

# Estimation of (static or dynamic) games under equilibrium multiplicity

Taisuke Otsu\*, Martin Pesendorfer<sup>†</sup>, Yuya Sasaki<sup>‡</sup> and Yuya Takahashi<sup>§</sup>

January 14, 2020

## Abstract

We propose a multiplicity-robust estimation method for (static or dynamic) games. The method allows for distinct behaviors and strategies across markets by treating market specific behaviors as correlated latent variables, with their conditional probability measure treated as an infinite-dimensional nuisance parameter. Instead of solving the intermediate problem which requires optimization over the infinite dimensional set, we consider the equivalent dual problem which entails optimization over only a finite-dimensional Euclidean space. This property allows for a practically feasible characterization of the identified region for the structural parameters. We apply the estimation method to newspaper market previously studied in Gentzkow et al. (2014) to characterize the identified region of marginal costs.

## 1 Introduction

This paper explores inference under equilibrium multiplicity. A multiplicity-robust estimation method is proposed and illustrated for a class of economic models that can be written as a fixed point problem. The class includes static games and dynamic Markov games. Actions can be discrete or continuous. Unlike the traditional two-step methods, researchers are not required to assume a single equilibrium is played in the cross-section. Equilibrium multiplicity and unobserved market heterogeneity are accommodated. Our method allows a researcher to identify and estimate the structural parameters of interest even if data of long-time series are not available.

To the best of our knowledge, the methods developed are the first to provide partial identification results for dynamic Markov games under equilibrium multiplicity. In the literature, partial identification of games with multiple equilibria has been limited to static contexts, e.g.

---

\*Taisuke Otsu: Department of Economics, London School of Economics and Political Science, Houghton Street, London UK WC2A 2AE. Email: t.otsu@lse.ac.uk.

<sup>†</sup>Martin Pesendorfer: Department of Economics, London School of Economics and Political Science, Houghton Street, London UK WC2A 2AE. Email: m.pesendorfer@lse.ac.uk.

<sup>‡</sup>Yuya Sasaki: Department of Economics, 415 Calhoun Hall, Vanderbilt University, Nashville, TN 37240, U.S.A.. Email: yuya.sasaki@vanderbilt.edu.

<sup>§</sup>Yuya Takahashi: Department of Economics The University of Washington. Savery Hall 329, Seattle, WA 98195. Email: ytakahas@uw.edu.

Ciliberto and Tamer (2009) and Molinari (2010). Our method provides a practically attractive computational procedure to estimate the identified region which applies to dynamic games and also static games.

A number of papers proposed two-step estimation methods for games under varying assumptions including Guerre et al. (2000) and Hortacsu and Kastl (2012) for static games and Jofre-Bonet and Pesendorfer (2003), Aguirregabiria and Mira (2007), Bajari et al. (2007), Pesendorfer and Schmidt-Dengler (2008), and Arcidiacono and Miller (2011) for dynamic games. These methods led to numerous empirical papers that analyze interactions among multiple players. Two-step methods assume that the first-stage estimators of players' policies are consistent. In typical microeconomic applications where a long time-series data set may not be available, researchers often pool data from different markets to construct first-stage estimates. However, unless the data are generated from a single (or identical) equilibrium in every market (or auction), the first-stage estimate of players' policies is not consistent for true policies, and hence the second-stage estimate of structural parameters generally fails to be consistent. The assumption of single equilibrium is indeed restrictive as multiplicity of equilibria is an innate feature of games, but many research papers effectively make this restrictive assumption for empirical analysis. Otsu et al. (2016) proposed several tests to examine the null hypothesis of a single equilibrium. They find that the assumption of a single equilibrium is not satisfied in the empirical application of Ryan (2012).

Our proposed estimator for identified regions allows for distinct behaviors and strategies across markets. We treat market specific behaviors as correlated latent variables, with their conditional probability measure treated as an infinite-dimensional nuisance parameter. Based on Schennach (2014), instead of solving the intermediate problem which requires optimization over the infinite-dimensional set, we consider the equivalent dual problem which entails optimization over only a finite-dimensional Euclidean space. This property allows for a practically feasible characterization of the identified region for the structural parameters.

Our application studies dynamic competition in the US newspaper market, which has been rigorously studied in Gentzkow et al. (2014). Equilibrium multiplicity is likely to arise due to a number of features inherent to this market. Most notably, the market is two-sided with revenues coming potentially from two sides: readers and advertisers. Newspapers strategically price their product in these two dimensions. We conduct statistical tests for equilibrium multiplicity and find that multiplicity is indeed a concern.

We formulate a dynamic Markovian game which shares key features of the static model of Gentzkow et al. (2014) but additionally allows for (i) entry and exit and (ii) dynamic forward looking behavior by newspapers. We restrict attention to a parametric class of strategies, which simplifies the computational burden substantially, and use the estimation method to estimate the identified region of structural parameters.

This paper is organized as follows. Section 2 introduces notations and definitions for the games considered. Section 3 illustrates the multiplicity concerns in examples. Section 4 provides the econometric results. Section 5 formulates a dynamic game in the newspaper industry and describes our estimation approach. Section 6 discusses our estimates, and Section 7 concludes.

## 2 Framework

This section describes our framework. We consider economic situations that can be characterized by a fixed point problem in an unobservable variable  $\mathbf{p}$ :

$$\mathbf{p} \in \Psi(\mathbf{p}; \theta), \quad (1)$$

where  $\Psi$  is a mapping from a bounded set  $\mathcal{P} \subseteq \mathbb{R}^l$  to  $\mathbb{R}^l$  and parameterized by a vector of parameters  $\theta \in \mathbb{R}^q$ .

The fixed point restriction (1) arises naturally in a number of economic settings. A leading example is Nash equilibria of game theoretic models. In game theoretic models,  $\mathcal{P}$  can be interpreted as the strategy set and the mapping  $\Psi$  is the best response correspondence. For example, the  $i$ -th component of the best response mapping  $\Psi$  in a static complete information  $N$ -player game with payoff functions  $(U_i)_{i=1}^N$  is defined as  $\arg \max_x U_i(x, \mathbf{p}_{-i}; \theta)$  where  $\mathbf{p}_{-i}$  denotes the strategies of players other than player  $i$ . A fixed point of this mapping  $\Psi$  is a Nash equilibrium. Our application considers a dynamic pricing game in the newspaper market in which the profit function is the discounted sum of future newspaper profits. The data consist of newspaper subscription price, advertising price and the newspaper circulation information.

The strategies  $\mathbf{p}$  are usually not observed in the data. Instead, the econometrician observes variables  $(Z_j)_{j=1}^M$  for a cross section of  $M$  markets. The observable  $Z_j = (\mathbf{a}_j^t, \mathbf{s}_j^t)_{t=1}^T$  consist of a sequence of action profiles  $\mathbf{a}_j^t \in \mathbf{A}$ , state and market variables  $\mathbf{s}_j^t \in \mathbf{S}$  for periods  $t = 1, \dots, T$  generated by the equilibrium play of some underlying economic model. We assume that the same model applies to distinct markets but allow for observed and unobserved heterogeneity across markets which we shall illustrate below. The action profiles  $\mathbf{a}$  can be viewed as the 'outcome' of the unknown strategy profiles  $\mathbf{p}$ . A key difficulty in the estimation arises as there can be multiple equilibria that could rationalize observed actions. In terms of our equation (1), this corresponds to multiple solutions to the fixed point problem. We do not constrain the set of solutions to the fixed point problem. In particular, we allow that  $\mathbf{p}_j \neq \mathbf{p}_{j'}$  for  $j \neq j'$ . For a large- $T$  setting the strategies  $\mathbf{p}_j$  can be estimated with the root- $T$  convergence rate for each market  $j$ , and these preliminary estimates can be plugged in to estimate  $\theta$  in the second step. This procedure parallels the large- $T$  nonlinear panel methods, where individual fixed effects are estimated with the root- $T$  convergence rate and asymptotic biases are corrected by analytic expansions (e.g., Hahn and Newey (2004)).

We focus on the asymptotic setting where  $T$  is fixed at a small number but  $M \rightarrow \infty$ . We denote this setting as short panels of a large number of markets. Because  $T$  does not grow in this asymptotic setting, the market-specific choice probabilities  $\mathbf{p}_j$  cannot be consistently estimated by a frequency estimator  $\hat{\mathbf{p}}_j$  in the first step, and hence the two-step methods are not guaranteed to work. Furthermore, we relax point identification as such, and pursue set identification of the structural parameters  $\theta$  and propose to estimate the identified region by the feasible computational method based on the approach proposed by Schennach (2014).

Let  $\pi$  denote the probability measure of the observable data  $Z$ . Let  $\lambda$  denote the conditional probability measure of the latent variables  $\mathbf{p}$  conditional on the observables  $Z$ . The joint distribution  $\lambda \times \pi$  is defined as the product of the conditional probability  $\lambda$  and the marginal  $\pi$ . We assume that  $(\mathbf{p}_j, Z_j)_{j=1}^M$  is an iid sequence, although we do not specify or estimate the

conditional distribution  $\lambda$  of the latent variable. This implies the observables  $(Z_j)_{j=1}^M$  are an iid sample from a cross-section of  $M$  markets. Our sampling assumption follows Assumption 1 in Tamer (2003) and Ciliberto and Tamer (2009) and requires the existence of an equilibrium selection function. We leave the equilibrium selection function unspecified and examine the question of what can be learned when researchers remain agnostic about this selection function. One could condition this function on the various equilibria as a function of observables in which the statistical model becomes one of a mixture. Without assumptions on equilibrium selection, the model is partially identified as is shown in Tamer (2003). Bjorn and Vuong (1984) and Bajari et al. (2010) adopt a related approach in which the equilibrium selection rule is parameterized and estimated.

Let  $\mathbf{N}$  denote the set of  $N$  players,  $\mathbf{S}$  denote the state space with the cardinality  $m_s$ , and  $\mathbf{A}_i$  denote the action space of player  $i \in \mathbf{N}$  with the cardinality  $K + 1$ . The equilibrium condition (1) is a fixed point problem which can be equivalently formulated as a problem of finding zeros of the following one-dimensional equation:

$$\sum_{i \in \mathbf{N}} \sum_{\mathbf{s} \in \mathbf{S}} \sum_{a_i \in \mathbf{A}_i} \{\mathbf{p}^{(a_i, \mathbf{s})} - \Psi^{(a_i, \mathbf{s})}(\mathbf{p}; \theta)\}^2 = 0, \quad (2)$$

where  $\mathbf{p}^{(a_i, \mathbf{s})}$  means the element of the vector  $\mathbf{p}$  corresponding to action-state profile  $(a_i, \mathbf{s})$ , and  $\Psi^{(a_i, \mathbf{s})}$  is the corresponding element of  $\Psi$ . We combine the aggregated equilibrium restrictions with the restrictions placed on the strategies by the empirical data. Let  $d = m_s \cdot N \cdot K + 1$ . In particular, we consider the following  $d$  dimensional function

$$g(\mathbf{p}, Z; \theta) = \left[ \begin{array}{c} \sum_{i \in \mathbf{N}} \sum_{\mathbf{s} \in \mathbf{S}} \sum_{a_i \in \mathbf{A}_i} \{\mathbf{p}^{(a_i, \mathbf{s})} - \Psi^{(a_i, \mathbf{s})}(\mathbf{p}; \theta)\}^2 \\ [f^{(\mathbf{s})} \mathbf{p}^{(a_i, \mathbf{s})} - f^{(a_i, \mathbf{s})}]_{i \in \mathbf{N}, \mathbf{s} \in \mathbf{S}, a_i \in \mathbf{A}_i} \end{array} \right], \quad (3)$$

consisting of the equilibrium condition aggregated into one equation and the  $d - 1$  restrictions placed on the choice probabilities by the empirical data. Here  $f^{(a_i, \mathbf{s})} = \sum_{t=1}^T 1\{a_i^t = a_i, \mathbf{s}^t = \mathbf{s}\}$  is the frequency of action-state profile  $(a_i, \mathbf{s})$ , and  $f^{(\mathbf{s})} = \sum_{t=1}^T 1\{\mathbf{s}^t = \mathbf{s}\}$  is the frequency of state  $\mathbf{s}$ . Note that the function  $g$  takes on values in  $\mathbb{R}^d$ .

Following Schennach (2014), we define the identified region for the structural parameters  $\theta$  as

$$\Theta_0 = \left\{ \theta \in \Theta : \inf_{\lambda \in \Lambda} |E_{\lambda \times \pi}[g(\mathbf{p}, Z; \theta)]| = 0 \right\}, \quad (4)$$

where  $\Lambda$  denotes the set of all regular conditional probability measures supported on the set of choice probabilities. The expectation  $E_{\lambda \times \pi}[\cdot]$  is infeasible to compute, because the true distribution  $\lambda$  of the equilibrium choice probabilities is unobserved. Instead, we assume that the market-specific choice probabilities  $\mathbf{p}$  are correlated latent variables, with their conditional probability measure  $\lambda$  treated as an infinite-dimensional nuisance parameter. In the empirical method that we plan to pursue as described below, we will make an inference for the structural parameters  $\theta$  without having to specify this measure  $\lambda$ .

Indeed, Schennach (2014) shows that

$$\theta \in \Theta_0 \text{ if and only if } \inf_{\gamma \in \mathbb{R}^d} |E_{\pi}[\tilde{g}(Z, \theta, \gamma)]| = 0, \quad (5)$$

for some  $\pi$ , where

$$\tilde{g}(Z, \theta, \gamma) = \frac{\int g(\mathbf{p}, Z; \theta) \cdot \exp(\gamma' g(\mathbf{p}, Z; \theta)) dF(\mathbf{p})}{\int \exp(\gamma' g(\mathbf{p}, Z; \theta)) dF(\mathbf{p})},$$

with a user-specified probability measure  $F$ . We let  $F$  be the probability measure of the random variable uniformly distributed on the set of feasible choice probabilities. The following proposition establishes that the conditions in Schennach (2014) apply under the continuity of  $\Psi(\cdot; \theta)$  in  $\theta \in \Theta$ . A proof is provided in Appendix B.

**Proposition S.** *Suppose that  $(\mathbf{p}_j, Z_j)_{j=1}^M$  is iid with support  $\mathcal{P} = [0, 1]^l$  for  $\mathbf{p}_j$  and some finite set  $\mathcal{Z}$  for  $Z_j$ ,  $F$  is the uniform distribution on  $\mathcal{P}$ , and  $\Psi(\cdot; \theta)$  is continuous at each  $\theta \in \Theta$ . Then the equivalence in (5) holds true.*

Note that, while the intermediate problem (4) requires optimization over the infinite-dimensional set  $\Lambda$ , the equivalent problem (5) entails optimization over only a finite-dimensional Euclidean space. This property allows for a practically feasible characterization of the identified region  $\Theta_0$  for the structural parameters.

Our exposition focuses on unobservable strategies as nuisance parameters, but readily accommodates unobserved payoff elements as nuisance parameters as well. Suppose the period payoff has additionally an additive payoff shock  $v_{ij}$ , which is time-invariant and market-specific. The equilibrium equation system (1) becomes  $\mathbf{p} \in \Psi(\mathbf{p}, \mathbf{v}; \theta)$  and the function  $g$  is redefined accordingly. With these modifications in place, the structural parameters can be partially identified in the presence of unobserved heterogeneity as before. The model in Tamer (2003) specifies a payoff function for player  $i$  of the form  $u_i(x, \mathbf{p}_{-i}, v_i; \theta)$  where the term  $v_i$  enters additively, and is known to market participants but not to the econometrician. Example 2 in Tamer (2003) assumes  $v_i$  is iid across markets and uniformly distributed on the unit interval. Interestingly, in our framework the true distribution of  $v_{ij}$  need not be known, which is an important advantage of our method. Furthermore, the distribution can be a discrete or continuous. This is a notable difference to the approach of Arcidiacono and Miller (2011) which requires the researcher to use a multinomial distribution for unobserved heterogeneity.

Next, we illustrate a simple game theoretic example that readily fits to our framework in (1) and has received considerable attention in the empirical literature. It is a static simultaneous move game. We shall consider the Cournot model in which output is a choice variable as it is commonly used in empirical works. Later-on we shall consider a dynamic game of the newspaper market in Markovian strategies which is also popular in empirical studies. Our dynamic model includes price as a continuous choice variable.

### 3 Examples: Equilibrium multiplicity in games

Suppose there are  $N$  players. Each player  $i$  chooses  $p_i$  to maximize payoff  $U_i(p_1, \dots, p_i, \dots, p_N)$ . The best response function  $p_i = \arg \max_x U_i(p_1, \dots, x, \dots, p_N)$  gives rise to the above fixed point constraint.

In the Cournot case, firm  $i$  produces  $p_i$  units of output with marginal cost  $c_i(p_i)$  and fixed cost  $C$ . Demand is given by the inverse demand function  $D(\sum_j p_j)$ . Profit of firm  $i$  is given

by  $\pi_i(p_i, \Sigma_{j \neq i} p_j) = [D(\Sigma_j p_j) - c_i(p_i)]p_i - C$ . Necessary and sufficient conditions for a unique equilibrium are given in Gaudet and Salant (1991). Multiplicity of equilibria can arise in the Cournot model when one or more of these conditions are violated. For example, when there are externalities in demand or in costs. A cost externalities may arise due to learning by doing with full diffusion, see Fudenberg and Tirole (1983). To illustrate this case briefly, assume marginal cost is piece-wise constant in which the marginal cost depends additionally (and fully) on competitors' output:  $c_i(p_i, \Sigma_{j \neq i} p_j) = c_1$ , if  $p_i + \Sigma_{j \neq i} p_j < \bar{P}$ , and  $c_i(p_i, \Sigma_{j \neq i} p_j) = c_2$ , if  $p_i + \Sigma_{j \neq i} p_j \geq \bar{P}$ . For  $c_1 > c_2$  two types of Cournot equilibria may exist: (i) a high marginal cost equilibrium in which total industry output is below the learning threshold  $\bar{P}$ , and (ii) a low marginal cost equilibrium in which total industry output is above the learning threshold  $\bar{P}$ . As before, equilibria can be found using the fixed point equation defined by the best response mapping.

Multiplicity of equilibria can arise because of firm heterogeneity and fixed costs. The identity of active firms is then not determined. For example, marginal costs may be such that in one equilibrium firm 1 is the sole producer, but another equilibrium may have firms 2 and 3 be the sole producers. Observe also that due to firm heterogeneity in marginal costs, the total output need not be the same across equilibria.

A static model of the newspaper market is studied in Gentzkow et al. (2014). We take their model and extend it to a dynamic duopoly game of the news paper market explained in more detail in Section 5. Importantly, we additionally allow for entry and exit, and dynamic forward-looking behavior by newspapers. The newspaper market has a rich potential for equilibrium multiplicity for a number of reasons. First, the newspaper market is two-sided, with revenues coming from subscribers or advertisers. Newspapers can thus position their product on advertising revenues, on subscriber revenues or something in between.

Second, entry and exit play a key role in the newspaper market which can give rise to multiplicity as discussed in Tamer (2003).

Third, newspaper readership is loyal to their favorite newspaper, with readers being reluctant to switch their already familiar newspaper. Over time a race to become the leading newspaper in subscription numbers can emerge which resembles a war of attrition game with well known multiplicity concerns, see Hendricks et al. (1988). The benefits of being the high circulation newspaper can be reaped by the winner.

Finally, multiplicity may also arise from unobserved market features. For example, the proportion of people interested in political news may differ across markets giving rise to newspapers making distinct pricing and advertising decisions across markets.

## 4 Econometrics

In this section we discuss inference methods for the identified set  $\Theta_0$  using the characterization in (5).

First of all, as described in Appendix F of Schennach (2014), conventional set inference methods (e.g., Chernozhukov et al. (2007)) may be applied based on a criterion function constructed from (5). For example, based on an iid sample  $(Z_j)_{j=1}^M$ , the GMM-type criterion

can be defined as

$$Q(\theta) = \sup_{\gamma \in \mathbb{R}^d} -\frac{1}{2} \bar{g}(\theta, \gamma)' \hat{V}(\theta, \gamma)^{-1} \bar{g}(\theta, \gamma), \quad (6)$$

where  $\bar{g}(\theta, \gamma) = M^{-1} \sum_{j=1}^M \tilde{g}(Z_j, \theta, \gamma)$  and  $\hat{V}(\theta, \gamma)$  is some estimator of  $Var(\tilde{g}(Z, \theta, \gamma))$ . From Appendix F.4 of Schennach (2014), a simple but conservative confidence set is obtained as

$$\hat{\mathcal{C}} = \{\theta \in \Theta : -2MQ(\theta) \leq \chi_{d, \alpha}^2\}, \quad (7)$$

where  $\chi_{d, \alpha}^2$  is the  $(1 - \alpha)$ -th quantile of the  $\chi^2$  distribution with degree of freedom  $d$ . Theorem F.1 of Schennach (2014) showed its asymptotic validity in the sense that  $\lim_{n \rightarrow \infty} \Pr\{\theta \notin \hat{\mathcal{C}}\} \leq \alpha$  for all  $\theta \in \Theta_0$ .

A drawback of the confidence set  $\hat{\mathcal{C}}$  is: when the dimension  $d$  of the moment function  $\tilde{g}$  (or Lagrange multiplier  $\gamma$ ) is high, the critical value tends to be large. Recall that  $d = m_s \cdot N \cdot K + 1$ . Thus, if the number of states, players, or actions is large, the confidence set may be too large to obtain a meaningful conclusion.

If  $Q(\theta)$  is a conventional (continuous updating) GMM criterion function, then one may employ Kleibergen (2005) statistic that takes a quadratic form of the gradient  $\partial Q(\theta)/\partial \theta$ . However, in our setup the criterion  $Q(\theta)$  involves the supremum over  $\gamma \in \mathbb{R}^d$  which may be achieved at  $|\gamma| \rightarrow \infty$ . As discussed in Schennach (2014), p. 355, the solution ‘‘at infinity’’ ( $|\gamma| \rightarrow \infty$ ) often happens when the infimum in (4) is achieved by some discrete distribution.

Here we propose an adapted version of Kleibergen (2005) statistic to the moment function  $\tilde{g}$  defined by the entropic latent variable integration. For each  $\theta$ , let  $\tilde{\gamma}(\theta)$  be a regularized estimator of the solution in (6) that satisfies Assumption K (ii) in Appendix B. Then our test statistic is written as

$$\begin{aligned} K(\theta) = & M \bar{g}(\theta, \tilde{\gamma}(\theta))' \hat{V}(\theta, \tilde{\gamma}(\theta))^{-1} \hat{D}(\theta) \left[ \hat{D}(\theta)' \hat{V}(\theta, \tilde{\gamma}(\theta))^{-1} \hat{D}(\theta) \right]^{-1} \\ & \times \hat{D}(\theta)' \hat{V}(\theta, \tilde{\gamma}(\theta))^{-1} \bar{g}(\theta, \tilde{\gamma}(\theta)), \end{aligned} \quad (8)$$

where  $\hat{D}(\theta)$  is a  $d \times q$  matrix with the  $l$ -th column

$$\frac{\partial \bar{g}(\theta, \tilde{\gamma}(\theta))}{\partial \theta_l} - \hat{G}_l(\theta, \tilde{\gamma}(\theta))' \hat{V}(\theta, \tilde{\gamma}(\theta))^{-1} \bar{g}(\theta, \tilde{\gamma}(\theta)),$$

and  $\hat{G}_l(\theta, \tilde{\gamma}(\theta)) = M^{-1} \sum_{j=1}^M \tilde{g}(Z_j, \theta, \tilde{\gamma}(\theta)) \partial \tilde{g}(Z_j, \theta, \tilde{\gamma}(\theta)) / \partial \theta_l$ . The limiting distribution of this statistic is obtained as follows.

**Proposition K.** *Suppose Assumption K in Appendix B and the assumptions of Proposition S hold. Then  $K(\theta_0) \xrightarrow{d} \chi_{q, \alpha}^2$  for each  $\theta_0 \in \Theta_0$ .*

Based on this statistic, the confidence set is obtained as

$$\tilde{\mathcal{C}} = \{\theta \in \Theta : K(\theta) \leq \chi_{q, \alpha}^2\}, \quad (9)$$

where  $\chi_{q, \alpha}^2$  is the  $(1 - \alpha)$ -th quantile of the  $\chi^2$  distribution with degree of freedom  $q$ . Note that the critical value  $\chi_{q, \alpha}^2$  depends only on the dimension of structural parameters  $\theta$  and is robust to the dimension of the moment function  $\tilde{g}$ .

## 5 Newspaper market

This section describes the newspaper market, tests for equilibrium multiplicity, and presents an estimable dynamic model.

The US newspaper market has been studied in the influential work by Gentzkow et al. (2014). We obtained the publicly available data on circulation, annual subscription price and advertising price for local US newspaper markets at four-year intervals between 1869 and 2004.<sup>1</sup> We select all 538 duopoly markets that are markets with at most two active local newspapers in any given year.

### 5.1 Descriptive sample statistics

Table 1 provides summary statistics for selected variables. All dollar values are measured in constant 1984 dollars. On average 0.84 newspapers are active in a given year. No newspaper exists on 31% of time periods, one newspaper is circulated on 54% of time periods and two newspapers exist on 15% of time periods. The average newspaper has a circulation of about 10,000. The average annual subscription price equals \$50 per year. The average advertising price equals \$25 which is measured per line or per inch. Accounting for circulation numbers, the advertising price per circulated newspaper is small and equals \$0.0025. Note that the subscription price is missing on 15% and the advertising price is missing on 47% of occasions.

Table 1: Descriptive Summary Statistics

Variable	No of Obs.	Mean	Std Dev	Min	Max
Number of Newspapers	18,830	0.84	0.66	0	2
Circulation	14,747	10,030	22,395	100	762,639
Annual Subscription Price	12,575	49.38	16.99	2.39	269.06
Advertising Price	7,664	24.58	142.17	0.27	3,027.25

Entry and exit are common in this industry. There are in total 1,029 occasions of newspaper entries and 1,448 occasions of newspaper exits. When there is no active newspaper in the market, there is about a 5% chance that an entry occurs during a time period. The exit rate when there is only one active newspaper is of the same magnitude with 5%. The entry and exit rates are higher when there is one respectively two newspapers in the market. The entry rate increases to 7% and the exit rate to 32%. That is, when two newspapers are active, which occurs on 15% of time periods, then there is about a one-third chance that at least one of the papers exits.

Before proceeding with our model and estimator, we consider tests for equilibrium multiplicity based on market share data.

### 5.2 Equilibrium multiplicity

Otsu et al. (2016) propose a statistical test for equilibrium multiplicity which we shall apply. The null hypothesis of the test is that the steady state distribution of market shares is identical

<sup>1</sup>The data are publicly available at <https://people.stanford.edu/gentzkow/data>.

across markets. The alternative is the negation of the null. To construct the test statistic we partition the market share data into ten bins. The critical value for the test statistic is obtained by bootstrapping. The Q-statistic equals 36.83 while the critical value equals 12.18. We can reject the null hypothesis of identical market share distributions.

We looked at several robustness checks. First, we considered finer partitions consisting of 20 bins. The resulting Q-statistic equals 37.66 while the critical value equals 12.91. Again, we can reject the null hypothesis of identical market share distributions.

Second, we divided the data into small and large markets based on the maximum circulation achieved in any year. Small markets are those below the median circulation level. The Q-statistic for small markets equals 23.69 while the critical value equals 5.93. The Q-statistic for large markets equals 12.53 while the critical value equals 6.22. We can reject the null hypothesis of identical market share distributions for both, small and large markets.

Next, we formulate an estimable dynamic game.

### 5.3 Dynamic game played by newspapers

Our dynamic duopoly game shares key features with the static model of Gentzkow et al. (2014), but additionally allows for (i) entry and exit and (ii) dynamic forward-looking behavior by newspapers. We make the following assumptions.

#### 5.3.1 Model assumptions

Time is discrete,  $t = 1, 2, \dots$ . Every period  $t$  the following stage game is played: Initially, newspapers (and the econometrician) observe publicly a vector of last period market shares  $\mathbf{s}_{t-1}$ . Then newspapers make decisions about entry/exit  $\mathbf{e}_t = (e_{1t}, e_{2t})$  with  $e_{it} \in \{0, 1\}$ , subscription price  $\mathbf{p}_t = (p_{1t}, p_{2t})$ , and advertising price  $\mathbf{A}_t = (A_{1t}, A_{2t})$ , simultaneously. At the end of the period, current period market shares  $\mathbf{s}_t = (s_{1t}, s_{2t})$  are realized. The circulation of newspaper  $i$  in period  $t$  is given by  $s_{it} \cdot S$ , where the total market size  $S$  is market specific and constant over time. The firms then collect their period payoffs. Following Gentzkow et al. (2014), the period payoff realization of newspaper  $i$  is specified as

$$\pi_i(\mathbf{e}_t, \mathbf{p}_t, \mathbf{A}_t, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) = e_{it} \cdot S \cdot [s_{it}(\mathbf{e}_t, \mathbf{p}_t, \mathbf{s}_{t-1}, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) \cdot \{p_{it} - \theta_2 + A_{it} \cdot \varphi(A_{it}, \boldsymbol{\xi}^A; \boldsymbol{\theta}_3)\} + \xi_i^m],$$

where  $\boldsymbol{\xi} = (\xi_1^n, \xi_2^n, \xi_1^A, \xi_2^A, \xi_1^m, \xi_2^m)$  is a vector of subscription demand  $\boldsymbol{\xi}^n = (\xi_1^n, \xi_2^n)$ , advertising demand  $\boldsymbol{\xi}^A = (\xi_1^A, \xi_2^A)$ , and market shocks  $\boldsymbol{\xi}^m = (\xi_1^m, \xi_2^m)$  for newspapers. The additive term  $\xi_i^m$  denotes a market-specific unobservable element observed by both firms but not the econometrician, such as operating fixed costs, as in Tamer (2003), which can be positive or negative. Total revenues consist of the sum of subscription and advertising revenues. The functional form of subscription and advertising demand and their respective shocks are explained below. The production costs exhibit constant marginal costs  $\theta_2$ . Notice that the period payoff realization of an in-active newspaper  $i$ ,  $e_{it} = 0$ , is equal to zero. The parameters of interest are described by the vector  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \theta_2, \boldsymbol{\theta}'_3)'$ .

Following the literature on dynamic game estimation, see Aguirregabiria and Nevo (2013) for a survey, we assume that every newspaper  $i$  maximizes the sum of expected discounted

future period payoffs with the common discount factor  $\beta < 1$ ,

$$E \left[ \sum_{t=1}^{\infty} \beta^{t-1} \pi_i(\mathbf{e}_t, \mathbf{p}_t, \mathbf{A}_t, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) \middle| \mathbf{s}_0, S, \boldsymbol{\xi}_i \right],$$

where the expectation is taken over realizations of future demand shocks, and  $\boldsymbol{\xi}_i$  are the elements of the period shock  $\boldsymbol{\xi}$  which are observable to newspaper  $i$  when it makes its decisions. Decisions are made in pure type-symmetric *Markovian strategies*. A pure type-symmetric Markovian strategy for newspaper  $i$  is a tuple of functions  $a_i = (e_i, p_i, A_i)$ , where  $e_i(\mathbf{s}, S, \boldsymbol{\xi}_i)$  denotes the entry/exit strategy,  $p_i(\mathbf{s}, S, \boldsymbol{\xi}_i)$  the price choice, and  $A_i(\mathbf{s}, S, \boldsymbol{\xi}_i)$  the advertising choice of newspaper  $i$  when market shares equal  $\mathbf{s}$  and the observable market attribute is  $S$ . The time subscript is omitted for simplicity of exposition. Let  $\sigma_i$  denote player  $i$ 's belief about the probability distribution of actions and state transitions. For example,  $\sigma_i(\mathbf{a}|\mathbf{s}, S, \boldsymbol{\xi}_i)$  denotes newspaper  $i$ 's conditional *ex ante* belief that action profile  $\mathbf{a} = (a_1, a_2)$  is taken when the information available to newspaper  $i$  is given by  $(\mathbf{s}, S, \boldsymbol{\xi}_i)$ .

A Markov Perfect Equilibrium (MPE) is a pair of strategies and beliefs  $(\mathbf{a}, \boldsymbol{\sigma}) = (a_1, a_2, \sigma_1, \sigma_2)$  that satisfies two conditions: (i) each player's strategy  $a_i$  is a best response to  $a_{-i}$  given beliefs  $\sigma_i$ , and (ii) each player's beliefs  $\sigma_i$  are consistent with strategies  $\mathbf{a}$  and the underlying shocks  $\boldsymbol{\xi}$ .

### 5.3.2 Law of motion for states and demand specification

The law of motion of the market shares is stochastic and governed by consumers' demand decisions. Newspaper demand is a standard differentiated products logit model while incorporating a key feature that readers can be loyal to their favorite newspaper, which is a new and empirically important feature. Let  $i'$  denote the newspaper purchased by a household  $b$  in period  $t-1$ . We model the indirect utility of household  $b$  purchasing newspaper  $i$  in period  $t$  at price  $p_{it}$  as

$$u_b(i, i', p_{it}) = \theta_{10} + \theta_{11} \cdot 1\{i = i'\} - \theta_{1p} \cdot p_{it} + \xi_i^n + \varepsilon_{ib},$$

where the random utility component  $\xi_i^n + \varepsilon_{ib}$  captures unobserved newspaper shock  $\xi_i^n$ , such as the quality shocks to the editorial team which affects all households jointly, and household specific preference shocks  $\varepsilon_{ib}$ . The newspaper shock  $\xi_i^n$  is drawn from a mean zero normal pdf  $\phi(\cdot|\sigma_p^2)$  and observed after period pricing decisions have been made, and  $\varepsilon_{ib}$  is a log Weibull distributed random shock observed by household  $b$ . The parameters of interest are  $\boldsymbol{\theta}_1 = (\theta_{10}, \theta_{11}, \theta_{1p}, \sigma_p^2)'$ , where  $\theta_{10}$  is an intercept,  $\theta_{11}$  denotes the loyalty or satisfaction parameter,  $\theta_{1p}$  is the subscription price sensitivity parameter, and  $\sigma_p^2$  is the variance of the unobserved newspaper shocks  $\xi_i^n$ . We normalize the utility of not buying a newspaper to zero,  $u_b(\emptyset, i') = 0$ . The choice set for household  $b$  is determined by newspaper availability  $\mathbf{e}$ , as newspaper  $i$  is available for purchase for household  $b$  in period  $t$  only if the newspaper is active,  $e_i = 1$ . With many households per market the purchase probability is given by the usual logit form

$$\Pr\{i, i'|\mathbf{p}, \mathbf{e}, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1\} = \frac{e_i \cdot \exp(\theta_{10} + \theta_{11} \cdot 1\{i = i'\} - \theta_{1p} \cdot p_i + \xi_i^n)}{1 + \sum_{i'' \in \{1,2\}} e_{i''} \cdot \exp(\theta_{10} + \theta_{11} \cdot 1\{i'' = i'\} - \theta_{1p} \cdot p_{i''} + \xi_{i''}^n)}.$$

The loyalty parameter  $\theta_{11}$  reflects the utility increment of buying a newspaper that is already familiar to the consumers. Notice that consumers make a static (myopic) choice every period and do not take the effect of their choices on the future into account. We believe that the assumption of myopic behavior by consumers is reasonable for this market.

The logit choice probabilities enable us to derive the law of motion of market shares by taking into account the editorial shocks  $\boldsymbol{\xi}^n$  and last period market shares  $\mathbf{s}_{t-1}$ . The market share realization is given by

$$s_i(\mathbf{s}_{t-1}, \mathbf{p}, \mathbf{e}, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) = \sum_{i' \in \{0,1,2\}} \Pr\{i, i' | \mathbf{p}, \mathbf{e}, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1\} \cdot s_{i', t-1}.$$

The expected market share  $s_i$ , with expectations taken over the unobserved newspaper shocks, equals

$$s_i(\mathbf{s}_{t-1}, \mathbf{p}, \mathbf{e}) = \iint s_i(\mathbf{s}_{t-1}, \mathbf{p}, \mathbf{e}, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) \phi(\xi_1^n | \sigma_p^2) \phi(\xi_2^n | \sigma_p^2) d\xi_1^n d\xi_2^n.$$

The advertisers' demand is static and given by the following reduced-form linear demand function:

$$\varphi(A_i, \boldsymbol{\xi}^A; \boldsymbol{\theta}_3) = \psi \cdot \{2 \exp(\xi_i^A) - A_i\},$$

where  $\xi_i^A$  is a random intercept variable observed prior to the choice of advertising price  $A_i$ . The random variable  $\xi_i^A$  is drawn from a normal pdf  $\phi(\cdot | \theta_{30}, \sigma_A^2)$  with mean  $\theta_{30}$  and variance  $\sigma_A^2$ . The advertising parameter vector is given by  $\boldsymbol{\theta}_3 = (\theta_{30}, \sigma_A^2)'$ . Advertising profits, which are measured per inch, per copy and per subscriber, are multiplied by a factor  $\psi$  to obtain advertising profits for the newspaper per subscriber and per year. The traditional size of US newspapers is half a broadsheet or 15 by 22 $\frac{3}{4}$  inches. We assume a maximum of 32 advertising pages per newspaper copy and calibrate  $\psi$  to be about 3.3 million 1 inch advertisements a reader could see per year.<sup>2</sup> Gentzkow et al. (2014) consider a related specification in which the mass of advertisers per newspaper equals one.

With the payoffs and the law of motion for the state variables defined we can now examine the equilibrium conditions.

## 5.4 Equilibrium conditions

In equilibrium players follow Markovian strategies  $(\mathbf{a}_i)_{i=1}^2 = (e_i, p_i, A_i)_{i=1}^2$ , which maximize the sum of expected discounted future period payoffs given the beliefs. The strategies determine entry/exit  $\mathbf{e}$ , subscription price  $\mathbf{p}$ , and advertising price  $\mathbf{A}$ . The decisions are made after observing the own advertising demand  $\xi_i^A$  and market characteristics  $\xi_i^m$ , but before editorial shocks  $\boldsymbol{\xi}^n$  are observed. The law of motion of the state variables depends on the choices made by newspapers, consumers and the editorial shocks  $\boldsymbol{\xi}^n$ . The informational environment is a 'private information' setting in which newspaper  $i$  observes the own advertising shock but not competitor's shock. Although not the focus of our analysis, our estimation approach can be readily extended to encompass flexibility in the underlying informational environments as in

---

<sup>2</sup>The maximum number appears in line with estimates. E.S. Turner, *The Shocking History of Advertising* 1965, page 124) calculates that in the mid 19th century New York's newspapers published 1,456,416 ads in a year.

Bergemann et al. (2017). What newspapers know about the advertising demand shocks  $\xi^A$  when making decisions can be captured by an additional nuisance parameter.

Following the dynamic estimation literature we impose a parametric functional form on price and entry strategies, commonly referred to as a policy function. The parameters for the strategies are denoted with  $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4, \tau_5) \in \mathbf{T}$  for a compact set  $\mathbf{T}$ . Using a parametric functional form facilitates numerical calculations and thus estimation considerably. The unrestricted (non-parametric) case is discussed in Appendix A. The specified policy function assumes price is non-negative, and when positive, linear in marginal costs plus an advertising intercept,

$$p_i(\mathbf{s}, \xi^A; \boldsymbol{\theta}, \boldsymbol{\tau}) = \max\{(\tau_1 + \tau_2 s_i + \tau_3 s_{3-i}) \cdot \theta_2 + \tau_4 \exp(\xi_i^A)^2, 0\},$$

where the parameters  $(\tau_1, \tau_2, \tau_3)$  describe the markup over costs attributable to a constant, own and rival market shares, and the parameter  $\tau_4$  measures variations due to the advertising intercept. The entry/exit strategy is postulated as

$$e_i(\mathbf{s}, \xi^A, \xi^m; \boldsymbol{\theta}, \boldsymbol{\tau}) = 1\{s_i \cdot \{p_i(\mathbf{s}, \xi_i) - \theta_2 + \exp(\xi_i^A)^2\} - \tau_5 \xi_i^m \geq 0\},$$

where the parameter  $\tau_5$  measures the proportional factor of fixed costs that variable profits need to cover for the newspaper to remain active. Our model assumes that the market size variable  $S$  enters multiplicatively, multiplying both variable profit and cost terms, and thus does not affect strategies.

The advertising price does not have dynamic linkages and can be solved by using a static first order condition

$$2 \exp(\xi_i^A) - 2A_i = 0, \quad \text{or} \quad A_i(\mathbf{s}, \xi_i) = \exp(\xi_i^A).$$

The parametric class of strategies can be viewed as a linear approximation to the unrestricted class over the relevant range. Observe that the class could readily be made richer, say by including more parameters capturing higher order terms of state variables, and thus eventually allowing accurate approximation to the unrestricted class.

Strategies are in equilibrium if strategy  $\mathbf{a}_i$  maximizes newspaper  $i$ 's sum of expected future payoffs. As payoffs are differentiable, a necessary condition for a local maximum in terms of the policy function parameter vector  $\boldsymbol{\tau}$  is given by the following first order condition

$$\frac{\partial E_{\sigma_i} [\sum_{t=1}^{\infty} \beta^{t-1} \pi_i(a_i, a_{3-i}, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) | \mathbf{s}_0, S, \boldsymbol{\xi}_i]}{\partial a_i} \frac{\partial a_i}{\partial \tau_j} = 0 \quad \text{for all } i, j, (\mathbf{s}_0, S, \boldsymbol{\xi}_i). \quad (10)$$

Adding  $\tau_j$  on the left and right hand side of the first order condition for  $\tau_j$  gives the  $j$ -th equation of the fixed point problem (1). Notice that this equation system may have multiple solutions because there can be multiple equilibria. Pesendorfer and Schmidt-Dengler (2008) analyze a simple dynamic entry-exit game with two players and find five equilibria. Doraszelski and Satterthwaite (2010) also provide several examples of games with multiple equilibria. Our estimator for the identified region explicitly allows for equilibrium multiplicity. Notice also that the choice of the set  $\mathbf{T}$  may include parameter vectors for which the sum of expected future payoffs is zero, because the policy function requires all firms to be inactive. Such parameter vectors constitute a minimum and are not desirable. In our implementation of the

estimator, we adjust the policy parameter space  $\mathbf{T}$  by dropping parameter values  $\boldsymbol{\tau}$  where zeros to the condition (10) correspond to a point where the sum of expected future payoffs is zero. Also note that we approximate the sum of expected future payoffs numerically by forward simulation using a finite number of periods. We evaluate the equilibrium conditions on a grid  $\mathbf{S}'$  of selected states.

To allow for equilibrium multiplicity and unobserved market-level heterogeneity in the estimation, we consider an estimation approach based on moment condition (3) that explicitly allows for distinct strategies across markets. We describe our estimator next.

## 5.5 Moment condition and estimator

While the moment condition (3) considers the unrestricted case, our application defines the moment condition using a parametric functional form for the policy function, which facilitate the numerical calculations and performance of the estimator for our data. The moment function  $g$  in our application consists of the equilibrium condition (10) defined on the compact parameter set  $\mathbf{T}$  plus two empirical moment conditions consisting of the first and second moments of subscription prices at the market level. The first empirical moment condition is the difference between the theoretical subscription price and the observed analogue,  $p_i(\mathbf{s}, \boldsymbol{\xi}^m; \boldsymbol{\theta}, \boldsymbol{\tau}) - p_{is}$ , summed across observed states and players. The second moment is in terms of the difference between the squared theoretical subscription prices and its observed counterpart. The function  $g$  for a newspaper market is given by,

$$g(Z, \boldsymbol{\tau}, S, \boldsymbol{\xi}; \boldsymbol{\theta}, \boldsymbol{\tau}) = \left[ \begin{array}{c} \sum_{i,j,s_0} \left( \frac{\partial E_{\sigma_i} [\sum_{t=1}^{\infty} \beta^{t-1} \pi_t(a_i, a_{3-i}, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) | \mathbf{s}_0, S, \boldsymbol{\xi}_i] \frac{\partial a_i}{\partial \tau_j}}{\sum_{i,s} \{p_i(\mathbf{s}, \boldsymbol{\xi}^A; \boldsymbol{\theta}, \boldsymbol{\tau}) - p_{is}\}} \right)^2 \\ \sum_{i,s} \{p_i(\mathbf{s}, \boldsymbol{\xi}^A; \boldsymbol{\theta}, \boldsymbol{\tau})^2 - p_{is}^2\} \end{array} \right]. \quad (11)$$

The first part vanishes when an equilibrium is reached. The second and third part link the theoretical strategies to the observables. They vanish when the parameter vector  $(\boldsymbol{\theta}, \boldsymbol{\tau})$  achieves a good fit.

We employ the GMM criterion function  $Q(\boldsymbol{\theta})$ , defined in (6), where

$$\tilde{g}(Z, \boldsymbol{\theta}, \gamma) = \frac{\int g(Z, \boldsymbol{\tau}, S, \boldsymbol{\xi}; \boldsymbol{\theta}, \boldsymbol{\tau}) \cdot \exp(\gamma' g(Z, \boldsymbol{\tau}, S, \boldsymbol{\xi}; \boldsymbol{\theta}, \boldsymbol{\tau})) dU(\boldsymbol{\tau}, \boldsymbol{\xi})}{\int \exp(\gamma' g(Z, \boldsymbol{\tau}, S, \boldsymbol{\xi}; \boldsymbol{\theta}, \boldsymbol{\tau})) dU(\boldsymbol{\tau}, \boldsymbol{\xi})}.$$

The conservative confidence region  $\hat{\mathcal{C}}$  is calculated by using (7). With three moment conditions, the 1% critical value equals 11.345. The confidence region  $\tilde{\mathcal{C}}$  defined in (9) has the critical value 6.635.

A key element in our estimator is that the policy functions may differ across markets due to two reasons: (i) market heterogeneity as captured by observable and unobservable elements in  $\boldsymbol{\xi}^m$  and (ii) equilibrium multiplicity as captured by parameters  $\boldsymbol{\tau}$ . Notice though that we assume newspapers within a market use identical strategies. The latter assumption is made to simplify the problem and could easily be relaxed by allowing the parameters  $\boldsymbol{\tau}$  to differ across newspapers.

Next, we report and illustrate our estimation results.

## 6 Estimation results

This section reports our estimation results. Our estimation approach explicitly allows for distinct strategies and unobserved heterogeneity across markets.

We use a two step approach to estimate the parameters of interest. The first step recovers the demand parameters  $(\theta_1, \theta_3)$  from the market share and price data. The second step characterizes the identified region of marginal cost parameters  $\theta_2$ .

In the first step we estimate the subscription and advertising demand parameters. The subscription demand parameter vector  $\theta_1$  is estimated using the method of moments by comparing the mean and variance of the observed and predicted market shares conditional on state variables. The data include information on the market shares  $(s_j^t)_{j,t}$  and subscription prices  $(p_j^t)_{j,t}$  observed every four years. The subscription price is assumed exogenous for readers. The advertising demand parameter vector  $\theta_3$  is estimated from the observed advertising prices  $(A_{it})_{i,t}$  using the method of moments based on the first order condition,  $\exp(\xi_i^A) - A_i = 0$ . The data do not include information on the number of advertisings per newspaper, and our formulation of advertising profits is driven by this functional form assumption. The following table summarizes our first stage coefficient estimates with standard errors in parenthesis.

Table 2: Estimates of subscribers and advertisers demand

Parameter	$\theta_{10}$	$\theta_{11}$	$\theta_{1p}$	$\sigma_p$	$\theta_{30}$	$\sigma_A$
Estimate	-2.8504	22.586	0.0204	1.3146	-7.5594	0.9644
Standard Deviation	(0.021)	(6.519)	(0.001)	(0.008)	(0.050)	(0.083)

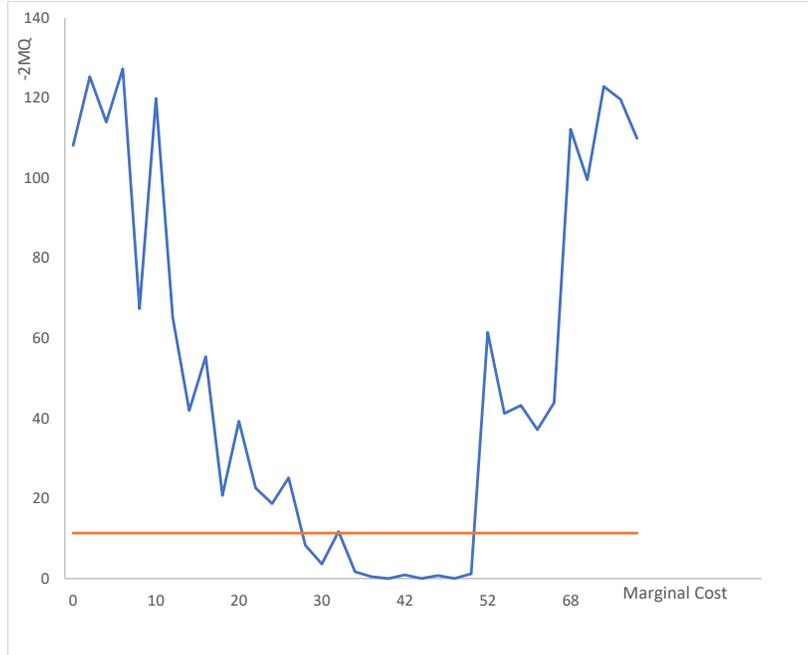
The subscription demand estimates imply a high degree of readership persistence. Readers are loyal to their newspaper and reluctant to switch. The demand elasticity implied from the estimates at the average subscription price of 50 at a market share of 1/2 amounts to  $-0.51$ . Notice that the standard errors are small. In part this is attributable to the fact that the data are very rich. In practice, it enables us to ignore the first stage estimation error in the second step.

The second step recovers the underlying cost parameter  $\theta_2$  by maximizing the GMM-type criterion function  $Q(\theta)$ , defined in (6), and using moment condition (11). The sum of expected future payoffs is calculated using 100 sample paths,  $r = 1, \dots, 100$ , where each path forward simulates 80 years using a sequence of independently drawn shocks  $\tilde{\xi}_t = (\xi_{1t}^n, \xi_{2t}^n, \xi_{1t}^A, \xi_{2t}^A)$  for  $t = 1, \dots, 80$ . The sequences of demand side error draws  $\tilde{\xi} = ((\tilde{\xi}_t^r)_{t=1}^{80})_{r=1}^{100}$  are held fixed throughout the procedure. The derivative of the value function is calculated numerically by using a 10% change in the  $\tau$  parameters. The annual discount factor is calibrated at 0.95.

Computation of the criterion function  $Q(\theta)$  takes a uniformly drawn set of parameters  $\tau$  with  $\tau_1 \in [0.9, 1.5]$ ,  $\tau_2 \in [-0.2, 0.2]$ ,  $\tau_3 \in [-0.2, 0.2]$ ,  $\tau_4 \in [-0.1, 1]$ ,  $\tau_5 \in [0, 2]$ , and  $\xi^m \in [-1, 1]$  and finds the three-dimensional parameter vector  $\gamma$  that maximizes the right hand side of (6) for a given value of the parameter  $\theta_2$ . The moment condition  $\tilde{g}$  is calculated by integrating the moment condition  $g$  for each market over the parameter values  $\tau, \xi^m$  by using Monte

Carlo integration and then summing across all markets. We use 1,000 uniform draws to approximate the integral and the draws are held fixed in the procedure. Draws for which the expected discounted sum of payoffs is zero are omitted. Individual markets have identical weight in our procedure. We omit data points with missing price or market share data, or where there are fewer than 10 observations in a market. We assume that the observability of data is exogenous to our data generating process. In total we use 4,596 data points for 144 markets.

Figure 1: Objective function  $-2MQ(\theta_2)$  evaluated at different values of marginal cost



Figures 1 and 2 plot the value of the criterion function stemming from the statistics (6) and (8) along with critical values for a grid of marginal cost parameter values  $\theta_2$ . Figure 1 considers the objective function  $-2MQ(\theta_2)$  with critical value  $c_{3,\alpha} = 11.345$ . We performed a grid search over the marginal cost parameter from 0 to 80. Figure 1 illustrates that the interval  $[28, 50]$  constitutes the 99% confidence region for the statistic based on (6).

Figure 2: Objective function  $K(\theta_2)$  evaluated at different values of marginal cost

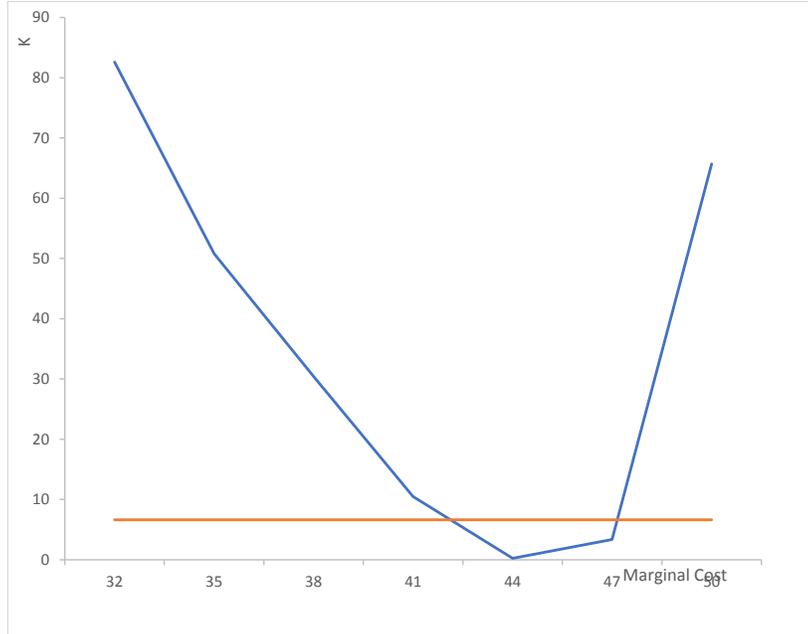


Figure 2 zooms in on the relevant cost range from 30 to 50, and considers the criterion  $K(\theta_2)$  based on equation and (8) with critical value  $c_{1,\alpha} = 6.635$ . The 99% confidence region for marginal costs is now narrower and becomes the interval  $[42, 47]$ . Figure 2 shows that the adapted criterion, which depends only on the dimension of structural parameters  $\theta$  and not on the dimension of the moment function  $g$ , gives a much sharper prediction in our case.

Our measure of marginal costs, which excludes operating fixed costs and sunk costs, is between 84% and 94% of marginal revenue. The implied per-period mark-up estimates are in line with the ones found in the static model of Gentzkow et al. (2014). A key difference is that we consider a dynamic framework with loyal readers newspapers in which newspapers have an additional incentive to price low in order to increase the customer base for the future. This effect is absent in a static model.

## 7 Conclusions

A novel multiplicity-robust inference method has been proposed for a class of economic models that can be written as a fixed point problem. The class includes static games and dynamic Markov games. Our proposed method allows for distinct behaviors and strategies across markets. The main idea is to treat market specific behaviors as correlated latent variables. Instead of solving the primal problem, we consider the equivalent dual problem which entails optimization over only a finite-dimensional Euclidean space. This property allows for a practically feasible characterization of the identified region for the structural parameters.

Our application illustrates the feasibility of the inference method by considering dynamic competition in the US newspaper market. Statistical tests of equilibrium multiplicity confirm that multiplicity is a concern in these data. In implementing the estimator, we restrict attention to a class of parametric strategies, or policy functions, which allow us to reduce the complexity of the numerical problem and permits estimation in practice. As was shown in the static contexts, e.g. Ciliberto and Tamer (2009), multiplicity of equilibria results in partial identification. We find that the identified region of marginal costs amounts to costs being between 84% and 94% of the newspaper revenues. Our method provides a practically attractive computational procedure to obtain the identification region in dynamic games and also static games.

## References

- Aguirregabiria, V. and P. Mira (2007). Sequential estimation of dynamic discrete games. *Econometrica* 75(1), 1–53.
- Aguirregabiria, V. and A. Nevo (2013). Recent developments in empirical io: Dynamic demand and dynamic games; in: *Advances in economics and econometrics: Tenth world congress*; cambridge university press. pp. 53–122.
- Arcidiacono, P. and R. A. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79(6), 1823–1867.
- Bajari, P., L. C. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica* 75(5), 571–599.
- Bajari, P., H. Hong, and S. P. Ryan (2010). Incomplete simultaneous discrete response model with multiple equilibria. *Econometrica* 78(5), 1529–1568.
- Bergemann, D., B. Brooks, and S. Morris (2017). First-price auctions with general information structures: Implications for bidding and revenue. *Econometrica* 85(1), 107–143.
- Bjorn, P. A. and Q. Vuong (1984). Simultaneous equations models for dummy endogenous variables: A game theoretic formulation with an application to labor force participation. *Caltech Working Paper 537*.
- Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75(5), 1243–1284.
- Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in airline markets. *Econometrica* 77, 1791–1828.
- Doraszelski, U. and M. Satterthwaite (2010). Computable markov-perfect industry dynamics. *The RAND Journal of Economics* 41(2), 215–243.
- Fudenberg, D. and J. Tirole (1983). Learning-by-doing and market performance. *Bell Journal of Economics* 14(2), 522–530.
- Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2014). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review* 104(10), 3073–3114.
- Guerre, E., I. Perrigne, and Q. Vuong (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica* 68(3), 525–574.
- Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72(4), 1295–1319.
- Hendricks, K., A. Weiss, and C. Wilson (1988). The war of attrition in continuous time with complete information. *International Economic Review* 29(4), 663–680.

- Hortacsu, A. and J. Kastl (2012). Valuing dealers' informational advantage: A study of canadian treasury auctions. *Econometrica* 80, 2511–2542.
- Jofre-Bonet, M. and M. Pesendorfer (2003). Estimation of a dynamic auction game. *Econometrica* 71, 1443–1489.
- Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica* 73(4), 1103–1123.
- Molinari, F. (2010). Econometric issues in the presence of multiple equilibria. In *The New Palgrave Dictionary of Economics*. Palgrave Macmillan.
- Otsu, T., M. Pesendorfer, and Y. Takahashi (2016). Pooling data across markets in dynamic markov games. *Quantitative Economics* 7(2), 523–559.
- Pesendorfer, M. and P. Schmidt-Dengler (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies* 75, 901–928.
- Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica* 80(3), 1019–1061.
- Schennach, S. M. (2014). Entropic latent variable integration via simulation. *Econometrica* 82(1), 345–385.
- Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies* 70(1), 147–165.

## A Characterization of equilibrium

This appendix characterizes the equilibrium for the case in which no parametric restrictions are placed on the equilibrium. The resulting equilibrium equations can be used for a non-parametric approach. We shall consider entry, advertising, and subscription price choices.

First, observe that it is optimal in state  $\mathbf{s}$  to be active if doing so achieves the high continuation value

$$e_i(\mathbf{s}_0, \boldsymbol{\xi}_i) = \begin{cases} 1 & \text{if } S \cdot [s_i(\mathbf{e}, \mathbf{p}, \mathbf{s}_0, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) \cdot \{p_i(\mathbf{s}_0, \boldsymbol{\xi}_i) - \theta_2 + A_i(\mathbf{s}_0, \boldsymbol{\xi}_i) \cdot \varphi(A_i(\mathbf{s}_0, \boldsymbol{\xi}_i), \boldsymbol{\xi}^A; \boldsymbol{\theta}_3)\} + \xi_i^m] \\ & + E [\sum_{t=2}^{\infty} \beta^{t-1} \pi_i(\mathbf{e}_t, \mathbf{p}_t, \mathbf{A}_t, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) | \mathbf{s}_1, S, \boldsymbol{\xi}_i] \\ & > 0 + E [\sum_{t=2}^{\infty} \beta^{t-1} \pi_i(\mathbf{e}_t, \mathbf{p}_t, \mathbf{A}_t, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) | (0, s'_{-i,1}), S, \boldsymbol{\xi}_i]; \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

Second, observe that the equilibrium advertising price is characterized by the first order condition with respect to  $A_i$  which implies

$$A_i(\mathbf{s}_0, \boldsymbol{\xi}_i) = \exp(\xi_i^A) \quad (13)$$

Notice that the continuation value does not enter here as advertising affects only current period payoffs.

Finally, the equilibrium subscription price is characterized by the necessary first-order condition consisting of the usual static components plus the change in the future sum of expected payoffs divided by the change in expected market shares of a one period price change

$$p_i(\mathbf{s}_0, \boldsymbol{\xi}_i) = \theta_2 - A_i(\mathbf{s}_0, \boldsymbol{\xi}_i) \varphi_i(A_i(\mathbf{s}_0, \boldsymbol{\xi}_i)) - \frac{s_i(\mathbf{e}, \mathbf{p}, \mathbf{s}_0, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1)}{\partial s_i(\mathbf{e}, \mathbf{p}, \mathbf{s}_0, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) / \partial \mathbf{p}_i} \quad (14)$$

$$+ \frac{1}{\partial s_i(\mathbf{e}, \mathbf{p}, \mathbf{s}_0, \boldsymbol{\xi}^n; \boldsymbol{\theta}_1) / \partial \mathbf{p}_i} \frac{\partial E [\sum_{t=2}^{\infty} \beta^{t-1} \pi_i(\mathbf{e}_t, \mathbf{p}_t, \mathbf{A}_t, \mathbf{s}_{t-1}, S, \boldsymbol{\xi}; \boldsymbol{\theta}) | \mathbf{s}_1, S, \boldsymbol{\xi}_i]}{\partial \mathbf{p}_i}.$$

Equations (12)-(14) characterize the set of equilibria. The equations give rise to a fixed point problem for equilibrium choices. Equation (13) is a static condition which can be solved separately.

## B Mathematical appendix

### B.1 Proof of Proposition S

To apply Schennach (2014, Theorem 2.1) in our setup, it is enough to verify that  $\kappa(\gamma, \theta) = E[\log \int \exp(\gamma' g(\mathbf{p}, Z; \theta)) dF(\mathbf{p})]$  exists and is twice continuously differentiable with respect to  $\gamma$  for all  $\gamma \in \mathbb{R}^d$ .

First, we show existence of  $\kappa(\gamma, \theta)$ . Pick any  $\gamma \in \mathbb{R}^d$  and  $\theta \in \Theta$  and denote  $\bar{\gamma} = \sum_{i=1}^L |\gamma_i|$ . Since

$$\exp(-\bar{\gamma}) \leq \exp(\gamma' g(\mathbf{p}, z; \theta)) \leq \exp(\bar{\gamma}),$$

for all  $(\mathbf{p}, z) \in \mathcal{P} \times \mathcal{Z}$ , it holds

$$-\bar{\gamma} \leq \log \int \exp(\gamma' g(\mathbf{p}, z; \theta)) F(\mathbf{p}) \leq \bar{\gamma},$$

for all  $z \in \mathcal{Z}$ . Therefore,  $\kappa(\gamma, \theta)$  exists as claimed.

Next, we show that  $\kappa(\gamma, \theta)$  is twice continuously differentiable with respect to  $\gamma$  for all  $\gamma \in \mathbb{R}^d$ . Since  $\Psi(\cdot, \theta)$  is continuous by assumption, so is  $g(\cdot, z; \theta)$  for each  $z \in \mathcal{Z}$ . Then the derivatives  $\frac{\partial}{\partial \gamma} \exp(\gamma' g(\cdot, z; \theta)) = \exp(\gamma' g(\cdot, z; \theta)) g(\cdot, z; \theta)$  and  $\frac{\partial^2}{\partial \gamma \partial \gamma'} \exp(\gamma' g(\cdot, z; \theta)) = \exp(\gamma' g(\cdot, z; \theta)) g(\cdot, z; \theta) g(\cdot, z; \theta)'$  are also continuous for each  $z \in \mathcal{Z}$ . Thus, the following integrals exist for each  $z \in \mathcal{Z}$

$$\begin{aligned} \frac{\partial}{\partial \gamma} \int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p}) &= \int \exp(\gamma' g(\mathbf{p}, z; \theta)) g(\mathbf{p}, z; \theta) dF(\mathbf{p}), \\ \frac{\partial^2}{\partial \gamma \partial \gamma'} \int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p}) &= \int \exp(\gamma' g(\mathbf{p}, z; \theta)) g(\mathbf{p}, z; \theta) g(\mathbf{p}, z; \theta)' dF(\mathbf{p}). \end{aligned}$$

Now consider the functions  $\delta_k(z) = \frac{\partial}{\partial \gamma_k} \log \int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})$  and  $\delta_{kl}(z) = \frac{\partial^2}{\partial \gamma_l \partial \gamma_k} \log \int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})$  for all  $k, l = 1, \dots, d$ . By the above integrals, we obtain

$$\begin{aligned} \delta_k(z) &= \frac{\int g_k(\mathbf{p}, z; \theta) \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})}{\int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})}, \\ \delta_{kl}(z) &= \frac{\int g_l(\mathbf{p}, z; \theta) \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})}{\int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})} \cdot \frac{\int g_k(\mathbf{p}, z; \theta) \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})}{\int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})} \\ &\quad - \frac{\int g_l(\mathbf{p}, z; \theta) g_k(\mathbf{p}, z; \theta) \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p})}{\left[ \int \exp(\gamma' g(\mathbf{p}, z; \theta)) dF(\mathbf{p}) \right]^2}. \end{aligned}$$

Thus, we can deduce  $|\delta_k(z)| \leq \exp(2\bar{\gamma})$  and  $|\delta_{kl}(z)| \leq \exp(4\bar{\gamma}) + \exp(3\bar{\gamma}) < \infty$  for all  $z \in \mathcal{Z}$  and  $k, l = 1, \dots, d$ . Therefore,  $\kappa(\gamma, \theta)$  is twice continuously differentiable and the conclusion follows.

## B.2 Assumptions and proof for Proposition K

Let  $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$  and define  $\gamma_0 \in \bar{\mathbb{R}}^d$  such that  $\inf_{\gamma \in \bar{\mathbb{R}}^d} |E_\pi[\tilde{g}(Z, \theta, \gamma)]| = \lim_{\gamma \rightarrow \gamma_0} |E_\pi[\tilde{g}(Z, \theta, \gamma)]|$ . Note that elements of  $\gamma_0$  may be infinity. Also let  $\tilde{G}_k(z, \theta, \gamma) = \frac{\partial}{\partial \theta_k} \tilde{g}(z, \theta, \gamma)$ ,  $\tilde{G}(\cdot, \theta, \gamma) = [\tilde{G}_1(\cdot, \theta, \gamma), \dots, \tilde{G}_q(\cdot, \theta, \gamma)]$ ,  $d_z$  be the cardinality of  $\mathcal{Z}$ , and  $\mathbb{G}_M f(\cdot) = \sqrt{M} \{M^{-1} \sum_{j=1}^M f(Z_j) - E[f(Z)]\}$  for a function  $f$  on  $\mathcal{Z}$ .

For every  $\theta_0 \in \Theta_0$ , we impose the following assumptions.

### Assumption K.

- (i) *The limits  $\tilde{g}(z, \theta_0, \gamma_0) = \lim_{\gamma \rightarrow \gamma_0} \tilde{g}(z, \theta_0, \gamma)$  and  $\tilde{G}_k(z, \theta_0, \gamma_0) = \lim_{\gamma \rightarrow \gamma_0} \tilde{G}_k(z, \theta_0, \gamma)$  exist for each  $z \in \mathcal{Z}$  and  $k \in \{1, \dots, q\}$ .*

(ii) There exist  $\tilde{\gamma}(\theta_0)$  such that as  $M \rightarrow \infty$ ,

$$\zeta_M(\theta_0) = \begin{bmatrix} \mathbb{G}_M \tilde{g}(\cdot, \theta_0, \gamma_0) \\ \text{vec}(\sqrt{M} \{ \tilde{g}(z, \theta_0, \tilde{\gamma}(\theta_0)) - \tilde{g}(z, \theta_0, \gamma_0) \}_{z \in \mathcal{Z}}) \\ \text{vec}(\mathbb{G}_M \tilde{G}(\cdot, \theta_0, \gamma_0)) \\ \text{vec}(\sqrt{M} \{ \tilde{G}_1(\cdot, \theta_0, \tilde{\gamma}(\theta_0)) - \tilde{G}_1(\cdot, \theta_0, \gamma_0) \}_{z \in \mathcal{Z}}) \\ \vdots \\ \text{vec}(\sqrt{M} \{ \tilde{G}_q(\cdot, \theta_0, \tilde{\gamma}(\theta_0)) - \tilde{G}_q(\cdot, \theta_0, \gamma_0) \}_{z \in \mathcal{Z}}) \end{bmatrix} \xrightarrow{d} N(0, \Omega(\theta_0)),$$

for some variance matrix  $\Omega(\theta_0)$ .

(iii) There is a consistent estimator  $\hat{\Omega}(\theta_0)$  of  $\Omega(\theta_0)$ .

(iv) The matrix

$$[I_d, (\text{vec}P\{Z = z\}_{z \in \mathcal{Z}})'] \otimes I_d \Omega^\dagger(\theta_0) [I_d, (\text{vec}P\{Z = z\}_{z \in \mathcal{Z}})'] \otimes I_d$$

has the full rank, where  $\Omega^\dagger(\theta_0)$  denotes the top left  $d(1 + d_z)$ -by- $d(1 + d_z)$  portion of  $\Omega(\theta_0)$ .

Under these assumptions and the ones imposed for Proposition S, we can show that  $K(\theta_0)$  defined in (8) converges in distribution to the  $\chi_q^2$  distribution.

For the rest of this subsection, we suppress  $\theta_0$  in the arguments of functions. For example,  $\tilde{g}(\cdot, \tilde{\gamma})$  means  $\tilde{g}(\cdot, \theta_0, \tilde{\gamma}(\theta_0))$ . For Proposition K, it is enough to show the following lemmas.

**Lemma K1.** Suppose Assumption K (i)-(ii) and the assumptions for Proposition S hold. Then

$$\mathbb{G}_M \{ \tilde{g}(\cdot, \tilde{\gamma})', \tilde{G}_1(\cdot, \tilde{\gamma})', \dots, \tilde{G}_q(\cdot, \tilde{\gamma})' \} = \mathbb{G}_M \{ \tilde{g}(\cdot, \gamma_0)', \tilde{G}_1(\cdot, \gamma_0)', \dots, \tilde{G}_q(\cdot, \gamma_0)' \} + o_p(1).$$

**Lemma K2.** Suppose Assumption K (i)-(ii) and the assumptions for Proposition S hold. Then

$$\begin{pmatrix} \frac{1}{\sqrt{M}} \sum_{j=1}^M \{ \tilde{g}(Z_j, \tilde{\gamma}) - E[\tilde{g}(Z, \gamma_0)] \} \\ \frac{1}{\sqrt{M}} \sum_{j=1}^M \{ \tilde{G}_1(Z_j, \tilde{\gamma}) - E[\tilde{G}_1(Z, \gamma_0)] \} \\ \vdots \\ \frac{1}{\sqrt{M}} \sum_{j=1}^M \{ \tilde{G}_q(Z_j, \tilde{\gamma}) - E[\tilde{G}_q(Z, \gamma_0)] \} \end{pmatrix} \xrightarrow{d} N(0, R\Omega R'),$$

where

$$R = \begin{pmatrix} I_d & (\text{vec}P\{Z = z\}_{z \in \mathcal{Z}})' \otimes I_{d_g} & 0 & 0 \\ 0 & 0 & I_q \otimes I_d & I_q \otimes ((\text{vec}P\{Z = z\}_{z \in \mathcal{Z}})' \otimes I_d) \end{pmatrix}.$$

The above variance matrix  $R\Omega R'$  can be estimated by  $\hat{R}\hat{\Omega}\hat{R}'$ , where  $\hat{\Omega}$  is given by Assumption K (iii) and  $\hat{R}$  is obtained by replacing  $P\{Z = z\}$  in the definition of  $R$  with the sample

frequency  $M^{-1} \sum_{j=1}^M \mathbb{I}\{Z_j = z\}$ . Based on the definition, we decompose

$$R\Omega R' = \begin{pmatrix} V & V^{gG} \\ V^{Gg} & V^{GG} \end{pmatrix}, \quad \hat{R}\hat{\Omega}\hat{R}' = \begin{pmatrix} \hat{V} & \hat{V}^{gG} \\ \hat{V}^{Gg} & \hat{V}^{GG} \end{pmatrix},$$

where  $\hat{V}$  is a  $d \times d$  matrix,  $\hat{V}^{Gg} = [\hat{V}^{G_1g'}, \dots, \hat{V}^{G_qg'}]'$  is a  $dq \times d$  matrices,  $\hat{V}^{G_1g'}$  is a  $d \times d$  matrix, and so on. Then we define

$$\begin{aligned} D &= \left[ E[\tilde{G}_1(Z, \gamma_0)] - V^{G_1g}V^{-1}E[\tilde{g}(Z, \gamma_0)], \dots, E[\tilde{G}_q(Z, \gamma_0)] - V^{G_qg}V^{-1}E[\tilde{g}(Z, \gamma_0)] \right], \\ \hat{D} &= \left[ \frac{1}{M} \sum_{j=1}^M \tilde{G}_1(Z_j, \tilde{\gamma}) - \hat{V}^{G_1g}\hat{V}^{-1} \frac{1}{M} \sum_{j=1}^M \tilde{g}(Z_j, \tilde{\gamma}), \right. \\ &\quad \left. \dots, \frac{1}{M} \sum_{j=1}^M \tilde{G}_q(Z_j, \tilde{\gamma}) - \hat{V}^{G_qg}\hat{V}^{-1} \frac{1}{M} \sum_{j=1}^M \tilde{g}(Z_j, \tilde{\gamma}) \right]. \end{aligned}$$

Note that  $D$  is well-defined under Assumption K (iv). We obtain the following orthogonal decomposition of the joint asymptotic normality result.

**Lemma K3.** *Suppose Assumption K and the assumptions for Proposition S hold. Then*

$$\begin{aligned} \frac{1}{\sqrt{M}} \sum_{j=1}^M \tilde{g}(Z_j, \tilde{\gamma}) &\xrightarrow{d} \xi \sim N(0, V), \\ \sqrt{M} \text{vec}(\hat{D} - D) &\xrightarrow{d} \xi_D \sim N(0, V_D), \end{aligned}$$

where  $\xi$  and  $\xi_D$  are independent and  $V_D = V^{GG} - V^{Gg}V^{-1}V^{gV}$ .

Based on these lemmas, we obtain

$$\left( \frac{1}{\sqrt{M}} \sum_{j=1}^M \tilde{g}(Z_j, \tilde{\gamma})' \hat{V}^{-1} \hat{D} \right) (\hat{D}' \hat{V}^{-1} \hat{D})^{-1/2} \xrightarrow{d} N(0, I_q),$$

and the conclusion of Proposition K follows. Below we provide proof for the above lemmas.

### Proof of Lemma K1

Pick any  $j = 1, \dots, d$ . Let  $\tilde{\mathcal{G}}_j = \{\tilde{g}_j(\cdot, \gamma) : \gamma \in \mathbb{R}^d\}$ . Since  $\mathcal{Z}$  is a finite set yielding the bracketing number  $N_{[]}(\varepsilon, \tilde{\mathcal{G}}_j, L^2(\pi)) \leq (2/\varepsilon)^{dz}$ ,  $\tilde{\mathcal{G}}_j$  is  $\pi$ -Donsker. Also note that  $\tilde{g}_j(\cdot, \gamma_0) \in L^2(\pi)$  is true because  $\tilde{g}_j$  is bounded on the finite set  $\mathcal{Z}$ . Furthermore, Assumption K (ii) implies  $[\int \{\tilde{g}_j(z, \tilde{\gamma}) - \tilde{g}_j(z, \gamma_0)\}^2 d\pi(z)]^{\frac{1}{2}} \xrightarrow{p} 0$ . Therefore, we have  $\mathbb{G}_M\{\tilde{g}_j(\cdot, \tilde{\gamma}) - \tilde{g}_j(\cdot, \gamma_0)\} \xrightarrow{p} 0$  by van der Vaart (1998, Lemma 19.24).

Similarly, Assumption K (ii) implies  $\left[ \int \left( \frac{\partial \tilde{g}_j(z, \tilde{\gamma})}{\partial \theta_k} - \frac{\partial \tilde{g}_j(z, \gamma_0)}{\partial \theta_k} \right)^2 d\pi(z) \right]^{\frac{1}{2}} \xrightarrow{p} 0$ , and we have  $\mathbb{G}_M \left( \frac{\partial \tilde{g}_j(\cdot, \tilde{\gamma})}{\partial \theta_k} - \frac{\partial \tilde{g}_j(\cdot, \gamma_0)}{\partial \theta_k} \right) \xrightarrow{p} 0$  for each  $k = 1, \dots, q$  by van der Vaart (1998; Lemma 19.24). The conclusion therefore follows.

## Proof of Lemma K2

By Lemma K1,

$$\begin{aligned} \frac{1}{\sqrt{M}} \sum_{j=1}^M \{\tilde{g}(Z_j, \tilde{\gamma}) - E[\tilde{g}(Z, \gamma_0)]\} &= \mathbb{G}_M \tilde{g}(\cdot, \gamma_0) + \sqrt{M} E[\tilde{g}(Z, \tilde{\gamma}) - \tilde{g}(Z, \gamma_0)] + o_p(1), \\ \frac{1}{\sqrt{M}} \sum_{j=1}^M \{\tilde{G}_1(Z_j, \tilde{\gamma}) - E[\tilde{G}_1(Z, \gamma_0)]\} &= \mathbb{G}_M \tilde{G}_k(\cdot, \gamma_0) + \sqrt{M} E[\tilde{G}_k(\cdot, \tilde{\gamma}) - \tilde{G}_k(\cdot, \gamma_0)] + o_p(1), \end{aligned}$$

for each  $k = 1, \dots, q$ . Thus, we can write

$$\begin{pmatrix} \frac{1}{\sqrt{M}} \sum_{j=1}^M \{\tilde{g}(Z_j, \tilde{\gamma}) - E[\tilde{g}(Z, \gamma_0)]\} \\ \frac{1}{\sqrt{M}} \sum_{j=1}^M \{\tilde{G}_1(Z_j, \tilde{\gamma}) - E[\tilde{G}_1(Z, \gamma_0)]\} \\ \vdots \\ \frac{1}{\sqrt{M}} \sum_{j=1}^M \{\tilde{G}_q(Z_j, \tilde{\gamma}) - E[\tilde{G}_q(Z, \gamma_0)]\} \end{pmatrix} = R\zeta_M + o_p(1),$$

and the conclusion follows by Assumption K (ii).

## Proof of Lemma K3

By Assumption K (iii), we have

$$\begin{pmatrix} I_d & 0 \\ -\hat{V}^{Gg}\hat{V}^{-1} & I_{dq} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} I_d & 0 \\ -V^{Gg}V^{-1} & I_{dq} \end{pmatrix}.$$

Therefore, the claimed result follows from Lemma K2 and Slutsky's theorem.

## B.3 Computation of $\tilde{\gamma}(\theta)$

Take any  $\theta_0 \in \Theta$ . To implement statistical inference based on the statistic  $K(\theta_0)$  in (8), we need to compute  $\tilde{\gamma}(\theta_0)$ . If the infimum  $\inf_{\gamma \in \mathbb{R}^d} |E[\tilde{g}(Z, \theta_0, \gamma)]|$  is achieved by an interior point, we can employ the solution of (6) as  $\tilde{\gamma}(\theta_0)$ , which satisfies Assumption K (ii) under mild regularity conditions. However, in our setup, it is often the case that the infimum is achieved “at infinity”, i.e.,  $\gamma_0 \in \bar{\mathbb{R}}^d$  with  $|\gamma_0| = \infty$ . In this section, we propose a method to compute  $\tilde{\gamma}(\theta_0)$  in such a scenario. More precisely, we consider the following setup.

### Assumption G.

- (i) *There exist positive constants  $\alpha$  and  $\beta$  such that  $\inf_{\gamma: |\gamma| \leq r} |E[\tilde{g}(Z, \theta_0, \gamma)]|^2 \geq \frac{1}{\alpha + r^\beta}$  for all  $r > 0$ .*
- (ii) *There exists a positive constant  $\varphi$  such that*

$$|E[\tilde{g}_l(\cdot, \theta_0, \gamma)]| \geq \varphi \sum_{z \in \mathcal{Z}} |\tilde{g}_l(z, \theta_0, \gamma) - \tilde{g}_l(z, \theta_0, \gamma_0)|$$

for all  $\gamma \in \mathbb{R}^d$  and  $l = 1, \dots, d$ .

(iii) There exist positive constants  $a$  and  $b$  such that for any stochastic sequence  $\{\gamma_M\}$  satisfying  $\tilde{g}(z, \theta_0, \gamma_M) \xrightarrow{p} \tilde{g}(z, \theta_0, \gamma_0)$  over  $z \in \mathcal{Z}$ , it holds:

$$\begin{aligned} |\tilde{g}_l(z, \theta_0, \gamma_M) - \tilde{g}_l(z, \theta_0, \gamma_0)| &\leq \frac{a}{|\gamma_M|^b} \text{ w.p.a.1,} \\ \left| \frac{\partial \tilde{g}_l(z, \theta_0, \gamma_M)}{\partial \theta_k} - \frac{\partial \tilde{g}_l(z, \theta_0, \gamma_0)}{\partial \theta_k} \right| &\leq \frac{a}{|\gamma_M|^b} \text{ w.p.a.1,} \end{aligned}$$

for all  $z \in \mathcal{Z}$ ,  $k = 1, \dots, q$ , and  $l = 1, \dots, d$ .

Assumption G (i) requires that  $\gamma_0$  is at infinity in the sense that the objective function is bounded away from zero for small  $|\gamma|$ . Assumption G (ii) guarantees that  $\gamma_0$  uniquely achieves the moment equality  $E[\tilde{g}(\cdot, \theta_0, \gamma)] = 0$ . Under this assumption, we propose to choose  $\hat{\gamma}(\theta_0)$  as

$$\hat{\gamma}_M(\theta_0) = \arg \min_{\gamma: r_M \leq |\gamma| \leq R_M} \frac{|\bar{g}(\theta_0, \gamma)|^2}{|\gamma|}, \quad (15)$$

where  $r_M$  and  $R_M$  are tuning constants satisfying:  $M^{-(1+\beta-b)/2b} r_M \rightarrow \infty$ ,  $M^{-(1+\beta)/2b} R_M \rightarrow c \in (0, \infty)$ , and  $r_M/R_M \rightarrow 0$  as  $M \rightarrow \infty$ . For example, one can choose  $r_M \propto M^{(1+\beta-b/2)/2b}$  and  $R_M \propto M^{(1+\beta)/2b}$ .

Under Assumption G, we can show that the following result on  $\hat{\gamma}_M(\theta_0)$ .

**Proposition G.** *Under Assumption G,*

$$\begin{aligned} \sqrt{M} |\tilde{g}_l(z, \theta_0, \hat{\gamma}_M(\theta_0)) - \tilde{g}_l(z, \theta_0, \gamma_0)| &\xrightarrow{p} 0, \\ \sqrt{M} \left| \frac{\partial \tilde{g}_l(z, \theta_0, \hat{\gamma}_M(\theta_0))}{\partial \theta_k} - \frac{\partial \tilde{g}_l(z, \theta_0, \gamma_0)}{\partial \theta_k} \right| &\xrightarrow{p} 0, \end{aligned}$$

for all  $z \in \mathcal{Z}$ ,  $k = 1, \dots, q$ , and  $l = 1, \dots, d$ .

This proposition implies that the limiting distribution of  $\zeta_M(\theta_0)$  in Assumption K (ii) is determined by  $\mathbb{G}_M \tilde{g}(\cdot, \theta_0, \gamma_0)$  and  $\text{vec}(\mathbb{G}_M \tilde{G}(\cdot, \theta_0, \gamma_0))$  and Assumption K (ii) is easily satisfied.

### Proof of Proposition G

Pick any  $B > 0$ ,  $z \in \mathcal{Z}$ , and  $l = 1, \dots, d$ . By Assumption G (iii) and Lemma G below, we have

$$\begin{aligned} &\Pr\{\sqrt{M} |\tilde{g}_l(z, \theta_0, \hat{\gamma}_M(\theta_0)) - \tilde{g}_l(z, \theta_0, \gamma_0)| \geq B\} \\ &\leq \Pr\{|\hat{\gamma}_M(\theta_0)| \leq B_1 M^{\frac{1}{2b}}\} + o(1) \\ &\leq \frac{C}{\sqrt{M}} \frac{\frac{R_M}{d} \left( \alpha + B_1^\beta M^{\frac{\beta}{2b}} \right) B_1 M^{\frac{1}{2b}}}{r_M \left[ \frac{R_M}{d} - \left( \alpha + B_1^\beta M^{\frac{\beta}{2b}} \right) B_1 M^{\frac{1}{2b}} \right]} + o(1), \end{aligned}$$

for some  $C > 0$ , where  $B_1 = (a/B)^{1/b}$ . Since the last expression converges to zero due to the conditions of  $r_M$  and  $R_M$ , we obtain the first statement of this proposition. The second statement follows by similar arguments.

It remains to show the following lemma.

**Lemma G.** Under Assumption G (i)-(ii), there exists a positive constant  $C$  such that

$$\Pr\{|\hat{\gamma}_M(\theta_0)| \leq c\} \leq \frac{C}{\sqrt{M}} \frac{\frac{R_M}{d}(\alpha + c^\beta)c}{r_M \left[\frac{R_M}{d} - (\alpha + c^\beta)c\right]},$$

for all  $c \in (r_M, R_M)$ ,  $r_M > 0$ , and  $R_M > r_M$  satisfying  $R_M/d > (\alpha + c^\beta)c$ , where  $C$  is universal across  $r_M$ ,  $R_M$ , and  $r$ .

### Proof of Lemma G

Let  $\check{g}_M(\theta, \gamma) = \sqrt{\frac{r_M}{|\gamma|}} \check{g}(\theta, \gamma)$ . Pick any  $\theta_0 \in \Theta_0$  and  $c \in (r_M, R_M)$ . Observe that

$$\begin{aligned} \Pr\{|\hat{\gamma}_M(\theta_0)| \leq c\} &\leq \Pr\left\{\inf_{\gamma: r_M \leq |\gamma| \leq c} |\check{g}_M(\theta_0, \gamma)|^2 \leq |\check{g}_M(\theta_0, \gamma_c)|^2\right\} \\ &\leq \Pr\left\{\inf_{\gamma: r_M \leq |\gamma| \leq c} |E[\check{g}_M(\theta_0, \gamma)]|^2 - |E[\check{g}_M(\theta_0, \gamma_c)]|^2 \leq 2 \sum_{l=1}^d \sup_{\gamma: r_M \leq |\gamma|} |\check{g}_{M,l}(\theta_0, \gamma)^2 - (E[\check{g}_{M,l}(\theta_0, \gamma)])^2|\right\} \\ &\leq \Pr\left\{\inf_{\gamma: r_M \leq |\gamma| \leq c} |E[\check{g}_M(\theta_0, \gamma)]|^2 - |E[\check{g}_M(\theta_0, \gamma_c)]|^2 \leq 4 \sum_{l=1}^d \sup_{\gamma: r_M \leq |\gamma|} |\check{g}_{M,l}(\theta_0, \gamma) - E[\check{g}_{M,l}(\theta_0, \gamma)]|\right\} \\ &\leq \frac{4}{\sqrt{M}} \frac{\sum_{l=1}^d E\left[\sup_{\gamma: r_M \leq |\gamma|} \sqrt{M} |\check{g}_{M,l}(\theta_0, \gamma) - E[\check{g}_{M,l}(\theta_0, \gamma)]|\right]}{\inf_{\gamma: r_M \leq |\gamma| \leq c} |E[\check{g}_M(\theta_0, \gamma)]|^2 - |E[\check{g}_M(\theta_0, \gamma_c)]|^2}, \end{aligned}$$

for any  $\gamma_c$  satisfying  $c \leq |\gamma_c| \leq R_M$ , where the first inequality follows from the definition of  $\hat{\gamma}_M(\theta_0)$ , the second inequality follows from the triangle inequality, the third inequality follows from  $|X^2 - Y^2| \leq \max\{2|X|, 2|Y|\} \cdot |X - Y|$ , and the last inequality follows from the Markov inequality. Since it holds for any  $\gamma_c$  satisfying  $c \leq |\gamma_c| \leq R_M$ , we can take  $\gamma_c$  with  $|\gamma_c| = R_M$  to satisfy

$$|E[\check{g}_M(\theta_0, \gamma_c)]|^2 \leq d \frac{r_M}{R_M}.$$

By Assumption G (ii), we also have

$$\inf_{\gamma: r_M \leq |\gamma| \leq c} |E[\check{g}_M(\theta_0, \gamma)]|^2 \geq \frac{r_M}{(\alpha + c^\beta)c}.$$

Furthermore, note that the class  $\mathcal{G}_l = \left\{ \sqrt{\frac{r_M}{|\gamma|}} \check{g}_l(\cdot, \theta, \gamma) : |\gamma| \geq r_M \right\}$  of functions has the uniform covering number bounded by  $(1/\varepsilon)^{d_z}$  and the envelop function of  $\mathcal{G}_l$  has a bounded  $L^2(P)$  norm because  $\check{g}_l(\cdot, \theta, \gamma)$  is bounded. Therefore, by van der Vaart and Wellner (1996, Theorem 2.14.1), there exists a positive constant  $C_0$  such that

$$E\left[\sup_{\gamma: r_M \leq |\gamma|} \sqrt{\frac{r_M}{|\gamma|}} \sqrt{M} |\check{g}_l(\theta_0, \gamma) - E[\check{g}_l(\theta_0, \gamma)]|\right] \leq C_0,$$

where  $C_0$  is universal across  $r_M$  and  $R_M$ . Combining these results, we obtain the conclusion.