Taisuke Otsu, Martin Pesendorfer, Yuya Takahashi

Pooling data across markets in dynamic Markov games

Article (Accepted version) (Refereed)

Pooling Data across Markets in Dynamic Markov Games *

Taisuke Otsu† Martin Pesendorfer‡ Yuya Takahashi§

First Draft: October 2012

This Draft: December 2015

Abstract

This paper proposes several statistical tests for finite state Markov games to examine whether data from distinct markets can be pooled. We formulate homogeneity tests of (i) the conditional choice and state transition probabilities, (ii) the steady-state distribution, and (iii) the conditional state distribution given an initial state. The null hypotheses of these homogeneity tests are necessary conditions (or maintained assumptions) for poolability of the data. Thus rejections of these null imply that the data cannot be pooled across markets. Acceptances of these null are considered as supporting evidences for the maintained assumptions of estimation using pooled data. In a Monte Carlo study we find that the test based on the steady-state distribution performs well and has high power even with small numbers of markets and time periods. We apply the tests to the empirical study of Ryan (2012) that analyzes dynamics of the U.S. Portland Cement industry and assess if the data across markets can be pooled.

Keywords: Dynamic Markov game, Poolability, Multiplicity of equilibria, Hypothesis testing.

Jel Classification: C12, C72, D44.

*We thank Stephen Ryan, Pierre Dubois, and conference participants of the 14th CEPR-JIE Applied IO Conference and seminar audiences in Alicante, Barcelona (Pompeu Fabra and Autonoma), Johns Hopkins, Kyoto, NYU, Paris School of Economics, Princeton, Texas Austin, UCL, UBC, and Wisconsin for thoughtful comments. We thank André Stenzel for excellent research assistance. Takahashi's work was supported by the Deutsche Forschungsgemeinschaft through SFB/TR 15.

†Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, t.otsu@lse.ac.uk.

‡Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, m.pesendorfer@lse.ac.uk.

§Department of Economics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, ytakahashi@jhu.edu.

1 Introduction

This paper proposes several statistical tests for finite state Markov games to examine whether data from distinct markets can be pooled. Data pooling is employed in a number of empirical applications of the two-step estimation methods for dynamic games recently developed.[1] These two-step estimators estimate players' policies and state transition probabilities in a first stage directly from the data as functions of observable state variables. The second stage conducts a search for the structural model parameters which best rationalize observed behaviors of players and state transitions using the first stage policy estimates as estimates for the equilibrium beliefs. A typical application may not have long time series data for a single market. Researchers are tempted to pool data from different markets (or games) to perform the first stage policy function estimation. To do so, researchers assume that the data are generated from a single and identical equilibrium in every market. This assumption has become popular in a number of recent papers.[2] To be more precise, the assumption commonly imposed requires that the game describing players' behavior is identical in all markets and that a single and identical equilibrium of that game is played in all markets. It also requires that the econometric model controls for all observable or unobservable market-level elements. A violation of the assumption results in inconsistent policy estimates and inconsistent structural parameter estimates. A violation of the assumption can arise because of equilibrium multiplicity. The single and identical equilibrium assumption may be very restrictive even if the markets are identical as multiplicity of equilibria is a well known feature inherent to games. Incorrectly imposing this assumption leads to erroneous inference.

A maintained assumption for estimation based on the pooled data is that the data generating processes are identical across markets. We propose three tests to assess homogeneity of the data generating processes. The first test compares directly the set of conditional choice or state transition probabilities estimated from the pooled (across markets) sample with the ones estimated from each market separately. The second test is based on the result that there is a unique steady-state distribution associated with a transition matrix of states under the assumption of communicating states. Based on this result, the second test compares the steady-state distribution estimated from the pooled sample with

[1]Several papers, including Jofre-Bonet and Pesendorfer (2003), Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008), Arcidiacono and Miller (2011), Kasahara and Shimotsu (2012), and Srisuma and Linton (2012), proposed two-step estimation methods for dynamic Markov games under varying assumptions. They led to a number of empirical papers that apply these methods to empirically analyze dynamic interactions between multiple players.

[2]Examples include Beresteanu, Ellickson and Misra (2010), Collard-Wexler (2013), Dunne, Klimek, Roberts and Xu (2013), Fan and Xiao (2014), Jeziorski, (2014), Lin (2014), Maican and Orth (2014), Minamihashi (2012), Nishiwaki (2015), Ryan (2012), Sanches and Silva Junior (2013), Snider (2009), Suzuki (2013), and Sweeting (2013). They impose the assumption of a single and identical equilibrium in all markets either explicitly or implicitly. The empirical sections of Aguirregabiria and Mira (2007) and Arcidiacono, Bayer, Blevins and Ellickson (2015) and the Monte Carlo exercise in Arcidiacono and Miller (2011) also impose the same assumption.

the one from each market. Our third test statistic is based on the conditional state distribution given the initial (observed) state. We contrast the observed relative frequencies of states to the theoretical predictions given the initial state. It turns out that the third test does not require several assumptions on Markov chains that are imposed for other tests. Each test has its own advantage. One advantage across all three tests is that we do not need to impose any mixing structure.

Since the null hypotheses of our homogeneity tests are necessary conditions or maintained assumptions for estimation based on pooled data across markets, a rejection of the null suggests that the data cannot be pooled. A violation can arise because: (i) multiple equilibria are played across markets; (ii) the game form describing players' behavior and interactions differs across markets; and (iii) the specified model is not sufficiently rich as it does not control for all observable or unobservable market-level heterogeneity adequately. It is difficult to distinguish these alternative explanations although we shall illustrate tests accounting for unobservable market-level heterogeneity as in Arcidiacono and Miller (2011) in more detail below. Our test is aimed at checking the validity of the maintained assumption for data pooling commonly imposed in the literature. A rejection of the null points to an inconsistency of the first stage estimates that arises from pooling different markets. Naturally, since the framework of this paper nests single agent settings as a special case with only one player, our tests can also be thought of as testing whether data can be pooled in the single agent case.

To illustrate the finite sample performance of our tests, we first apply the tests to simulated data using an example of multiple equilibria in Pesendorfer and Schmidt-Dengler (2008). Our tests, particularly the one based on the steady-state distribution, perform well and have high power even with small numbers of markets and time periods. We then apply our tests to the empirical study of Ryan (2012) that analyzes dynamics of the U.S. Portland Cement industry. Our tests reject the null hypothesis that the data from distinct markets are generated from an identical data generating process.

To the best of our knowledge, this is the first paper that proposes tests to assess the validity of data pooling in a general class of dynamic Markov games. Our tests may give a researcher guidance on whether she can pool different markets to estimate policy functions in the first stage. A rejection of the null hypothesis suggests that one or more modeling assumption differs across markets. In the context of static games with incomplete information, de Paula and Tang (2011) propose a test of multiplicity of equilibria that requires conditional independence between players' actions. Since our tests exploit the panel structure of the data and rely on the way that the game and states evolve, our tests are fundamentally different from theirs. One notable difference is that while de Paula and Tang (2011) maintain the assumption of independent-across-players private shocks, we can allow for within-period correlation in players' actions and for unobserved state variables.

This paper is organized as follows. Section 2 lays out a class of general dynamic Markov games we work with and provides some background on Markov chains. Section 3 proposes several test statistics. In Sections 4 we conduct a Monte Carlo study to examine finite sample properties. Section 5 applies our tests to data of Ryan (2012). Section 6 concludes. Appendix A contains technical details.

2 Model

This section describes elements of a general dynamic Markov game with discrete time $t = 1, 2, \ldots$. We focus on the description of players' state variables and actions. These states and actions are the observable outcome variables for some underlying dynamic game which we do not observe. We leave the details of the game unspecified. Instead we shall focus on testable implications of the observed outcomes. Our setting includes the single agent case as a special case when there is one agent per market. We first describe the framework which applies for all markets $j = 1, \ldots, M$.

Players. A typical player is denoted by $i = 1, \ldots, N$. The single agent case arises when $N = 1$. The number of players is fixed and does not change over time. Every period the econometrician observes a profile of states and actions described as follows.

States. Each player is endowed with state variables $s_i^t \in \{1, \ldots, L\}$ in finite support. The state variable s_i^t is publicly observed by all players. We maintain the assumption that the econometrician also observes s_i^t. The vector of all players' public state variables is denoted by $\mathbf{s}^t = (s_1^t, \ldots, s_N^t) \in \mathbf{S} = \{1, \ldots, L\}^N$ whose cardinality is $m_s = L^N$. In Section 3.5, we discuss the case where some of the public state variables are unobservable by the econometrician.

Actions. Each player chooses an action $a_i^t \in \{0, 1, \ldots, K\}$ in finite support. The decisions are made after the state is observed. The decisions can be made simultaneously or sequentially. The decision may also be taken after an idiosyncratic random utility (or a random profit shock) is observed. We leave the details of the decision process unspecified. Our specification encompasses the random-utility modeling assumptions, and allows for within-period correlation in the random utility component across actions and across players. The vector of joint actions in period t is denoted by $\mathbf{a}^t = (a_1^t, \ldots, a_N^t) \in \mathbf{A} = \{0, 1, \ldots, K\}^N$ whose cardinality is $m_a = (K+1)^N$. We assume actions are publicly observed by all players and the econometrician.

Choice probability matrix. Let $\sigma(\mathbf{a}|\mathbf{s}) = \Pr\{\mathbf{a}^t = \mathbf{a}|\mathbf{s}^t = \mathbf{s}\}$ denote the conditional probability that an action profile $\mathbf{a}$ will be chosen conditionally on a state $\mathbf{s}$. Throughout the paper, we assume that σ is time invariant and is conditionally independent from other past actions and states. The matrix of conditional choice probabilities is denoted by σ, which has dimension $m_s \times (m_a m_s)$. It consists of conditional probabilities $\sigma(\mathbf{a}|\mathbf{s})$ in row $\mathbf{s}$, column $(\mathbf{a}, \mathbf{s})$, and zeros in row $\mathbf{s}$, column $(\mathbf{a}, \mathbf{s}')$ with $\mathbf{s}^{'} \neq \mathbf{s}$.

State-action transition matrix. Let $g(\mathbf{s}'|\mathbf{a}, \mathbf{s}) = \Pr\{\mathbf{s}^{t+1} = \mathbf{s}'|\mathbf{a}^t = \mathbf{a}, \mathbf{s}^t = \mathbf{s}\}$ denote the state-action transition probability that a state $\mathbf{s}'$ is reached when the current action profile and state are given by $(\mathbf{a}, \mathbf{s})$. We also assume that

g is time invariant and is conditionally independent from other past actions and states. We use the symbol $\mathbf{G}$ to denote the $(m_a m_s) \times m_s$ dimensional state-action transition matrix in which column $\mathbf{s}' \in \mathbf{S}$ consists of the vector of probabilities $\{g(\mathbf{s}'|\mathbf{a}, \mathbf{s})\}_{\mathbf{a}\in\mathbf{A},\mathbf{s}\in\mathbf{S}}$.

State transition matrix. Under the above assumptions on σ and $\mathbf{G}$, the state variables $\mathbf{s}^t$ obey a (first-order) Markov chain with the (stationary) state transition matrix $\mathbf{P} = \sigma\mathbf{G}$ whose dimension is $m_s \times m_s$. A typical element $p(\mathbf{s}'|\mathbf{s}) = \sum_{\mathbf{a}\in\mathbf{A}} \sigma(\mathbf{a}|\mathbf{s})g(\mathbf{s}'|\mathbf{a}, \mathbf{s})$ of $\mathbf{P}$ equals the probability that state $\mathbf{s}'$ is reached when the current state is given by $\mathbf{s}$. Hereafter we focus on the first-order Markov chain. However, our testing procedures can be extended to higher-order Markov chains since higher-order Markov chains can be reformulated as first-order ones by modifying the state space (see, e.g., Billingsley, 1961).

Limiting steady-state distribution. When the limit exists, let $Q(\mathbf{s}', \mathbf{s}) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbf{1}\{\mathbf{s}^t = \mathbf{s}', \mathbf{s}^0 = \mathbf{s}\}$ denote the long run proportion of time that the Markov chain $\mathbf{P}$ spends in state $\mathbf{s}'$ when starting at the initial state $\mathbf{s}^0 = \mathbf{s}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. Suppose the unconditional long run proportion of time $Q(\mathbf{s}') = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \mathbf{1}\{\mathbf{s}^t = \mathbf{s}'\}$ that the Markov chain $\mathbf{P}$ spends in state $\mathbf{s}'$ satisfies $Q(\cdot) = Q(\cdot, \mathbf{s})$ for all initial states $\mathbf{s}$. Then the m_s dimensional row vector of probabilities $\mathbf{Q} = \{Q(\mathbf{s})\}_{\mathbf{s}\in\mathbf{S}}$ is called the steady-state distribution of the Markov chain. Observe that the state space is finite and $\mathbf{Q}$ describes a multinomial distribution.

The properties of Markov chains are well known. We next describe some property useful for our purpose. To do so, we introduce the concept of communicating states.

Communicating states. We say that a state $\mathbf{s}'$ is reachable from $\mathbf{s}$ if there exists an integer T so that the chain $\mathbf{P}$ will be at state $\mathbf{s}'$ after T periods with positive probability. If $\mathbf{s}'$ is reachable from $\mathbf{s}$, and $\mathbf{s}$ is reachable from $\mathbf{s}'$, then the states $\mathbf{s}$ and $\mathbf{s}'$ are said to communicate.

Lemma 1 *Suppose all states of the Markov chain* $\mathbf{P}$ *communicate.*[3] *Then the steady-state distribution* $\mathbf{Q}$ *exists and is unique. It satisfies* $Q(\mathbf{s}) > 0$ *for all* $\mathbf{s} \in \mathbf{S}$ *and* $\mathbf{Q} = \mathbf{QP}$.

This lemma guarantees existence and uniqueness of the steady-state distribution and states that the long run proportion of time that the Markov chain $\mathbf{P}$ spends in state $\mathbf{s}$ is strictly positive for any state $\mathbf{s} \in \mathbf{S}$ and the equation $\mathbf{Q} = \mathbf{QP}$ must hold. A proof of the above properties is given in Levin, Peres and Wilmer (2009, Proposition 1.14 and Corollary 1.17) for example.

Communicating states are typically invoked in applied work, see Ericson and Pakes (1995). Communicating states naturally emerge in dynamic discrete choice models using a random utility specification, see McFadden (1973). The random component having full support in the real numbers implies that all actions arise with strictly positive probability for any state $\mathbf{s} \in \mathbf{S}$. Thus, states will communicate if the state-action transition matrix allows that state $\mathbf{s}'$, or $\mathbf{s}$,

[3] This is also called that the Markov chain $\mathbf{P}$ is ergodic or irreducible.

can in principle be reached when starting from state $\mathbf{s}$, respectively $\mathbf{s}'$, for any pair of states $\mathbf{s}, \mathbf{s}' \in \mathbf{S}$.

The feature that all states communicate may also emerge when actions are chosen with probability one for some (or all) states. Our set-up includes these settings as well. What is required for states to communicate in this case is that there exists a sequence of state-action profiles $\{(\mathbf{a}^1, \mathbf{s}^1), \ldots, (\mathbf{a}^t, \mathbf{s}^t)\}$ so that the chain starting at state $\mathbf{s}$ will be at state $\mathbf{s}'$ after t periods for any $\mathbf{s}, \mathbf{s}' \in \mathbf{S}$.

3 Homogeneity tests for poolability

This section describes hypotheses that we aim at testing and proposes statistical tests for those hypotheses. For each market j, a sequence of action-state profiles $(\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1,\ldots,T}$ is observed, where T is the length of time periods in the data set. Our null hypothesis is that the observed profiles are generated from an identical data generating process in all markets, and the alternative is that the data generating process is distinct for some markets. This null hypothesis is a maintained assumption for estimation based on pooled data. Based on the set-up described in the previous section, the data generating process of the profiles $(\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1,\ldots,T}$ is characterized by the conditional choice probability matrix σ_j and state-action transition matrix $\mathbf{G}_j$ that imply the transition matrix of states $\mathbf{P}_j = \sigma_j \mathbf{G}_j$. In particular, we focus on homogeneity of σ_j and $\mathbf{P}_j$ across markets, and test the following null hypotheses:

$$\begin{aligned} \mathrm{H}_0^{\sigma} &: \quad \sigma_1 = \cdots = \sigma_M, \\ \mathrm{H}_0^{\mathbf{P}} &: \quad \mathbf{P}_1 = \cdots = \mathbf{P}_M, \end{aligned} \tag{1}$$

and the alternatives are their negations. The null hypothesis H_0^{σ} is based on the idea that the equilibrium choice probabilities are identical across markets. The null $\mathrm{H}_0^{\mathbf{P}}$ has a similar motivation given that the state-action transition is identical across markets. Economic models may have the feature that the state-action transition matrix $\mathbf{G}$ is exogenously given and by construction identical across markets. In such cases, testing the conditional choice probabilities has the same interpretation to testing the state transition probabilities. However, in general, the tests may not be equivalent. A rejection of the null $\mathrm{H}_0^{\mathbf{P}}$ could arise either because of non-identical choice probabilities σ_j or because of heterogeneous state-action transition matrices $\mathbf{G}_j$. Which test is most suitable depends on the economic application at hand and each test has its own rationale.

If all states of the Markov chain $\mathbf{P}$ communicate, then by Lemma 1, there exists a unique steady-state distribution $\mathbf{Q}$ and the identical equilibrium hypothesis may be tested by homogeneity of the steady-state distribution,

$$\mathrm{H}_0^{\mathbf{Q}} : \mathbf{Q}_1 = \cdots = \mathbf{Q}_M. \tag{2}$$

As discussed in the next subsection, if the cardinality of the action or state space is large, then the power of the test for H_0^{σ} or $\mathrm{H}_0^{\mathbf{P}}$ tends to be low relative to that for $\mathrm{H}_0^{\mathbf{Q}}$ because a decrease in the degrees of freedom can be expected. Thus, the power of the homogeneity test can be increased by testing the steady-state distribution.

Lemma 1 says that the null $\mathrm{H}_0^{\mathbf{P}}$ of equal transition matrices implies the null $\mathrm{H}_0^{\mathbf{Q}}$ of equal steady-state distributions. Thus, a rejection of $\mathrm{H}_0^{\mathbf{Q}}$ provides a strong evidence for a rejection of $\mathrm{H}_0^{\mathbf{P}}$. By testing $\mathrm{H}_0^{\mathbf{Q}}$ first, we may exploit the property that the power of testing the null $\mathrm{H}_0^{\mathbf{Q}}$ is typically higher than the power of testing the null $\mathrm{H}_0^{\mathbf{P}}$. However, it should be noted that the converse is not true: the equivalence of the steady-state distribution across markets does not necessarily imply that of the transition matrix.

To test the above hypotheses, we consider the situation where for each market j, we observe the action-state profiles $(\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1,\ldots,T}$ with sufficiently large T. The test procedures discussed in the next subsection are theoretically justified when the time length T increases to infinity. However, the researcher may face the situation where the length of time periods T is relatively short compared to the number of markets M. In such a scenario, it would be natural to treat the action-state profiles with fixed T across markets as an i.i.d. sample (over $j = 1, \ldots, M$) from the distribution parametrized by a common choice probability σ or a common transition matrix $\mathbf{P}$. For example, testing may be based on the conditional state distribution $\mathbf{s}^t|\mathbf{s}^1 = \mathbf{s}$ given the initial state $\mathbf{s}$ for $t = 2, \ldots, T$. By conditioning on the initial state we do not require that states communicate so that the industry at hand can reach the steady-state distribution. This situation arises naturally in new or growing industries. Using the transition matrix $\mathbf{P}$, the conditional distribution $\mathbf{s}^t|\mathbf{s}^1 = \mathbf{s}$ is described by $\iota_{\mathbf{s}}'\mathbf{P}^t$, where $\iota_{\mathbf{s}}$ takes one at the element corresponding to $\mathbf{s}$ and zero otherwise. There are many ways to compare the vector of conditional probabilities $\{\Pr\{\mathbf{s}^t = \mathbf{s}'|\mathbf{s}^1 = \mathbf{s}\}\}_{\mathbf{s}'\in\mathbf{S}}$ with the theoretical prediction $\iota_{\mathbf{s}}'\mathbf{P}^t$. For example, at a given initial state $\mathbf{s}$, we can consider the null hypothesis in the form of

$$\mathrm{H}_0^{\mathbf{s}} : \left\{ \frac{1}{T-1} \sum_{t=2}^{T} \Pr\{\mathbf{s}^t = \mathbf{s}'|\mathbf{s}^1 = \mathbf{s}\} \right\}_{\mathbf{s}'\in\mathbf{S}} = \frac{1}{T-1} \sum_{t=2}^{T} \iota_{\mathbf{s}}'\mathbf{P}^t. \tag{3}$$

The left hand side is a vector of model-free conditional probabilities. The right hand side is the model-based prediction for those probabilities. Note that the hypothesis $\mathrm{H}_0^{\mathbf{s}}$ is implied from two assumptions: (i) the data $(\mathbf{s}_j^t)_{t=1,\ldots,T}$ for $j = 1, \ldots, M$ are i.i.d. over j which allows us to express the hypothesis $\mathrm{H}_0^{\mathbf{s}}$ without using a market index j, and (ii) the Markov chain is first-order and time-homogeneous. Thus, a rejection of $\mathrm{H}_0^{\mathbf{s}}$ may be interpreted as violation of the i.i.d. assumption (perhaps associated with multiplicity of equilibrium) or misspecification of the Markov chain (such as time-inhomogeneity or higher-order).

The left hand side denotes the empirical frequency (across markets) of visiting state $\mathbf{s}'$ in periods $t = 2, \ldots, T$ conditional on the initial state $\mathbf{s}^1 = \mathbf{s}$. The right hand side is the theoretical predicted counterpart under the null of homogeneity across markets. A violation of (3) would indicate that the empirical frequency distribution (across markets) differs from the one predicted by the theoretical model. Hypothesis (3) focuses on the average probabilities of visiting each state given the initial state $\mathbf{s}$. We may do so for selected initial states. Alternatively, one may consider all possible initial states jointly by testing the null $\mathrm{H}_0 : \Pr\{\mathbf{s}^t = \mathbf{s}'|\mathbf{s}^1 = \mathbf{s}\} = \iota_{\mathbf{s}}'\mathbf{P}^t$ for all $\mathbf{s} \in \mathbf{S}$ and t or its linear combinations. We note that the null $\mathrm{H}_0^{\mathbf{s}}$ tests the validity of the i.i.d. parametric model for $(\mathbf{s}_j)_{j=1,\dots,M}$ with $\mathbf{s}_j = (\mathbf{s}_j^1, \dots, \mathbf{s}_j^T)$ for fixed T.

As mentioned above, a rejection of the null can arise from multiple equilibria, the game form differing across markets, and/or unobservable market-level heterogeneity. Our framework nests single agent settings as a special case. In case of rejection, the first possibility (multiple equilibria) is naturally excluded so the interpretation of the rejection would be simpler. Therefore, our tests can be thought of as testing whether richer heterogeneity among agents should be considered in the single agent case.

3.1 Testing choice and transition probabilities

Let us first consider testing for H_0^σ and $\mathrm{H}_0^{\mathbf{P}}$ in (1) based on the conditional choice and transition probabilities, respectively. We form a generally applicable chi-squared test statistic based on the conditional choice or transition probability, that is

$$\mathcal{T}_P = \sum_{j=1}^{M} \sum_{\mathbf{d}\in\mathbf{D}} W_j(\mathbf{d})\{\widehat{P}_j(\mathbf{d}) - \widehat{P}(\mathbf{d})\}^2, \tag{4}$$

where $\widehat{P}_j(\mathbf{d})$ is a nonparametric estimator of the probability of interest for a market j without imposing the null hypothesis of interest, $\widehat{P}(\mathbf{d})$ is another nonparametric estimator under the null of homogeneity of $\widehat{P}_j(\mathbf{d})$ across markets, and $W_j(\mathbf{d})$ is a weight or standardization to obtain a standard limiting distribution.

For example, to test homogeneity of the conditional choice probabilities H_0^σ, we set $\mathbf{d} = (\mathbf{a}, \mathbf{s})$ and $\mathbf{D} = \mathbf{A} \times \mathbf{S}$. Let $f_j(\mathbf{a}, \mathbf{s}) = \sum_{t=1}^{T} \mathbf{1}\{\mathbf{a}_j^t = \mathbf{a}, \mathbf{s}_j^t = \mathbf{s}\}$ be the frequency of action state profile $(\mathbf{a}, \mathbf{s})$ in market j and $f_j(\mathbf{s}) = \sum_{t=1}^{T} \mathbf{1}\{\mathbf{s}_j^t = \mathbf{s}\}$ be the frequency of state $\mathbf{s}$ in market j. Then we estimate the conditional choice probabilities for the action profile $\mathbf{a}$ given the current state $\mathbf{s}$ in market j, $\sigma_j(\mathbf{a}|\mathbf{s})$, by the relative frequencies

$$\widehat{P}(\mathbf{d}) = \frac{\sum_{j=1}^{M} f_j(\mathbf{a}, \mathbf{s})}{\sum_{j=1}^{M} f_j(\mathbf{s})}, \qquad \widehat{P}_j(\mathbf{d}) = \frac{f_j(\mathbf{a}, \mathbf{s})}{f_j(\mathbf{s})}, \tag{5}$$

with and without imposing H_0^σ, respectively. To obtain the chi-squared limiting distribution, we set the weight as $W_j(\mathbf{d}) = f_j(\mathbf{s})/\widehat{P}(\mathbf{d})$.

Also, to test the equivalence of the transition matrices $\mathrm{H}_0^{\mathbf{P}}$, we set $\mathbf{d} = (\mathbf{s}', \mathbf{s})$ and $\mathbf{D} = \mathbf{S} \times \mathbf{S}$. Let $f_j^1(\mathbf{s}', \mathbf{s}) = \sum_{t=1}^{T-1} \mathbf{1}\{\mathbf{s}_j^{t+1} = \mathbf{s}', \mathbf{s}_j^t = \mathbf{s}\}$ and $f_j^1(\mathbf{s}) = \sum_{t=1}^{T-1} \mathbf{1}\{\mathbf{s}_j^t = \mathbf{s}\}$. Then we estimate the transition probability $p_j(\mathbf{s}'|\mathbf{s})$ by

$$\widehat{P}(\mathbf{d}) = \frac{\sum_{j=1}^{M} f_j^1(\mathbf{s}', \mathbf{s})}{\sum_{j=1}^{M} f_j^1(\mathbf{s})}, \qquad \widehat{P}_j(\mathbf{d}) = \frac{f_j^1(\mathbf{s}', \mathbf{s})}{f_j^1(\mathbf{s})}, \tag{6}$$

with and without imposing $\mathrm{H}_0^{\mathbf{P}}$, respectively. The weight is set as $W_j(\mathbf{d}) = f_j^1(\mathbf{s})/\widehat{P}(\mathbf{d})$.

The limiting null distribution of the statistic $\mathcal{T}_P$ is obtained in the following proposition (see Appendix A.1 for the proof).

Proposition 1 *Consider the set-up of Section 2. Suppose that all states of the Markov chain $\mathbf{P}_j$ communicate for each $j = 1, \ldots, M$ and that the observations $(\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1,\ldots,T}$ are mutually independent over $j = 1, \ldots, M$. Then under H_0^{σ} (or respectively $\mathrm{H}_0^{\mathbf{P}}$), the statistic $\mathcal{T}_P$ converges in distribution to the chi-squared distribution with degrees of freedom $(M-1)m_s(m_a-1)$ (or respectively $(M-1)m_s(m_s-1)$) as the length of time periods T increases to infinity.*

Bootstrap critical value. The chi-squared limiting distributions of the statistic $\mathcal{T}_P$ gives us critical values to control the asymptotic null rejection probabilities. Alternatively one may compute critical values by some bootstrap method.

For example, to test the null $\mathrm{H}_0^{\mathbf{P}}$, we can randomly pick an initial state $\mathbf{s}_0 \in \mathbf{S}$ and then draw the bootstrap counterpart $f_j^{1,b}(\mathbf{s}', \mathbf{s})$ of $f_j^1(\mathbf{s}', \mathbf{s})$ from the estimated conditional probability $\widehat{P}(\mathbf{d})$ in (6) for $\mathbf{s}, \mathbf{s}' \in \mathbf{S}$, $j = 1, \ldots, M$, and $b = 1, \ldots, B$. Note that we start the sampling process only after a certain number of time periods in order to neutralize the effect of the arbitrary choice of the initial state. Then the bootstrap counterpart $\mathcal{T}_P^b$ of the statistic $\mathcal{T}_P$ is given by replacing $f_j^1(\mathbf{s}', \mathbf{s})$ and $f_j^1(\mathbf{s})$ in (6) with $f_j^{1,b}(\mathbf{s}', \mathbf{s})$ and $f_j^{1,b}(\mathbf{s})$, respectively.

Also, to test the null H_0^{σ}, we can use the fact that action profiles $\mathbf{a} \in \mathbf{A}$ conditional on a state $\mathbf{s} \in \mathbf{S}$ are multinomially distributed with probabilities $\sigma_j(\mathbf{a}|\mathbf{s})$ in market j. State $\mathbf{s} \in \mathbf{S}$ occurs with frequency $f_j(\mathbf{s})$ and the probability of observing action state profiles $(\mathbf{a}, \mathbf{s})$ from $f_j(\mathbf{s})$ trials is given by the multinomial

$$\{f_j(\mathbf{a}, \mathbf{s})\}_{\mathbf{a} \in \mathbf{A}} | f_j(\mathbf{s}) \sim \mathrm{Multinomial}(f_j(\mathbf{s}), \{\sigma_j(\mathbf{a}|\mathbf{s})\}_{\mathbf{a} \in \mathbf{A}}),$$

for each $j = 1, \ldots, M$. We can use this distribution to implement a parametric bootstrap. More precisely, we fix $\mathbf{s} \in \mathbf{S}$ and draw the bootstrap counterpart $\{f_j^b(\mathbf{a}, \mathbf{s})\}_{\mathbf{a} \in \mathbf{A}}$ of $\{f_j(\mathbf{a}, \mathbf{s})\}_{\mathbf{a} \in \mathbf{A}}$ for $b = 1, \ldots, B$ from the multinomial distribution with the number of trials $f_j(\mathbf{s})$ and the weight vector $\{\widehat{P}(\mathbf{a}, \mathbf{s})\}_{\mathbf{a} \in \mathbf{A}}$ in (5). Then the bootstrap counterpart $\mathcal{T}_P^b$ is given by replacing $f_j(\mathbf{a}, \mathbf{s})$ in (5) with $f_j^b(\mathbf{a}, \mathbf{s})$. Here we only resample $f_j^b(\mathbf{a}, \mathbf{s})$ and the number of trials $f_j(\mathbf{s})$ is held fixed by the original sample. Based on a similar argument to Andrews (1997, Corollary 1), we can see that the $(1-\alpha)$-th quantile of $\mathcal{T}_P^1, \ldots, \mathcal{T}_P^B$ is an asymptotically valid critical value.

Similarly, to test the null $\mathrm{H}_0^{\mathbf{P}}$ on the transition matrices, we draw a bootstrap counterpart $\{f_j^{1,b}(\mathbf{s}', \mathbf{s})\}_{\mathbf{s}' \in \mathbf{S}}$ of $\{f_j^1(\mathbf{s}', \mathbf{s})\}_{\mathbf{s}' \in \mathbf{S}}$ from the multinomial distribution with the number of trials $f_j^1(\mathbf{s})$ and weight vector $\{\widehat{P}(\mathbf{s}', \mathbf{s})\}_{\mathbf{s}' \in \mathbf{S}}$ in (6). Then the bootstrap counterpart $\mathcal{T}_P^b$ is given by replacing $f_j^1(\mathbf{s}', \mathbf{s})$ in (6) with $f_j^{1,b}(\mathbf{s}', \mathbf{s})$.

Optimal test statistic. The test statistic $\mathcal{T}_P$ is constructed by measuring the chi-squared distance between the nonparametric estimators $\widehat{P}(\mathbf{d})$ and $\widehat{P}_j(\mathbf{d})$

for the discrete distribution over $\mathbf{D}$ with and without imposing the null hypothesis, respectively. There are many other ways to measure the discrepancy between the single market and full-sample estimates. For example, we can measure discrepancy of conditional probabilities by the (weighted) Kullback-Leibler divergence

$$\mathcal{T}_P^* = 2\sum_{j=1}^{M}\sum_{\mathbf{d}\in\mathbf{D}} W_j(\mathbf{d})\widehat{P}_j(\mathbf{d})\log\frac{\widehat{P}_j(\mathbf{d})}{\widehat{P}(\mathbf{d})}. \tag{7}$$

In order to test the null hypothesis H_0^{σ} on the conditional choice probabilities, we can set as $\mathbf{d} = (\mathbf{a}, \mathbf{s})$ and $\mathbf{D} = \mathbf{A} \times \mathbf{S}$ and estimate $\widehat{P}_j(\mathbf{d})$ and $\widehat{P}(\mathbf{d})$ as in (5). Also, to test the null hypothesis $\mathrm{H}_0^{\mathbf{P}}$ on the transition probabilities, we set as $\mathbf{d} = (\mathbf{s}', \mathbf{s})$ and $\mathbf{D} = \mathbf{S} \times \mathbf{S}$ and then estimate $\widehat{P}_j(\mathbf{d})$ and $\widehat{P}(\mathbf{d})$ as in (6). For both cases, we set the weight as $W_j(\mathbf{d}) = f_j(\mathbf{s})$ to obtain the chi-squared limiting distribution. The test statistic $\mathcal{T}_P^*$ is a likelihood-ratio version of the chi-squared statistic $\mathcal{T}_P$. These statistics are asymptotically equivalent under the null and local alternative hypotheses (e.g. van der Vaart, 1998, Lemma 17.3).

On the other hand, in the literature of hypothesis testing for multinomial distributions, Hoeffding (1965) discovered that the likelihood ratio statistic for the simple hypothesis on multinomials enjoys some global power optimality which is not shared by the chi-squared statistic. In particular, under some restriction on the convergence rate of the type I error probability, the likelihood ratio statistic achieves the highest power under fixed alternatives. This optimality is called the generalized Neyman-Pearson optimality and has been extended to several contexts (see, Gutman, 1989). By extending the argument in Gutman (1989) to our set-up, we derive the following optimality for $\mathcal{T}_P^*$ (see Appendix A.2 for the proof).

Proposition 2 *Under the same set-up of Proposition 1 with fixed initial states* $(\mathbf{s}_1^0, \ldots, \mathbf{s}_M^0)$, *consider the statistic* $\mathcal{T}_P^*$ *with* $W_j(\mathbf{d}) = f_j(\mathbf{s})$ *and (5) to test* H_0^{σ}. *There exists a positive sequence* $\delta_T = O(T^{-1}\log T)$ *such that*

$$\lim_{T\to\infty}\frac{1}{T}\log\Pr\{\mathcal{T}_P^* \geq 2T(\alpha - \delta_T) : \mathrm{H}_0^{\sigma}\} \leq -\alpha, \tag{8}$$

for $\alpha > 0$, *and that for any test statistic* $\mathcal{T}_A$ *for* H_0^{σ} *satisfying*

$$\lim_{T\to\infty}\frac{1}{T}\log\Pr\{\mathcal{T}_A \textit{ rejects } \mathrm{H}_0^{\sigma} : \mathrm{H}_0^{\sigma}\} \leq -\alpha, \tag{9}$$

it holds

$$\Pr\{\mathcal{T}_P^* \geq 2T(\alpha - \delta_T) : \mathrm{H}_1^{\sigma}\} \geq \Pr\{\mathcal{T}_A \textit{ rejects } \mathrm{H}_0^{\sigma} : \mathrm{H}_1^{\sigma}\}, \tag{10}$$

for all T *large enough.*

Also the same result holds for the statistic $\mathcal{T}_P^*$ *with* $W_j(\mathbf{d}) = f_j(\mathbf{s})$ *and (6) to test* $\mathrm{H}_0^{\mathbf{P}}$ *by replacing* H_0^{σ}, H_1^{σ}, *and* T *with* $\mathrm{H}_0^{\mathbf{P}}$, $\mathrm{H}_1^{\mathbf{P}}$, *and* $T-1$, *respectively.*

This proposition says that in the class of test statistics satisfying the restriction on the exponential decay rate of the type I error probability in (9), the Kullback-Leibler statistic $\mathcal{T}_P^*$ attains the highest power. This optimality result is a natural extension of the generalized Neyman-Pearson optimality analysis to homogeneity testing of conditional choice or transition probabilities.

Parametric model for σ and $\mathbf{P}$. Suppose we parametrize the choice probability σ_j or transition matrix $\mathbf{P}_j$ by a parametric model $\sigma(\mathbf{a}|\mathbf{s};\theta_j)$ or $p(\mathbf{s},\mathbf{s}';\theta_j)$, such as logit. We assume that the functional forms are identical across markets and the different equilibria are characterized by different parameter values of θ_j. In this case, the null hypothesis of interest can be written as

$$\mathrm{H}_0^\theta : \theta_1 = \cdots = \theta_M.$$

Since this is a parameter hypothesis for a discrete parametric model, standard maximum likelihood theory applies. In particular, the score test would be convenient since the test statistic requires only the full sample estimator.

Comment on the large-T asymptotics. The asymptotic analysis for the test based on $\mathcal{T}_P$ (and $\mathcal{T}_Q$ in the next subsection) is conducted under the framework of $T \to \infty$ while M fixed. As far as the researcher is interested in consistency of the parameter estimates as $T \to \infty$, we do not need to pool the data across markets. However, if the researcher is concerned with efficiency of the estimates, the test based on $\mathcal{T}_P$ would be a useful diagnostic to decide whether she can increase the sample size by pooling. Indeed there are some empirical examples where T is not small but the researchers are tempted to pool the data across markets, such as Ryan (2012) and Collard-Wexler (2013).

Also, we note that the large-T asymptotic analysis above is basically for convenience to obtain the critical value. Although it is computationally too expensive to implement in our typical examples, in principle it is possible to conduct exact (i.e., fixed T) test based on $\mathcal{T}_P$ by adapting the simulation-based approach of Besag and Mondal (2013).

Large number of markets M. The asymptotic distribution of the test statistic $\mathcal{T}_P$ is derived under the assumption that the number of markets M is fixed. However, there are some cases where M is large relative to the length of time T; e.g., Collard-Wexler (2013) and Dunne, Klimek, Roberts and Xu (2013). When M is large, it may be useful to investigate the limiting behavior of the statistic $\mathcal{T}_P$ as both M and T diverge to infinity. Let $\{M_T\}$ be a sequence satisfying $M_T \to \infty$ and $M_T/T \to 0$ as $T \to \infty$. In this case, intuitively, the degree of freedom for the limiting distribution of $\mathcal{T}_P$ grows to infinity. Thus after standardization, the limiting distribution of $\mathcal{T}_P$ is characterized by the standard normal. For example, the test statistic for H_0^σ based on (5) satisfies

$$\frac{\mathcal{T}_P - (M_T - 1)m_s(m_a - 1)}{\sqrt{2(M_T - 1)m_s(m_a - 1)}} \xrightarrow{d} N(0,1),$$

as $T \to \infty$ under H_0^σ. A similar result applies for the test of $\mathrm{H}_0^{\mathbf{P}}$.

Comparison with de Paula and Tang (2011). Note that our test can allow for within-period correlation. In the context of static games with incomplete information, de Paula and Tang (2011) test conditional independence between players' actions to check if there are more than one equilibria in the data generating process. This test relies on the assumption of independent-across-players actions conditional on state variables. For example, this may arise if there is a utility component in payoffs unobserved by the econometrician but known to players.[4] Our test is more flexible and permits within-period correlation in players' actions conditional on state variables. The permissible information structure and set of games our framework can deal with is more general. Our tests explore the way that the game and states evolve and require repeated observations for each market.[5]

3.2 Testing steady-state distribution

We now consider testing of $\mathrm{H}_0^{\mathbf{Q}}$ in (2), which examines the steady-state distribution in individual markets and compares it to the average (across markets) steady-state distribution. Under the null hypothesis of identical steady-state distributions, the market specific and average market distributions are close to each other. The test statistic is more intuitive in the sense that it compares two steady-state distributions directly. However, the test requires that the steady-state distributions exist and that the Markov chain is in the steady-state, see Lemma 1. That is, regardless of which hypothesis is true, we assume that all states in the chain $\mathbf{P}_j$ communicate for all markets j. The relative frequencies $\widehat{\mathbf{Q}}_j = \{T^{-1} f_j(\mathbf{s})\}_{\mathbf{s}\in\mathbf{S}}$ are nonparametric estimates of the steady-state distribution $\mathbf{Q}_j$. By Billingsley (1961, Theorem 3.3), the limiting distribution of $\widehat{\mathbf{Q}}_j$ is obtained as

$$T^{1/2}(\widehat{\mathbf{Q}}_j - \mathbf{Q}_j) \xrightarrow{d} N(0, \mathbf{V}_j), \tag{11}$$

where the asymptotic variance $\mathbf{V}_j$ is defined in Appendix A.3. Since $\mathrm{rank}(\mathbf{V}_j) = m_s - 1$, we can obtain a test statistic for $\mathrm{H}_0^{\mathbf{Q}}$ as

$$\mathcal{T}_Q = T\sum_{j=1}^{M}(\widehat{\mathbf{Q}}_j - \widehat{\mathbf{Q}})'\widehat{\mathbf{V}}^{-}(\widehat{\mathbf{Q}}_j - \widehat{\mathbf{Q}}) \xrightarrow{d} \chi^2((M-1)(m_s-1)), \tag{12}$$

---

[4] For example, suppose the random profit shocks are correlated across players within a time period. Then, $\sigma(a_1|\mathbf{s})\cdots\sigma(a_N|\mathbf{s}) \neq \sigma(\mathbf{a}|\mathbf{s})$ even under H_0^{σ}. Two-step methods work in a similar manner as in the case of i.i.d. profit shocks; a researcher would have to estimate the choice probability of action profile $\mathbf{a}$ instead of each player's CCPs separately. The inference of the underlying structural parameters can then be based on the joint choice probability estimates and the appropriate equilibrium conditions.

[5] Tests of independence are used in various contexts to find evidence for unobserved variations in data that non-trivially affect agents' actions. For example, Chiappori and Salanié (2000) test the conditional independence of the choice of better coverage and the occurrence of an accident using data of automobile insurance, and attributes a violation of the conditional independence to the existence of asymmetric information between customers and insurance companies. de Paula and Tang (2011) assume independent private shocks in games with incomplete information and regard additional variations (after controlling for observable covariates) as coming from multiple equilibria being played in data. On the other hand, Navarro and Takahashi (2012) interpret a violation of the conditional independence as a rejection of models of pure private shocks.

under $\mathrm{H}_0^{\mathbf{Q}}$, where $\widehat{\mathbf{Q}} = M^{-1}\sum_{j=1}^{M}\widehat{\mathbf{Q}}_j$ and $\widehat{\mathbf{V}}^{-}$ means a generalized inverse of $\widehat{\mathbf{V}}$, which is defined in Appendix A.3. Although this statistic validates the use of the chi-squared critical value for the asymptotic test, the estimator $\widehat{\mathbf{V}}$ may not be easy to compute and requires a bandwidth choice. Thus in our simulation and empirical studies below, we replace $\widehat{\mathbf{V}}$ in (12) with the identity matrix and employ some bootstrap critical value.

3.3 Testing conditional state distribution given the initial state

Our final test does not require that the Markov chain has a unique steady-state distribution or that all states communicate. Such situations may arise in new or growing industries when the steady-state has not been reached yet. It may also arise in situations when there is no unique steady-state distribution. For example, it may arise when some states are absorbing. These situations share the feature that the limiting state distributions may depend on the initial state. To develop a test for this case we consider the conditional state distribution given the initial state. We assume that the number of markets M is large (and the length of time periods T can be short).

To describe a suitable test statistic, we treat the state profiles across markets as an i.i.d. sample from the distribution parametrized by the transition matrix $\mathbf{P}$, and propose a test for the null hypothesis $\mathrm{H}_0^{\mathbf{s}}$ in (3). Let $\widehat{\mathbf{P}}$ be the frequency estimator of the state transition matrix based on the whole state profiles. Also let $\widehat{\mathbf{Q}}_{\mathbf{s}}^t = \left\{\frac{\sum_{j=1}^{M}\mathbf{1}\{\mathbf{s}_j^t=\mathbf{s}',\mathbf{s}_j^1=\mathbf{s}\}}{\sum_{j=1}^{M}\mathbf{1}\{\mathbf{s}_j^1=\mathbf{s}\}}\right\}_{\mathbf{s}'\in\mathbf{S}}$ be the relative frequency estimator for the vector of conditional probabilities $\left\{\Pr\{\mathbf{s}^t=\mathbf{s}'|\mathbf{s}^1=\mathbf{s}\}\right\}_{\mathbf{s}'\in\mathbf{S}}$ for $t = 2,\dots,T$ for a given initial state $\mathbf{s}$. If our model parametrized by $\mathbf{P}$ is correct, the contrast between $\widehat{\mathbf{Q}}_{\mathbf{s}}^t$ and $\iota_{\mathbf{s}}'\widehat{\mathbf{P}}^t$ should be close to zero for all $t = 2,\dots,T$. We evaluate the contrast $\mathbf{C}_{\mathbf{s}}' = (T-1)^{-1}(\sum_{t=2}^{T}\widehat{\mathbf{Q}}_{\mathbf{s}}^t - \iota_{\mathbf{s}}'\sum_{t=2}^{T}\widehat{\mathbf{P}}^t)$. The test statistic satisfies

$$\mathcal{T}_{\mathbf{s}} = M\mathbf{C}_{\mathbf{s}}'\widehat{\mathbf{V}}_{\mathbf{s}}^{-}\mathbf{C}_{\mathbf{s}} \xrightarrow{d} \chi^2(m_s-1), \tag{13}$$

as $M\to\infty$ with fixed T under $\mathrm{H}_0^{\mathbf{s}}$, where $\widehat{\mathbf{V}}_{\mathbf{s}}^{-}$ is a generalized inverse of an estimator of the asymptotic variance of $\sqrt{M}\mathbf{C}_{\mathbf{s}}$ under $\mathrm{H}_0^{\mathbf{s}}$.[6] As in (12), the estimator $\widehat{\mathbf{V}}_{\mathbf{s}}$ may not be easy to compute. Thus in our simulation and empirical studies below, we replace $\widehat{\mathbf{V}}_{\mathbf{s}}$ in (13) with the identity matrix and employ some bootstrap critical value.

The test based on $\mathcal{T}_{\mathbf{s}}$ requires $T\geq 3$. The standard argument implies that it has non-trivial power against local alternatives approaching to the null at

[6] We can also consider the hypothesis

$$\mathrm{H}_0^{\mathbf{s},T} : \Pr\left\{\mathbf{s}^t=\mathbf{s}'|\mathbf{s}^1=\mathbf{s}\right\} = \iota_{\mathbf{s}}'\mathbf{P}^t \quad \text{for all } t=2,\dots,T.$$

Under $\mathrm{H}_0^{\mathbf{s},T}$, the Wald statistic for this hypothesis will converge to $\chi^2(T(m_s-1))$ as $M\to\infty$ with fixed T. Also its normalized version converges to the standard normal distribution as $M,T\to\infty$ but $T/M\to 0$.

the $\sqrt{M}$-rate. The local power function is characterized by a non-central χ^2 distribution. As T increases, both the non-centrality parameter and degree-of-freedom increase. Thus, overall the effect of T on local power is indeterminate.

We note that a rejection by the statistic $\mathcal{T}_{\mathbf{s}}$ occurs typically in two scenarios. First, even though the state profile $(\mathbf{s}_j^t)_{t=1,\dots,T}$ is i.i.d. over $j = 1, \dots, M$, a violation of the first-order time-homogeneous Markov chain assumption yields a large value of $\mathcal{T}_{\mathbf{s}}$. Second, if the state profile $(\mathbf{s}_j^t)_{t=1,\dots,T}$ is not i.i.d. over $j = 1, \dots, M$, then there is no guarantee that $\widehat{\mathbf{Q}}_{\mathbf{s}}^t$ and $\iota_{\mathbf{s}}'\widehat{\mathbf{P}}^t$ converge to the same limit[7] and the statistic $\mathcal{T}_{\mathbf{s}}$ tends to be large.[8] Although we cannot distinguish these sources of rejection, we argue that the second type of rejection can be associated with multiplicity of equilibria.

3.4 Relationships among test statistics

The three test statistics provided in the previous subsections have different advantages depending on the application and the type of data. Given that in standard dynamic discrete models, player's behavior is described in the form of conditional choice probabilities, the test based on H_0^σ ($\mathcal{T}_P$ and $\mathcal{T}_P^*$ using (5)) would be a natural starting point. It is also reasonable to use the test based on $\mathrm{H}_0^{\mathbf{P}}$. Under the assumption that $\mathbf{G}_j$ is identical for all markets j, testing $\mathrm{H}_0^{\mathbf{P}}$ plays a similar role to testing H_0^σ. In general, however, rejecting the null $\mathrm{H}_0^{\mathbf{P}}$ may also arise because of differences in $\mathbf{G}_j$ even if H_0^σ holds.

Homogeneity of data generating processes across markets can also be tested by the null hypothesis $\mathrm{H}_0^{\mathbf{Q}}$ using the steady-state distribution test statistic $\mathcal{T}_Q$. Since the dimension of the hypothesis decreases, we expect it to have higher power compared to $\mathcal{T}_P$. It should also be emphasized, however, that there is a region where homogeneity is violated but the test based on $\mathcal{T}_Q$ is not able to detect. Put differently, if the test based on $\mathcal{T}_Q$ rejects the null hypothesis $\mathrm{H}_0^{\mathbf{Q}}$, we conclude that the maintained assumption for poolability is violated; on the other hand, if it does not rejects the null, there may still be multiple equilibria or some misspecification that would invalidate pooling (as two distinct transition matrices may yield the same steady-state distribution). Therefore, we recommend the following procedure in practice. First, the test of $\mathrm{H}_0^{\mathbf{Q}}$ based on $\mathcal{T}_Q$ is applied to take advantage of its desirable power property. If the null $\mathrm{H}_0^{\mathbf{Q}}$ is rejected, then we stop and conclude that the maintained assumption for estimation using pooled data is violated. If the test does not reject the null $\mathrm{H}_0^{\mathbf{Q}}$, then we proceed to apply $\mathcal{T}_P$ for H_0^σ or $\mathrm{H}_0^{\mathbf{P}}$. By proceeding in this way, it can

[7]Similarly to existing specification test statistics with the parametric rate, it is possible that the probability limits of $\widehat{\mathbf{Q}}_{\mathbf{s}}^t$ and $\iota_{\mathbf{s}}'\widehat{\mathbf{P}}^t$ coincide under the alternative. This issue is known as the implicit null hypothesis (see, e.g., Mizon and Richard, 1986). The statistic $\mathcal{T}_{\mathbf{s}}$ has no power in such a situation. To alleviate this issue, the statistic $\mathcal{T}_{\mathbf{s}}$ may be calculated for different initial states.

[8]For example, consider the case where $\{(\mathbf{s}_j^t)_{t=1,\dots,T}\}_{j=1}^{M/2}$ and $\{(\mathbf{s}_j^t)_{t=1,\dots,T}\}_{j=M/2+1}^{M}$ are i.i.d. samples following distinct Markov chains. In this case, we can see that $\widehat{\mathbf{Q}}_{\mathbf{s}}^t$ and $\iota_{\mathbf{s}}\widehat{\mathbf{P}}^t$ converge to distinct limits.

be made sure that the tests are consistent and the power property of $\mathcal{T}_Q$ can be exploited.

There are also situations where states do not communicate or initial conditions matter. In such cases, the conditional state distribution test $\mathcal{T}_{\mathbf{s}}$ can be used. It is also worth emphasizing that $\mathcal{T}_{\mathbf{s}}$ is suitable when M becomes large, while T is fixed (i.e., short panel). Some empirical applications in IO have this data structure; e.g., Collard-Wexler (2013) and Dunne, Klimek, Roberts and Xu (2013).

3.5 Unobservable state variables

This subsection considers the situation in which the public state variables have two components, $\mathbf{s}^t = (\mathbf{s}_1^t, \mathbf{s}_2^t)$. The component $\mathbf{s}_1^t$ is observed by the econometrician but the component $\mathbf{s}_2^t$ is not. Arcidiacono and Miller (2011) consider this framework with a parametric model of unobserved state variables with known finite support. They propose an estimation strategy based on the EM-type algorithm. Our testing procedures presented above can be extended to this framework as follows.

First, we note that the null of homogeneity of the joint transition $\Pr\{\mathbf{s}^{t+1} = \mathbf{s}'|\mathbf{s}^t = \mathbf{s}\}$ implies the homogeneity of the marginal transition $\Pr\{\mathbf{s}_1^{t+1} = \mathbf{s}_1'|\mathbf{s}_1^t = \mathbf{s}_1\}$. Thus, if the researcher is willing to assume that all states of the Markov chain $\mathbf{P}_j$ communicate, then our tests for $\mathrm{H}_0^{\mathbf{P}}$, H_0^{σ}, and $\mathrm{H}_0^{\mathbf{Q}}$ presented in Sections 3.1-3.3 can be applied to the observable component $\mathbf{s}_1^t$. These tests using only $\mathbf{s}_1^t$ can be interpreted as the ones for homogeneity of the marginal transition $\Pr\{\mathbf{s}_1^{t+1} = \mathbf{s}_1'|\mathbf{s}_1^t = \mathbf{s}_1\}$, which is implied from homogeneity of the joint transition $\Pr\{\mathbf{s}^{t+1} = \mathbf{s}'|\mathbf{s}^t = \mathbf{s}\}$. Therefore, a rejection by $\mathcal{T}_P$ using only $\mathbf{s}_1^t$ implies a rejection of homogeneity of $\Pr\{\mathbf{s}^{t+1} = \mathbf{s}'|\mathbf{s}^t = \mathbf{s}\}$ even though the econometrician does not observe $\mathbf{s}_2^t$. On the other hand, an acceptance by $\mathcal{T}_P$ using only $\mathbf{s}_1^t$ does not necessarily imply an acceptance of homogeneity of $\Pr\{\mathbf{s}^{t+1} = \mathbf{s}'|\mathbf{s}^t = \mathbf{s}\}$. Similar comments apply to the tests of H_0^{σ} and $\mathrm{H}_0^{\mathbf{Q}}$.

Second, a researcher may be interested in situations in which the unobserved component has a permanent time-invariant market level variable (sometimes called market level unobserved heterogeneity). We shall illustrate how our test statistics can be modified to allow for an unobservable time-invariant state variable $\mathbf{s}_2$ with a known finite support. Suppose $\mathbf{s}_2$ is binary for simplicity. We can modify the null hypothesis $\mathrm{H}_0^{\mathbf{s}}$ in (3) by setting the transition matrix as a mixture $\mathbf{P} = \pi\mathbf{P}^{(a)} + (1-\pi)\mathbf{P}^{(b)}$. Under $\mathrm{H}_0^{\mathbf{s}}$ with certain regularity conditions, we can estimate $(\pi, \mathbf{P}^{(a)}, \mathbf{P}^{(b)})$ based on the pooled sample across markets by applying the methods in Arcidiacono and Miller (2011) and Kasahara and Shimotsu (2009). Based on these estimates, we obtain an estimator of $\mathbf{P}$, say $\tilde{\mathbf{P}}$. Then we can apply the test statistic $\mathcal{T}_{\mathbf{s}}$ in (13) by replacing $\widehat{\mathbf{P}}$ with $\tilde{\mathbf{P}}$, i.e.,

$$\tilde{\mathcal{T}}_{\mathbf{s}} = M\tilde{\mathbf{C}}_{\mathbf{s}}'\tilde{\mathbf{V}}_{\mathbf{s}}^{-}\tilde{\mathbf{C}}_{\mathbf{s}}, \tag{14}$$

where $\tilde{\mathbf{C}}_{\mathbf{s}}' = (T-1)^{-1}(\sum_{t=2}^{T}\widehat{\mathbf{Q}}_{\mathbf{s}}^t - \iota_{\mathbf{s}}'\sum_{t=2}^{T}\tilde{\mathbf{P}}^t)$ and $\tilde{\mathbf{V}}_{\mathbf{s}}^{-}$ is a generalized inverse of an estimator of the asymptotic variance of $\sqrt{M}\tilde{\mathbf{C}}_{\mathbf{s}}$ under $\mathrm{H}_0^{\mathbf{s}}$. Similar to $\mathcal{T}_{\mathbf{s}}$, this statistic converges to a χ^2 distribution under $\mathrm{H}_0^{\mathbf{s}}$ as $M \to \infty$ while T is fixed.

Third, we illustrate how to extend the test for $\mathrm{H}_0^{\mathbf{P}}$ in (1) to accommodate unobservable time-invariant state variables. Again, for simplicity of exposition suppose $\mathbf{s}_2$ is binary. We can modify the null hypothesis as

$$\tilde{\mathrm{H}}_0^{\mathbf{P}} : \mathbf{s}_j \text{ is a Markov chain from } \mathbf{P}^{(a)} \text{ or } \mathbf{P}^{(b)} \text{ for all } j.$$

As $M \to \infty$, we can consistently estimate $\mathbf{P}^{(a)}$ and $\mathbf{P}^{(b)}$ using the pooled sample across markets by applying Arcidiacono and Miller (2011) or Kasahara and Shimotsu (2009). Let $\tilde{\mathbf{P}}^{(a)}$ and $\tilde{\mathbf{P}}^{(b)}$ be such estimators. On the other hand, as $T \to \infty$, the estimator $\hat{\mathbf{P}}_j$ defined in (6) consistently estimates the transition for each market j and thus converges to $\mathbf{P}^{(a)}$ or $\mathbf{P}^{(b)}$ under $\tilde{\mathrm{H}}_0^{\mathbf{P}}$. Based on these observations, a test statistic for $\tilde{\mathrm{H}}_0^{\mathbf{P}}$ may be constructed as $\tilde{\mathcal{T}}_P = \sum_{j=1}^M \tilde{\mathcal{T}}_{P,j}$, where

$$\tilde{\mathcal{T}}_{P,j} = \min \left\{ \begin{array}{c} \left(\sum_{\mathbf{s}',\mathbf{s}\in\mathbf{S}} \frac{f_j^1(\mathbf{s})}{\tilde{P}^{(a)}(\mathbf{s}',\mathbf{s})}\{\widehat{P}_j(\mathbf{s}',\mathbf{s}) - \tilde{P}^{(a)}(\mathbf{s}',\mathbf{s})\}^2\right), \\ \left(\sum_{\mathbf{s}',\mathbf{s}\in\mathbf{S}} \frac{f_j^1(\mathbf{s})}{\tilde{P}^{(b)}(\mathbf{s}',\mathbf{s})}\{\widehat{P}_j(\mathbf{s}',\mathbf{s}) - \tilde{P}^{(b)}(\mathbf{s}',\mathbf{s})\}^2\right) \end{array} \right\}. \tag{15}$$

This construction of the test statistic (i.e., aggregate the statistic $\tilde{\mathcal{T}}_{P,j}$ over cross-section units $j = 1, \ldots, M$) appears often in the literature of large-T panel data analysis (see, e.g., Baltagi, 2008, ch. 12). In this literature, it is common to take the sequential limits (i.e., take $T \to \infty$ first to derive the limiting distribution of $\tilde{\mathcal{T}}_{P,j}$ for each j, and then take $M \to \infty$ to establish the limiting distribution of $\tilde{\mathcal{T}}_P$) to analyze the asymptotic properties of test statistics, such as panel unit root tests. Phillips and Moon (1999) provided additional requirements to strengthen the sequential limit theory to the joint one, where T and M can grow in an arbitrary way. However, in our setup, the statistic $\tilde{\mathcal{T}}_{P,j}$ for market j depends on both M (for $\tilde{P}^{(a)}$ and $\tilde{P}^{(b)}$) and T (for $\widehat{P}_j(\mathbf{s})$). Therefore, the existing techniques of large-T panel data analysis are not directly applicable. Although the complete analysis of the asymptotic theory for $\tilde{\mathcal{T}}_P$ is beyond the scope of this paper, we can adjust the construction of the test statistic to fit into the sequential asymptotic framework. To this end, we choose the sample size to estimate $\tilde{\mathbf{P}}^{(a)}$ and $\tilde{\mathbf{P}}^{(b)}$ as a function of T, say C_T. Also we assume $C_T/T \to \infty$ as $T \to \infty$, which guarantees that the estimation errors $\tilde{P}^{(a)} - P^{(a)}$ and $\tilde{P}^{(b)} - P^{(b)}$ are negligible. Since $\tilde{\mathbf{P}}^{(a)}$ and $\tilde{\mathbf{P}}^{(b)}$ are typically computed by a pooled sample across markets, the requirement $C_T/T \to \infty$ is mild. Under these additional requirements, the statistic $\tilde{\mathcal{T}}_{P,j}$ depends only on T and satisfies

$$\begin{aligned} \tilde{\mathcal{T}}_{P,j} &= \min \left\{ \begin{array}{c} \left(\sum_{\mathbf{s}',\mathbf{s}\in\mathbf{S}} \frac{f_j^1(\mathbf{s})}{\tilde{P}^{(a)}(\mathbf{s}',\mathbf{s})}\{\widehat{P}_j(\mathbf{s}',\mathbf{s}) - P^{(a)}(\mathbf{s}',\mathbf{s})\}^2\right), \\ \left(\sum_{\mathbf{s}',\mathbf{s}\in\mathbf{S}} \frac{f_j^1(\mathbf{s})}{\tilde{P}^{(b)}(\mathbf{s}',\mathbf{s})}\{\widehat{P}_j(\mathbf{s}',\mathbf{s}) - P^{(b)}(\mathbf{s}',\mathbf{s})\}^2\right) \end{array} \right\} + o_p(1) \\ &\overset{d}{\to} \chi^2_{m_s(m_s-1)} \quad \text{as } T \to \infty \text{ under } \tilde{\mathrm{H}}_0^{\mathbf{P}}, \end{aligned}$$

for every j. Therefore, we can obtain the limiting distribution of $\tilde{\mathcal{T}}_P = \sum_{j=1}^M \tilde{\mathcal{T}}_{P,j}$

under the sequential limit, i.e.,

$$\frac{\tilde{\mathcal{T}}_P - Mm_s(m_s-1)}{\sqrt{2Mm_s(m_s-1)}} \xrightarrow{d} N(0,1),$$

as $T \to \infty$ followed by $M \to \infty$ sequentially. This sequential limiting result may be strengthened to the joint one by verifying additional conditions in Phillips and Moon (1999, Lemma 6).

In practice, the test for $\tilde{\mathrm{H}}_0^{\mathbf{P}}$ based on $\tilde{\mathcal{T}}_P$ is used as follows. If we reject $\tilde{\mathrm{H}}_0^{\mathbf{P}}$, the maintained assumption for pooling the whole data is violated and it is recommended to look for a subset that preserves homogeneity. On the other hand, acceptance of the null $\tilde{\mathrm{H}}_0^{\mathbf{P}}$ is considered as a supporting evidence for the researcher to pool the data across markets to implement two-step estimation for parameters, where the first-step estimates are constructed by using $\tilde{\mathbf{P}}^{(a)}$ and $\tilde{\mathbf{P}}^{(b)}$.

4 Monte Carlo

This section examines the practical aspects of the proposed tests in a Monte Carlo study. We consider a simple and transparent dynamic oligopoly game with multiple equilibria. The game was illustrated and analyzed in more detail in Pesendorfer and Schmidt-Dengler (2008). It has the following features.

There are two players, binary actions $a_i^t \in \{0,1\}$, and binary states $s_i^t \in \{0,1\}$. The distribution of the profitability shocks is the standard normal. The discount factor is fixed at 0.9. The state transition law is given by $s_i^{t+1} = a_i^t$. Period payoffs are symmetric and parametrized as follows:

$$\pi(a_i, a_j, s_i) = \begin{cases} 0 & \text{if } a_i = 0;\ s_i = 0 \\ 0.1 & \text{if } a_i = 0;\ s_i = 1 \\ \pi^1 - 0.2 & \text{if } a_i = 1;\ a_j = 0;\ s_i = 0 \\ \pi^2 - 0.2 & \text{if } a_i = 1;\ a_j = 1;\ s_i = 0 \\ \pi^1 & \text{if } a_i = 1;\ a_j = 0;\ s_i = 1 \\ \pi^2 & \text{if } a_i = 1;\ a_j = 1;\ s_i = 1 \end{cases}$$

where $\pi^1 = 1.2$; and $\pi^2 = -1.2$. The period payoffs can be interpreted as stemming from a game with switching costs and/or as entry/exit game. A player that selects action 1 receives monopoly profits π^1 if she is the only active player, and she receives duopoly profits π^2 otherwise. Additionally, a player that switches states from 0 to 1 incurs the entry cost 0.2; while a player that switches from 1 to 0 receives the exit value 0.1.

Multiplicity. The game illustrates the possibility of multiple equilibria which is a feature inherent to games. The following analysis focuses on two asymmetric equilibria of the three equilibria described in Pesendorfer and Schmidt-Dengler (2008). In equilibrium (i), player two is more likely to choose action 0 than player one in all states. The ex ante probability vectors for both players are given by $\sigma(a_1 = 0|s_1, s_2) = (0.27, 0.39, 0.20, 0.25)$, $\sigma(a_2 = 0|s_2, s_1) =$

$(0.72, 0.78, 0.58, 0.71)$, where the order of the elements in the probability vectors corresponds to the state vector $(s_1, s_2) \in \{(0,0),(0,1),(1,0),(1,1)\}$.

In equilibrium (ii), player two is more likely to choose action 0 than player one in all states with the exception of state $(1,0)$. The probability vectors are given by $\sigma(a_1 = 0|s_1, s_2) = (0.38, 0.69, 0.17, 0.39)$, $\sigma(a_2 = 0|s_2, s_1) = (0.47, 0.70, 0.16, 0.42)$.

Design. The simulated data are generated by randomly drawing a time series of actions from the calculated equilibrium choice probabilities described above for each of the equilibria (i)-(ii) respectively. The initial state is taken as $(0,0)$ and we start the sampling process after 100 periods. The number of markets and the length of the time series is varied in the experiment with the aim at staying close to typical industry applications. We choose $M = 20, 40, \ldots, 640$ and $T = 5, 10, \ldots, 640$. The parameter λ denotes the fraction of markets that adopt equilibrium (i) while $1 - \lambda$ denotes the fraction of markets that adopt equilibrium (ii).

Implementation. The Monte Carlo study considers the conditional choice probability multiplicity test by $\mathcal{T}_P$, its optimal version by $\mathcal{T}_P^*$, the steady-state distribution test by $\mathcal{T}_Q$, and the conditional state distribution test by $\mathcal{T}_{\mathbf{s}}$ as described in Section 3. In this example, $\mathbf{a}^t = \mathbf{s}^{t+1}$ and the state transition probabilities $\mathbf{P}$ equal the conditional choice probabilities σ. Therefore, the null hypotheses H_0^σ and $\mathrm{H}_0^{\mathbf{P}}$ and their tests are identical. To implement $\mathcal{T}_P$ in (4) and $\mathcal{T}_P^*$ in (7), we employ the formula in (6).[9] The steady-state probabilities $\mathbf{Q}$ are estimated by the relative frequencies. For the steady-state distribution test by $\mathcal{T}_Q$, we use the identity matrix for the variance matrix in (12). For the conditional state distribution test by $\mathcal{T}_{\mathbf{s}}$, we consider the sum $\mathcal{T}_{\mathbf{S}} = \sum_{\mathbf{s} \in \mathbf{S}} \mathcal{T}_{\mathbf{s}}$ instead of focusing on a particular initial state. To compute $\mathcal{T}_{\mathbf{s}}$, we replace the variance matrix $\widehat{\mathbf{V}}_{\mathbf{s}}$ in (13) with the identity matrix.

The critical values of these test statistics are calculated using a bootstrap procedure. For every bootstrap iteration b, we simulate choice/state profiles $\{\mathbf{s}_j^b\}$ from the transition matrix based on $\widehat{P}(\mathbf{d})$ defined in (6) for every market j. For the first three tests (i.e., the tests by $\mathcal{T}_P$, $\mathcal{T}_P^*$, and $\mathcal{T}_Q$), as in the data generating process, the initial state is taken as $(0,0)$ and we start the sampling process after 100 periods. For the test by $\mathcal{T}_{\mathbf{s}}$, for each market, we use the same initial state as is observed in the simulated sample and start the game from that state. The bootstrap counterparts of the test statistics are calculated for $b = 1, \ldots, B$. The critical values are obtained by the 95th percentile of the bootstrapped statistics.

Results. The experiment is based on $B = 999$ repetitions for the bootstrap sample and 1,000 Monte Carlo repetitions. Tables 1-4 report the results of the experiments. These tables report the percentages of rejections of our tests for selected values of M, T, and λ.

We first study the size properties of our tests. Tables 1 and 2 consider the cases of $\lambda = 1$ and $\lambda = 0$, respectively. For these cases, there is a unique

[9] When $\sum_{t=1}^{T} \mathbf{1}\{\mathbf{s}_j^t = \mathbf{s}\} = 0$ (or $\sum_{j=1}^{M} \sum_{t=1}^{T} \mathbf{1}\{\mathbf{s}_j^t = \mathbf{s}\} = 0$), we remove such states from the summand of the test statistics.

equilibrium and the null hypotheses are satisfied. All tests perform reassuringly well leading to a five percent rejection frequency as T and/or M increase.

We next assess the power properties of our tests. Table 3 considers the case of $\lambda = 0.5$, where the first and second equilibria arise with equal probability. It shows that as the number of time periods T and/or markets M increases, all the tests typically reject the null more frequently. The two conditional choice probability tests ($\mathcal{T}_P$ and $\mathcal{T}_P^*$) and the steady-state distribution test ($\mathcal{T}_Q$) perform better than the conditional state distribution test ($\mathcal{T}_{\mathbf{s}}$) for moderate values of M (e.g., $M = 20$ or 40). When M becomes large ($M = 320$ or 640), $\mathcal{T}_{\mathbf{s}}$ dominates $\mathcal{T}_P$ and $\mathcal{T}_P^*$ especially when T is relatively small. Comparing the conditional choice probability tests and the steady-state distribution test, we find that $\mathcal{T}_Q$ performs better than $\mathcal{T}_P$ and $\mathcal{T}_P^*$. A possible reason is that $\mathcal{T}_Q$ uses fewer cells than $\mathcal{T}_P$ and $\mathcal{T}_P^*$. $\mathcal{T}_Q$ is based on m_s cells while $\mathcal{T}_P$ and $\mathcal{T}_P^*$ are based on $(m_s m_a)$ cells. Table 3 also illustrates that for a typical industry application with about 40 markets and 20 time periods the performance of $\mathcal{T}_Q$ is satisfying. Also the test by $\mathcal{T}_P$ and the optimal test by $\mathcal{T}_P^*$ have similar performance. For a better comparison based on the result in Proposition 2, we compute the size-adjusted power for $\mathcal{T}_P$ and $\mathcal{T}_P^*$. We find that the size-adjusted power for $\mathcal{T}_P^*$ is higher than that for $\mathcal{T}_P$ in most cases.[10] To further investigate the power properties of these tests, Table 4 considers the case of $\lambda = 0.9$. That is, the first equilibrium is played in 90% of M markets. While all the tests have lower power than in Table 3, the relative performances of these tests appear the same. $\mathcal{T}_Q$ has still the best performance among all tests.

Overall, our Monte Carlo illustrates that the steady-state distribution test by $\mathcal{T}_Q$ performs well for moderate sample sizes of T and M. It seems well suited for typical industry applications.[11][12]

[10] For example, when $M = 40$, $T = 20$, and $\lambda = 0.5$, Table 3 suggests that the power for $\mathcal{T}_P$ is higher than the power for $\mathcal{T}_P^*$. On the other hand, the size-adjusted power for $\mathcal{T}_P$ is 16.1, while the size-adjusted power for $\mathcal{T}_P^*$ is 21.8.

[11] The number of markets M and time periods T in several important papers in the literature are $(M = 1600, T = 24)$ in Collard-Wexler (2013), $(M = 639, T = 5)$ in Dunne, Klimek, Roberts, and Xu (2013), $(M = 23, T = 19)$ in Ryan (2012), and $(M = 102, T = 4)$ in Sweeting (2012).

[12] We also check the performance of $\tilde{\mathcal{T}}_{\mathbf{s}}$ and $\tilde{\mathcal{T}}_P$ using the following simple simulation design. Suppose there are only two states. Consider the following three state transition matrices:

$$\mathbf{P}^{(a)} = \begin{pmatrix} 0.3 & 0.7 \\ 0.3 & 0.7 \end{pmatrix}, \mathbf{P}^{(b)} = \begin{pmatrix} 0.7 & 0.3 \\ 0.7 & 0.3 \end{pmatrix}, \mathbf{P}^{(c)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix}.$$

Under the null, each market follows $\mathbf{P}^{(a)}$ or $\mathbf{P}^{(b)}$ with equal probability. Under the alternative, each market follows $\mathbf{P}^{(a)}$ or $\mathbf{P}^{(b)}$ with probability of 0.25, and follows $\mathbf{P}^{(c)}$ with probability of 0.5. We compute the size and power of the two test statistics with varying numbers of markets and time periods. Overall, the size approaches 5 percent quickly for both statistics. For the power, $\tilde{\mathcal{T}}_P$ performs better than $\tilde{\mathcal{T}}_{\mathbf{s}}$. For example, the power of $\tilde{\mathcal{T}}_P$ when $(M = 500, T = 10)$, $(M = 500, T = 30)$, and $(M = 500, T = 50)$ is 9.3, 80.1, and 99.7 percent, respectively. On the other hand, the corresponding figures for $\tilde{\mathcal{T}}_{\mathbf{s}}$ are 5.0, 7.0, and 9.1 percent, respectively. The details of this exercise are available upon request.

Table 1. Monte Carlo Results: $\lambda = 1$

M	T	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_{\mathbf{s}}$
20	5	13.2	5.9	1.3	3.2
20	10	7.0	4.5	2.5	3.9
20	20	4.4	5.0	3.5	4.9
20	40	5.1	6.2	4.3	4.0
20	80	5.7	6.6	5.0	2.9
20	320	4.4	4.4	4.8	3.4
20	640	6.1	5.3	4.9	3.5
40	5	6.5	2.3	1.3	3.7
40	10	3.8	2.7	2.9	5.0
40	20	4.3	3.4	3.5	4.1
40	40	4.5	5.3	3.4	4.8
40	80	5.3	5.3	5.7	3.0
40	320	5.2	5.4	4.5	5.3
40	640	5.3	5.4	4.9	4.4
80	5	5.3	1.5	1.2	4.8
80	10	3.2	1.2	2.5	4.8
80	20	5.2	3.5	2.5	4.3
80	40	3.9	3.9	3.5	5.7
80	80	4.7	4.6	5.0	4.7
80	320	4.9	5.5	5.4	5.1
80	640	4.2	4.1	5.3	5.7
160	5	4.9	0.6	2.0	5.1
160	10	3.4	0.9	2.1	4.2
160	20	3.3	2.4	4.1	4.5
160	40	4.8	4.8	3.9	3.3
160	80	4.5	4.6	5.4	4.7
160	320	5.4	5.7	6.2	3.8
160	640	4.7	4.2	5.3	4.4
320	5	5.0	0.5	1.4	4.5
320	10	3.6	0.8	3.2	4.5
320	20	4.3	1.9	3.8	5.3
320	40	4.5	4.6	3.9	5.4
320	80	4.8	4.1	4.3	5.3
320	320	4.8	5.6	5.0	5.1
320	640	6.0	5.8	5.6	5.7
640	5	4.3	0.4	0.7	4.2
640	10	3.2	0.9	1.9	4.3
640	20	4.7	2.9	3.6	5.3
640	40	4.8	4.3	4.4	3.4
640	80	5.4	4.8	4.0	5.0
640	320	5.1	4.9	4.4	4.5
640	640	5.6	5.5	5.7	4.3

Table 2. Monte Carlo Results: $\lambda = 0$

M	T	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_{\mathbf{s}}$
20	5	13.5	12.7	0.5	4.8
20	10	7.9	8.4	2.5	4.3
20	20	5.8	7.1	3.4	5.3
20	40	4.8	5.4	3.6	3.8
20	80	5.1	5.2	4.4	5.5
20	320	5.0	4.9	5.9	5.4
20	640	3.5	3.6	4.7	4.4
40	5	8.0	6.4	1.2	4.7
40	10	5.2	4.9	2.3	5.2
40	20	6.2	6.9	3.5	4.0
40	40	3.9	5.3	3.6	3.7
40	80	5.6	4.5	3.8	3.9
40	320	5.0	4.7	5.2	5.7
40	640	5.1	5.2	4.2	4.0
80	5	4.6	3.7	1.6	5.2
80	10	5.6	5.1	1.9	4.7
80	20	4.9	5.7	3.8	5.5
80	40	4.7	5.1	3.1	4.4
80	80	5.3	5.0	4.8	4.8
80	320	3.3	3.7	4.5	4.7
80	640	4.0	4.1	4.3	3.8
160	5	4.0	1.4	1.3	5.1
160	10	4.7	3.9	2.6	6.0
160	20	4.5	4.5	2.9	3.7
160	40	6.3	5.5	3.3	5.1
160	80	5.5	5.2	3.4	4.9
160	320	4.6	4.8	3.3	5.1
160	640	5.3	5.1	4.1	4.2
320	5	4.1	1.7	0.8	4.9
320	10	4.8	2.8	1.4	5.6
320	20	4.6	3.6	3.8	4.4
320	40	5.0	4.4	3.8	6.0
320	80	5.8	6.1	4.5	6.1
320	320	6.4	6.5	4.2	6.0
320	640	5.2	5.5	5.3	4.7
640	5	4.2	2.2	1.2	5.8
640	10	4.9	2.0	1.7	5.0
640	20	4.0	3.9	3.2	4.8
640	40	5.3	4.6	3.7	6.0
640	80	4.4	4.8	4.9	6.0
640	320	4.7	4.5	3.5	6.3
640	640	5.2	5.6	5.1	5.3

Table 3. Monte Carlo Results: $\lambda = 0.5$

M	T	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_{\mathbf{s}}$
20	5	10.3	8.3	2.9	5.9
20	10	6.5	7.4	20.2	13.7
20	20	27.8	27.4	63.9	23.5
20	40	79.7	76.1	97.9	47.7
20	80	99.9	99.8	100.0	72.4
20	320	100.0	100.0	100.0	97.1
20	640	100.0	100.0	100.0	98.2
40	5	4.7	4.1	6.9	8.1
40	10	7.4	5.5	37.8	15.8
40	20	44.6	36.2	89.0	36.5
40	40	97.4	94.3	99.9	64.4
40	80	100.0	100.0	100.0	83.8
40	320	100.0	100.0	100.0	98.1
40	640	100.0	100.0	100.0	99.8
80	5	3.3	2.3	12.4	10.3
80	10	10.8	5.8	64.3	27.9
80	20	68.5	55.5	99.1	53.2
80	40	100.0	99.9	100.0	84.3
80	80	100.0	100.0	100.0	95.8
80	320	100.0	100.0	100.0	99.9
80	640	100.0	100.0	100.0	99.9
160	5	2.9	0.9	22.8	18.7
160	10	12.4	5.8	89.5	48.9
160	20	92.3	78.6	100.0	82.6
160	40	100.0	100.0	100.0	95.9
160	80	100.0	100.0	100.0	99.5
160	320	100.0	100.0	100.0	100.0
160	640	100.0	100.0	100.0	100.0
320	5	2.2	1.1	44.9	35.2
320	10	20.8	6.8	99.5	77.1
320	20	99.7	96.3	100.0	98.0
320	40	100.0	100.0	100.0	100.0
320	80	100.0	100.0	100.0	100.0
320	320	100.0	100.0	100.0	100.0
320	640	100.0	100.0	100.0	100.0
640	5	1.5	0.6	78.0	69.3
640	10	33.2	10.5	100.0	98.0
640	20	100.0	100.0	100.0	100.0
640	40	100.0	100.0	100.0	100.0
640	80	100.0	100.0	100.0	100.0
640	320	100.0	100.0	100.0	100.0
640	640	100.0	100.0	100.0	100.0

Table 4. Monte Carlo Results: $\lambda = 0.9$

M	T	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_\mathbf{s}$
20	5	10.7	6.0	2.4	6.4
20	10	6.5	4.8	11.4	14.7
20	20	11.7	12.8	30.1	20.0
20	40	32.7	35.3	64.6	29.2
20	80	75.8	76.5	94.2	44.6
20	320	100.0	100.0	100.0	71.5
20	640	100.0	100.0	100.0	82.9
40	5	4.5	2.5	3.1	10.0
40	10	5.4	4.2	19.0	17.0
40	20	16.0	14.8	45.5	26.0
40	40	49.0	50.1	87.1	41.5
40	80	93.5	92.5	99.9	58.8
40	320	100.0	100.0	100.0	89.4
40	640	100.0	100.0	100.0	94.2
80	5	3.4	1.7	4.3	10.7
80	10	5.9	3.2	28.9	21.1
80	20	23.3	19.7	71.3	33.9
80	40	72.8	73.2	98.0	53.5
80	80	99.7	99.6	100.0	73.2
80	320	100.0	100.0	100.0	95.2
80	640	100.0	100.0	100.0	98.4
160	5	4.0	0.9	8.8	16.8
160	10	6.0	2.1	46.6	28.9
160	20	38.2	30.6	92.5	47.0
160	40	93.4	92.4	100.0	68.1
160	80	100.0	100.0	100.0	85.7
160	320	100.0	100.0	100.0	98.9
160	640	100.0	100.0	100.0	99.5
320	5	2.9	1.0	14.7	22.2
320	10	9.1	3.7	73.3	45.0
320	20	60.2	49.7	99.9	68.6
320	40	99.8	99.7	100.0	88.4
320	80	100.0	100.0	100.0	96.7
320	320	100.0	100.0	100.0	99.7
320	640	100.0	100.0	100.0	100.0
640	5	2.7	0.5	32.1	40.9
640	10	12.3	4.8	93.3	69.9
640	20	86.6	77.2	100.0	91.0
640	40	100.0	100.0	100.0	98.2
640	80	100.0	100.0	100.0	100.0
640	320	100.0	100.0	100.0	100.0
640	640	100.0	100.0	100.0	100.0

5 Empirical Application

Recently, a number of empirical papers apply a dynamic game to data and estimate parameters of the game using two step methods. These papers include Ryan (2012), Collard-Wexler (2013), Sweeting (2013), Beresteanu, Ellickson and Misra (2010), and the empirical section of Aguirregabiria and Mira (2007), among others. Panel data frequently contain a number of markets over a relatively short time period. Researchers tend to pool different markets together to estimate policy functions in the first stage. To do this pooling, an important assumption is that a single equilibrium is played in every market. This section tests the homogeneity hypotheses for poolability using the data of Ryan (2012). We chose Ryan (2012) because it is one of a few papers already published and because the number of state variables is relatively small so that it fits well our illustrative purpose.

To evaluate the welfare costs of the 1990 Amendments to the Clean Air Act on the Portland cement industry in the U.S., Ryan (2012) develops a dynamic oligopoly model based on Ericson and Pakes (1995) and estimates the model using a two-step method developed by Bajari, Benkard and Levin (2007). In his application, there are 23 geographically separated markets. To estimate firms' policy functions in the first stage, Ryan (2012) assumes that the data are generated by a single Markov Perfect Equilibrium. We apply our test to check this assumption. One caveat is that we use a discrete state space framework, while Ryan (2012) uses a continuous state space. Thus, we have to discretize the state variables in Ryan (2012)'s application to perform the test. For a fine grid, however, little differences between the two frameworks are expected in practice.

We first summarize Ryan (2012)'s model. Then, we explain the procedure of our test in this context.

5.1 Ryan (2012)'s model

Ryan (2012) assumes that N firms play a dynamic oligopoly game in each regional cement market. Firms make decisions to maximize the discounted sum of expected profits. The timing of the decisions is as follows. At the beginning of each period, incumbent firms draw a private scrap value and decides whether to exit the market or not. Then, potential entrants receive a private draw of entry costs and investment costs. At the same time, incumbent firms who have not decided to exit the market draw private costs of investment and divestment. Then, all entry and investment decisions are made simultaneously. Firms compete in the product market and profits realize. Finally, firms enter and exit, and their capacity levels change according to the investment/divestment decisions in this period.

Let $\mathbf{s} = (s_1, \ldots, s_N) \in \mathbf{S}$ be the capacity levels of N firms and let ε_i be a vector of all private shocks to firm i. Assuming that ε_i is iid over time and focusing on pure Markovian strategies, firm i's strategy is a mapping from states and private shocks to actions. The game payoff for firm i is defined as the discounted sum of expected period payoffs given the beliefs now and in the future. The collection of strategies and beliefs is a MPE if (i) for all i, firm i's strategy is a best response to its rivals' strategies given the beliefs at all states $\mathbf{s} \in \mathbf{S}$ and (ii) for all i, the beliefs of firm i are consistent with the strategies. The existence of pure strategy equilibria in a class of dynamic games is provided in Doraszelski and Satterthwaite (2010). The model of Ryan (2012) also falls in this class. Furthermore, multiplicity of equilibria is prevalent.

Ryan (2012) follows the two-step method developed by Bajari, Benkard and Levin (2007). In the first stage, Ryan (2012) estimates the entry, exit, and investment policies as a function of states. Because of the issue of multiplicity, different equilibria may be played in different markets. However, since Ryan (2012) has only 19 years of time series compared to a large state space, estimating policy functions market by market is not practical. Thus, he imposes the following assumption:

Assumption 1 *The same equilibrium is played in all markets.*

Based on this assumption Ryan pools all markets when estimating policy functions. Our aim is to test the validity of this assumption.

In addition to Assumption 1, Ryan (2012) assumes flexible functional forms for the policy functions. First, the probability of entry is modeled as a probit regression,

$$\begin{aligned} & \Pr\{\text{firm } i \text{ enters in period } t | \ s_i = 0, \mathbf{s}\} \\ = \ & \Phi\left(\psi_1 + \psi_2(\textstyle\sum_{j \neq i} s_j^t) + \psi_3 \mathbf{1}\{t > 1990\}\right), \end{aligned} \tag{16}$$

where $\Phi(\cdot)$ is the cdf of the standard normal. The dummy $\mathbf{1}\{t > 1990\}$ is introduced to account for the change in firms' behavior after the introduction of the 1990 Amendments.

Second, the exit probability is also modeled as probit,

$$\begin{aligned} & \Pr\{\text{firm } i \text{ exits in period } t | \ s_i > 0, \mathbf{s}\} \\ = \ & \Phi\left(\psi_4 + \psi_5 s_i^t + \psi_6(\textstyle\sum_{j \neq i} s_j^t) + \psi_7 \mathbf{1}\{t > 1990\}\right). \end{aligned} \tag{17}$$

Finally, the investment policy is modeled using the empirical model of the (S,s) rule by Attanasio (2000). Specifically, firms adjust the current capacity level to a target level of capacity when current capacity exceeds one of the bands around the target level. The target level s_i^{*t} is given by

$$\ln s_i^{*t} = \lambda_1' b_1(s_i^t) + \lambda_2' b_2(\textstyle\sum_{j \neq i} s_j^t) + u_i^{*t}, \tag{18}$$

where u_i^{*t} is iid normal with zero mean and a homoscedastic variance, the functions $b_1(\cdot)$ and $b_2(\cdot)$ denote cubic b-spline, which is to capture flexible functional forms in the variables s_i^t and $\sum_{j \neq i} s_j^t$, respectively. The lower and upper bands are given by

$$\underline{s}_i^t = s_i^{*t} - \exp\left(\lambda_3' b_1(s_i^t) + \lambda_4' b_2(\textstyle\sum_{j \neq i} s_j^t) + \underline{u}_i^{bt}\right) \tag{19}$$

and

$$\overline{s}_i^t = s_i^{*t} + \exp\left(\lambda_3' b_1(s_i^t) + \lambda_4' b_2(\sum\nolimits_{j\neq i} s_j^t) + \overline{u}_i^{bt}\right), \tag{20}$$

where $\underline{u}_i^{bt}$ and $\overline{u}_i^{bt}$ are assumed iid normal with zero mean and equal variance. It is assumed that the upper and lower bands are symmetric functions of the target capacity. To estimate (18), Ryan (2012) simply replaces $\ln s_i^{*t}$ with $\ln s_i^{t+1}$ and runs OLS using the sample with $s_i^t \neq s_i^{t+1}$. To estimate parameters in (19) and (20), Ryan (2012) regresses $\ln|s_i^{t+1} - s_i^t|$ on b_1 and b_2 using the sample with $s_i^t \neq s_i^{t+1}$. The implicit assumption here is that the level of capacity observed before the change (i.e., s_i^t) is equal to either the lower or the upper bands depending on whether the investment is positive or negative.[13] To estimate the variances of u_i^{*t}, $\underline{u}_i^{bt}$, and $\overline{u}_i^{bt}$, Ryan (2012) calculates the sum of the squared residuals at the estimated parameters and divide it by $(n - k_\lambda)$, where n is the sample size used in least squares and k_λ is the number of parameters in λ for each equation.

Once all these parameters are estimated, the value functions can be computed by forward simulation. If Assumption 1 holds and the functional forms are flexible enough, the first stage delivers consistent estimates of choice probabilities associated with the equilibrium that is played in the data. However, if there are more than one equilibria in the data, estimates of choice probabilities are not consistent, and estimates of structural parameters in the second stage are not consistent either.

The model specified above implies the Markov transition probability $\mathbf{P}$ and the corresponding steady-state distribution $\mathbf{Q}$. Although Ryan (2012) uses a parametric specification in his first stage estimate for the feasibility reason, we apply our test directly to $\mathbf{P}$ and $\mathbf{Q}$. It is a major advantage of our tests that the model's details do not have to be specified.

5.2 Data

We download the data from the Econometrica webpage. The dataset contains information on all the Portland cement plants in the United States from 1980 to 1998. Following Ryan (2012), we assume that every plant is owned by different firms. For each plant, we observe the name of company that owns the plant and the location of the plant. A plant consists of several kilns. For each kiln, we observe the fuel type, process type, and the year when the kiln was installed. We organize the data in the following way. The capacity of a plant is simply defined as the sum of capacity of all kilns that are installed in the plant. Plants sometimes change their company name. One reason is that plants are sold to a different company. Another possibility is that two or more firms merge and names change accordingly. In such cases, it appears as if the old plant exits the market and a new firm (plant) enters the market at the same time. To deal with such spurious entry/exit, we check information of kilns (fuel type, process type, year of installation) installed in the plant that changed the company name, and

[13] For an interpretation and justification of this implicit assumption, see Attanasio (2000).

if those information have not changed at all, we assume that the plant stays in the market (we assume that no entry and exit took place associated with this name change).

As a result, we obtained the same plant-level data as Ryan (2012). Table 5 shows its summary statistics.

Table 5. Summary Statistics of Plant-Level Data

	Min	Mean	Max	Std. Dev.	Sample size
Quantity (1,000 tons)	177	699	2348	335	2233
Capacity (1,000 tons)	196	797	2678	386	2233
Investment (1,000 tons)	-728	2.19	1140	77.60	2077

5.3 Homogeneity tests for poolability

Ryan (2012)'s panel data contain states and actions over 19 years for 23 different markets.[14] Since our Monte Carlo study indicates that the steady-state distribution test by $\mathcal{T}_Q$ performs better than the other tests when the number of markets is small, we first apply the steady-state distribution test to Ryan (2012)'s data. Then to account for the possibility that homogeneity is violated but the test based on $\mathcal{T}_Q$ is not able to detect, we also apply the state transition probability test by $\mathcal{T}_P$ and its optimal version $\mathcal{T}_P^*$. For the sake of completeness, we apply the test based on $\mathcal{T}_{\mathbf{s}}$ as well. The original state space of Ryan (2012) consists of firm-level capacities. We focus on a lower-dimensional state variable consisting of the total market-level capacity s_j^t obtained by summing capacity levels across firms, i.e. $s_j^t = \sum_i s_j^{it}$. Hereafter we consider testing the null hypotheses $\mathrm{H}_0^{\mathbf{Q}}$, $\mathrm{H}_0^{\mathbf{P}}$, and $\mathrm{H}_0^{\mathbf{s}}$ based on s_j^t. Note that a rejection of the null based on the market-level capacity implies a rejection of the null for the full model with firm-level capacities, but that the converse is not true.

Our test proceeds as follows. Ryan (2012) assumed that the same equilibrium was played in all markets before 1990 and that another identical equilibrium was played in all markets after 1990. We test these hypotheses in different time periods by the statistics $\mathcal{T}_Q$, $\mathcal{T}_P$, $\mathcal{T}_P^*$, and $\mathcal{T}_{\mathbf{s}}$.

To implement these tests, we discretize the support of s_j^t into 50 bins with equal intervals of 250 thousand tons (0-250 thousand tons, 250-500 thousand tons, and so on). Figure 1 depicts the discretized state distributions before and after 1990.

[14] Ryan (2012)'s Java code available at the Econometrica website generates only 22 markets, while his first-stage estimation appears to be using 23 markets (23 markets times 18 years equals 414 observations). One natural way to increase the number of markets is to disaggregate one large market into two. In California, we can observe two clusters of plants; one in Northern California around the San Francisco area and another in Southern California around the Los Angeles area. These two clusters are remote by more than 350 miles. Thus, we believe that Northern and Southern California can be considered two separate markets.

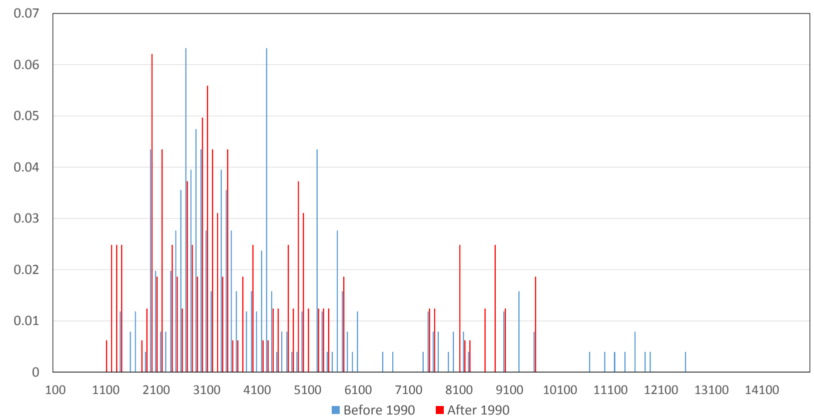

Figure 1: Steady-State Distribution of Total Capacity (1,000 ton)

For these samples, the steady-state distributions are estimated by the relative frequencies

$$\widehat{Q}_j^{\text{before}}(s) = \frac{1}{T^{\text{before}}}\sum_{t=1980}^{1990}\mathbf{1}\{s_j^t = s\} \quad \text{for } s \in \{1,\dots,50\} \text{ and } j = 1,\dots,23,$$

$$\widehat{Q}_j^{\text{after}}(s) = \frac{1}{T^{\text{after}}}\sum_{t=1991}^{1998}\mathbf{1}\{s_j^t = s\} \quad \text{for } s \in \{1,\dots,50\} \text{ and } j = 1,\dots,23.$$

Then the test statistic $\mathcal{T}_Q$ is obtained as

$$\mathcal{T}_Q^l = T^l \sum_{j=1}^{23}\sum_{s=1}^{50}\{\widehat{Q}_j^l(s) - \widehat{Q}^l(s)\}^2, \tag{21}$$

for $l = \{\text{before}, \text{after}\}$. Also for $\mathbf{d} = (s', s) \in \{1,\dots,50\}^2$, the state transition probabilities are estimated by

$$\widehat{P}_j^{\text{before}}(\mathbf{d}) = \frac{\sum_{t=1980}^{1989}\mathbf{1}\{s_j^{t+1} = s', s_j^t = s\}}{\sum_{t=1980}^{1989}\mathbf{1}\{s_j^t = s\}} \quad \text{for } j = 1,\dots,23,$$

$$\widehat{P}_j^{\text{after}}(\mathbf{d}) = \frac{\sum_{t=1991}^{1997}\mathbf{1}\{s_j^{t+1} = s', s_j^t = s\}}{\sum_{t=1991}^{1997}\mathbf{1}\{s_j^t = s\}} \quad \text{for } j = 1,\dots,23,$$

and $\widehat{P}^{\text{before}}(\mathbf{d})$ and $\widehat{P}^{\text{after}}(\mathbf{d})$ are defined as in (6). The test statistic $\mathcal{T}_P$ is obtained as

$$\mathcal{T}_P^l = \sum_{j=1}^{23} \sum_{\mathbf{d} \in \mathbf{D}} W_j^l(\mathbf{d}) \{\widehat{P}_j^l(\mathbf{d}) - \widehat{P}^l(\mathbf{d})\}^2 \quad \text{for } l = \{\text{before}, \text{after}\}, \tag{22}$$

where $W_j^l(\mathbf{d}) = \sum_{t=1}^{T^l-1} \mathbf{1}\{s_j^t = s\}/\widehat{P}^l(\mathbf{d})$. The test statistic $\mathcal{T}_P^*$ is given by

$$\mathcal{T}_P^{*l} = 2 \sum_{j=1}^{23} \sum_{\mathbf{d} \in \mathbf{D}} W_j^l(\mathbf{d}) \widehat{P}_j^l(\mathbf{d}) \log \frac{\widehat{P}_j^l(\mathbf{d})}{\widehat{P}^l(\mathbf{d})} \quad \text{for } l = \{\text{before}, \text{after}\}. \tag{23}$$

Finally, the test statistic $\mathcal{T}_{\mathbf{s}}$ is defined accordingly as in (13).[15]

The bootstrap critical values for the first three tests are computed as follows. For each bootstrap iteration b, we simulate the game for 19 years and 23 markets. More precisely, we draw an initial state from the distribution $M^{-1} \sum_{j=1}^{M} \widehat{Q}_j^{\text{before}}(\cdot)$ and generate Markov chains by the transition matrix $\widehat{P}^{\text{before}}(\cdot)$ for $t = 1980, \ldots, 1990$. In the same way, we use $M^{-1} \sum_{j=1}^{M} \widehat{Q}_j^{\text{after}}(\cdot)$ and $\widehat{P}^{\text{after}}(\cdot)$ to generate a sequence of states for $t = 1990, \ldots, 1998$. For the simulated b-th bootstrap sample, we estimate $\{\widehat{Q}_j^{l,b}(s),\ \widehat{P}_j^{l,b}(\mathbf{d}),\ \widehat{P}^{l,b}(\mathbf{d}),\ f_j^{l,b}(s)\}$ for all $j = 1, \ldots, 23$, $l = \{\text{before}, \text{after}\}$, $s = 1, \ldots, 50$, and $\mathbf{d} \in \{1, \ldots, 50\}^2$. We then compute the bootstrap counterparts $\mathcal{T}_Q^b$, $\mathcal{T}_P^b$, and $\mathcal{T}_P^{*b}$ using (21), (22) and (23), respectively. For the test by $\mathcal{T}_{\mathbf{s}}$, all steps are the same as other tests. The number of bootstrap iterations is $B = 999$.

Table 6. Baseline Results

Before 1990				
	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_{\mathbf{s}}$
Test statistics	199.481	159.426	101.549	273.867
5% critical value	174.548	144.663	113.454	292.766
p-value	0.009	0.010	0.330	0.125
After 1990				
	$\mathcal{T}_P$	$\mathcal{T}_P^*$	$\mathcal{T}_Q$	$\mathcal{T}_{\mathbf{s}}$
Test statistics	89.430	90.579	81.032	131.867
5% critical value	93.275	91.780	95.543	179.406
p-value	0.089	0.055	0.599	0.619

Table 6 summarizes the test results. $\mathcal{T}_P$ and $\mathcal{T}_P^*$ imply that we reject the homogeneity hypothesis at the 1% significance level for the period before 1990 and at the 10% significance level for the period after 1990. The fact that the test by $\mathcal{T}_Q$ does not reject the null, while the tests by $\mathcal{T}_P$ and $\mathcal{T}_P^*$ reject it may suggest that distinct conditional choice probabilities have similar (or perhaps

[15] We replace the variance matrix with the identity matrix.

identical) steady-state distributions. The result of the test by $\mathcal{T}_{\mathbf{s}}$ may be because the power is low under the current sample size.

To perform the above tests, we implicitly assume that the relevant state variable is firm-level cement capacity only. However, regional markets differ significantly in their size. Therefore, the rejection of the null hypothesis of homogeneity may simply have come from the large amount of observable heterogeneity. To capture such market-level heterogeneity, we control for the size of population of each market following Ryan (2012).[16] Specifically, we calculate the average (over 19 time periods) population size by market, and divide 23 markets into 7 "small" markets, 8 "medium" markets, and 8 "large" markets. For each of these subgroups of markets, we apply the tests by $\mathcal{T}_P$ and $\mathcal{T}_P^*$.

Table 7. Test conditional on market size

	$\mathcal{T}_P$		$\mathcal{T}_P^*$	
Small markets	Before 1990	After 1990	Before 1990	After 1990
Test statistics	25.269	12.122	20.455	15.955
5% critical value	26.778	21.435	26.589	22.532
p-value	0.074	0.469	0.241	0.293
Medium markets	Before 1990	After 1990	Before 1990	After 1990
Test statistics	47.910	10.902	37.288	11.708
5% critical value	51.765	16.691	43.010	17.449
p-value	0.089	0.285	0.144	0.311
Large markets	Before 1990	After 1990	Before 1990	After 1990
Test statistics	25.918	13.409	26.353	15.150
5% critical value	23.601	15.550	20.963	16.577
p-value	0.023	0.121	0.003	0.093

Table 7 summarizes the results of the tests on subgroups of markets. This suggests that while our tests do not reject the hypothesis $\mathrm{H}_0^{\mathbf{P}}$ of homogeneity for small and medium markets, they still reject the null hypothesis for the group of large markets, especially for the period before 1990.

Finally, we use $\tilde{\mathcal{T}}_{\mathbf{s}}$ in (14) and $\tilde{\mathcal{T}}_P$ in (15) to account for potential unobserved heterogeneity. The parameters in $\tilde{\mathbf{P}}$, that is, π and the elements in $\mathbf{P}^{(a)}$ and $\mathbf{P}^{(b)}$ are estimated by the MLE. To compute the critical value, the parametric bootstrap is employed as before. Table 8 summarizes the results. The upper panel of the table shows the test results based on the full sample. The tests based on transition probability matrices do not reject the null of homogeneity across markets at the 5% significance level. However, if we control for the size of population, a slightly different picture emerges. The lower panel shows the test results when we focus on the subsample of large markets. As in Table 7, the test based on $\tilde{\mathcal{T}}_P$ rejects the null hypothesis of homogeneity at the 5% significance

[16] Ryan (2012) tries controlling for regional population in one of his specifications of policy function estimation, but does not include it in his preferred specification.

level for the period before 1990. In addition, the test based on $\tilde{\mathcal{T}}_{\mathbf{s}}$ supports the same conclusion.

Table 8. Test accounting for unobserved types

	$\tilde{\mathcal{T}}_{\mathbf{s}}$		$\tilde{\mathcal{T}}_{P}$	
Full sample	Before 1990	After 1990	Before 1990	After 1990
Test statistics	295.754	153.537	76.702	29.306
5% critical value	308.187	190.933	83.030	33.576
p-value	0.094	0.412	0.084	0.150
Large markets	Before 1990	After 1990	Before 1990	After 1990
Test statistics	89.083	51.096	8.170	0.375
5% critical value	84.475	61.145	8.051	4.085
p-value	0.018	0.324	0.046	0.675

Our result suggests that the data should not be pooled even if a researcher accounts for two unobservable market types and uses an appropriate method (e.g., Arcidiacono and Miller, 2011). Since we assume that unobserved heterogeneity follows a finite-mixture model with only two components, the rejection of our tests may point to the existence of more general types of unobserved heterogeneity, the presence of multiple equilibria within a group of markets of the same type, or both. One caveat is that the size of the tests may not have converged quickly enough. Therefore, given the small sample size in this application, our results should be treated as suggestive.

6 Conclusion

This paper proposes several statistical tests for finite state Markov games to examine the data from distinct markets can be pooled. The tests are based on homogeneity (across markets) of the conditional choice and state transition probabilities, the steady-state distribution, and the conditional state distribution. We perform a Monte Carlo study and find that the steady-state distribution test works well and has high power even with a small number of markets and time periods. We apply our tests to the empirical application of Ryan (2012) and reject the null hypothesis of homogeneity, which is a maintained assumption for estimation using pooled data.

Two caveats need to be emphasized. First, in case of rejection, researchers may be tempted to apply the tests repeatedly to subsamples until the null hypothesis is no longer rejected. While this exercise may be informative for identifying the cause of the rejection, it is not statistically justified. In general, if the same test is applied to the subsample after a rejection based on the full sample, the test statistic should be modified to incorporate the fact that the test rejects the null with the full sample. Such a sequential testing procedure would involve more sophisticated statistical theory and is beyond the scope of our paper.

Second, our test statistics are proposed within the finite state discrete time Markov class. The theory of finite state Markov chains is well developed and allows us to borrow well known results from the probability theory literature. To extend the tests to a richer state space, we would need to borrow results from a more involved statistical literature making the tests perhaps less accessible to researchers. However, we believe that our tests cover a wide class of dynamic games that are used in the empirical IO literature. With a bounded state space, as is typical the case in IO applications, the observable difference between games with finite state and games with a continuous state space seem superficial and not essential as in practice the data are finite. Researchers may use a finer grid when the data become richer.

A Appendix

A.1 Proof of Proposition 1

We first consider the statistic $\mathcal{T}_P$ for H_0^σ defined by (5), that is

$$\mathcal{T}_P = \sum_{j=1}^{M} \sum_{(\mathbf{a},\mathbf{s})\in\mathbf{A}\times\mathbf{S}} \frac{\{f_j(\mathbf{a},\mathbf{s}) - f_j(\mathbf{s})\hat{\sigma}(\mathbf{a}|\mathbf{s})\}^2}{f_j(\mathbf{s})\hat{\sigma}(\mathbf{a}|\mathbf{s})},$$

where $\hat{\sigma}(\mathbf{a}|\mathbf{s}) = \frac{\sum_{j=1}^M f_j(\mathbf{a},\mathbf{s})}{\sum_{j=1}^M f_j(\mathbf{s})}$. Let $\xi_j(\mathbf{a},\mathbf{s}) = \{f_j(\mathbf{a},\mathbf{s}) - f_j(\mathbf{s})\sigma_j(\mathbf{a}|\mathbf{s})\}/f_j(\mathbf{s})^{1/2}$ and define the $(m_a m_s)$-dimensional vector $\xi_j = \{\xi_j(\mathbf{a},\mathbf{s})_{\mathbf{a}\in\mathbf{A}}\}_{\mathbf{s}\in\mathbf{S}}$. Since $\mathbf{a}_j^t|\mathbf{s}_j^t$ is conditionally independent from past values, the Markov chain $\mathbf{P}$ is stationary, and all states of $\mathbf{P}$ communicate, the same argument to the proof of Billingsley (1961, Theorem 3.1) implies

$$\xi_j \xrightarrow{d} N(0, \mathrm{diag}\{V_j(\mathbf{s})\}_{\mathbf{s}\in\mathbf{S}}),$$

for each $j = 1,\ldots,M$, where $[V_j(\mathbf{s})]_{(k,l)} = \mathbf{1}\{k = l\}\sigma_j(\mathbf{a}_k|\mathbf{s}) - \sigma_j(\mathbf{a}_k|\mathbf{s})\sigma_j(\mathbf{a}_l|\mathbf{s})$ for $k,l = 1,\ldots,m_a$. Thus, we obtain

$$\sum_{(\mathbf{a},\mathbf{s})\in\mathbf{A}\times\mathbf{S}} \frac{\{f_j(\mathbf{a},\mathbf{s}) - f_j(\mathbf{s})\sigma_j(\mathbf{a}|\mathbf{s})\}^2}{f_j(\mathbf{s})\sigma_j(\mathbf{a}|\mathbf{s})} \xrightarrow{d} \chi^2(m_s(m_a-1)), \tag{24}$$

for each $j = 1,\ldots,M$. Note that under the set-up of Section 2, $\hat{\sigma}(\mathbf{a}|\mathbf{s})$ is the maximum likelihood estimator of $\sigma(\mathbf{a}|\mathbf{s})$ under $\mathrm{H}_0^\sigma : \sigma_1 = \cdots = \sigma_M = \sigma$ using the full-sample $(\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1,\ldots,T}$ for $j = 1,\ldots,M$. Therefore, based on (24), the asymptotic theory of the chi-squared statistic (e.g., Lemma 17.3 of van der Vaart, 1998) implies the conclusion.

We now consider the statistic $\mathcal{T}_P$ for $\mathrm{H}_0^{\mathbf{P}}$ defined by (6), that is

$$\mathcal{T}_P = \sum_{j=1}^{M} \sum_{(\mathbf{s}',\mathbf{s})\in\mathbf{S}\times\mathbf{S}} \frac{\{f_j^1(\mathbf{s}',\mathbf{s}) - f_j^1(\mathbf{s})\hat{p}(\mathbf{s}'|\mathbf{s})\}^2}{f_j^1(\mathbf{s})\hat{p}(\mathbf{s}'|\mathbf{s})},$$

where $\hat{p}(\mathbf{s}'|\mathbf{s}) = \frac{\sum_{j=1}^M f_j^1(\mathbf{s}',\mathbf{s})}{\sum_{j=1}^M f_j^1(\mathbf{s})}$. In this case, Billingsley (1961, Theorem 3.1) directly implies the asymptotic normality of $\{f_j^1(\mathbf{s}',\mathbf{s}) - f_j^1(\mathbf{s})p_j(\mathbf{s}'|\mathbf{s})\}/f_j^1(\mathbf{s})^{1/2}$. Thus, a similar argument yields the conclusion.

A.2 Proof of Proposition 2

We prove the optimality for $\mathcal{T}_P^*$ to test $\mathrm{H}_0^{\mathbf{P}}$. The case for testing H_0^σ is shown in the same manner although the notation becomes more complicated.

Let $\omega_j = (\mathbf{s}_j^1,\ldots,\mathbf{s}_j^T) \in \Omega_j$ and $\Omega = \Omega_1 \times \cdots \times \Omega_M$ be the sample space of the observables $\omega = (\omega_1,\ldots,\omega_M)$. The sample space Ω is partitioned into different types $\{\Lambda_l\}_{l=1,\ldots,L}$, where $\{\Lambda_l\}_{l=1,\ldots,L}$ is a collection of disjoint subsets of Ω satisfying $\Omega = \cup_{l=1}^{L} \Lambda_l$ and any element in Λ_l yields the same joint counts $\{f_j^1(\cdot,\cdot)\}_{j=1,\dots,M}$. A test is defined as a partition (Ω_A, Ω_R) of Ω, where Ω_A and Ω_R mean the acceptance and rejection regions, respectively.

First, we show that for any test (Ω_A, Ω_R), there exists a test $(\tilde{\Omega}_A, \tilde{\Omega}_R)$ based only on the joint counts $\{f_j^1(\cdot,\cdot)\}_{j=1,\dots,M}$ such that

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T-1} \log \Pr\{\tilde{\Omega}_R : \mathrm{H}_0^{\mathbf{P}}\} &\leq \lim_{T\to\infty} \frac{1}{T-1} \log \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\}, \\ \lim_{T\to\infty} \frac{1}{T-1} \log \Pr\{\tilde{\Omega}_A : \mathrm{H}_1^{\mathbf{P}}\} &\leq \lim_{T\to\infty} \frac{1}{T-1} \log \Pr\{\Omega_A : \mathrm{H}_1^{\mathbf{P}}\}, \end{aligned} \tag{25}$$

Note that the subset Ω_A or Ω_R contains at least half of the elements in Λ_l for each $l = 1, \dots, L$. Thus, for any (Ω_A, Ω_R), we can define $(\tilde{\Omega}_A, \tilde{\Omega}_R)$ as follows. For each $l = 1, \dots, L$, if Ω_A (or respectively Ω_R) contains at least half of the elements in Λ_l, then let $\tilde{\Omega}_A$ (or respectively $\tilde{\Omega}_R$) include all elements in Λ_l. Observe that $(\tilde{\Omega}_A, \tilde{\Omega}_R)$ depends only on $\{f_j^1(\cdot,\cdot)\}_{j=1,\dots,M}$ by construction. Now, pick any type Λ_l such that $\Lambda_l \subset \tilde{\Omega}_R$. It holds

$$\begin{aligned} \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} &\geq \Pr\{\Omega_R \cap \Lambda_l : \mathrm{H}_0^{\mathbf{P}}\} \geq \frac{1}{2} \Pr\{\Lambda_l : \mathrm{H}_0^{\mathbf{P}}\} \\ &= \frac{1}{2} \prod_{j=1}^{M} \Pr\{\Lambda_{l,j} : \mathrm{H}_0^{\mathbf{P}}\}, \end{aligned} \tag{26}$$

where the first inequality follows from the set inclusion relationship, the second inequality follows from the facts that at least half of elements of Λ_l is contained in Ω_R (due to $\Lambda_l \subset \tilde{\Omega}_R$) and that all elements in Λ_l occur with same probability, and the equality follows from independence of $(\omega_1, \dots, \omega_M)$ and $\Lambda_l = \Lambda_{l,1} \times \cdots \times \Lambda_M$. By Gutman (1989, Lemma 1), if the initial values $(\mathbf{s}_1^0, \dots, \mathbf{s}_M^0)$ are fixed, for any probability measure P on Ω given by a Markov chain, there exists a positive sequence $\delta_T = O(T^{-1} \log T)$ such that

$$\exp\left(-(T-1)\{K(q_{j,l}, p) + \delta_T\}\right) \leq \Pr\{\Lambda_{l,j} : P\} \leq \exp\left(-(T-1)\{K(q_{j,l}, p) - \delta_T\}\right), \tag{27}$$

where $q_{j,l}(\cdot,\cdot)$ is the two-period joint empirical measure given by the type $\Lambda_{l,j}$, $p(\cdot,\cdot)$ is the two-period joint measure given by P, and

$$K(q_{j,l}, p) = \sum_{\mathbf{s}\in\mathbf{S}} q_{j,l}(\mathbf{s}) \sum_{\mathbf{s}'\in\mathbf{S}} q_{j,l}(\mathbf{s}'|\mathbf{s}) \log \frac{q_{j,l}(\mathbf{s}'|\mathbf{s})}{p(\mathbf{s}'|\mathbf{s})}$$

is the Kullback-Leibler divergence for $q_{j,l}$ and p. Combining (26) and (27),

$$\Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} \geq \frac{1}{2} \exp\left(-(T-1)\left\{\sum_{j=1}^{M} K(q_{j,l}, p) + \delta_{1T}\right\}\right), \tag{28}$$

for some $\delta_{1T} = O(T^{-1} \log T)$. Here p is the common joint measure under $\mathrm{H}_0^{\mathbf{P}}$.

Thus, we have

$$\begin{aligned}
\Pr\{\tilde{\Omega}_R : \mathrm{H}_0^{\mathbf{P}}\} &= \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} + \sum_{l:\Lambda_l \subset \tilde{\Omega}_R} \Pr\{\Omega_A \cap \Lambda_l : \mathrm{H}_0^{\mathbf{P}}\} \\
&\leq \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} + \sum_{l:\Lambda_l \subset \tilde{\Omega}_R} \exp\left(-(T-1)\left\{\sum_{j=1}^{M} K(q_{j,l}, p) - \delta_{2T}\right\}\right) \\
&\leq \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} + \sum_{l:\Lambda_l \subset \tilde{\Omega}_R} \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\} \exp((T-1)\delta_{3T}) \\
&= \Pr\{\Omega_R : \mathrm{H}_0^{\mathbf{P}}\}\{1 + L_R \exp((T-1)\delta_{3T})\},
\end{aligned}$$

for some $\delta_{2T}, \delta_{3T} = O(T^{-1}\log T)$, where the first equality follows from the construction of $\tilde{\Omega}_R$, the first inequality follows from $\Pr\{\Omega_A \cap \Lambda_l : \mathrm{H}_0^{\mathbf{P}}\} \leq \Pr\{\Lambda_l : \mathrm{H}_0^{\mathbf{P}}\}$ and (27), the second inequality follows from (28), and the last equality follows from the definition of $L_R = \sum_{l=1}^{L} \mathbf{1}\{\Lambda_l \subset \tilde{\Omega}_R\}$. Therefore, the first inequality in (25) follows by $(T-1)^{-1}\log L_R \to 0$. The second inequality in (25) is obtained in the same manner (by replacing $\tilde{\Omega}_R$, Ω_R, and $\mathrm{H}_0^{\mathbf{P}}$ with $\tilde{\Omega}_A$, Ω_A, and $\mathrm{H}_1^{\mathbf{P}}$, respectively). By (25), we can focus on the test defined by the joint counts $\{f_j^1(\cdot,\cdot)\}_{j=1,\ldots,M}$.

Next, we show (10). Pick any test $(\tilde{\Omega}_A, \tilde{\Omega}_R)$ based only on $\{f_j^1(\cdot,\cdot)\}_{j=1,\ldots,M}$ that satisfies (9). Then there exists $\delta_{4T} = O(T^{-1}\log T)$ such that

$$\begin{aligned}
e^{-\alpha(T-1)} &\geq \Pr\{\tilde{\Omega}_R : \mathrm{H}_0^{\mathbf{P}}\} = \sum_{l:\Lambda_l \subset \tilde{\Omega}_R} \prod_{j=1}^{M} \Pr\{\Lambda_{l,j} : \mathrm{H}_0^{\mathbf{P}}\} \\
&\geq \exp\left(-(T-1)\left\{\sum_{j=1}^{M} K(q_{j,l}, p) + \delta_{4T}\right\}\right), \qquad (29)
\end{aligned}$$

for any l satisfying $\Lambda_l \subset \tilde{\Omega}_R$ and all T large enough, where the first inequality follows from (9), the equality follows from independence of $(\omega_1, \ldots, \omega_M)$ and $\Lambda_l = \Lambda_{l,1} \times \cdots \times \Lambda_M$ and the fact that $\tilde{\Omega}_R$ depends only on the types, and the second inequality follows from (27). Thus, if the rejection by $\tilde{\Omega}_R$ occurs, then the observed empirical joint empirical measure $\{q_j\}_{j=1,\ldots M}$ satisfies (29) and setting p as the joint empirical measure $q_{\mathrm{total}}(\cdot,\cdot) = \frac{1}{M(T-1)}\sum_{j=1}^{M} f_j^1(\cdot,\cdot)$ in (29) implies

$$\alpha - \delta_{4T} \leq \sum_{j=1}^{M} K(q_j, q_{\mathrm{total}}) = \frac{\mathcal{T}_P^*}{2(T-1)},$$

for all T large enough, and (10) follows.

Finally, we show (8). Define the entropy of a two-period joint measure $q(\cdot,\cdot)$ as

$$H(q) = -\sum_{\mathbf{s}\in\mathbf{S}} q(\mathbf{s}) \sum_{\mathbf{s}'\in\mathbf{S}} q(\mathbf{s}'|\mathbf{s}) \log q(\mathbf{s}'|\mathbf{s}).$$

Then by the definition of $K(\cdot,\cdot)$, the test statistic is written as

$$\frac{\mathcal{T}_P^*}{2(T-1)} = MH(q_{\text{total}}) - \sum_{j=1}^{M} H(q_j). \tag{30}$$

Let Ω_R^* be the rejection region of the test $\mathbf{1}\{\mathcal{T}_P^* \geq 2(T-1)(\alpha-\delta_{4T})\}$. Also, let $q_j^{\omega_j}(\cdot,\cdot)$ be the two-period joint empirical measure based on ω^j and $q_{\text{total}}^{\omega}(\cdot,\cdot) = M^{-1}\sum_{j=1}^{M} q_j^{\omega_j}(\cdot,\cdot)$. We have

$$\begin{aligned}
\Pr\{\Omega_R^* : \mathrm{H}_0^{\mathbf{P}}\} &= \sum_{\omega\in\Omega_R^*}\prod_{j=1}^{M}\Pr\{\omega_j : \mathrm{H}_0^{\mathbf{P}}\} \\
&\leq \sum_{\omega\in\Omega_R^*}\exp(-(T-1)MH(q_{\text{total}}^{\omega})) \\
&\leq \exp(-(T-1)(\alpha-\delta_{4T}))\sum_{\omega\in\Omega_R^*}\exp\left(-(T-1)\sum_{j=1}^{M}H(q_j^{\omega_j})\right) \\
&\leq \exp(-(T-1)(\alpha-\delta_{4T}))\prod_{j=1}^{M}\sum_{\omega_j\in\Omega_j}\exp(-(T-1)H(q_j^{\omega_j})) \\
&\leq \exp(-(T-1)(\alpha-\delta_{4T})+(T-1)O(T^{-1}\log T)),
\end{aligned}$$

where the equality follows from independence of $(\omega_1,\ldots,\omega_M)$, the first inequality follows from the fact that under $\mathrm{H}_0^{\mathbf{P}}$ the log likelihood $\sum_{j=1}^{M}\log\Pr\{\omega_j : \mathrm{H}_0^{\mathbf{P}}\}$ of observed ω is maximized by $q_{\text{total}}^{\omega}$ with maximum $-M(T-1)H(q_{\text{total}}^{\omega})$, the second inequality follows from $\omega\in\Omega_R^*$ and (30) (i.e., $MH(q_{\text{total}}^{\omega})-\sum_{j=1}^{M}H(q_j^{\omega_j}) \geq 2(\alpha-\delta_{4T})$), the third inequality follows from the Jensen inequality and $\Omega_R^*\subset\Omega$, and the last inequality follows from the upper bounds of the entropy and number of types of Markov chains in Davisson, Longo and Sgarro (1981, Theorem 1 combined with eq. (4)). Therefore, (8) follows.

A.3 Detail for the test statistic T_Q

The asymptotic variance $\mathbf{V}^j$ in (11) has the (k,l)-th element

$$v_{kl}^j = \mathbf{1}\{k=l\}q_k^j - q_k^j q_l^j + q_k^j\sum_{m=1}^{\infty}(p_{kl}^{j(m)}-q_l) + q_l^j\sum_{m=1}^{\infty}(p_{lk}^{j(m)}-q_k),$$

q_k^j is the k-th element of $\mathbf{Q}^j$, $p_{kl}^{j(m)}$ is the (k,l)-th element of $(\mathbf{P}^j)^m$. It should be noted that $\mathrm{rank}(\mathbf{V}^j) = m_s-1$ due to the linear constraint $(1,\ldots,1)'\mathbf{F}^j = T-1$. Under $\mathrm{H}_0^{\mathbf{Q}}$, it holds $\mathbf{V}=\mathbf{V}^1=\cdots=\mathbf{V}^M$ and the common asymptotic variance $\mathbf{V}$ can be estimated by e.g. Newey and West's (1987) estimator $\widehat{\mathbf{V}}$ whose (k,l)-th element is defined as

$$\widehat{v}_{kl} = \mathbf{1}\{k=l\}\widehat{q}_k - \widehat{q}_k\widehat{q}_l + \widehat{q}_k\sum_{m=1}^{b_T}(\widehat{p}_{kl}^{(m)}-\widehat{q}_l) + \widehat{q}_l\sum_{m=1}^{b_T}(\widehat{p}_{lk}^{(m)}-\widehat{q}_k),$$

where $\widehat{q}_k$ is the k-th element of $\frac{1}{M(T-1)}\sum_{j=1}^{M}\mathbf{F}^j$, $\widehat{p}_{kl}^{(m)}$ is the (k,l)-th element of $\widehat{\mathbf{P}}^m$, and $\widehat{\mathbf{P}} = \left\{\frac{1}{M(T-1)}\sum_{j=1}^{M} f_j^1(\mathbf{s}', \mathbf{s})\right\}_{\mathbf{s},\mathbf{s}'\in\mathbf{S}}$. Also the bandwidth b_T satisfies $b_T \to \infty$ and $T^{-1/2}b_T \to 0$.

References

[1] Aguirregabiria, V. and P. Mira (2007) Sequential simulation-based estimation of dynamic discrete games, *Econometrica*, 75, 1-53.

[2] Andrews, D. W. K. (1997) A conditional Kolmogorov test, *Econometrica*, 65, 1097-1128.

[3] Arcidiacono, P., Bayer, P., Blevins, J. and P. Ellickson (2015) Estimation of dynamic discrete choice models in continuous time with an application to retail competition, Working paper.

[4] Arcidiacono, P. and R. Miller (2011) Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity, *Econometrica*, 79, 1823-1867.

[5] Attanasio, O. (2000) Consumer durables and inertial behavior: estimation and aggregation of (S,s) rules for automobile purchases, *Review of Economic Studies*, 67, 667-696.

[6] Bajari, P., Benkard, L. and J. Levin (2007) Estimating dynamic models of imperfect competition, *Econometrica*, 75, 1331-1370.

[7] Baltagi, B. H. (2008) *Econometric Analysis of Panel Data*, fourth edition, West Sussex, UK: Wiley.

[8] Beresteanu, A., Ellickson, P. and S. Misra (2010) The dynamics of retail oligopoly, Working paper.

[9] Besag, J. and D. Mondal (2013) Exact goodness-of-fit tests for Markov chains, *Biometrics*, 69, 488-496.

[10] Billingsley, P. (1961) Statistical methods in Markov chains, *Annals of Mathematical Statistics*, 32, 12-40.

[11] Chiappori, P. A. and B. Salanié (2000) Testing for asymmetric information in insurance markets, *Journal of Political Economy*, 108, 56-78.

[12] Collard-Wexler, A. (2013) Demand fluctuations in the ready-mix concrete industry, *Econometrica*, 81, 1003-1037.

[13] Davisson, L. D., Longo, G. and A. Sgarro (1981) The error exponent for the noiseless encoding of finite ergodic Markov sources, *IEEE Transactions on Information Theory*, 27, 431-438.

[14] de Paula, A. and X. Tang (2011) Inference of signs of interaction effects in simultaneous games with incomplete information, *Econometrica*, 80, 143-172.

[15] Doraszelski, U. and M. Satterthwaite (2010) Computable Markov-perfect industry dynamics, *RAND Journal of Economics*, 41, 215-243.

[16] Dunne, T., Klimek, S., Roberts, M. and D. Xu (2013) Entry, exit and the determinants of market structure, *RAND Journal of Economics*, 44, 462-487.

[17] Ericson, R. and A. Pakes (1995) Markov perfect industry dynamics: a framework for empirical work, *Review of Economic Studies*, 62, 53-82.

[18] Fan, Y. and M. Xiao (2014) Competition and subsidies in the deregulated U.S. local telephone industry, Working paper.

[19] Gutman, M. (1989) Asymptotically optimal classification for multiple tests with empirically observed statistics, *IEEE Transactions on Information Theory*, 35, 401-408.

[20] Hoeffding, W. (1965) Asymptotically optimal tests for multinomial distributions, *Annals of Mathematical Statistics*, 36, 369-401.

[21] Jeziorski, P. (2014) Estimation of cost synergies from mergers: application to U.S. radio, *RAND Journal of Economics*, 45, 816-846.

[22] Jofre-Bonet, M. and M. Pesendorfer (2003) Estimation of a dynamic auction game, *Econometrica*, 71, 1443-1489.

[23] Kasahara, H. and K. Shimotsu (2009) Nonparametric identification of finite mixture models of dynamic discrete choices, *Econometrica*, 77, 135-175.

[24] Kasahara, H. and K. Shimotsu (2012) Sequential estimation of structural models with a fixed point constraint, *Econometrica*, 80, 2303-2319.

[25] Levin, D. A., Peres, Y. and E. L. Wilmer (2009) *Markov Chains and Mixing Times*, American Mathematical Society.

[26] Lin, H. (2014) Quality choice and market structure: a dynamic analysis of nursing home oligopolies, *International Economic Review*, forthcoming.

[27] Maican, F. and M. Orth (2014) Entry regulations, product differentiation and determinants of market structure, Working paper.

[28] McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, Academic Press: New York.

[29] Minamihashi, N. (2012) Natural monopoly and distorted competition: evidence from unbundling fiber-optic networks, Working paper.

[30] Mizon, G. E. and J.-F. Richard (1986) The encompassing principle and its application to testing non-nested hypothesis, *Econometrica*, 54, 657-678.

[31] Navarro, S. and Y. Takahashi (2012) A semiparametric test of agent's information sets for games of incomplete information, Working paper.

[32] Newey, W. K. and K. D. West (1987) A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55, 703-708.

[33] Nishiwaki, M. (2015) Horizontal mergers and divestment dynamics in a sunset industry, Working paper.

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

[35] Pesendorfer, M. and P. Schmidt-Dengler (2008) Asymptotic least squares estimators for dynamic games, *Review of Economic Studies*, 75, 901-928.

[36] Phillips, P. C. B. and H. R. Moon (1999) Linear regression limit theory for nonstationary panel data, *Econometrica*, 67, 1057-1111.

[37] Ryan, S. (2012) The costs of environmental regulation in a concentrated industry, *Econometrica*, 80, 1019-1061.

[38] Sanches, F. and D. Silva-Junior (2013) Bank privatization and the supply of banking services: evidence from a dynamic structural model, Working paper.

[39] Snider, C. (2009) Predatory incentives and predation policy: the American Airlines case, Working paper.

[40] Srisuma, T. and O. Linton (2012) Semiparametric estimation of Markov decision processes with continuous state space: discrete control, *Journal of Econometrics*, 166, 320-341.

[41] Suzuki, J. (2013) Land use regulation as a barrier to entry: evidence from the Texas lodging industry, *International Economic Review*, 54, 495-523.

[42] Sweeting, A. (2013) Dynamic product positioning in differentiated product industries: the effect of fees for musical performance rights on the commercial radio industry, *Econometrica*, 81, 1763-1803.

[43] van der Vaart, A. W. (1998) *Asymptotic Statistics*, Cambridge, UK: Cambridge University Press.