GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 423

Testing for Equilibrium Multiplicity in Dynamic Markov Games

Taisuke Otsu *
Martin Pesendorfer **
Yuya Takahashi ***

* London School of Economics
** London School of Economics
*** University of Mannheim

April 2013

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

Testing for Equilibrium Multiplicity in Dynamic Markov Games*

Taisuke Otsu† Martin Pesendorfer‡ Yuya Takahashi§

First Draft: October 2012

This Draft: April 2013

Abstract

This paper proposes several statistical tests for finite state Markov games to examine the null hypothesis that the data are generated from a single equilibrium. We formulate tests of (i) the conditional choice probabilities, (ii) the steady-state distribution of states and (iii) the conditional distribution of states conditional on an initial state. In a Monte Carlo study we find that the chi-squared test of the steady-state distribution performs well and has high power even with a small number of markets and time periods. We apply the chi-squared test to the empirical application of Ryan (2012) that analyzes dynamics of the U.S. Portland Cement industry and test if his assumption of single equilibrium is supported by the data.

Keywords: Dynamic Markov Game, Multiplicity of Equilibria, Testing.
Jel Classification: C12, C72, D44.

---

*We wish to thank Stephen Ryan and seminar audiences in Alicante, Barcelona (Pompeu Fabra and Autonoma), Princeton, Wisconsin, UCL and UBC for thoughtful comments. 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, University of Mannheim, ytakahas@mail.uni-mannheim.de.

1 Introduction

While a class of dynamic Markov games was formalized many years ago, see Maskin and Tirole (1988) and Ericson and Pakes (1995) for an empirical framework, empirical applications have been limited until recently. 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 Linton and Srisuma (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.

The basic idea of two-step methods is as follows[1]. In the first stage, players' policies and state transition probabilities are estimated directly from the data as functions of observable state variables. These functions are reduced-form in that the estimated parameters are not the parameters of the underlying economic model. In the second stage, a search for the structural model parameters which best rationalizes observed behaviors of players and state transitions is conducted. The second stage uses the estimated policies as estimates for the equilibrium beliefs, since these two should coincide in Markov Perfect equilibria. In this approach structural model parameters can be estimated without solving an equilibrium even once.

Two step methods significantly broadened the research scope on dynamic problems that can be empirically addressed. In practice, some of the necessary conditions for these methods to work are not easily satisfied. To obtain reasonable estimates of policy functions and state transition probabilities, the data need to contain rich information on actions and state transitions for every observable state which are generated from the same equilibrium. In a typical IO application, a long time-series data may not be available. Researchers are tempted to pool data from different markets (games) to perform the first stage policy function estimation. To do so, researchers assume that the data are generated from a single (identical) equilibrium in every market. This assumption has become popular in a number of recent papers.[2] If this assumption is violated, then the estimated policies are a mixture of different policies, each of which corresponds to different equilibria. The assumption may be very restrictive as multiplicity of equilibria is a well known feature inherent to games. Incorrectly imposing this assumption leads to erroneous inference.

This paper proposes several test statistics to test the null hypothesis that the data is generated from a single equilibrium in a class of finite-state Markov games. Specifically, we test multiplicity of equilibria in three ways. The first

[1] A two-step method itself was pioneered by Hotz and Miller (1993).

[2] Examples include Beresteanu, Ellickson, and Misra (2010), Collard-Wexler (2010), Dunne, Klimek, Roberts, and Xu (2011), Fan and Xiao (2012), Jeziorski, (2012), Lin (2011), Maican and Orth (2012), Minamihashi (2012), Nishiwaki (2010), Ryan (2012), Sanches and Silva Junior (2012), Snider (2009), Suzuki (2012), and Sweeting (2011). They impose this assumption either explicitly or implicitly. The empirical sections of Aguirregabiria and Mira (2007) and Arcidiacono, Bayer, Blevins, and Ellickson (2012) also impose the same assumption.

test directly compares the set of conditional choice probabilities estimated from the pooled (across markets) sample with the one estimated from each market separately. The second test is based on the result that a steady-state distribution of states is associated with a generically unique transition matrix of states under the assumption of communicating Markov chains. Based on this result, the second set of tests compares the steady-state distribution estimated from the pooled sample with the one from each market. This test has two variants, the Chi-squared test and the Kolmogorov test, depending on either L^2 norm or sup norm being used, respectively. The third test uses the distribution of states conditional on the initial (observed) state. We apply the conditional Kolmogorov test developed by Andrews (1997). It turns out that the third test does not require several assumptions on Markov chains that are imposed for the second test. The third test has wider applicability.

To illustrate the performance of our tests, we first apply our tests to a simulated data using an example of multiple equilibria in Pesendorfer and Schmidt-Dengler (2008) to investigate finite-sample properties. Our tests, particularly the chi-squared test based on the steady-state distribution, perform well and have high power even with a small number of markets and time periods. Then, we apply the chi-squared test to the empirical application of Ryan (2012) that analyzes dynamics of the U.S. Portland Cement industry and test if his assumption of single equilibrium is supported by the data. We find that the hypothesis of single equilibrium is rejected at the 5% level. A further investigation shows that several outlier markets appear to be generating the rejection. If we exclude outlier markets, the test does not reject the null hypothesis.

To the best of our knowledge, this is the first paper that proposes tests of multiple equilibria 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. Furthermore, as a by-product, our tests work as specification tests. One common practice in the literature of estimating dynamic games is to impose parametric functional forms on policy functions estimated in the first stage. Under the assumption of stationary MPE, the steady-state distribution of states implied by the estimated policy functions and observed states should be consistent. Thus, if functional forms are not flexible enough, the test rejects the null hypothesis.

Our test based on the steady-state distribution provides a natural and formal way to check goodness of fit of policy functions for dynamic Markov models. Typically researchers check goodness of fit for their model in somewhat arbitrary ways. One common practice is to simulate the model at the estimated parameter values and compare several simulated key endogenous variables with the observed counterparts by eyeballing.[3] Another way is to look at the R^2

[3] Collard-Wexler (2010) compares various moments including entry/exit rates, the number of plants per market, and so on. Fan and Xiao (2012) examine the percentage of markets with n firms, where $n = 0, 1, 2$, and above. Jeziorski (2012) examines the average likelihood of observing a merger. Jofre-Bonet and Pesendorfer (2003) compare the bid levels (mean and standard deviation) and the probability of observing a regular bid. Lin (2011) compares several moments such as entry/exit rates, the percentage of low quality nursing homes, etc. Nishiwaki

or the pseudo R^2 obtained from estimated policy functions.[4] While these are informative to some extent, it is difficult to know precisely how well the model fits the data based on these figures. There are several formal specification tests, for example the overidentification test when an economic model is estimated with GMM. Our test based on the steady-state distribution fully exploits the conditions imposed by MPE. It nicely serves as a specification test in dynamic Markov games.

It should be emphasized that multiplicity of equilibria is observationally equivalent to time-invariant unobservable market-level heterogeneity in our framework. Our tests apply when there exist multiple equilibria, unobservable heterogeneity, or both.[5] Thus, a rejection of our tests points to an inconsistency of the first stage estimates that arise 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 the existence of unobservable types in single agent dynamic models.

There is a close link between our analysis and the recent literature on identification of finite mixture models. Since the number of equilibria is generically finite, dynamic games with multiple equilibria and a well-defined selection rule can be regarded as one class of finite mixture models. Kasahara and Shimotsu (2009) consider identifiability of finite mixture models of dynamic discrete choices. Among other things, they provide a condition to identify the lower bound of the number of mixture components. Theoretically, if the identified lower bound is larger than one, it implies multiplicity of equilibria. While this identification result may potentially be useful, it is not obvious to build on this to construct an implementable test.

Our paper also relates to de Paula and Tang (2011) that use tests of conditional independence between players' actions to test multiplicity of equilibria in the context of static games with incomplete information. Since our tests exploit the panel structure of the data and rely on the way that the game and states evolve, their tests and our tests are fundamentally different. 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' private shocks.

This paper is organized as follows. Section 2 lays out a class of general dynamic Markov games we work with and proves several important properties on

(2010) compares the number of divestment. Sanches and Silva Junior (2012) compare entry probabilities. Snider (2009) plots the model time series for prices and capacities versus the actual series. Sweeting (2011) investigates the share of radio stations that switch formats and changes in station revenues over time.

[4] Ryan (2012) compares the R^2, the pseudo R^2, and the value of likelihood in first stage policy function estimations among several specifications. The empirical section of Aguirregabiria and Mira (2007) look at the R^2 of the number of entries and exits.

[5] Aguirregabiria and Mira (2012) discuss difficulties of identifying and estimating models with both unobservable heterogeneity and multiple equilibria. Arcidiacono and Miller (2011) develop a two-step method that can account for unobservable heterogeneity with finite support. Since the knowledge of the number of points in the support (and values of those variables) is required, this method is not directly applicable to the case where multiple equilibria are present in the data.

Markov chains, steady-state distributions, and multiplicity of equilibria. Based on these properties, Section 3 proposes several test statistics. In Sections 4 we conduct a Monte Carlo study to examine finite-sample properties. Section 5 applies one of our tests to data of Ryan (2012). Section 6 concludes.

2 Model

This section describes elements of a general dynamic Markov game with discrete time, $t = 1, 2 \ldots, \infty$. 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 properties of the observed outcomes. Our setting includes the single agent case as a special case when there is one agent per market.

2.1 Set-up

We first describe the framework which applies for all markets $j \in \mathbf{M} = \{1, \ldots, M\}$. We then expand on the cross market structure.

Players. The set of players is denoted by $\mathbf{N} = \{1, \ldots, N\}$ and a typical player is denoted by $i \in \mathbf{N}$. The single agent case arises when $N = 1$. The number of players is fixed and does not change over time. Every period the following variables are observed:

States. Each player is endowed with a state variable $s_i^t \in \mathbf{S}_i = \{1, \ldots, L\}$. The state variable s_i^t is publicly observed by all players and the econometrician. The vector of all players' public state variables is denoted by $\mathbf{s}^t = (s_1^t, \ldots, s_N^t) \in \mathbf{S} = \times_{j=1}^N \mathbf{S}_j$. The cardinality of the state space $\mathbf{S}$ is finite and equals $m_s = L^N$.

Actions. Each player chooses an action $a_i^t \in A_i = \{0, 1, \ldots, K\}$. 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 modelling assumptions, and allows for within-period correlation in the random utility component across actions and across players. An action profile $\mathbf{a}^t$ denotes the vector of joint actions in period t, $\mathbf{a}^t = (a_1^t, \ldots, a_N^t) \in A = \times_{j=1}^N A_j$. The cardinality of the action space A equals $m_a = (K+1)^N$.

Choice probability matrix. Let $\sigma(\mathbf{a}|s)$ denote the conditional probability that action profile $\mathbf{a}$ will be chosen conditional on state s. The matrix of conditional choice probabilities is denoted by $\boldsymbol{\sigma}$. It has dimension $m_s \times (m_a \cdot m_s)$. 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. The state-action transition is described by an indicator function $g : A \times \mathbf{S} \times \mathbf{S} \longrightarrow \{0, 1\}$ where a typical element $g\left(\mathbf{a}^t, \mathbf{s}^t, \mathbf{s}^{t+1}\right)$

denotes the probability that state $\mathbf{s}^{t+1}$ is reached when the current action profile and state are given by $(\mathbf{a}^t, \mathbf{s}^t)$. We require $\sum_{\mathbf{s}' \in \mathbf{S}} g(\mathbf{a}, \mathbf{s}, \mathbf{s}') = 1$ for all $(\mathbf{a}, \mathbf{s}) \in A \times \mathbf{S}$. We use the symbol $\mathbf{G}$ to denote the $(m_a \cdot m_s) \times m_s$ dimensional state-action transition matrix in which column $s^{'} \in \mathbf{S}$ consists of the vector of probabilities $\left[g(\mathbf{a}, \mathbf{s}, \mathbf{s}')_{\mathbf{a} \in \mathbf{A}, \mathbf{s} \in \mathbf{S}}\right]$.

State transition matrix. Let $\mathbf{P} = \boldsymbol{\sigma}\mathbf{G}$ denote the $m_s \times m_s$ dimensional state transition matrix induced by the choice probability matrix $\boldsymbol{\sigma}$ and state-action transition matrix $\mathbf{G}$. A typical element $p(\mathbf{s}, \mathbf{s}')$ equals the probability that state $\mathbf{s}'$ is reached when the current state is given by $\mathbf{s}$, $p(\mathbf{s}, \mathbf{s}') = \sum_{\mathbf{a} \in \mathbf{A}} \sigma(\mathbf{a}|\mathbf{s}) \cdot g(\mathbf{a}, \mathbf{s}, \mathbf{s}')$. Since the elements in each row of $\mathbf{P}$ sum to 1, $\sum_{\mathbf{s}' \in \mathbf{S}} p(\mathbf{s}, \mathbf{s}') = 1$ for all $\mathbf{s} \in \mathbf{S}$, the matrix $\mathbf{P}$ is a right stochastic matrix. The matrix $\mathbf{P} \in \mathcal{P}$ is called a Markov chain. The set $\mathcal{P}$ denotes the set of all right stochastic matrices.

Limiting steady-state distribution. When the limit exists, let $Q_{\mathbf{s}}(\mathbf{s}')$ denote the long run proportion of time that the Markov chain $\mathbf{P}$ spends in state $\mathbf{s}$ when starting at the initial state $\mathbf{s}^0 = \mathbf{s}'$

$$Q_{\mathbf{s}}(\mathbf{s}') = \lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{t=1}^{n} \mathbf{1}(\mathbf{s}^t = \mathbf{s} | \mathbf{s}^0 = \mathbf{s}')$$

The unconditional long run proportion of time that the Markov chain $\mathbf{P}$ spends in state $\mathbf{s}$ is given by

$$Q_{\mathbf{s}} = Q_{\mathbf{s}}(\mathbf{s}') \text{ with probability 1 for all initial states } \mathbf{s}'.$$

If for all $\mathbf{s} \in \mathbf{S}$, $Q_{\mathbf{s}}$ exists and is independent of the initial state $\mathbf{s}'$, satisfies $\sum_{\mathbf{s}' \in \mathbf{S}} Q_{\mathbf{s}'} = 1$, then the $1 \times m_s$ dimensional 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.

2.2 Some properties

The properties of Markov chains are well known. We next describe properties 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 and a sequence of states $\left(\mathbf{s}^1, \ldots, \mathbf{s}^T\right)$ so that the chain $\mathbf{P}$ will be at state $\mathbf{s}'$ after T periods. 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* $\mathbf{s}, \mathbf{s}' \in \mathbf{S} \times \mathbf{S}$ *of the chain* $\mathbf{P} \in \mathcal{P}$ *communicate with themselves. Then the following properties hold:*
(i) The steady-state distribution $\mathbf{Q}$ *exists. It satisfies* $Q_{\mathbf{s}} > 0$ *for all* $\mathbf{s} \in \mathbf{S}$ *and*

$$\mathbf{Q} = \mathbf{QP}.$$

(ii) The steady-state distribution $\mathbf{Q}$ *is unique.*

Property (i) 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. Property (ii) states that the steady-state distribution is unique. A proof of the above properties is given in Proposition 1.14 and Corollary 1.17 in Levin, Peres and Wilmer (2009).

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 $\mathbf{s}$, 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 full support assumption is made in standard dynamic discrete choice models, see Arcidiacono and Miller (2011) for a recent formulation.

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 $\left(\left(\mathbf{s}^1, \mathbf{a}^1\right), \ldots, \left(\mathbf{s}^n, \mathbf{a}^n\right)\right)$ so that the chain starting at state $\mathbf{s}$ will be at state $\mathbf{s}'$ after n periods for any $\mathbf{s}, \mathbf{s}' \in \mathbf{S}$.

Next, we highlight that a sequence of realizations of a steady state Markov chain represent random draws from the steady state distribution $\mathbf{Q}$. This property is commonly used to generate random numbers drawn from a distribution.

Lemma 2 *Suppose* $\left(\mathbf{s}^0, \ldots, \mathbf{s}^T\right)$ *are realizations of a Markov chain* $\mathbf{P} \in \mathcal{P}$ *with steady-state distribution* $\mathbf{Q}$ *and with the property that all states communicate with themselves. If* $\mathbf{s}^0 \sim \mathbf{Q}$*, then* $\mathbf{s}^t \sim \mathbf{Q}$ *for all* $t = 1, \ldots, T$*.*

Proof. Given any distribution $\mathbf{Q}$ on the state space $\mathbf{S}$, from the definition of the Markov chain transition matrix $\mathbf{P}$ it follows that if $\mathbf{s}^0 \sim \mathbf{Q}$, then $\mathbf{s}^1 \sim \mathbf{QP}, \mathbf{s}^2 \sim \mathbf{QP}^2, ..., \mathbf{s}^T \sim \mathbf{QP}^T$. From Lemma 1 property (i), we know that if $\mathbf{Q}$ is the steady-state distribution, then it must satisfy $\mathbf{Q} = \mathbf{QP}$. Multiplying both sides of the equation (on the right) by $\mathbf{P}$ yields $\mathbf{QP} = \mathbf{QP}^2$ and so on yielding $\mathbf{Q} = \mathbf{QP}^t$ for $t = 1, \ldots, T$. Thus, we can conclude that if $\mathbf{s}^0 \sim \mathbf{Q}$, then $\mathbf{s}^t \sim \mathbf{Q}$ for all $t = 1, \ldots, T$. ■

The Lemma illustrates that a sequence of realizations from a Markov chain in steady-state are random draws from the steady-state distribution.

2.3 Multiplicity

This section discusses some implications on the steady state distribution when there are multiple markets and individual markets are governed by possibly distinct transition matrices. A typical data set contains a collection of outcomes of the game independently played in M different markets.

As discussed in Pesendorfer and Schmidt-Dengler (2008), even though multiplicity of equilibria are prevalent in the class of games we study, the Markovian assumption implies that a single equilibrium is played in a market-level time series. Let $\mathbf{P}^j$ be the transition matrix induced by the equilibrium choice probabilities $\boldsymbol{\sigma}^j$ played in market j and let $\mathbf{Q}^j = (Q^j_{\mathbf{s}})_{\mathbf{s}\in\mathbf{S}}$ be the steady-state distribution that is associated with that transition matrix. This section requires that the steady-state distribution $\mathbf{Q}^j$ exists and is unique, see Lemma 1.

The possibility of distinct transition matrices in distinct markets arises naturally in games as multiplicity of equilibria is a well known feature inherent to games. Distinct transition matrices for distinct markets may also arise if some of the unobserved elements in the market are distinct. This section considers the case in which at least two markets have distinct transition matrices $\mathbf{P}^1 \neq \mathbf{P}^2$. We illustrate implications on the steady state distribution.

The following Lemma establishes that with probability one the market specific steady state distributions will differ $\mathbf{Q}^1 \neq \mathbf{Q}^2$.

Lemma 3 *Consider any two arbitrary right stochastic matrices* $\mathbf{P}^1, \mathbf{P}^2 \in \mathcal{P}$ *with* $\mathbf{P}^1 \neq \mathbf{P}^2$*. With probability one the steady state distributions differ,* $\mathbf{Q}^1 \neq \mathbf{Q}^2$*.*

Proof. Consider the steady state distribution property (i) for the matrix $\mathbf{P}^2$ in Lemma 1. If $\mathbf{Q^1}$ is also the steady state distribution for the transition matrix $\mathbf{P}^2$, then this places $(m_s - 1)$ linear restrictions on $\mathbf{P}^2$ of the form $\mathbf{Q^1} = \mathbf{Q^1} \cdot \mathbf{P}^2$. The event that an arbitrary matrix $\mathbf{P}^2$ satisfies these restrictions has Lebesgue measure zero on the set of all $m_s \times m_s$ dimensional right stochastic matrices. ∎

The proof argument is based on property (i) in Lemma 1 which requires that a steady state distribution satisfies the linear relationship $\mathbf{Q} = \mathbf{Q} \cdot \mathbf{P}$. The probability that an arbitrary transition matrix $\mathbf{P}$ satisfies these linear restrictions for a specific steady state distribution $\mathbf{Q}$ has probability zero.

3 Tests for multiplicity

This section describes the hypothesis that we aim at testing and introduces three types of test statistics for this purpose. Under the null hypothesis the observed data are generated from an identical data generating processes in all markets. The alternative is that the data generating process is distinct for some markets.

For each market $j \in \mathbf{M}$ a sequence of action and state profiles $(\mathbf{a}^{tj}, \mathbf{s}^{tj})_{t=1,\ldots,T}$ is observed, where T is the length of time periods in the data set. These observables allow us to make inference about the conditional choice probability matrix $\boldsymbol{\sigma}$, the transition matrix $\mathbf{P}$ and the distribution of states $\mathbf{Q}$ by using a suitable estimator. Our null hypothesis is whether the individual markets $j \in \mathbf{M}$ yield the same estimator. The alternative is the negation of the null.

The null and alternative hypotheses can be written in terms of conditional choice probabilities $\boldsymbol{\sigma}$:

$$\begin{aligned} H_0 &: \boldsymbol{\sigma}^j = \boldsymbol{\sigma} \text{ for all } j \in \mathbf{M}; \\ H_1 &: \boldsymbol{\sigma}^j \neq \boldsymbol{\sigma}^{j'} \text{ for some } j, j' \in \mathbf{M}. \end{aligned}$$

Assuming that all the states in the Markov chain $\mathbf{P} = \boldsymbol{\sigma}\mathbf{G} \in \mathcal{P}$ communicate with themselves, then by Lemma 1- 3, these hypotheses are generically equivalent to a test of identical steady-state distributions:

$$\begin{aligned} H_0' &: \mathbf{Q}^j = \mathbf{Q} \text{ for all } j \in \mathbf{M}; \\ H_1' &: \mathbf{Q}^j \neq \mathbf{Q}^{j'} \text{ for some } j, j' \in \mathbf{M}. \end{aligned}$$

By Lemma 1 and 3 we also know that the null and the alternative in the above test cannot be observationally equivalent.

Finally, there may be circumstances in which the researcher does not know whether all states in the Markov chain $\mathbf{P}$ communicate with themselves, or when there is a concern that the initial conditions matter. In this case, the conditional state distribution may form the basis of the test:

$$\begin{aligned} H_0^{''} &: \mathbf{Q}^j(\mathbf{s}') = \mathbf{Q}(\mathbf{s}') \text{ for all } j \in \mathbf{M}, \mathbf{s}' \in \mathbf{S}; \\ H_1^{''} &: \mathbf{Q}^j(\mathbf{s}') \neq \mathbf{Q}^{j'}(\mathbf{s}') \text{ for some } j, j' \in \mathbf{M} \text{ and } \mathbf{s}' \in \mathbf{S}. \end{aligned}$$

Our first test, the conditional choice probability test, is based on the first set of hypotheses. Our second test, the (unconditional) steady-state distribution test is based on the second set of hypothesis and the third test, the conditional state distribution test, is based on the third set of hypotheses.

If more than one equilibrium is played in the data, then $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\mathbf{P}}$ would be estimates of a mixture of different choice and transition probabilities. Yet, the theoretical distribution of choices and states implied by the mixture model differs from the theoretical distribution with $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\mathbf{P}}$. Each of our test statistics is aimed at detecting whether such a difference is present in the data.

Before discussing each test statistic, we comment on estimators that are used for testing. In principle, any consistent estimator for $\boldsymbol{\sigma}$, $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{Q}(\mathbf{s}')$ can be used in our tests and there is no unique way to estimate them. Thus, which estimator should be used in practice depends on the application in question. We consider three types of estimators in this paper: First, a frequency estimator which is a natural non-parametric estimator for distributions and probabilities on finite spaces. Second, when states communicate so that a unique $\mathbf{Q}$ exists for any $\mathbf{P}$ (Lemma 1), instead of directly using a frequency estimator for $\mathbf{Q}$, the steady-state distribution implied by estimated transition probabilities could be used. If the equilibrium Markov chain is a-periodic, then $\mathbf{Q}$ can be calculated as $\mathbf{Q}(\widehat{\mathbf{P}}) = \frac{1}{m_s} \cdot I \cdot (\lim_{n\to\infty} \widehat{\mathbf{P}}^n)$ where $\widehat{\mathbf{P}}$ is a consistent estimator for $\mathbf{P}$ and $\mathbf{I}$ denotes a row vector of one.[6] Third, the conditional choice probabilities can

[6] In general finding the steady-state probabilities $\mathbf{Q}$ amounts to finding the eigenvectors of P. Gallager (1996) shows that the largest real eigenvalue of $\mathbf{P}$ is $\lambda = 1$ with associated right eigenvector $e = (1, 1, ..., 1)'$ unique up to a scale factor. Furthermore, $\mathbf{Q}$ is the unique left eigenvector of the eigenvalue $\lambda = 1$.

be parametrized and form the basis to calculate the implied $\mathbf{P}$ and $\mathbf{Q}$. Let $\boldsymbol{\theta} \in \Theta \subset \mathbf{R}^k$ be a parameter vector and $\boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$ be a parametric conditional choice probability estimator. Then, $\boldsymbol{\sigma}(\widehat{\boldsymbol{\theta}})$ can be used to calculate $\mathbf{P}$ and $\mathbf{Q}$.[7] An advantage of this last approach is that k can be a much smaller number than the dimensionality of $\boldsymbol{\sigma}$. As we mentioned in the Introduction, this is a common practice in the literature, as the sample size is typically small compared to the dimensionality of $\boldsymbol{\sigma}$.

3.1 Testing the conditional choice probabilities

We first form a generally applicable test statistic based directly on the conditional choice probabilities. The test does not require that states communicate. It holds for general Markov models even if the steady-state distribution is not unique. The test is based on the idea that equilibrium choice probabilities are unique. Any two equilibria will have distinct choice probabilities. Now, if there are multiple equilibria played across markets, then this will result in variation in choice probabilities across markets. Thus, the cross sectional variation can be exploited to detect multiplicity.

Let $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ be an estimator for $\sigma(\mathbf{a}|\mathbf{s})$, the probability that action profile $\mathbf{a}$ is chosen when the current state is given by $\mathbf{s}$. When H_0 is true, we let $\sigma_0(\mathbf{a}|\mathbf{s})$ denote the true value of $\sigma(\mathbf{a}|\mathbf{s})$. We assume that under H_0, $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ is a $\sqrt{T}-$consistent estimator of $\sigma_0(\mathbf{a}|\mathbf{s})$. The estimate $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ can use all available data, that is all data for all markets. Let $\widetilde{\sigma}^j_{\mathbf{a}|\mathbf{s}}$ be the observed frequency of action profile $\mathbf{a}$ in state $\mathbf{s}$ for market j

$$\widetilde{\sigma}^j_{\mathbf{a}|\mathbf{s}} = \begin{cases} \frac{\sum_{t=1}^T \mathbf{1}\left(\mathbf{a}^{tj}=\mathbf{a}, \mathbf{s}^{tj}=\mathbf{s}\right)}{\sum_{t=1}^T \mathbf{1}(\mathbf{s}^{tj}=\mathbf{s})} & \text{if } \sum_{t=1}^T \mathbf{1}(\mathbf{s}^{tj}=\mathbf{s}) > 0; \\ 0 & \text{otherwise.} \end{cases}$$

It should be emphasized that $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ can be any consistent estimator including a frequency estimator, while $\widetilde{\sigma}^j_{\mathbf{a}|\mathbf{s}}$ is the observed frequency. If conditional choice probabilities are parameterized by $\boldsymbol{\theta}$, then $\sigma(\mathbf{a}|\mathbf{s})$ is replaced with $\sigma(\mathbf{a}|\mathbf{s};\boldsymbol{\theta})$ in the following discussion. In what follows, we keep $\sigma(\mathbf{a}|\mathbf{s})$ for notational simplicity.

Our conditional choice probabilities chi-squared test statistic is

$$CCP = T \cdot \sum_{j\in\mathbf{M}} \sum_{\mathbf{s}\in\mathbf{S}} \sum_{\mathbf{a}\in\mathbf{A}} \left[\frac{\left[\widetilde{\sigma}^j_{\mathbf{a}|\mathbf{s}} - \widehat{\sigma}(\mathbf{a}|\mathbf{s})\right]^2}{\widehat{\sigma}(\mathbf{a}|\mathbf{s})} \cdot \mathbf{1}\left(\sum_{t=1}^T \mathbf{1}(\mathbf{s}^{tj}=\mathbf{s}) > 0\right)\right], \qquad (1)$$

which counts the squared distance between the predicted probability $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ and the observed market-j frequency across all actions, states and markets. We arbitrarily count only those observations from markets j in which state $\mathbf{s}$ is indeed observed and omitted all others. As T increases, this selection should

[7] A natural alternative is a hybrid of the second and third possibilities. That is, the chain $\mathbf{P}$ could be parametrized instead of $\boldsymbol{\sigma}$. The implied steady-state distribution can be calculated as $\mathbf{Q}\left(\mathbf{P}\left(\widehat{\theta}\right)\right)$.

not matter, but it may matter for finite sample properties. This selection is made arbitrarily in the sense that alternative specifications can be used as well.

If $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ is a frequency estimator, then the CCP will be a standard Pearson statistic. As the number of observations increases, the CCP statistic approaches a chi-squared distribution with the appropriate degrees of freedom provided state realizations are independent across different time periods within each market. In Markovian games state realizations tend to be serially correlated implying that the CCP will not approach the chi-squared distribution. In addition, deriving an asymptotic distribution that accounts for serial correlation is not straightforward, since unless $\widehat{\sigma}(\mathbf{a}|\mathbf{s})$ is a frequency estimator estimated with iid samples, the test statistic will be an intractable mixture of chi-squared distributions as shown in Andrews (1988). Therefore, we use a parametric bootstrap to calculate the critical region of the test statistic.

We bootstrap using the estimated choice probabilities $\widehat{\boldsymbol{\sigma}}$. For every bootstrap sample b, we take M sample draws $\left\{(\mathbf{a}^{tj}, \mathbf{s}^{tj})_{t=1,\ldots,T}\right\}_{j\in\mathbf{M}}$ by simulating a state path from $\widehat{\boldsymbol{\sigma}}$ and $\mathbf{G}$ for every market j. We estimate $\widehat{\boldsymbol{\sigma}}^b$ from this bootstrap sample and calculate the associated test statistic CCP^b. We define the critical value as the 95th percentile in the distribution of CCP^b.

Note that our test can allow for within-period correlation in the random utility component across actions and across players. In the context of static games with incomplete information, de Paula and Tang (2011) test conditional independence between players' actions, $\widehat{\sigma}(a_i|\mathbf{s}) \cdot \widehat{\sigma}(a_j|\mathbf{s}) \neq \widehat{\sigma}(a_i, a_j|\mathbf{s})$, to check if there are more than one equilibria in the data generating process. This test relies on the assumption of independent-across-players private shocks. Our test is more flexible and permits within-period correlation in players' shocks. 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 evolves and requires repeated observations for each market.[8]

3.2 Testing the (unconditional) steady-state distribution

Our next test builds on the assumptions imposed on the Markov structure. It 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.

[8] 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 the equilibrium selection rule. On the other hand, Navarro and Takahashi (2012) assume a non-degenerate selection rule and interpret a violation of the conditional independence as a rejection of models of pure private shocks.

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.

We consider two distinct test statistics of the hypothesis of identical steady-state distributions across markets. The first is based on the Chi-Squared criterion while the second uses the Kolmogorov statistic.

3.2.1 Chi-Squared test statistic

Let $\widehat{\mathbf{Q}}$ be an estimator for the steady-state distribution $\mathbf{Q}$. When H_0 is true, we let $\mathbf{Q}_0$ denote the true value of $\mathbf{Q}$. We assume that under H_0, $\widehat{\mathbf{Q}}$ is a $\sqrt{T}-$consistent estimator of $\mathbf{Q}_0$. The estimate $\widehat{\mathbf{Q}}$ can use all available data, that is all data for all markets. Let $\widetilde{\mathbf{Q}}^j = \left(\widetilde{Q}^j_{\mathbf{s}}\right)_{\mathbf{s}\in\mathbf{S}}$ count the relative frequencies of observing states in market j with element

$$\widetilde{Q}^j_{\mathbf{s}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\left(\mathbf{s}^{tj} = \mathbf{s}\right) \qquad \text{for } \mathbf{s}\in\mathbf{S} \text{ and } j \in \mathbf{M}.$$

Our steady-state distribution chi-squared statistic is given by

$$SSC = T \cdot \sum_{j\in\mathbf{M}} \sum_{\mathbf{s}\in\mathbf{S}} \frac{\left[\widetilde{Q}^j_{\mathbf{s}} - \widehat{Q}_{\mathbf{s}}\right]^2}{\widehat{Q}_{\mathbf{s}}}, \tag{2}$$

which counts the squared distance between the predicted probability $\widehat{Q}_{\mathbf{s}}$ and the observed market-j frequency $\widetilde{Q}^j_{\mathbf{s}}$ across all states and markets. Any consistent estimator $\widehat{\mathbf{Q}} = \left(\widehat{Q}_{\mathbf{s}}\right)_{\mathbf{s}\in\mathbf{S}}$ can be used in the test statistic SSC. An intuitive nonparametric estimator for the steady-state distribution is the frequency estimators $\widetilde{\mathbf{Q}}$ counting the relative frequency of states in all markets with element,

$$\widetilde{Q}_{\mathbf{s}} = \frac{1}{TM}\sum_{t=1}^{T}\sum_{j\in\mathbf{M}} \mathbf{1}\left(\mathbf{s}^{tj} = \mathbf{s}\right) \qquad \text{for } \mathbf{s}\in\mathbf{S}.$$

Observe, that pooling both time periods and different markets is justified by Lemma 2.

An alternative estimator of $\widehat{\mathbf{Q}}$ can be based on suitable estimator of the transition probability matrix $\mathbf{P}$ as briefly discussed before. Let $p\left(\mathbf{s},\mathbf{s}';\widehat{\boldsymbol{\theta}}\right)$ be a $\sqrt{T}$ consistent estimator for $p\left(\mathbf{s},\mathbf{s}'\right)$, the probability that state $\mathbf{s}'$ is reached when the current state is given by $\mathbf{s}$ for some k-dimensional real-valued parameter vector $\boldsymbol{\theta}\in\mathbb{R}^k$. The estimate $p\left(\mathbf{s},\mathbf{s}';\widehat{\boldsymbol{\theta}}\right)$ can use all available data, that is all data for all markets. The estimator for the state transition matrix is $\widehat{\mathbf{P}} = \left(p\left(\mathbf{s},\mathbf{s}';\widehat{\boldsymbol{\theta}}\right)\right)_{\mathbf{s}\in\mathbf{S},\mathbf{s}'\in\mathbf{S}}$. The induced estimator of the steady-state distribution is $\mathbf{Q}(\widehat{\mathbf{P}}) = [Q_{\mathbf{s}}(\widehat{\mathbf{P}})]_{\mathbf{s}\in\mathbf{S}}$ which are the steady-state probabilities induced by the choice probability estimator $\widehat{\mathbf{P}}$.

For the same reason as the previous subsection, we use a parametric bootstrap to calculate the critical region of the test statistic. We bootstrap using the estimated transition probabilities $\widehat{\mathbf{P}}$ (and not by using $\widehat{\mathbf{Q}}$ directly). For every bootstrap sample b, we take M sample draws $\{(\mathbf{s}^{tj})_{t=1,\ldots,T}\}_{j\in\mathbf{M}}$ by simulating a state path from $\widehat{\mathbf{P}}$ for every market $j \in \mathbf{M}$. We estimate $\widehat{\mathbf{Q}}^b$ from this bootstrap sample and calculate the associated chi-squared test statistic SSC^b. We define the critical value as the 95th percentile in the distribution of SSC^b.

If the cardinality of the state space m_s is large relative to the number of observations, then the SSC statistic may be uninformative and perform poorly. To see this, observe that when m_s is large the probability $\widehat{Q}_{\mathbf{s}}$, which enters the denominator in the SSC test statistic, may become small for some states. Even small differences between $\widetilde{Q}^j_{\mathbf{s}}$ and $\widehat{Q}_{\mathbf{s}}$, when multiplied with a large number $1/\widehat{Q}_{\mathbf{s}}$, imply a large SSC test statistic. In such cases, the SSC test statistic may not be informative. To overcome this difficulty, we replace the predicted probability in the denominator with constant probabilities for all states $1/m_s$ and propose the following SSC' statistic

$$SSC' = (T \cdot m_s) \cdot \sum_{j\in\mathbf{M}} \sum_{\mathbf{s}\in\mathbf{S}} \left[\widetilde{Q}^j_{\mathbf{s}} - \widehat{Q}_{\mathbf{s}}\right]^2, \tag{3}$$

which sums the squared deviation between the predicted probability $\widehat{Q}_{\mathbf{s}}$ and the observed market-j frequency $\widetilde{Q}^j_{\mathbf{s}}$ across all states and markets. The critical region of the SSC' test statistic is calculated by using a bootstrap as described above.

3.2.2 Kolmogorov test statistic

The Kolmogorov statistics considers the maximal deviation of individual cells from the market average. It is based on the sup norm. Note that there is some arbitrariness in the way we label states and there is no unique way to label the multinomial cdf. The following test applies to any arbitrarily chosen labelling.

Our steady-state distribution Kolmogorov test statistic is

$$SSK = \sqrt{T} \max_{\mathbf{s}\in\mathbf{S}, j\in\mathbf{M}} \left|\widetilde{F}^j_{\mathbf{s}} - F\left(\mathbf{s}; \widehat{\mathbf{Q}}\right)\right|, \tag{4}$$

where $\widetilde{F}^j_{\mathbf{s}}$ denotes the empirical cumulative frequency distribution of observing state $\mathbf{s}$ in market j

$$\widetilde{F}^j_{\mathbf{s}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}(\mathbf{s}^t_j \leq \mathbf{s}) \qquad \text{for } \mathbf{s} \in \mathbf{S}; \tag{5}$$

and $F\left(\mathbf{s}; \widehat{\mathbf{Q}}\right)$ denotes the cdf of the multinomial distribution of observing state $\mathbf{s}$ based on a $\sqrt{T}$ consistent estimator of the steady state distribution $\widehat{\mathbf{Q}}$ under H_0.

The asymptotic null distribution of single-sample Kolmogorov's tests for parametric families of distributions in which parameters are estimated is given by Pollard (1984) and Beran and Millar (1989). This result applies for the test statistic $\sqrt{T}\max_{\mathbf{s}\in\mathbf{S}}\left|\widetilde{F}_{\mathbf{s}}^{j}-F\left(\mathbf{s};\widehat{\mathbf{Q}}\right)\right|$ when a single sample from market j is used in constructing the empirical cumulative distribution function $\widetilde{F}^{j}$ and when $\widehat{\mathbf{Q}}$ is a $\sqrt{T}$ consistent estimator possibly using data from all markets.

There is no obvious way of exploiting the cross market deviation of the single market statistic. Our SSK statistics in equation (4) is modified from the single market expression by taking the maximum deviation across all markets $j \in M$ jointly. The aim is to detect at least one deviating market. We could also modify equation (4) by taking average over market:

$$SSK' = \frac{1}{M}\sum_{j\in\mathbf{M}}\sqrt{T}\max_{s\in\mathbf{S}}\left|\widetilde{F}_{s}^{j}-F\left(s;\widehat{\mathbf{Q}}\right)\right|, \tag{6}$$

where $\widetilde{F}_{s}^{j}$ and $F\left(s;\widehat{\mathbf{Q}}\right)$ are defined in the same way as (4). One potential advantage of the SSK' statistic is that all markets are likely to contribute to increasing the value of the test statistic under the alternative, which may not be the case with SSK. We could imagine some situation in which the SSK test performs poorly due to the fact that it exploits only the maximal difference across cells and markets. On the other hand, we could also think of a situation where the SSK test would have high power. For example, suppose that there are two equilibria in the data generating process and that the first equilibrium is played in 90% of all the markets. Then, while $\sqrt{T}\max_{\mathbf{s}\in\mathbf{S}}\left|\widetilde{F}_{\mathbf{s}}^{j}-F\left(\mathbf{s};\widehat{\mathbf{Q}}\right)\right|$ tends to be small for markets in which the first equilibrium is played, it would be large for markets with another equilibrium. The SSK test would reject the null based on such large deviations even though the frequency of observing such markets is low. This is a potential advantage of the SSK test compared to the chi-squared tests as well as the SSK' test, all of which average out those deviations and may have low power in this example.

Given that the data generating process is ex-ante unknown to the researcher and that we aim at dealing with a variety of data generating processes, we select the first SSK statistic in (4), as the advantage of the SSK' test may be similar to that of our chi-squared tests. For the sake of comparisons, our Monte Carlo study also examined the SSK' test.

The critical region of our test statistic in (4) is calculated using a bootstrap. We bootstrap using the same procedure as above. For every bootstrap sample b, we take M sample draws $\{(\mathbf{s}^{tj})_{t=1,\ldots,T}\}_{j\in\mathbf{M}}$ by simulating a state path from $\widehat{\mathbf{P}}$ for every market $j \in \mathbf{M}$. Importantly, the bootstrap is defined by using the estimated choice probabilities $\widehat{\mathbf{P}}$ and not by using $\mathbf{Q}(\widehat{\mathbf{P}})$ directly. We estimate $\widehat{\mathbf{Q}}^{b}$ from this bootstrap sample and calculate the associated chi-squared test statistic SSK^{b}. We define the critical value as the 95th percentile in the distribution of SSK^{b}.

3.3 Conditional state distribution test

Next, we propose a test based on the conditional state distribution $Q_{\mathbf{s}}(\mathbf{s}')$. The conditional state distribution is defined as the frequency distribution of states conditional on the initial state being $\mathbf{s}'$. This test is more general than the previous one in the sense that it does not require that all states communicate nor does it require a unique steady-state distribution.

The conditional Kolmogorov test developed by Andrews (1997) is suitable for this hypothesis. The test considers a parametric model that specifies the conditional distribution of states given the initial state variable $\mathbf{s}'$. For each market, we observe the distribution of states that the game visits. Let $Q^j_{\mathbf{s}}(\mathbf{s}')$ denote the relative frequency that the game in market j visits state $\mathbf{s}$ when the initial state equals $\mathbf{s}'$. Let $\mathbf{Z}^j = \left(\left\{Q^j_{\mathbf{s}}(\mathbf{s}^{j\prime})\right\}_{\mathbf{s}\in\mathbf{S}}, \mathbf{s}^{j\prime}\right) \in \mathbf{R}^{m_s} \times \mathbf{S}$ denote a random variable for market j consisting of the pair of $\mathbf{Q}^j(\mathbf{s}^{j\prime})$ and the initial state in the market $\mathbf{s}^{j\prime} \in \mathbf{S}$. Assume that $\left\{\mathbf{Z}^j\right\}_{j\in\mathbf{M}}$ are iid with conditional distribution function $H(\cdot|\mathbf{s}^{j\prime})$ of $\mathbf{Q}^j$ given $\mathbf{s}^{j\prime}$. Note that if the Markov chain has a unique steady-state distribution, then the marginal distribution of $\mathbf{s}^{j\prime}$ is simply $\mathbf{Q}$.

We consider a parametric family of conditional distributions of the response variables $\mathbf{Q}^j$ given the covariate $\mathbf{s}^{j\prime}$. The parametric family of conditional distribution functions is denoted with

$$\{F(\mathbf{Q}|\mathbf{s},\mathbf{P}) : \mathbf{P} \in \mathcal{P}\}$$

where $F(\mathbf{Q}|\mathbf{s},\mathbf{P})$ is the distribution function of $\mathbf{Q}$ when the state transition matrix equals $\mathbf{P}$ and the initial state is $\mathbf{s}$.

The null hypothesis is

$$H_0 : H(\cdot|\cdot) = F(\cdot|\cdot,\mathbf{P}) \text{ for some } \mathbf{P} \in \mathcal{P}.$$

The alternative hypothesis is the negation of H_0.

Let $\widehat{\mathbf{P}}$ be an estimator of $\mathbf{P}$. When H_0 is true, we let $\mathbf{P}_0$ denote the true value of $\mathbf{P}$. We assume that under H_0, $\widehat{\mathbf{P}}$ is a $\sqrt{M}-$consistent estimator of $\mathbf{P}_0$.

Define the conditional-state distribution Andrews test as

$$CSA = \max_{k\in\mathbf{M}} \left| \frac{1}{\sqrt{M}} \sum_{j\in\mathbf{M}} \left[\mathbf{1}\left(\mathbf{Q}^j \le \mathbf{Q}^k\right) - F\left(\mathbf{Q}^k|\mathbf{s}^{j\prime},\widehat{\mathbf{P}}\right)\right] \mathbf{1}\left(\mathbf{s}^{j\prime} \le \mathbf{s}^{k\prime}\right) \right| \tag{7}$$

The CSA statistic is based on the difference between the empirical df and the semi-parametric/semi-empirical distribution function $\sum_{j\in\mathbf{M}} F\left(\mathbf{Q}^k|\mathbf{s}^{j\prime},\widehat{\mathbf{P}}\right)\mathbf{1}\left(\mathbf{s}^{j\prime} \le \mathbf{s}^{k\prime}\right)$. The reason for using the semi-parametric/semi-empirical distribution function is that the parametric model does not specify the distribution function $\mathbf{Z}^j$. It only specifies the conditional distribution function of $\mathbf{Q}^j$ given the covariate $\mathbf{s}^{j\prime}$. Also, observe that the CSA statistic is not defined by taking the supremum over all points $\mathbf{Z}$ in $\mathbf{R}^{m_s} \times \mathbf{S}$. Rather, the CSA statistic is defined by taking the maximum over points $\mathbf{Z}$ in the sample $\left\{\mathbf{Z}^j\right\}_{j\in\mathbf{M}}$.

The asymptotic null distribution of CSA depends on $\mathbf{P}_0$ as well as the distribution function of the covariate, and thus asymptotic critical values for CSA cannot be tabulated. Instead, a parametric bootstrap can be used to obtain critical values and p values for the CSA statistic. Andrews (1997) demonstrates that the following bootstrap is valid.

We simulate B bootstrap samples each of size M: $\left\{\mathbf{Z}^{jb}\right\}_{j\in\mathbf{M}}$ for $b = 1, ..., B$, where $\mathbf{Z}^{jb} = (\mathbf{Q}^{jb}, \mathbf{s}^{j\prime})$ for $j \in \mathbf{M}$. Note that for each bootstrap sample, we have the same covariate $\left\{\mathbf{s}^{j\prime}\right\}_{j\in\mathbf{M}}$ as the original sample. Given $\mathbf{s}^{j\prime}$, we simulate $\mathbf{Q}^{jb}$ using the parametric conditional density associated with the distribution function $F\left(\mathbf{Q}|\mathbf{s}, \widehat{\mathbf{P}}\right)$. This is repeated M times to give $\left\{\mathbf{Q}^{jb}\right\}_{j\in\mathbf{M}}$. We repeat this procedure for $b = 1, ..., B$. Finally, we compute CSA for each of b bootstrap sample, denoted CSA^b. We use $\left\{CSA^b\right\}_{b=1,...,B}$ to calculate critical values and p values.

Note that the above test could also condition on a set of covariates instead of a scalar variable $\mathbf{s}^{j\prime}$. For example, if there are a finite number of time-invariant market types (and if that is observable), the distribution of $\mathbf{Q}^j$ will be conditional on $(\mathbf{s}^{j\prime}, m^j)$, where m^j denotes j's market type that has a finite support. The test statistic and procedure would remain the same. The Monte Carlo section focuses on the case with a scalar conditioning covariate.

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 $\{0, 1\}$ and binary states $\{0, 1\}$. The distribution of the profitability shocks F 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 are parametrized as follows:

$$\pi\left(a_i, a_j, s_i\right) = \begin{cases} 0 & \text{if } a_i = 0;\ s_i = 0 \\ x & \text{if } a_i = 0;\ s_i = 1 \\ \pi^1 + c & \text{if } a_i = 1;\ a_j = 0;\ s_i = 0 \\ \pi^2 + c & \text{if } a_i = 1;\ a_j = 1;\ s_i = 0 \\ \pi^1 & \text{if } a_i = 1;\ a_j = 0;\ s_i = 1 \\ \pi^2 & \text{if } a_i = 1;\ a_j = 1;\ s_i = 1 \end{cases}$$

where $x = 0.1$; $c = -0.2$; $\pi^1 = 1.2$; 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 c; while a player that switches from 1 to 0 receives the exit value x.

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 $(s_1, s_2) = ((0,0),(0,1),(1,0),(1,1))$.

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 Monte Carlo study considers the conditional choice probabilities test, the steady-state distribution chi-squared test and Kolmogorov test and the conditional state distribution Andrews test as described in section 3. 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 chose $M = 20, 40, ..., 320$ and $T = 4, 8, ..., 32$. 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 choice probabilities for each market are estimated by a frequency estimator

$$\widetilde{\sigma}^j_{\mathbf{a}|\mathbf{s}} = \begin{cases} \frac{\sum_{t=1}^T \mathbf{1}(\mathbf{a}^{jt}=\mathbf{a})\cdot\mathbf{1}(\mathbf{s}^{jt}=\mathbf{s})}{\sum_{t=1}^T \mathbf{1}(\mathbf{s}^{jt}=\mathbf{s})} & \text{for } \sum_{t=1}^T \mathbf{1}(\mathbf{s}^{jt} = \mathbf{s}) > 0; \\ 0 & \text{otherwise;} \end{cases}$$

and $\widetilde{\sigma}(\mathbf{a}|\mathbf{s})$ is calculated as

$$\widetilde{\sigma}(\mathbf{a}|\mathbf{s}) = \begin{cases} \frac{\sum_{j\in\mathbf{M}}\sum_{t=1}^T \mathbf{1}(\mathbf{a}^{jt}=\mathbf{a})\cdot\mathbf{1}(\mathbf{s}^{jt}=\mathbf{s})}{\sum_{j\in\mathbf{M}}\sum_{t=1}^T \mathbf{1}(\mathbf{s}^{jt}=\mathbf{s})} & \text{for } \sum_{j\in\mathbf{M}}\sum_{t=1}^T \mathbf{1}(\mathbf{s}^{jt} = \mathbf{s}) > 0; \\ 0 & \text{otherwise} \end{cases}$$

Observe that the estimators $\widetilde{\boldsymbol{\sigma}}$, $\widetilde{\boldsymbol{\sigma}}^j$ are the maximum likelihood estimators. They are consistent under H_0 as $T \to \infty$ and $\widehat{\boldsymbol{\sigma}}$ is additionally consistent as $M \to \infty$. In this example $\mathbf{a}^t = \mathbf{s}^{t+1}$ and the state transition probabilities equal the conditional choice probabilities $\sigma(\mathbf{a}|\mathbf{s}) = p(\mathbf{s}', \mathbf{s})$. Hence, $\widehat{\mathbf{P}}$, $\widehat{\mathbf{P}}^j$ are consistently estimated under H_0 as well. The steady-state probabilities $\mathbf{Q}(\widehat{\mathbf{P}}) = [Q_{\mathbf{s}}(\widehat{\mathbf{P}})]_{\mathbf{s}\in\mathbf{S}}$ are induced by the choice probability estimator $\widehat{\mathbf{P}}$. The induced steady state distribution is calculated by approximating the limit distribution as $\mathbf{Q}(\widehat{\mathbf{P}}) \approx \widehat{\mathbf{P}}^{200} \cdot \left(\mathbf{I}\cdot\frac{1}{4}\right)$ which is accurate up to about 10 decimals.[9]

Our CCP and SSC tests are implemented using equations (1) and (2) respectively. As explained in section 3 for a multinomial distribution there is

[9] With $R = 50$ already a very good approximation to about 6 decimals is achieved.

no unique way in which the Kolmogorov test is implemented. Our SSK test statistic is implemented based on equation (4) as

$$SSK = \sqrt{T} \max_{\mathbf{s}\in\mathbf{S}, j\in\mathbf{M}} \left| \widetilde{F}^j_{\mathbf{s}} - F\left(\mathbf{s}; \widetilde{\mathbf{Q}}\right) \right|$$

where we use a frequency estimator $\widetilde{\mathbf{Q}}$ to calculate the cdf of the multinomial distribution of observing state $\mathbf{s}$ and $\widetilde{F}^j_{\mathbf{s}}$ is calculated based on equation (5).

We base the CSA test on the random variable defined as

$$\mathbf{Y}^j = \left(\sum_t \mathbf{1}\left(\mathbf{s}^{jt} = (0,0)\right), \sum_t \mathbf{1}\left(\mathbf{s}^{jt} = (0,1)\right), \sum_t \mathbf{1}\left(\mathbf{s}^{jt} = (1,0)\right) \right),$$

which counts how often the game visits state one, two, and three.[10] Our CSA test statistic is implemented based on equation (7) as

$$CK_M = \max_{k\in\mathbf{M}} \left| \frac{1}{\sqrt{M}} \sum_{j\in\mathbf{M}} \left[\mathbf{1}\left(\mathbf{Y}^j \leq \mathbf{Y}^k\right) - F\left(\mathbf{Y}^k | \mathbf{s}^{j\prime}, \widehat{\mathbf{P}}\right)\right] \cdot \mathbf{1}\left(\mathbf{s}^{j\prime} \leq \mathbf{s}^{k\prime}\right) \right|$$

where $F\left(\mathbf{Y}^k | \mathbf{s}^{j\prime}, \widehat{\mathbf{P}}\right)$ is the cumulative distribution function for the random variable $\mathbf{Y}^k$ conditional on the initial state $\mathbf{s}^{j\prime}$ and with transition probabilities $\hat{\mathbf{P}}$.

The critical regions of the described test statistics are calculated using a bootstrap procedure. The bootstrap sample for the conditional choice probabilities test and steady state distribution test are identical. For every bootstrap b, M sample draws $\{\mathbf{s}^{jb}\}_{j\in\mathbf{M}}$ are obtained by simulating a choice/state path from $\widehat{\boldsymbol{\sigma}}$ for every market j. As in the data generating process, the initial state is taken as $(0,0)$ and we start the sampling process after 100 periods. For every bootstrap sample b the associated test statistic is calculated using the above given formula, respectively described in section 3. The critical value is defined as the 95th percentile in the distribution of bootstrapped test statistics. The bootstrap sample for the conditional state distribution test is generated in almost the same way, except that for each market j we use the same initial state as the original sample and draw a subsequent state path from the that initial state.

Results. Tables 1-4 report the results of the experiments. The Tables report the percentage of rejections of our tests for every value of M, T, λ. The experiment is based on 299 repetitions for the bootstrap sample and $1,000$ Monte Carlo repetitions. CCP denotes the conditional choice probabilities test, SSC denotes the steady-state distribution chi-squared test, SSC' is its variant given

[10] There are only four states and $\mathbf{Y}^j$ fully captures an outcome of the game. There are other ways to form the response variable $\mathbf{Y}^j$. For example, we could define $\mathbf{Y}^j = \left(\sum_t s_1^{jt}, \sum_t s_2^{jt}, \sum_t \left(s_1^{jt} \cdot s_2^{jt}\right)\right)$, which counts how often individual players have state equalling 1 and how often both players jointly have state equalling 1 in market j.

in equation (3), SSK denotes the steady-state distribution Kolmogorov test, and CSA denotes the conditional state distribution Andrews test as described in section 3.

Table 1. Monte Carlo Results: $\lambda = 0.5$

M	T	CCP	SSC	SSC'	SSK	CSA
20	4	10.8	1.7	2.4	2.8	2.9
20	8	5.8	12.5	11.7	8.8	3.1
20	12	11.9	27.8	29.9	10.3	3.7
20	16	19.9	43.6	45.8	14.6	4.2
20	20	29.9	61.6	63.4	14.6	4.6
20	24	43.9	74.6	77.4	16.9	3.8
20	28	54.3	84.0	86.3	22.3	3.5
20	32	65.4	90.7	90.4	25.9	4.8
40	4	5.6	2.3	3.3	2.4	2.9
40	8	4.1	21.6	20.6	10.5	3.7
40	12	14.7	48.8	50.6	12.8	3.6
40	16	30.6	70.5	73.9	16.2	5.2
40	20	52.6	86.4	88.9	19.1	5.4
40	24	69.3	94.7	95.9	20.9	6.0
40	28	78.4	97.5	98.3	24.0	7.3
40	32	90.4	99.2	99.4	28.3	7.8
80	4	1.5	5.4	4.8	3.3	3.0
80	8	3.9	40.9	41.0	8.8	4.4
80	12	21.5	75.3	78.1	14.6	5.7
80	16	52.1	92.8	93.8	17.1	7.4
80	20	77.1	98.9	99.2	22.1	11.4
80	24	92.4	99.5	99.9	26.3	11.4
80	28	98.1	100.0	100.0	26.8	15.8
80	32	99.6	100.0	100.0	32.9	21.4
160	4	1.9	9.9	10.9	4.1	3.0
160	8	5.4	68.0	68.9	4.1	5.4
160	12	35.6	96.4	97.2	14.0	8.2
160	16	80.7	99.8	99.9	20.5	16.7
160	20	96.8	100.0	100.0	24.7	25.0
160	24	99.9	100.0	100.0	27.1	34.1
160	28	100.0	100.0	100.0	32.1	44.6
160	32	100.0	100.0	100.0	33.8	57.7
320	4	1.4	23.6	24.7	5.1	3.0
320	8	5.1	93.3	93.9	1.1	5.9
320	12	57.5	100.0	100.0	15.1	15.7
320	16	98.0	100.0	100.0	21.3	36.6
320	20	99.9	100.0	100.0	26.5	58.1
320	24	100.0	100.0	100.0	30.4	80.7
320	28	100.0	100.0	100.0	35.7	89.4
320	32	100.0	100.0	100.0	42.8	96.2

CCP is the conditional choice probabilities test, SSC and SSC' are the steady-state distribution chi-squared tests, SSK is the steady-state distribution Kolmogorov test and CSA is the conditional state distribution Andrews test.

Table 2. Monte Carlo Results: $\lambda = 0.9$

M	T	CCP	SSC	SSC'	SSK	CSA
20	4	11.3	2.4	2.0	4.5	3.4
20	8	5.4	10.7	5.8	10.0	6.0
20	12	6.7	21.2	15.1	14.7	5.8
20	16	9.0	28.9	22.4	16.8	6.2
20	20	9.8	37.9	29.3	20.5	6.3
20	24	13.7	45.9	40.1	26.5	5.6
20	28	14.5	55.8	47.2	32.8	6.3
20	32	18.3	59.8	52.6	35.5	5.3
40	4	5.8	2.2	1.8	4.1	3.1
40	8	3.9	15.1	10.7	12.9	5.1
40	12	5.9	30.0	22.8	19.8	4.7
40	16	8.5	46.8	36.3	26.2	5.7
40	20	11.1	58.3	46.2	31.1	6.1
40	24	17.3	68.1	58.3	36.2	5.5
40	28	19.8	75.2	65.6	40.4	5.8
40	32	26.6	82.8	73.0	44.4	6.6
80	4	3.3	4.0	2.8	5.1	3.2
80	8	2.5	24.0	15.2	21.0	4.6
80	12	7.3	50.5	35.4	24.8	6.3
80	16	11.0	66.5	54.9	32.8	6.5
80	20	17.6	81.2	70.8	38.3	7.4
80	24	27.1	89.4	81.0	47.4	7.8
80	28	34.8	94.1	88.6	53.7	9.2
80	32	42.9	97.3	93.7	59.7	9.1
160	4	2.4	7.4	4.7	4.6	4.1
160	8	3.5	46.1	30.2	19.7	5.8
160	12	12.0	73.6	57.7	34.4	5.4
160	16	20.3	89.5	80.5	45.4	9.6
160	20	29.9	96.9	91.9	51.1	8.9
160	24	43.1	98.9	97.0	61.6	12.2
160	28	56.5	99.7	99.0	67.0	13.0
160	32	68.1	99.9	99.7	73.0	14.8
320	4	2,0	15.2	8.2	4.5	2.6
320	8	2.1	68.6	51.1	26.1	4.5
320	12	13.3	93.5	84.0	42.6	8.1
320	16	30.8	99.4	97.1	58.2	11.1
320	20	47.1	99.9	99.9	65.3	15.7
320	24	65.5	100.0	100.0	73.1	20.3
320	28	80.9	100.0	100.0	80.0	28.7
320	32	91.0	100.0	100.0	85.7	34.0

CCP is the conditional choice probabilities test, SSC and SSC' are the steady-state distribution chi-squared tests, SSK is the steady-state distribution Kolmogorov test and CSA is the conditional state distribution Andrews test.

Table 3. Monte Carlo Results: $\lambda = 1$

M	T	CCP	SSC	SSC'	SSK	CSA
20	4	13.0	2.1	1.4	2.3	3.3
20	8	6.1	2.4	2.5	3.0	5.0
20	12	5.8	3.9	3.4	4.3	4.2
20	16	5.0	3.3	3.9	4.3	4.5
20	20	3.6	3.0	3.1	4.3	5.3
20	24	4.0	3.9	4.2	4.9	4.1
20	28	3.0	4.0	4.0	4.0	5.6
20	32	3.9	3.9	4.6	4.0	4.7
40	4	6.7	1.2	0.6	3.5	3.7
40	8	4.4	1.8	1.8	4.1	4.9
40	12	3.9	3.3	3.2	4.8	5.6
40	16	3.1	3.8	3.3	4.2	4.7
40	20	3.4	2.8	2.9	4.5	4.4
40	24	4.7	3.0	3.6	4.4	4.7
40	28	3.1	3.6	4.1	5.0	4.0
40	32	5.2	4.4	5.9	4.1	5.1
80	4	4.3	1.4	1.1	4.7	2.7
80	8	3.5	1.7	2.0	3.9	4.1
80	12	4.1	2.9	2.3	5.4	4.6
80	16	3.6	2.8	4.5	5.1	5.2
80	20	5.0	2.8	3.0	4.4	4.8
80	24	4.0	4.1	4.4	5.2	6.0
80	28	3.9	3.5	3.6	5.1	5.3
80	32	4.4	4.0	4.4	4.4	4.6
160	4	4.6	1.2	1.3	4.3	3.8
160	8	3.9	2.9	2.8	5.7	5.2
160	12	4.6	4.1	3.0	6.6	4.0
160	16	3.9	3.0	3.2	5.0	4.2
160	20	4.0	4.3	4.5	3.9	4.4
160	24	3.3	4.2	3.9	6.6	4.9
160	28	3.3	4.3	4.4	5.4	5.7
160	32	4.2	4.0	4.3	4.3	4.2
320	4	4.0	0.9	1.1	4.6	3.5
320	8	2.4	2.6	2.3	3.9	4.0
320	12	3.6	2.5	2.4	6.0	4.8
320	16	4.2	2.6	2.7	4.6	4.1
320	20	4.8	4.1	3.6	4.9	5.0
320	24	4.5	3.4	2.8	5.8	4.8
320	28	3.6	4.2	4.3	4.2	5.1
320	32	5.2	4.3	4.2	4.2	4.3

CCP is the conditional choice probabilities test, SSC and SSC' are the steady-state distribution chi-squared tests, SSK is the steady-state distribution Kolmogorov test and CSA is the conditional state distribution Andrews test.

Table 4. Monte Carlo Results: $\lambda = 0$

M	T	CCP	SSC	SSC'	SSK	CSA
20	4	13.5	1.0	0.6	1.8	2.6
20	8	7.4	2.1	2.2	3.5	3.9
20	12	6.4	1.9	2.1	3.9	5.4
20	16	4.5	3.3	3.2	4.5	5.3
20	20	4.7	3.8	3.2	4.3	5.5
20	24	4.6	3.7	3.9	5.2	4.1
20	28	4.4	4.1	3.8	5.5	4.7
20	32	2.9	4.2	4.1	5.2	5.6
40	4	8.4	0.5	0.3	2.6	3.2
40	8	5.0	2.0	2.0	2.9	3.8
40	12	4.4	3.0	2.3	4.3	4.1
40	16	4.7	3.8	3.3	4.3	5.3
40	20	4.0	3.9	3.8	4.4	4.5
40	24	4.2	3.0	2.6	4.1	5.8
40	28	4.1	3.3	3.4	4.8	4.1
40	32	3.5	4.2	3.6	4.7	4.0
80	4	3.8	0.3	0.1	2.3	2.9
80	8	5.1	2.9	2.6	1.6	5.7
80	12	4.5	3.9	3.4	4.3	5.0
80	16	4.4	3.0	3.1	4.0	4.4
80	20	5.0	4.2	4.2	4.7	5.8
80	24	3.5	3.6	3.5	4.8	5.1
80	28	4.4	3.4	3.1	5.3	5.1
80	32	4.3	4.3	4.7	5.5	4.3
160	4	3.8	0.5	0.5	2.7	3.9
160	8	4.9	3.1	2.6	2.0	4.5
160	12	4.7	2.9	2.5	2.1	4.4
160	16	4.6	3.3	2.9	3.4	4.7
160	20	4.2	3.0	3.0	4.7	5.8
160	24	4.5	3.2	3.1	4.7	4.7
160	28	4.6	3.0	2.8	5.1	5.0
160	32	4.6	3.8	3.6	5.5	4.5
320	4	5.3	0.3	0.3	4.3	3.6
320	8	6.0	2.4	1.9	4.0	5.1
320	12	4.5	3.1	3.0	0.9	4.7
320	16	4.0	3.3	2.8	4.6	4.4
320	20	4.0	3.8	3.9	5.4	4.9
320	24	4.6	2.7	3.0	5.9	5.4
320	28	4.4	2.3	2.5	4.8	4.6
320	32	4.1	3.4	3.5	3.7	4.9

CCP is the conditional choice probabilities test, SSC and SSC' are the steady-state distribution chi-squared tests, SSK is the steady-state distribution Kolmogorov test and CSA is the conditional state distribution Andrews test.

Table 1 considers the case when $\lambda = 0.5$. It shows that as the number of time periods T and/or markets M increases, the tests reject the null with increasing frequency. The two chi-squared test statistics, CCP and SSC, perform better than the two Kolmogorov tests, SSK and CSA, for moderate values of T and/or M. A possible reason may be that the chi-squared tests use the information in all cells and not only the maximal difference across cells. Comparing the two chi-squared tests, we find that the SSC test performs better than the CCP test. A possible reason is that the SSC test uses fewer cells than the CCP test. The SSC test is based on m_s cells while the CCP test is based on $m_s \cdot m_a$ cells. The table also illustrates that for a typical industry application with about 40 markets and 32 time periods the SSC test may perform reasonably.

To further investigate power properties of these tests, Table 2 considers the case when $\lambda = 0.9$. That is, the first equilibrium is played in 90% of M markets. While CCP, SSC, and CSA all have lower power than the case of Table 1, the SSK test has higher power. Overall, the SSK test performs now better than the CCP test. The SSC has still the best performance among all four tests. This result is consistent with our argument in section 3.2.2.[11]

Tables 3 and 4 consider the cases when $\lambda = 1$ and when $\lambda = 0$, respectively. All tests perform reassuringly well leading to a five percent rejection frequency as T and/or M increase.

In all four tables, we also include the SSC' test. The SSC and SSC' tests both perform equally well; the SSC' performs slightly better when $\lambda = 0.5$, while the SSC has higher power when $\lambda = 0.9$. The finding appears natural since the potential problem of the SSC test discussed in the previous section is more relevant when the size of the state space is large. To see this, we construct an additional Monte Carlo example with two distinct Markov chains where the number of states is 16, instead of 4.[12] We perform the same simulation exercise as described above. Table 5 summarizes the power of the SSC and SSC' tests for the cases of $\lambda = 0.5$ and $\lambda = 0.9$. We observe that the SSC' test substantially outperforms the SSC test in almost all pairs of M and T, and both in the cases of $\lambda = 0.5$ and $\lambda = 0.9$. The table confirms our earlier discussion in the previous section. To be complete, Table 6 reports the simulation results for $\lambda = 1$ and $\lambda = 0$ (when the null hypothesis is true). Again, the SSC' is preferred in the sense that the SSC rejects the null too often when it is not supposed to do so for T small.

[11] To further support our argument, we also use the SSK' statistc defined in (6) both when $\lambda = 0.5$ and $\lambda = 0.9$. As we expected, when $\lambda = 0.5$, the SSK' test performs better than the SSK test and its power is about as high as that of the CCP statistic. On the other hand, when $\lambda = 0.9$, the SSK' test underperforms both the SSK test and the CCP test in almost all pairs of M and T.

[12] The row vector of the first transition matrix is (0.03, 0.04, 0.05, 0.06, 0.07, 0.25, 0.095, 0.085, 0.075, 0.065, 0.055, 0.045, 0.035, 0.025, 0.015, 0.005) for every row. That is, regardless of the current state, the game transits to the first state with probability of 0.03, to the second state with probability of 0.04, and so on. The row vector of the second transition matrix is (0.005, 0.015, 0.025, 0.035, 0.045, 0.055, 0.065, 0.075, 0.085, 0.095, 0.25, 0.07, 0.06, 0.05, 0.04, 0.03) for every row. The transition matrix is chosen arbitrarily and is not based on a specific game theoretic model. The transition matrix is aimed at illustrating the performance of tests with 16 states instead of 4.

Table 5. Monte Carlo Results: No. state 16

		$\lambda = 0.5$		$\lambda = 0.9$	
M	T	SSC	SSC'	SSC	SSC'
20	4	19.2	5.3	18.7	3.2
20	8	21.8	25.4	13.1	8.8
20	12	34.1	61.9	17.5	16.5
20	16	60.0	89.0	25.9	28.5
20	20	78.9	97.2	33.6	39.4
20	24	90.1	99.3	41.6	50.4
20	28	96.6	100.0	53.0	63.3
20	32	99.1	100.0	61.8	73.1
40	4	13.8	9.7	13.6	3.5
40	8	26.7	52.7	12.8	12.0
40	12	58.4	90.9	21.1	26.6
40	16	87.7	99.4	36.7	47.1
40	20	97.3	100.0	51.3	65.0
40	24	99.5	100.0	65.6	79.0
40	28	99.9	100.0	75.9	87.5
40	32	100.0	100.0	85.3	94.3
80	4	8.2	18.3	10.0	4.3
80	8	48.7	83.8	18.7	23.0
80	12	88.8	99.8	34.2	48.1
80	16	99.2	100.0	60.4	74.6
80	20	100.0	100.0	76.9	88.7
80	24	100.0	100.0	88.9	95.9
80	28	100.0	100.0	94.1	98.3
80	32	100.0	100.0	98.5	99.3
160	4	16.4	39.8	8.6	8.7
160	8	77.6	99.3	29.8	39.2
160	12	99.4	100.0	59.7	76.2
160	16	100.0	100.0	84.2	94.9
160	20	100.0	100.0	94.7	98.8
160	24	100.0	100.0	99.3	100.0
160	28	100.0	100.0	99.7	100.0
160	32	100.0	100.0	100.0	100.0
320	4	32.0	70.5	11.5	16.0
320	8	97.4	100.0	48.6	68.2
320	12	100.0	100.0	85.8	95.4
320	16	100.0	100.0	98.7	99.9
320	20	100.0	100.0	100.0	100.0
320	24	100.0	100.0	100.0	100.0
320	28	100.0	100.0	100.0	100.0
320	32	100.0	100.0	100.0	100.0

Table 6. Monte Carlo Results: No. State 16

		$\lambda = 1$		$\lambda = 0$	
M	T	SSC	SSC'	SSC	SSC'
20	4	16.6	3.4	7.3	1.5
20	8	6.4	3.5	3.5	1.7
20	12	5.1	3.5	3.1	1.6
20	16	4.9	4.2	4.3	3.2
20	20	4.1	4.0	2.8	2.9
20	24	4.2	3.6	4.0	2.7
20	28	4.8	3.7	6.1	4.3
20	32	4.0	4.0	4.8	4.0
40	4	11.8	2.5	7.7	1.4
40	8	5.2	2.5	3.9	2.5
40	12	4.9	3.3	3.6	2.4
40	16	4.8	5.1	5.1	4.5
40	20	3.0	2.3	3.6	3.0
40	24	3.9	3.6	4.6	4.0
40	28	4.2	4.0	5.9	4.2
40	32	4.7	4.8	4.1	4.3
80	4	10.4	1.4	7.0	1.1
80	8	4.6	2.4	3.3	3.5
80	12	4.6	4.0	4.5	3.9
80	16	3.7	4.5	4.0	4.8
80	20	3.0	3.8	4.3	3.3
80	24	4.4	3.8	4.9	5.3
80	28	5.5	4.9	4.7	4.0
80	32	4.5	3.8	4.5	4.4
160	4	6.6	1.8	5.0	1.8
160	8	4.0	3.0	3.3	3.9
160	12	3.4	2.5	3.4	3.9
160	16	4.0	5.4	4.0	3.4
160	20	4.3	4.3	5.7	4.8
160	24	5.0	3.8	4.4	5.0
160	28	5.5	5.3	5.0	3.2
160	32	4.6	4.6	5.1	4.3
320	4	2.3	1.8	2.5	1.9
320	8	3.8	3.0	3.4	3.6
320	12	3.7	3.7	3.8	3.5
320	16	3.8	4.3	4.1	3.9
320	20	3.7	3.3	4.3	3.9
320	24	4.9	4.5	4.3	4.6
320	28	5.0	5.3	3.5	3.8
320	32	4.6	5.5	5.1	4.9

Our Monte Carlo illustrates that our steady-state distribution chi-squared test SSC performs well for moderate sample sizes. It seems well suited for typical industry applications. The next section applies the test in an empirical application.

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 (2010), Sweeting (2011), Beresteanu, Ellickson, and Misra (2010), and the empirical section of Aguirregabiria and Mira (2007), among others. Panel data frequently contain a large 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 this single equilibrium assumption 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 $\mathbf{s} = (s_1, ..., s_N) \in \mathbf{S}$ be the set of 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 Markovian strategies, firm i's strategy is a mapping from states and private shocks to actions; $\widetilde{\sigma}_i : \mathbf{S} \times \mathcal{E} \to A$, where $\mathcal{E}$ is a domain of ε_i. By integrating over ε_i, we have a mapping from states to the probability distribution over the action space; $\sigma_i : \mathbf{S} \to \Delta A$. Let $V_i(\mathbf{s}; \boldsymbol{\sigma})$ denote the value of firm i if the current state is $\mathbf{s}$ and firms follow strategy $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_N)$ now and in the future. The profile $\boldsymbol{\sigma}^* = (\sigma_1^*, ..., \sigma_N^*)$ is a MPE if for all i, $V_i\left(\mathbf{s}; \sigma_i^*, \boldsymbol{\sigma}_{-i}^*\right) \geq V_i(\mathbf{s}; \sigma_i^{'}, \boldsymbol{\sigma}_{-i}^*)$ for all σ' and $\mathbf{s} \in \mathbf{S}$. 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,

$$\begin{aligned} & \Pr\left(\text{firm } i \text{ enters in period } t \mid s_i = 0, \mathbf{s}\right) \\ = \; & \Phi\left(\psi_1 + \psi_2\left(\sum\nolimits_{j \neq i} s_j^t\right) + \psi_3 \mathbf{1}\left(t > 1990\right)\right), \end{aligned} \tag{8}$$

where $\Phi(\cdot)$ is the cdf of the standard normal. The dummy $\mathbf{1}\left(t > 1990\right)$ 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,

$$\begin{aligned} & \Pr\left(\text{firm } i \text{ exits in period } t \mid ; s_i > 0, \mathbf{s}\right) \\ = \; & \Phi\left(\psi_4 + \psi_5 s_i^t + \psi_6\left(\sum\nolimits_{j \neq i} s_j^t\right) + \psi_7 \mathbf{1}\left(t > 1990\right)\right). \end{aligned} \tag{9}$$

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

$$\ln s_i^{*t} = \lambda_1' \cdot b_1\left(s_i^t\right) + \lambda_2' \cdot b_2\left(\sum\nolimits_{j \neq i} s_j^t\right) + u_i^{*t}, \tag{10}$$

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

$$\underline{s}_i^t = s_i^{*t} - \exp\left(\lambda_3' b_1\left(s_i^t\right) + \lambda_4' b_2\left(\sum_{j\neq i} s_j^t\right) + \underline{u}_i^{bt}\right) \tag{11}$$

and

$$\overline{s}_i^t = s_i^{*t} + \exp\left(\lambda_3' b_1\left(s_i^t\right) + \lambda_4' b_2\left(\sum_{j\neq i} s_j^t\right) + \overline{u}_i^{bt}\right), \tag{12}$$

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 (10), 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 (11) and (12), 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 reduced form 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.

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 1981 to 1999. 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,

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

year of installation) installed in the plant that changed the company name, and 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 7 shows its summary statistics.

Table 7. Summary Statistics of Plant-Level Data

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

We estimate (8) and (9) by a probit regression. Table 8 compares our estimates of these policy functions with those of Ryan (2012). Our entry and exit probit regressions give very close estimates as those of Ryan (2012). Small differences in those estimates may have come from either differences in the optimization routine or the market definition that may slightly differ between Ryan (2012) and our data.[14] We believe that these differences are sufficiently small so they would not distort the test result.

[14] Ryan (2012)'s Java code available at the Econometrica website generates only 22 markets, while his probit 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.

Table 8. Entry and Exit Policies

	Entry Policy Ryan (2012)	Entry Policy Reproduce	Exit Policy Ryan (2012)	Exit Policy Reproduce
Own capacity	–	–	−1.5661	−1.5179
	–	–	(0.268)	(0.262)
Competitors' capacity	0.0448	0.0595	0.0456	0.0633
	(0.0365)	(0.0421)	(0.0173)	(0.0190)
After 1990	−608.9773	−561.1106	−595.2687	−568.6195
	(263.9545)	(296.7534)	(161.6594)	(171.0794)
Constant	−1714.599	−1764.259	−1000.619	−1129.785
	(215.2315)	(263.2785)	(171.2286)	(180.5120)
Log-likelihood	−70.01	−76.27	−227.21	−226.49
Prob > χ^2	0.0177	0.0215		
Sample size	414	414	2233	2233

Note: Standard errors are in parentheses. Both coefficients and standard errors are scaled up by 1,000.

Replicating Ryan (2012)'s investment policy functions is less straightforward. Using OLS for only the sample with $s_i^t \neq s_i^{t+1}$ would generate inconsistent estimates of parameters. In particular, the variances of $\underline{u}_i^{bt}$ and $\overline{u}_i^{bt}$ would be inconsistently estimated as the sample are selected based on the realization of the errors. Our test-statistics require consistent estimates. A possible solution to obtain consistent estimates of the parameters would be to adopt a different inference method, based on maximizing the joint likelihood function including the sample with $s_i^t = s_i^{t+1}$ as well. However, this inference method did not work well in practice as the data are not rich enough to identify all the parameters of interest repeatedly in the Monte Carlo. Instead of doing so, we decided to modify the specification

$$\ln s_i^{*t} = \begin{cases} \beta' x_i^t + u_i^{*t} & \text{if } \gamma' x_i^t + u_i^{bt} > 0 \\ \ln s_i^t & \text{otherwise} \end{cases} \tag{13}$$

where

$$x_i^t = \left(1, s_i^t, \left(s_i^t\right)^2, \sum\nolimits_{j \neq i} s_j^t, \left(\sum\nolimits_{j \neq i} s_j^t\right)^2, s_i^t \left(\sum\nolimits_{j \neq i} s_j^t\right)\right).$$

We assume that $u_i^{*t} \sim N\left(0, \sigma^{*2}\right)$ and $u_i^{bt} \sim N(0,1)$ and estimate $\left(\beta, \gamma, \sigma^{*2}\right)$ by the maximum likelihood. To account for a possible structural change in 1990, instead of using a dummy variable, we allow β and γ to differ between the period before 1990 and the period after 1990.

This simple specification has three advantages. First, the maximization of the likelihood of this specification behaves well numerically, so it suits the case where one needs to repeat the same estimation procedure many times.

While the model described in (10)-(12) is theoretically identified, it is sometimes challenging numerically to obtain the parameters.[15] Second, our specification includes the interaction term between own capacity and competitors' capacity, which may be important in the policy function. We could add the third b-spline for such interaction term in Ryan (2012)'s specification, but with the cost of having many more additional parameters. Third, our specification still captures the fact that the plant does not adjust its capacity level frequently.

Table 9 shows the estimate results for the investment function in (13).

Table 9. Investment Policy

$\beta, \gamma, \sigma^{*2}$	Target Equ. Before 1990	Target Equ. After 1990	Adj. Prob. Before 1990	Adj. Prob. After 1990
Own capacity	223.7894	214.2676	-6.8058	-95.4399
	(7.2386)	(0.0001)	(39.3968)	(55.5917)
Own capacity	-0.0499	-0.0459	0.0036	0.0404
squared	(0.0026)	(0.0050)	(0.0160)	(0.0247)
Competitors' capacity	1.5962	2.7825	10.7577	-20.0321
	(1.3729)	(3.5565)	(6.9503)	(11.4533)
Competitors' capacity	0.0000	-0.0001	-0.0005	0.0010
squared	(0.0001)	(0.0004)	(0.0005)	(0.0011)
Own capacity times	-0.0018	-0.0024	-0.0053	0.0093
Competitors' capacity	(0.0008)	(0.0015)	(0.0040)	(0.0064)
Constant	518983.5	520433.3	-53911.7	53316.4
	(4375.9)	(6673.1)	(23565.5)	(31891.9)
Band σ^{*2}	0.02562		-	
	(0.00057)		-	

Note: Standard errors are in parentheses. Both coefficients and standard errors for β and γ are scaled up by 100,000.

5.3 Testing the Assumption of Unique Equilibrium

Ryan (2012)'s panel data contains states and actions over 19 years for 23 different markets. Since our Monte Carlo study indicates that the steady-state distribution chi-square test performs much better than the conditional choice probability test or conditional state distribution test when the number of markets is small, we apply the steady-state distribution chi-square test (and its variants robust to a large state space) to Ryan (2012)'s data.

[15] Attanasio (2000) places several restrictions on the model to make his estimation manageable. Even with these restrictions, one still needs numerical integration, which is costly. We provide further detail for the difficulty of estimating the full model in the current application upon request.

5.3.1 Baseline Result

Our test proceeds as follows. First, we estimate policy functions (8), (9), and (13). Entry and exit regressions have 3 and 4 parameters, respectively. For the investment equation, we have 25 parameters as shown in Table 9. Thus, we have 32 parameters in total. Let $\mathbf{P}(\widehat{\boldsymbol{\theta}})$ be the Markov chain implied by the estimated parameter vector $\widehat{\boldsymbol{\theta}}$.

Second, we approximate the steady-state distribution $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}))$ by forward simulation. Since Ryan (2012) assumes that the equilibrium played before 1990 is different from the one played after 1990. By Lemma 3, the steady-state distributions are also different between the periods before and after 1990. Thus, we separately simulate the game before and after 1990, each of which is done 10,000 times. Let $\mathbf{Q}^{before}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{before}))$ and $\mathbf{Q}^{after}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{after}))$ denote the steady-state distributions before and after 1990, respectively. To simulate them, as initial conditions, we set the number of plants (firms) around its sample average (6 for the period before 1990 and 5 for the period after 1990) and set the initial capacity level for each plant at its sample average (775 thousand tons for the before-1990 distribution and 831 thousand tons for the after-1990 distribution). We forward simulate the game for 400 years.

One difficulty of simulating a game is that one has to rely on interpolation and extrapolation of choice probabilities, when the game visits a state that is never observed in the data. This is prevalent especially when the state space is large relative to the sample size. Since there is no obvious way of dealing with this problem, we take a conservative route. First, we define

$$\begin{aligned} \overline{s_i^t} &= \max_i \left\{ s_i^t \right\} \\ \underline{s_i^t} &= \min_i \left\{ s_i^t \right\}, \end{aligned}$$

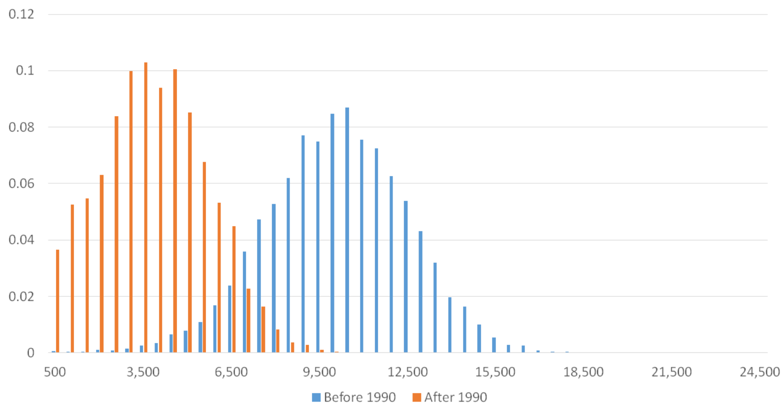
and whenever the equation (13) implies a larger (smaller) value of the target capacity than $\overline{s_i^t}$ ($\underline{s_i^t}$), we replace s_i^{*t} with $\overline{s_i^t}$ ($\underline{s_i^t}$). Second, we use the same procedure for the entry and exit probabilities. Third, we impose an upper bound on the number of plants in one market. In the standard Ericson-Pakes model, the researcher bounds the state space from above, based on primitives of the model and data. In this paper, as a baseline specification, we use the maximum number of plants observed in the data (20 plants in 1980 in Texas) as an upper bound and do not allow any entry if the market already has 20 plants.[16]

Let s_n^* denote the total capacity (sum of incumbents' capacity) at $t = 400$ for the n-th simulation. We obtain $\{s_n^*\}_{n=1}^{10,000}$ for each of $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{before}))$ and $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{after}))$ and discretize the support of these distributions into 75 bins with equal intervals of 500 thousand tons (0-500 thousand tons, 500-1,000 thousand tons, and so on).

Figure 1 depicts the discretized steady-state distributions before and after 1990, respectively.

[16] Later on, we change this number to see the test result is sensitive to this choice.

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



Third, letting $Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{before}))$ and $Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{after}))$ denote the s-th element of $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{before}))$ and $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{after}))$, respectively and

$$\begin{aligned} \widetilde{Q}_s^{j,before} &= \frac{1}{10}\sum_{t=1981}^{1990} \mathbf{1}\left(s^{tj}=s\right) \qquad \text{for } s\in\{1,...,75\} \text{ and } j=1,...,23 \\ \widetilde{Q}_s^{j,after} &= \frac{1}{9}\sum_{t=1991}^{1999} \mathbf{1}\left(s^{tj}=s\right) \qquad \text{for } s\in\{1,...,75\} \text{ and } j=1,...,23, \end{aligned}$$

we calculate several test statistics. We calculate the SSC test-statistic, as it performs best in the Monte Carlo study:

$$SSC = (19)\cdot\sum_{j=1}^{23}\sum_{s\in\{1,\ldots,75\}}\sum_{l\in\{before,after\}} \frac{\left[\widetilde{Q}_s^{j,l} - Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{l}))\right]^2}{Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{l}))}. \tag{14}$$

To deal with the problem that this statistic behaves poorly when the size of the state space is large compared to the sample size (see Section 3), we also calculate the modified SSC' statistic which omits the predicted probability from the denominator:

$$SSC' = (19\cdot 75)\cdot\sum_{j=1}^{23}\sum_{s\in\{1,\ldots,75\}}\sum_{l\in\{before,after\}} \left[\widetilde{Q}_s^{j,l} - Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{l}))\right]^2$$

For comparison purposes we also report the SSK statistic:

$$SSK = \sqrt{19}\max_{j\in\mathbf{M},s\in\{1,\ldots,75\},l\in\{before,after\}}\left|\widetilde{F}_s^{j,l} - F\left(s;Q_s(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{l}))\right)\right|, \tag{15}$$

where F is the cumulative distribution of state. The CSA is not suitable in the current context, as the number of market is small.

Finally, we use a bootstrap to calculate the critical region of the test statistic in the following way. For each bootstrap sample b, we simulate the game for 19 years and 23 markets. To neutralize the effect of arbitrary initial conditions, we run the game for 400 time periods before storing data (that is, we store the data from the 401th to 419th period for each market). For the simulated b-th bootstrap sample, we estimate policy functions using equations (8), (9), and (13). Let $\widehat{\boldsymbol{\theta}}^{b}$ be the set of parameters from b-th bootstrap sample. Then, we follow the same procedure as above to approximate the steady-state distributions $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{b,before}))$ and $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{b,after}))$. Finally, we compute SSC^{b}, SSC'^{b}, and SSK^{b} using (14)-(15) with $\widetilde{Q}_{s}^{j,b,before}$, $\widetilde{Q}_{s}^{j,b,after}$, $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{b,before}))$, and $\mathbf{Q}(\mathbf{P}(\widehat{\boldsymbol{\theta}}^{b,after}))$. We repeat this bootstrap procedure 199 times.

Table 10. Baseline Results

	SSC	SSC'	SSK
Test statistics	6,404.458	26.225	4.345
5% Critical value	77,481.761	23.339	4.356
p value	0.739	0.015	0.704

Note: We have not multiplied the normalizing constants for SSC and SSC' statistics to ease comparisons.

Table 10 summarizes the test results. SSC' implies that we reject the hypothesis that the equilibrium played in the data is unique at the 1.5% significance level.

To investigate where the rejection comes from, Figures 2 and 3 compare the observed distribution with the steady-state distribution for the sample before 1990 and after 1990, respectively. As we can easily see, the steady-state distribution and the observed distribution significantly differ from each other for the period before 1990. On the other hand, Figures 4 and 5 show the observed and steady-state distributions of one typical bootstrap sample, which is generated under the null hypothesis of unique equilibrium. Since the observed distribution is plotted using 414 observations (23 markets over 19 time periods), while the steady-state distribution is plotted using 10,000 simulations, it is natural that these two distributions do not match exactly even under the null hypothesis. However, these figures already deliver a hint about the source of the multiplicity.

Figure 2: Observed and Steady-State Distribution (Before 1990)

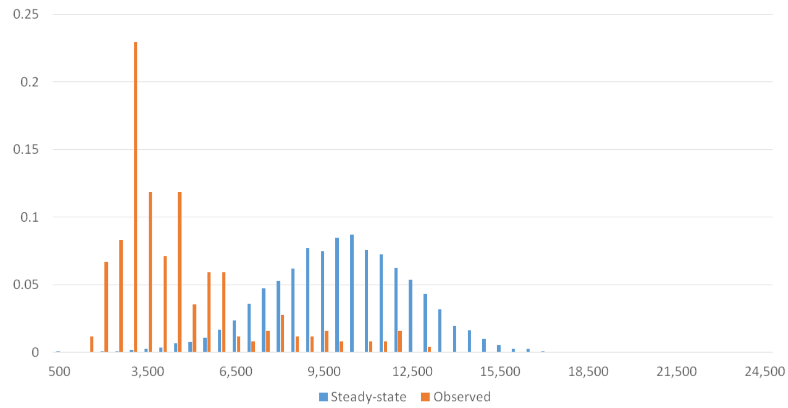


Figure 3: Observed and Steady-State Distribution (After 1990)

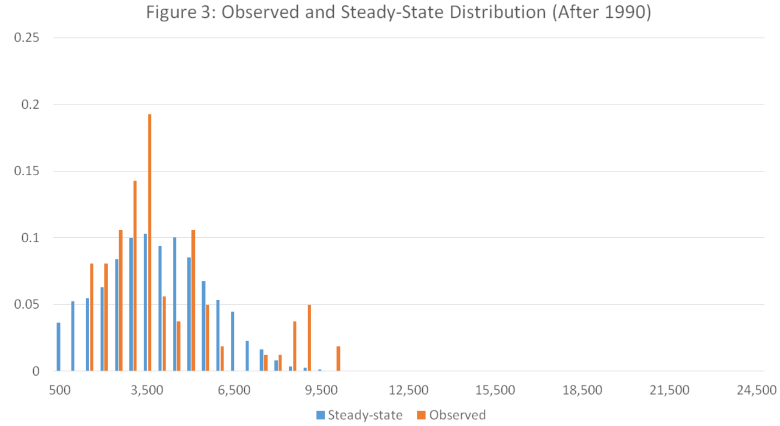
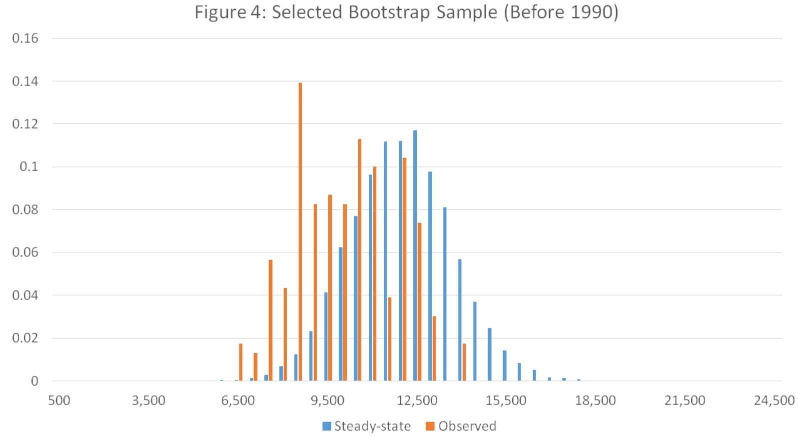
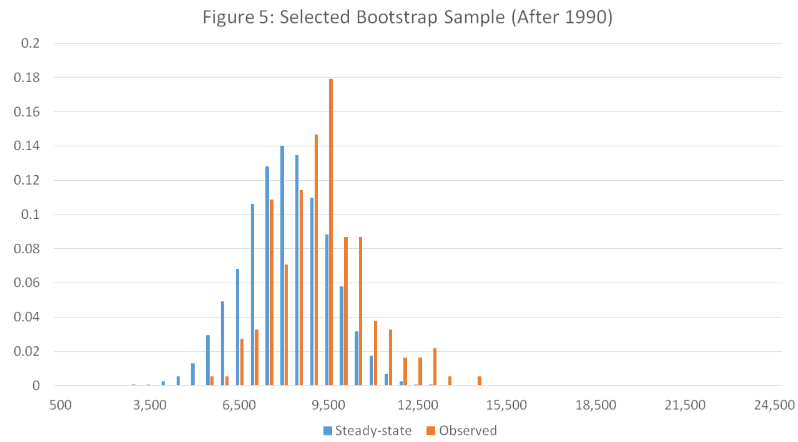


Figure 4: Selected Bootstrap Sample (Before 1990)





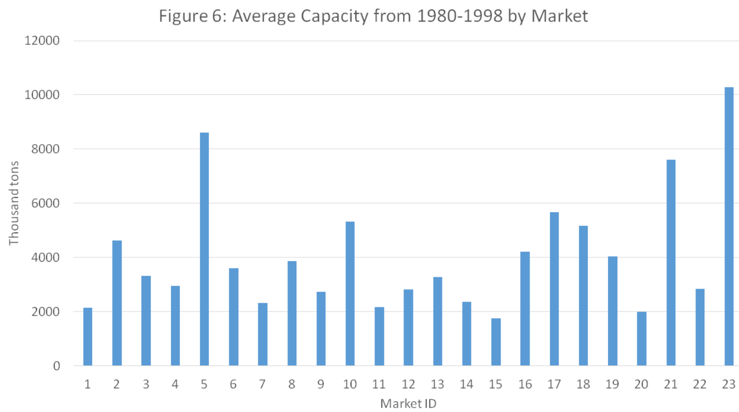
5.3.2 Other Specifications

Before investigating the baseline test result further, we perform several robustness checks.

First, Ryan (2012) assumes that firms had not anticipated the change in the environment before 1990. Thus, when the 1990 Amendments to the Clean Air Act was introduced, the game jumped to the new equilibrium. As can be seen in Figure 1, the steady-state distributions before and after 1990 are significantly different. Thus, it is likely that right after 1990, the game stayed in transient states for awhile. This may distort our test result. To see if this possibly produces the rejection of the null hypothesis, we implement the test using the distribution before 1990 only.

Next, we investigate if the test result is sensitive to discretization. Since the state space is continuous in Ryan (2012)'s model, a crude discretization of the steady state distribution may cause a problem. Differences in the distributions between the continuous state space and discrete state space tend to vanish as the grids become finer. Thus, we expect that differences in the test outcome stemming from discretization would also vanish. As a robustness check, we discretize the support of the steady state distributions into 375 bins with equal intervals of 100 thousand tons (0-100 thousand tons, 100-200 thousand tons, and so on) and implement the test by simulating the game 50,000 times to approximate the steady-state distribution.

Finally, we exclude seemingly outlier markets and apply the test. Figure 6 reports the average (over 19 time periods) total capacity, $\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{N} s_i^{jt}\right)$, by market.



As we can see in the figure, market 5 (Southern California), market 21 (Pennsylvania), and market 23 (Texas) appear to be outliers. As a third robustness check we perform the same test for the subsample without using data from these three markets.

Table 11. Robustness Checks (SSC')

	Before 1990	Fine Grid	Subsample
Test statistics	13.903	15.740	22.440
5% Critical value	10.028	10.382	26.958
p value	0.015	0.015	0.126

Note: We have not multiplied the normalizing constants for the SSC' statistic to ease comparisons.

Table 11 summarizes the results of robustness checks. We focus on the test-statistic SSC'. The first two robustness checks suggest that we still reject the hypothesis of a unique equilibrium.[17] It is interesting that our test delivers a different result once we exclude outlier markets. The p-value is 0.126. Thus, we cannot reject the hypothesis that data is generated by a single equilibrium.[18]

To investigate possible reasons for the difference in test results, in Figure 7 we plot the steady-state distributions from the full sample and subsample for the period before 1990, along with the observed distribution from the full sample. The steady state distribution from the subsample (with the mode around 5,500 thousand tons) is located on the left side of the full-sample distribution (with the mode around 10,500 thousand tons). This reflects the fact that we excluded three outliers (Southern California, Pennsylvania, and Texas) from the sample.

[17] As an additional check, we set the maximum number of plants in a market at 25 and perform the test. The p-value of the test statistic is 0.025, implying that the hypothesis of a unique equilibrium is rejected.

[18] We also perform the test using the subsample and the distribution before 1990 only. The Chi-squared statistic SSC' is $(19 \cdot 75) \cdot 11.962$, while the 95th percentile of the bootstrap distribution is $(19 \cdot 75) \cdot 15.482$. The p-value is 0.080, and hence we cannot reject the null hypothesis with the conventional test size of 5%.

Obviously, the steady-state distribution calculated from the subsample has a much larger overlap with the observed distribution.

The result from the period after 1990 shown in Figure 8 has a similar pattern as the one in Figure 7, although the difference in the steady-state distributions between the full sample and subsample is more subtle. It is worth noting that the steady-state distribution simulated with the subsample has a tiny spike around 20,000 thousand tons. This implies that with a small probability the total market size diverges, although it stops around 20,000 thousand tons because of the limit we placed on the total number of plants in a market. This may be due to the smaller sample size in the subsample, so the estimates of the policy functions are imprecise.

Figure 7: Comparing Two Steady-State Distributions (Before 1990)

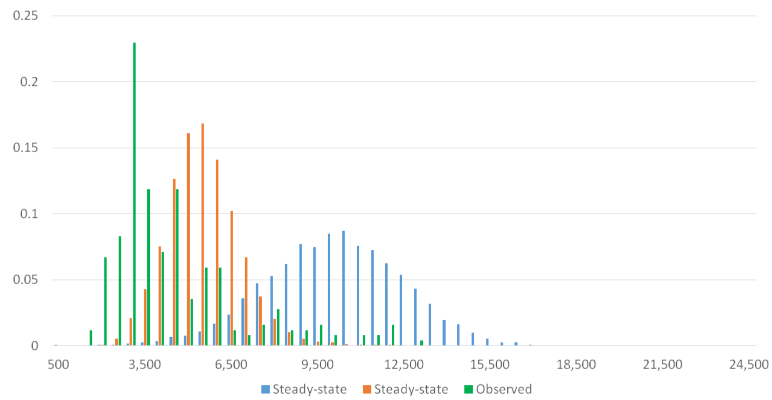
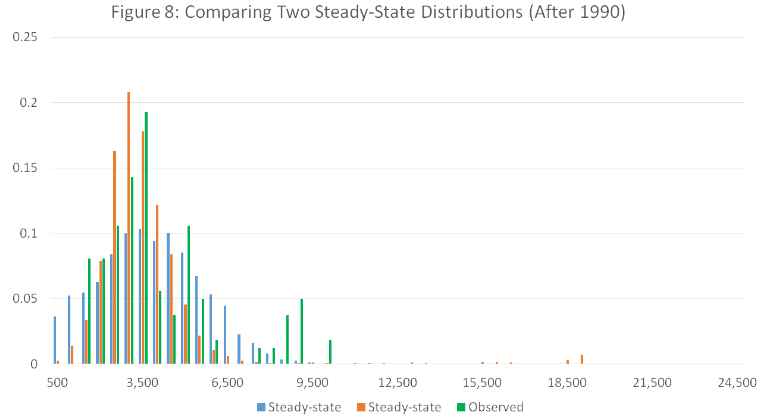


Figure 8: Comparing Two Steady-State Distributions (After 1990)



Thus, the evidence suggests that, although it is not conclusive, the equilibrium played in large markets is different from the equilibrium played in other average-sized markets.

6 Conclusion

This paper proposes several statistical tests for finite state Markov games to examine the null hypothesis that the data are generated by a single equilibrium. The tests are the conditional choice probabilities test, the steady-state distribution test, and the conditional state distribution test. 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 the steady-state distribution test to the empirical application of Ryan (2012) and reject the null hypothesis of single equilibrium in the data.

Three caveats need to be emphasized. First, multiplicity of equilibria and the existence of unobservable market level heterogeneity are observationally equivalent in our framework. Our tests detect both, multiple equilibria and unobservable heterogeneity. However, in case of a rejection, a researcher is left agnostic about causes of the rejection. Our framework gives no guidance for the researcher on a next step. In principle, unobservable heterogeneity and multiplicity of equilibria are different in that the former is payoff-relevant, while the latter is not. We could separate these two sources of mixing at the cost of fully specifying the payoff structure of the game[19]. This is left for future work.

Second, in addition to unobservable heterogeneity, a rejection of the null hypothesis could also point to a misspecification of policy functions. If the policy function is parametric, a rejection of the null could suggest that a more flexible functional form is required. Thus, our tests can also serve as a specification test. On one hand, given that researchers check the goodness of fit of their models in somewhat ad-hoc ways in the literature, it is another contribution of our paper to provide a formal specification test that researchers could use, even in a context where multiple equilibria do not cause problems in estimation (e.g., analysis based on one long time series). On the other hand, there are several specification tests available in the econometrics literature, and it is not clear if our tests work better than these available tests. Formally developing a specification test that best exploits properties of MPE is still left for future work.

Third, 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

[19] In a broader context, while leaving the details of the game unspecified makes our framework general, a researcher could be better off by exploiting some theoretical restriction implied by a specific model, and thereby having higher power. There is a trade-off between generality of the framework and power of the test. How much a researcher wants to specify her model for testing purposes depends on the application in question.

not essential as in practice the data are finite. Researchers may use a finer grid when the data become richer.

References

[1] Aguirregabiria, V. and Mira, P. (2007) "Sequential Simulation-Based Estimation of Dynamic Discrete Games," *Econometrica*, 75 (1), 1–53.

[2] Aguirregabiria, V. and Mira, P. (2012) "Structural Estimation and Counterfactual Experiments in Games when the Data Come from Multiple Equilibria," working paper.

[3] Andrews, D. (1988) "Chi-square Diagnostic Tests for Econometric Models: Introduction and Applications," *Journal of Econometrics*, 37 (1), 135-156.

[4] Andrews, D. (1997) "A Conditional Kolmogorov Test," *Econometrica*, 65 (5), 1097-1128.

[5] Arcidiacono, P., Bayer, P., Blevins, J., and Ellickson, P. (2012) "Estimation of Dynamic Discrete Choice Models in Continuous Time," working paper.

[6] Arcidiacono, P. and Miller, R. (2011), "Conditional Choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79 (6), 1823-1867.

[7] Attanasio, O. (2000) "Consumer Durables and Inertial Behavior: Estimation and Aggregation of (S,s) Rules for Automobile Purchases," *Review of Economic Studies*, 67 (4), 667-696.

[8] Bajari, P., Benkard, L. and Levin, J. (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5), 1331-1370.

[9] Beran, R. and P. W. Millar (1989) "A Stochastic Minimum Distance Test for Multivariate Parametric Models," *Annals of Statistics*, 17 (1), 125-140.

[10] Beresteanu, A., Ellickson, P., and Misra, S. (2010) "The Dynamics of Retail Oligopoly," working paper.

[11] Chiappori, P-A and Salanié, B. (2000) "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108 (1), 56-78.

[12] Collard-Wexler, A. (2010) "Demand Fluctuations in the Ready-Mix Concrete Industry," working paper.

[13] de Paula, A., and Tang, X. (2011) "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80 (1), 143-172.

[14] Doraszelski, U. and Satterthwaite, M. (2010) “Computable Markov-Perfect Industry Dynamics,” *RAND Journal of Economics*, 41 (2), 215–243.

[15] Dunne, T., Klimek, S., Roberts, M., and Xu, D. (2011) “Entry, Exit and the Determinants of Market Structure,” working paper.

[16] Ericson, R. and Pakes, A. (1995) “Markov Perfect Industry Dynamics: A Framework for Empirical Work,” *Review of Economic Studies*, 62 (1), 53–82.

[17] Fan, Y. and Xiao, M. (2012) “Competition and Subsidies in the Deregulated U.S. Local Telephone Industry,” working paper.

[18] Gallager, R. G. (1996), *Discrete Stochastic Processes*, Kluwer, Boston.

[19] Hotz, J. and Miller, R. (1993) “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60 (3), 497-529.

[20] Jeziorski, P. (2012) “Estimation of Cost Efficiencies from Mergers: Application to U.S. Radio,” working paper.

[21] Jofre-Bonet, M. and Pesendorfer, M. (2003) “Estimation of a Dynamic Auction Game,” *Econometrica*, 71 (5), 1443-1489.

[22] Kasahara, H. and Shimotsu, K. (2009) “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica*, 77 (1), 135-175.

[23] Kasahara, H. and Shimotsu, K. (2012) “Sequential Estimation of Structural Models with a Fixed Point Constraint,” *Econometrica*, 80 (5), 2303–2319.

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

[25] Lin, H. (2011) “Quality Choice and Market Structure: A Dynamic Analysis of Nursing Home Oligopolies,” working paper.

[26] Linton, O. and S.T. Srisuma (2012) “Semiparametric Estimation of Markov Decision Processes with Continuous State Space: Discrete Control,” *Journal of Econometrics*, 166 (2), 320-341.

[27] Maican, F. and Orth, M. (2012) “Store Dynamics, Differentiation and Determinants of Market Structure,” working paper.

[28] Maskin, E. and Tirole, J. (1988) “A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition With Large Fixed Costs,” *Econometrica*, 56 (3), 549-569.

[29] McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, Academic Press: New York.

[30] Minamihashi, N. (2012) "Natural Monopoly and Distorted Competition: Evidence from Unbundling Fiber-Optic Networks," working paper.

[31] Navarro, S. and Takahashi, Y. (2012) "A Semiparametric Test of Agent's Information Sets for Games of Incomplete Information," working paper.

[32] Nishiwaki (2010) "Horizontal Mergers and Divestment Dynamics in a Sunset Industry," working paper.

[33] Pakes, A., Ostrovsky, M., and Berry, S. (2007) "Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)," *RAND Journal of Economics*, 38 (2), 373-399.

[34] Pesendorfer, M. and Schmidt-Dengler, P. (2008) "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75 (3), 901–928.

[35] Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

[36] Ryan, S. (2012) "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80(3), 1019-1061.

[37] Sanches, F. and Silva-Junior, D. (2012) "Public Banks Improve Private Banks Performance: Evidence from a Dynamic Structural Model," working paper.

[38] Snider, C. (2009) "Predatory Incentives and Predation Policy: The American Airlines Case," working paper.

[39] Suzuki, J. (2012) "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," forthcoming in *International Economic Review*.

[40] Sweeting, A. (2011) "Dynamic Product Positioning in Differentiated Product Industries: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," working paper.