde{S},\tilde{A})$ and $(\breve{S},\breve{A})$ symmetrically. Therefore, we have $\tilde{S} \in R_S(I,\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},(\breve{S},\breve{A}))$, as desired. ■

Lemma B.4. For any $I \in \mathcal{I}$ and any $g \in \mathbf{G}(I)$, $\breve{X}$ and $g\breve{X}$ have the same sufficient statistic in (13), i.e., $U(\breve{X}) = U(g\breve{X})$.

Proof. Let $\breve{X} = (\breve{S},\breve{A})$ and $\tilde{X} = (\tilde{S},\tilde{A}) = g(\breve{S},\breve{A})$. By definition 4.1, this implies that $\tilde{S} \in R_S(\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},(\breve{S},\breve{A}))$. By (13), it then suffices to show the following statements:

1. $\breve{S}_{i,1} = \tilde{S}_{i,1}$ for all $i = 1,\ldots,n$,
2. $\sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t} = s, \tilde{A}_{i,t} = a, \tilde{S}_{i,t+1} = s'\} = \sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{A}_{i,t} = a, \breve{S}_{i,t+1} = s'\}$ for all $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}$,
3. $\sum_{i=1}^{n} 1\{\tilde{S}_{i,T} = s, \tilde{A}_{i,T} = a\} = \sum_{i=1}^{n} 1\{\breve{S}_{i,T} = s, \breve{A}_{i,T} = a\}$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$.

The first statement follows from $\tilde{S} \in R_S(\breve{S})$ and condition (a) in Definition 3.1. The second and third statements follow from $\tilde{A} \in R_A(\tilde{S},(\breve{S},\breve{A}))$ and conditions (a) and (b) in Definition 3.2, respectively. ■

Several upcoming results involve a Markov chain of transformations in $\mathbf{G}$, specified in Definition B.1.

Definition B.1. Let $(G^{(1)},\ldots,G^{(K)})$ denote a Markov chain of transformations of $\mathcal{X}$ onto itself that is defined as follows:

- $G^{(1)} : \mathcal{X} \to \mathcal{X}$ be equal to the identity transformation, i.e., $x = G^{(1)}x$ for any $x \in \mathcal{X}$
- For any $k = 2,\ldots,K$ and given $(G^{(1)},\ldots,G^{(k-1)},X)$, $G^{(k)} : \mathcal{X} \to \mathcal{X}$ is a random transformation distributed according to the following transition probability:

$$P(G^{(k)} = \tilde{g} \mid G^{(1)},\ldots,G^{(k-1)},X) = P(G^{(k)} = \tilde{g} \mid G^{(k-1)}) = \sum_{I\in\mathcal{I}}\sum_{g\in\mathbf{G}(I)} \frac{1\{\tilde{g} = g \circ (G^{(k-1)})\}}{|\mathcal{I}| \times |\mathbf{G}(I)|}. \quad \text{(B-22)}$$

Lemma B.5. Conditional on X, $(X^{(1)},\ldots,X^{(K)})$ generated by Algorithm 3.1 and $(G^{(1)}X,\ldots,G^{(K)}X)$ with $(G^{(1)},\ldots,G^{(K)})$ as in Definition B.1 have the same distribution.

Proof. We condition on X throughout this proof. First, note that Algorithm 3.1 and Definition B.1 imply that $X = X^{(1)} = G^{(1)}X$. Second, note that $(X^{(1)},\ldots,X^{(K)})$ and $(G^{(1)}X,\ldots,G^{(K)}X)$ are both Markov chains in $\mathcal{X}$. To complete the proof, it suffices to show that they have the same transition probabilities. The transition probability of $(X^{(1)},\ldots,X^{(K)})$ is specified in (9). It then suffices to show that, for any $k = 2,\ldots,K$, $\tilde{X} = (\tilde{S},\tilde{A}) \in \mathcal{X}$, and $G^{(k-1)}X = \breve{X} = (\breve{S},\breve{A}) \in \mathcal{X}$,

$$\begin{aligned}
&P(G^{(k)}X = \tilde{X} \mid G^{(1)}X,\ldots,G^{(k-2)}X, G^{(k-1)}X = \breve{X}, X) \\
&= P(G^{(k)}X = \tilde{X} \mid G^{(k-1)}X = \breve{X}, X) \\
&= \begin{cases} \sum_{I\in\mathcal{I}} \dfrac{1\{\tilde{S} \in R_S(I,\breve{S}), \tilde{A} \in R_A(\tilde{S},\breve{X})\}}{|\mathcal{I}| \times |R_S(I,\breve{S})| \times |R_A(\tilde{S},\breve{X})|} & \text{if } |R_S(I,\breve{S})| \times |R_A(\tilde{S},\breve{X})| > 0, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \quad \text{(B-23)}$$

For the rest of the proof, we fix $k = 2, \ldots, K$, and $\tilde{X} = (\tilde{S}, \tilde{A}), \breve{X} = (\breve{S}, \breve{A}) \in \mathcal{X}$ arbitrarily. To show (B-23), consider the following derivation.

$$\begin{aligned}
&P(G^{(k)}X = \tilde{X} \mid G^{(1)}X, \ldots, G^{(k-2)}X, G^{(k-1)}X = \breve{X}, X) \\
&\overset{(1)}{=} E[P(G^{(k)}X = \tilde{X} \mid G^{(1)}, \ldots, G^{(k-1)}, X) \mid G^{(1)}X, \ldots, G^{(k-2)}X, G^{(k-1)}X = \breve{X}, X] \\
&\overset{(2)}{=} E[P(G^{(k)}X = \tilde{X} \mid G^{(k-1)}, X) \mid G^{(1)}X, \ldots, G^{(k-2)}X, G^{(k-1)}X = \breve{X}, X],
\end{aligned} \tag{B-24}$$

where (1) holds by the law of total probability, and (2) holds by (B-22). From (B-24), (B-23) follows if we show that, for $G^{(k-1)}X = \breve{X}$,

$$P(G^{(k)}X = \tilde{X} \mid G^{(k-1)}, X) = \begin{cases} \sum_{I \in \mathcal{I}} \dfrac{1\{\tilde{S} \in R_S(I, \breve{S}), \tilde{A} \in R_A(\tilde{S}, \breve{X})\}}{|\mathcal{I}| \times |R_S(I, \breve{S})| \times |R_A(\tilde{S}, \breve{X})|} & \text{if } |R_S(I, \breve{S})| \times |R_A(\tilde{S}, \breve{X})| > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{B-25}$$

To show (B-25), consider the following derivation.

$$\begin{aligned}
P(G^{(k)}X = \tilde{X} \mid G^{(k-1)}, X) &\overset{(1)}{=} P(G^{(k)}(G^{(k-1)})^{-1}\breve{X} = \tilde{X} \mid G^{(k-1)}, X) \\
&= \sum_{g \in \mathbf{G}} P(G^{(k)} = g \mid G^{(k-1)}, X) 1\{g(G^{(k-1)})^{-1}\breve{X} = \tilde{X}\} \\
&\overset{(2)}{=} \sum_{g \in \mathbf{G}} \frac{\sum_{I \in \mathcal{I}} \frac{1}{|\mathcal{I}|} \sum_{\tilde{g} \in \mathbf{G}(I)} 1\{g = \tilde{g} \circ (G^{(k-1)})^{-1}\}}{|\mathbf{G}(I)|} 1\{g(G^{(k-1)})^{-1}\breve{X} = \tilde{X}\} \\
&= \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \frac{\sum_{\tilde{g} \in \mathbf{G}(I)} \sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} 1\{g(G^{(k-1)})^{-1}\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} \\
&\overset{(3)}{=} \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \frac{\sum_{\tilde{g} \in \mathbf{G}(I)} \sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} 1\{\tilde{g}\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} \\
&= \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \frac{\sum_{\tilde{g} \in \mathbf{G}(I)} 1\{\tilde{g}\breve{X} = \tilde{X}\} \sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\}}{|\mathbf{G}(I)|} \\
&\overset{(4)}{=} \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} \frac{\sum_{\tilde{g} \in \mathbf{G}(I)} 1\{\tilde{g}\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|},
\end{aligned} \tag{B-26}$$

where (1) holds by $G^{(k-1)}X = \breve{X}$ and the fact that $(G^{(k-1)})^{-1} \in \mathbf{G}$ since $\mathbf{G}$ is a group (by Lemma 4.2), (2) holds by (B-22), (3) holds because $\{g = \tilde{g} \circ (G^{(k-1)})\}$ occurs if and only if $\{g(G^{(k-1)})^{-1} = \tilde{g}\}$, and (4) holds because $\sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} = 1$, as we show in the next paragraph.

To show $\sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} = 1$, consider the following argument. Since $g, \tilde{g}, G^{(k-1)} \in \mathbf{G}$, and $\mathbf{G}$ is a group, $\tilde{g} \circ (G^{(k-1)}) \in \mathbf{G}$, and so $\exists g \in \mathbf{G}$ s.t. $g = \tilde{g} \circ (G^{(k-1)})$, i.e., $\sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} \geq 1$. Now, suppose that $\sum_{g \in \mathbf{G}} 1\{g = \tilde{g} \circ (G^{(k-1)})\} > 1$. This implies that $\exists g_1, g_2 \in \mathbf{G}$ with $g_1 \neq g_2$ s.t. $g_1 = \tilde{g} \circ (G^{(k-1)}) = g_2$. But using again that $\mathbf{G}$ is a group, $\exists g_1^{-1} \in \mathbf{G}$ and so $g_1^{-1}g_2 = g_1^{-1}g_1$ and $g_2 g_1^{-1} = g_1 g_1^{-1}$ and both equal to the identity transformation. This would imply that $g_1^{-1} = g_2^{-1}$, and since the inverse transformation is unique, reach a contradiction.

Fix $I \in \mathcal{I}$ arbitrarily. By (B-26), (B-25) then follows from showing that

$$\sum_{\tilde{g} \in \mathbf{G}(I)} \frac{1\{\tilde{g}\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} = \begin{cases} \dfrac{1\{\tilde{S} \in R_S(I, \breve{S}), \tilde{A} \in R_A(\tilde{S}, \breve{X})\}}{|R_S(I, \breve{S})| \times |R_A(\tilde{S}, \breve{X})|} & \text{if } |R_S(I, \breve{S})| \times |R_A(\tilde{S}, \breve{X})| > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{B-27}$$

We divide our argument into two cases. First, consider $|R_S(I,\breve{S})| \times |R_A(\tilde{S},\breve{X})| = 0$. In this case, we have $\nexists g \in \mathbf{G}(I)$ s.t. $g\breve{X} = \tilde{X}$, and therefore

$$\sum_{g\in\mathbf{G}(I)} \frac{1\{g\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} = 0,$$

which verifies (B-27).

Second, consider $|R_S(I,\breve{S})| \times |R_A(\tilde{S},\breve{X})| > 0$. Then, consider the following derivation.

$$\begin{aligned}
\frac{\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} &\overset{(1)}{=} \frac{\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\}}{|\mathbf{G}(I)|} 1\{\tilde{S} \in R_S(I,\breve{S}), \tilde{A} \in R_A(\tilde{S},\breve{X})\} \\
&\overset{(2)}{=} \frac{\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \breve{X}\}}{|\mathbf{G}(I)|} 1\{\tilde{S} \in R_S(I,\breve{S}), \tilde{A} \in R_A(\tilde{S},\breve{X})\} \\
&\overset{(3)}{=} \frac{1\{\tilde{S} \in R_S(I,\breve{S}), \tilde{A} \in R_A(\tilde{S},\breve{X})\}}{|R_S(I,\breve{S})| \times |R_A(\breve{S},\breve{X})|} \\
&\overset{(4)}{=} \frac{1\{\tilde{S} \in R_S(I,\breve{S}), \tilde{A} \in R_A(\tilde{S},\breve{X})\}}{|R_S(I,\breve{S})| \times |R_A(\tilde{S},\breve{X})|},
\end{aligned} \tag{B-28}$$

which verifies (B-27), where (1) follows from Definition 4.1, as it implies that $\{\tilde{g}\breve{X} = \tilde{X}\}$ with $\tilde{g} \in \mathbf{G}(I)$, $\breve{X} = (\breve{S},\breve{A})$, and $\tilde{X} = (\tilde{S},\tilde{A})$ implies that $\{\tilde{S} \in R_S(I,\breve{S})\}$ and $\{\tilde{A} \in R_A(\tilde{S},\breve{X})\}$, (2) follows from Lemma B.6, and (3) is shown in (B-29), and (4) follows from Lemma B.7 (which applies because the expression is multiplied by $1\{\tilde{S} \in R_S(I,\breve{S})\}$).

To show (3) in (B-28), consider the following argument.

$$\begin{aligned}
|\mathbf{G}(I)| &\overset{(1)}{=} \sum_{\tilde{X}\in\mathcal{X}} \left(\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\} \right) \\
&\overset{(2)}{=} \sum_{\tilde{S}\in R_S(I,\breve{S})} \sum_{\tilde{A}\in R_A(\tilde{S},\breve{X})} \left(\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\} \right) \\
&\overset{(3)}{=} \left(\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \breve{X}\} \right) \sum_{\tilde{S}\in R_S(I,\breve{S})} |R_A(\tilde{S},\breve{X})| \\
&\overset{(4)}{=} \left(\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \breve{X}\} \right) |R_A(\breve{S},\breve{X})| \times |R_S(I,\breve{S})|,
\end{aligned} \tag{B-29}$$

where (1) follows from partitioning $\mathbf{G}(I)$ into its possible range of outcomes when applied to $\breve{X} \in \mathcal{X} = \mathcal{S}^{nT} \times \mathcal{A}^{nT}$, (2) follows from Definition 4.1, as it implies that $\{\tilde{g}\breve{X} = \tilde{X}\}$ with $\tilde{g} \in \mathbf{G}(I)$, $\breve{X} = (\breve{S},\breve{A})$, and $\tilde{X} = (\tilde{S},\tilde{A})$ if and only if $\{\tilde{S} \in R_S(I,\breve{S})\}$ and $\{\tilde{A} \in R_A(\tilde{S},\breve{X})\}$, (3) follows from Lemma B.6, and (4) follows from Lemma B.7. ■

Lemma B.6. Fix $\breve{X} = (\breve{S},\breve{A}) \in \mathcal{X}$, $\tilde{X} = (\tilde{S},\tilde{A}) \in \mathcal{X}$, and $I \in \mathcal{I}$ arbitrarily. Then, $\tilde{S} \in R_S(I,\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},\breve{X})$ implies that $\sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\} = \sum_{g\in\mathbf{G}(I)} 1\{g\breve{X} = \breve{X}\}$.

Proof. Fix $\breve{X} = (\breve{S},\breve{A}) \in \mathcal{X}$, $\tilde{X} \in \mathcal{X}$, and $I \in \mathcal{I}$ arbitrarily, and assume that $\tilde{S} \in R_S(I,\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},\breve{X})$. By definition of $\mathbf{G}(I)$, $R_S(I,\breve{S})$, and $R_A(\tilde{S},\breve{X})$, $\tilde{S} \in R_S(I,\breve{S})$ and $\tilde{A} \in R_A(\tilde{S},\breve{X})$ implies that $\exists \breve{g} \in \mathbf{G}(I)$ s.t.

$\breve{g}\breve{X} = \tilde{X}$. Therefore,

$$\sum_{g \in \mathbf{G}(I)} 1\{g\breve{X} = \tilde{X}\} \overset{(1)}{=} \sum_{g \in \mathbf{G}(I)} 1\{g\breve{X} = \breve{g}\breve{X}\} \overset{(2)}{=} \sum_{g \in \mathbf{G}(I)} 1\{\breve{g}^{-1}g\breve{X} = \breve{X}\} \overset{(3)}{=} \sum_{g \in \mathbf{G}(I)} 1\{g\breve{X} = \breve{X}\},$$

where (1) holds by $\breve{g}\breve{X} = \tilde{X}$, (2) holds because $\breve{g} \in \mathbf{G}(I)$ and that $\mathbf{G}(I)$ is a group (by Lemma B.3), (3) holds by $\{\breve{g}^{-1}g : g \in \mathbf{G}(I)\} = \mathbf{G}(I)$, as $\mathbf{G}(I)$ is a group (again, by Lemma B.3). ■

Lemma B.7. Fix $\breve{X} = (\breve{S}, \breve{A}) \in \mathcal{X}$ and $I \in \mathcal{I}$ arbitrarily. Then, $\tilde{S} \in R_S(I, \breve{S})$ implies that $|R_A(\tilde{S}, \breve{X})| = |R_A(\breve{S}, \breve{X})|$.

Proof. Fix $\breve{X} = (\breve{S}, \breve{A}) \in \mathcal{X}$ and $I \in \mathcal{I}$ arbitrarily, and assume that $\tilde{S} \in R_S(I, \breve{S})$.

We first show that $|R_A(\breve{S}, (\breve{S}, \breve{A}))| \leq |R_A(\tilde{S}, (\breve{S}, \breve{A}))|$. Let $(A^1, \ldots, A^C)$ enumerate the (distinct) elements in $R_A(\breve{S}, (\breve{S}, \breve{A}))$. By $\tilde{S} \in R_S(I, \breve{S})$ and Lemma B.8, there is a permutation π s.t $\tilde{S} = \breve{S}_\pi$ and $A^c_\pi \in R_A(\tilde{S}, (\breve{S}, A^c))$ for each $c = 1, \ldots, C$. We now show that $(A^1_\pi, \ldots, A^C_\pi)$ are all distinct elements. To this end, suppose that $\exists c_1, c_2 \in \{1, \ldots, C\}$ s.t. $A^{c_1}_\pi = A^{c_2}_\pi$. If that were the case, and by the fact that a permutation is a bijective relationship, we conclude that $A^{c_1} = A^{c_2}$. Since $(A^1, \ldots, A^C)$ are distinct, we conclude that $c_1 = c_2$, as desired. To conclude the argument, it suffices to show that $A^c_\pi \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$ for all $c = 1, \ldots, C$. To this end, choose $c = 1, \ldots, C$ arbitrarily. Since $\breve{S} \in R_S(I, \breve{S})$ (trivially) and $A^c \in R_A(\breve{S}, (\breve{S}, \breve{A}))$, Definition 4.1 implies that $\exists g_1 \in \mathbf{G}(I)$ s.t. $g_1(\breve{S}, \breve{A}) = (\breve{S}, A^c)$. Since $\tilde{S} \in R_S(I, \breve{S})$ and $A^c_\pi \in R_A(\tilde{S}, (\breve{S}, A^c))$, Definition 4.1 implies that $\exists g_2 \in \mathbf{G}(I)$ s.t. $g_2(\breve{S}, A^c) = (\tilde{S}, A^c_\pi)$. Since $\mathbf{G}(I)$ is a group (by Lemma B.3), we conclude that $g_3 = g_2 \circ g_1 \in \mathbf{G}(I)$. Since $g_3(\breve{S}, \breve{A}) = (\tilde{S}, A^c_\pi)$ and $g_3 \in \mathbf{G}(I)$, Definition 4.1 implies that $A^c_\pi \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$, as desired.

We next show that $|R_A(\breve{S}, (\breve{S}, \breve{A}))| \geq |R_A(\tilde{S}, (\breve{S}, \breve{A}))|$. Let $(A^1, \ldots, A^C)$ enumerate the (distinct) elements in $R_A(\tilde{S}, (\breve{S}, \breve{A}))$. Since $\tilde{S} \in R_S(I, \breve{S})$ and by the fact that the Definition 3.1 treats $\tilde{S}$ and $\breve{S}$ symmetrically, we conclude that $\breve{S} \in R_S(I, \tilde{S})$. In turn, by $\breve{S} \in R_S(I, \tilde{S})$ and Lemma B.8, there is a permutation π s.t. $\breve{S} = \tilde{S}_\pi$ and $A^c_\pi \in R_A(\breve{S}, (\tilde{S}, A^c))$ for each $c = 1, \ldots, C$. By repeating the previous argument, we can show that $(A^1_\pi, \ldots, A^C_\pi)$ are all distinct elements. To conclude the proof, it suffices to show that $A^c_\pi \in R_A(\breve{S}, (\breve{S}, \breve{A}))$ for all $c = 1, \ldots, C$. To this end, choose $c = 1, \ldots, C$ arbitrarily. Since $\tilde{S} \in R_S(I, \breve{S})$ and $A^c \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$, Definition 4.1 implies that $\exists g_1 \in \mathbf{G}(I)$ s.t. $g_1(\breve{S}, \breve{A}) = (\tilde{S}, A^c)$. Since $\breve{S} \in R_S(I, \tilde{S})$ and $A^c_\pi \in R_A(\breve{S}, (\tilde{S}, A^c))$, Definition 4.1 implies that $\exists g_2 \in \mathbf{G}(I)$ s.t. $g_2(\tilde{S}, A^c) = (\breve{S}, A^c_\pi)$. Since $\mathbf{G}(I)$ is a group (by Lemma B.3), we conclude that $g_3 = g_2 g_1 \in \mathbf{G}(I)$. Since $g_3(\breve{S}, \breve{A}) = (\breve{S}, A^c_\pi)$ and $g_3 \in \mathbf{G}(I)$, Definition 4.1 implies that $A^c_\pi \in R_A(\breve{S}, (\breve{S}, \breve{A}))$, as desired. ■

Lemma B.8. For any $\breve{S} \in \mathcal{S}^{nT}$, $I \in \mathcal{I}$ and $\tilde{S} \in R_S(I, \breve{S})$, there exists a permutation $\pi : \{1, \ldots, n\} \times \{1, \ldots, T\} \to \{1, \ldots, n\} \times \{1, \ldots, T\}$ such that $\tilde{S} = \breve{S}_\pi$ and $\breve{A}_\pi \in R_A(\tilde{S}, (\breve{S}, \breve{A}))$ for every $\breve{A} \in \mathcal{A}^{nT}$.

Proof. Fix $\breve{S}, \in \mathcal{S}^{nT}$ and $I \in \mathcal{I}$ arbitrarily and assume that $\tilde{S} \in R_S(I, \breve{S})$. For every $s, s' \in \mathcal{S}$, let

$$\begin{aligned}
\text{Index}_1(s, s') &= \{(i, t) \in \{1, \ldots, n\} \times \{1, \ldots, T-1\} : (\breve{S}_{i,t}, \breve{S}_{i,t+1}) = (s, s')\}, \\
\text{Index}_2(s, s') &= \{(i, t) \in \{1, \ldots, n\} \times \{1, \ldots, T-1\} : (\tilde{S}_{i,t}, \tilde{S}_{i,t+1}) = (s, s')\}, \\
\text{Index}_1(s) &= \{(i, T) : i \in \{1, \ldots, n\}, \breve{S}_{i,T} = s\}, \\
\text{Index}_2(s) &= \{(i, T) : i \in \{1, \ldots, n\}, \tilde{S}_{i,T} = s\}.
\end{aligned}$$

We use

$$C(s,s') \equiv |\text{Index}_1(s,s')| \overset{(1)}{=} |\text{Index}_2(s,s')|,$$
$$C(s) \equiv |\text{Index}_1(s)| \overset{(2)}{=} |\text{Index}_2(s)|,$$

where (1) and (2) hold by $\tilde{S} \in R_S(I, \breve{S})$.

For every $s, s' \in \mathcal{S}$, we can enumerate $\text{Index}_1(s,s')$ by $(\nu_1(1,s,s'),\ldots,\nu_1(C(s,s'),s,s'))$, $\text{Index}_2(s,s')$ by $(\nu_2(1,s,s'),\ldots,\nu_2(C(s,s'),s,s'))$, $\text{Index}_1(s)$ by $(\nu_1(1,s),\ldots,\nu_1(C(s),s))$, and $\text{Index}_2(s)$ by $(\nu_2(1,s),\ldots,\nu_2(C(s),s))$. By definition, $(\nu_1(1,s,s'),\ldots,\nu_1(C(s,s'),s,s'))$ represent the (i,t) indices that satisfy $(\breve{S}_{i,t},\breve{S}_{i,t+1}) = (s,s')$ and $(\nu_1(1,s),\ldots,\nu_1(C(s),s))$ represent the (i,T) indices that satisfy $\breve{S}_{i,T} = s$, $(\nu_2(1,s,s'),\ldots,\nu_2(C(s,s'),s,s'))$ represent the (i,t) indices that satisfy $(\tilde{S}_{i,t},\tilde{S}_{i,t+1}) = (s,s')$, and $(\nu_2(1,s),\ldots,\nu_2(C(s),s))$ represent the (i,T) indices that satisfy $\tilde{S}_{i,T} = s$.

These enumerations allows us to interpret $\breve{S}$ as a permutation of the values of $\tilde{S}$. We denote this permutation by $\pi : \{1,\ldots,n\} \times \{1,\ldots,T\} \to \{1,\ldots,n\} \times \{1,\ldots,T\}$, and we characterize it next. For any $(i,t) \in \{1,\ldots,n\} \times \{1,\ldots,T-1\}$, there exists $(s,s') \in \mathcal{S}$ and $c = 1,\ldots,C(s,s')$ s.t. $(i,t) = \nu_1(c,s,s') \in \text{Index}_1(s,s')$. In this case, set $\pi(i,t) = \nu_2(c,s,s')$. By this construction,

$$\breve{S}_{i,t} = \breve{S}_{\nu_1(c,s,s')} = \tilde{S}_{\nu_2(c,s,s')} = \tilde{S}_{\pi(i,t)},$$

Similarly, for any $i \in \{1,\ldots,n\}$, there exists $s \in \mathcal{S}$ and $c = 1,\ldots,C(s)$ s.t. $(i,T) = \nu_1(c,s) \in \text{Index}_1(s)$. In this case, set $\pi(i,T) = \nu_2(c,s)$. By this construction,

$$\breve{S}_{i,T} = \breve{S}_{\nu_2(c,s,)} = \tilde{S}_{\nu_2(c,s)} = \tilde{S}_{\pi(i,T)}.$$

To show the second part, for any $\breve{A} \in \mathcal{A}^{nT}$, consider $\tilde{A} = \breve{A}_\pi$. For each $s, s' \in \mathcal{S}$, note that

$$\begin{aligned}\tilde{A}_{\nu_2(c,s,s')} &= \breve{A}_{\nu_1(c,s,s')} \quad \text{for } c = 1,\ldots,C(s,s')\\ \tilde{A}_{\nu_2(c,s)} &= \breve{A}_{\nu_1(c,s)} \quad \text{for } c = 1,\ldots,C(s).\end{aligned} \tag{B-30}$$

To complete the proof, it suffices to show that $\tilde{A} \in R_A(\tilde{S}, \breve{X})$. To this end, it suffices to verify conditions (a)-(b) in Definition 3.2. We only show condition (a), as condition (b) can be shown using an analogous argument. For any $s, s' \in \mathcal{S}$ and $a \in \mathcal{A}$, consider the following derivation.

$$\begin{aligned}\sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\breve{S}_{i,t} = s, \breve{A}_{i,t} = a, \breve{S}_{i,t+1} = s'\} &\overset{(1)}{=} \sum_{(i,t)\in\text{Index}_1(s,s')} 1\{\breve{A}_{i,t} = a\}\\ &= \sum_{c=1}^{C(s,s')} 1\{\breve{A}_{\nu_1(c,s,s')} = a\}\\ &\overset{(2)}{=} \sum_{c=1}^{C(s,s')} 1\{\tilde{A}_{\nu_2(c,s,s')} = a\}\\ &= \sum_{(i,t)\in\text{Index}_2(s,s')} 1\{\tilde{A}_{i,t} = a\}\\ &\overset{(3)}{=} \sum_{i=1}^{n}\sum_{t=1}^{T-1} 1\{\tilde{S}_{i,t} = s, \tilde{A}_{i,t} = a, \tilde{S}_{i,t+1} = s'\},\end{aligned} \tag{B-31}$$

where (1) follows from the fact that $\text{Index}_1(s,s')$ is the collection of all indices $(i,t) \in \{1,\ldots,n\} \times \{1,\ldots,T-1\}$ s.t. $(\breve{S}_{i,t}, \breve{S}_{i,t+1}) = (s,s')$, (2) holds by (B-30), and (3) follows from the fact that $\text{Index}_2(s,s')$ is the collection of all indices $(i,t) \in \{1,\ldots,n\} \times \{1,\ldots,T-1\}$ s.t. $(\breve{S}_{i,t}, \breve{S}_{i,t+1}) = (s,s')$. ■

Lemma B.9. The transition probability in (B-22) is symmetric, i.e., for any $g, \breve{g} \in \mathbf{G}$, $P(G^{(k+1)} = \breve{g}|G^{(k)} = g) = P(G^{(k+1)} = g|G^{(k)} = \breve{g})$.

Proof. Fix $g, \breve{g} \in \mathbf{G}$ arbitrarily and consider the following argument.

$$\begin{aligned} P(G^{(k+1)} = \breve{g}|G^{(k)} = g) &= \sum_{I \in \mathcal{I}} \frac{1}{|\mathcal{I}|} \sum_{\tilde{g} \in \mathbf{G}(I)} \frac{1\{\breve{g} = \tilde{g} \circ g\}}{|\mathbf{G}(I)|} \\ &\overset{(1)}{=} \sum_{I \in \mathcal{I}} \frac{1}{|\mathcal{I}|} \sum_{\tilde{g} \in \mathbf{G}(I)} \frac{1\{g = \tilde{g}^{-1} \circ \breve{g}\}}{|\mathbf{G}(I)|} \\ &\overset{(2)}{=} \sum_{I \in \mathcal{I}} \frac{1}{|\mathcal{I}|} \sum_{\tilde{g} \in \mathbf{G}(I)} \frac{1\{g = \tilde{g} \circ \breve{g}\}}{|\mathbf{G}(I)|} \\ &= P(G^{(k+1)} = g|G^{(k)} = \breve{g}), \end{aligned}$$

where (1) follows from the fact that $\mathbf{G}(I)$ is a group (by Lemma B.3), and so $\exists \tilde{g}^{-1} \in \mathbf{G}(I)$ for any $\tilde{g} \in \mathbf{G}(I)$, and that $1\{\breve{g} = \tilde{g} \circ g\} = 1\{\tilde{g}^{-1} \circ \breve{g} = g\}$, and (2) follows from defining $\mathbf{G}(I) = \{\tilde{g}^{-1} : \tilde{g} \in \mathbf{G}(I)\}$, which holds because $\mathbf{G}(I)$ is a group (again, by Lemma B.3). ■

References

Aguirregabiria, V. and A. Magesan (2020): "Identification and estimation of dynamic games when players' beliefs are not in equilibrium," *The Review of Economic Studies*, 87, 582–625.

Aguirregabiria, V. and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70, 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

Arcidiacono, P. and R. A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

Bajari, P., D. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Besag, J. and D. Mondal (2013): "Exact Goodness-of-Fit Tests for Markov Chains," *Biometrics*, 69, 488–496.

Collard-Wexler, A. (2013): "Demand fluctuations in the ready-mix concrete industry," *Econometrica*, 81, 1003–1037.

de Paula, A. and X. Tang (2012): "Inference of Signs of Interaction Effects in Simultaneous Games With Incomplete Information," *Econometrica*, 80, 143–172.

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2013): "Entry, exit, and the determinants of market structure," *The RAND Journal of Economics*, 44, 462–487.

Ericson, R. and A. Pakes (1995): "arkov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.

Hotz, J. V. and R. T. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economics Studies*, 60, 497–529.

Hotz, J. V., R. T. A. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economics Studies*, 61, 265–289.

Kandel, D., M. Yossi, R. Unger, and P. Winkler (1996): "Shuffling biological sequences," *Discrete Applied Mathematics*, 71, 171–185.

Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypothesis: Third edition*, Springer.

Maskin, E. and J. Tirole (2001): "Markov Perfect Equilibrium I: Observable Actions," *Journal of Economic Theory*, 100, 191–219.

Otsu, T., M. Pesendorfer, and Y. Takahashi (2016): "Pooling data across markets in dynamic Markov Games," *Quantitative Economics*, 7, 523–559.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*, 38, 373–399.

Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928.

——— (2010): "Sequential Estimation of Dynamic Discrete Games: A Comment," *Econometrica*, 78, 833–842.

Robert, C. and G. Casella (2004): *Monte Carlo Statistical Methods*, Springer Texts in Statistics.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

Ryan, S. (2012): "The costs of environmental regulation in a concentrated industry," *Econometrica*, 80, 1019–1061.

Sweeting, A. (2013): "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 81, 1763–1803.