Quant Mark Econ (2014) 12:167–207
DOI 10.1007/s11129-014-9144-8

Estimating discrete-choice games of incomplete information: Simple static examples

Che-Lin Su

Received: 24 October 2012 / Accepted: 11 February 2014 / Published online: 4 April 2014
© Springer Science+Business Media New York 2014

Abstract We investigate the computational aspect of estimating discrete-choice games under incomplete information. In these games, multiple equilibria can exist. Also, different values of structural parameters can result in different numbers of equilibria. Consequently, under maximum-likelihood estimation, the likelihood function is a discontinuous function of the structural parameters. We reformulate the maximum-likelihood estimation problem as a constrained optimization problem in the joint space of structural parameters and economic endogenous variables. Under this formulation, the objective function and structural equations are smooth functions. The constrained optimization approach does not require repeatedly solving the game or finding all the equilibria. We use two static-game models to demonstrate this approach, conducting Monte Carlo experiments to evaluate the finite-sample performance of the maximum-likelihood estimator, two-step estimators, and the nested pseudo-likelihood estimator.

Keywords Structural estimation · Discrete-choice games of incomplete information · Constrained optimization · Multiple equilibria

JEL Classification C13 · C57 · C61

1 Introduction

During the past decade, estimating empirical models of games of incomplete information has become an important and active research area in industrial organization, applied econometrics and quantitative marketing; see Seim (2006), Ellickson and

C.-L. Su (✉)
Booth School of Business, The University of Chicago, 5807 S. Woodlawn Avenue,
Chicago, IL 60637, USA
e-mail: che-lin.su@chicagobooth.edu

Misra (2008, 2011), Sweeting (2009), Zhu and Singh (2009), Bajari et al. (2010), Vitorino (2012), Misra (2013) as well as Orhun (2013). An important feature of empirical models of these games is that multiple equilibria can exist. This multiplicity leads to the issue of which equilibria are being played in the data. To date, with the exception of moment inequality estimators (see, for example, Ciliberto and Tamer 2009, and Pakes et al. 2011), it is commonly assumed that only one equilibrium is played in each market in the data. Even under this assumption, however, it is still conceivable that the number of equilibria can change for different values of structural parameters, which raises another potential issue for maximum-likelihood estimation of games: the likelihood function, which is defined as a function of structural parameters, can be discontinuous. When applying the nested fixed-point algorithm, researchers are confronted with two computational tasks: first, for each candidate vector of parameters considered, all Nash equilibria must be found in order to evaluate the corresponding likelihood value; second, the objective (likelihood) function is potentially discontinuous. For the first task, no computational methods can guarantee finding all the equilibria of a game, unless the equilibrium equations form a system of polynomial equations; see Judd et al. (2012). Without finding all the equilibria, the likelihood function can be mis-specified. For the second task, no reliable algorithm exists to maximize a discontinuous function. Although some heuristic approaches (for example, grid search or genetic algorithms) can be applied to find an approximated solution, in practice these methods tend to be very slow and, typically, do not find good approximations.

Because of the potential computational costs that make implementing the nested fixed-point algorithm impractical, researchers have proposed two-step estimators for estimating games; see Bajari et al. (2007), Pakes et al. (2007), Pesendorfer and Schmidt-Dengler (2008), and Arcidiacono and Miller (2011). The main advantage of two-step estimators is their computational simplicity: they do not require solving for an equilibrium. In a dynamic-game setting, where the cost of computing an equilibrium can increase drastically as the number of firms and/or state space increases, two-step estimators offer an attractive alternative to estimating structural models. However, the performance of two-step estimators depends largely on the accuracy of estimates in the first step and the criterion function used in the second step; see the discussion in Pakes et al. (2007). For dynamic games with many states, more data will be required to obtain an accurate estimate in the first step. Furthermore, if researchers wish to conduct counterfactual policy analysis after obtaining parameter estimates, then they still need to solve for an equilibrium (or the equilibria) of the game.

Aguirregabiria and Mira (2007) have proposed the nested pseudo-likelihood (NPL) estimator for estimating discrete-choice games. The NPL estimator aims to decrease the potentially large finite-sample biases with two-step estimators. They also proposed the NPL algorithm, a recursive iteration of a two-step pseudo maximum-likelihood estimator, to compute a solution of the NPL estimator. When the NPL algorithm converges, it solves the structural equations and, hence, produces an equilibrium of the game. It has, however, been shown by Pakes et al. (2007) as well as Pesendorfer and Schmidt-Dengler (2008) that the NPL estimator can perform worse than two-step estimators in finite samples. Moreover, to achieve convergence, the NPL algorithm requires an implicit assumption that the equilibria that generate the observed data are stable under best response iterations or Lyapunov stable. This requirement imposes an undesirable equilibrium selection criterion on the data generating process. Pesendorfer and Schmidt-Dengler (2010) has argued that best response stability is not a convincing assumption for games of incomplete information. When the best-response stable assumption is not satisfied, the NPL algorithm will either fail to converge or, worse, converge to the wrong equilibria, which leads to incorrect parameter estimates for the NPL estimator; see Pesendorfer and Schmidt-Dengler (2010) for such an example.

Su and Judd (2012) have proposed a constrained optimization approach to estimate structural models. They have demonstrated the use of their approach on the bus-engine replacement model in Rust (1987), and presented a general formulation that, in principle, can be applied to estimate games with multiple equilibria. Using the insight from Su and Judd (2012), Vitorino (2012) was the first to use the constrained optimization approach to estimate an empirical entry-and-exit game in the shopping center industry.

One goal in this paper is to investigate the computational aspect of using the constrained optimization approach to maximum-likelihood estimation of static discrete-choice games of incomplete information. Assuming that only one equilibrium is played in each market in the data, we present a constrained optimization formulation to maximum-likelihood estimation of discrete-choice games of incomplete information. Our formulation, defined over the joint space of both structural parameters and economic equilibrium, yields a smooth likelihood function as the objective, and smooth structural equations as the constraints. The constrained optimization approach does not require repeatedly solving for an equilibrium or all the equilibria at each guess of structural parameters. Thus, this approach reduces the perceived computational burden of implementing the maximum-likelihood estimator. While we do not claim that the constrained optimization approach can solve large-scale dynamic games with billions of states, variables, and equations, we believe it offers a valuable alternative for estimating structural models with up to 100,000 variables and constraints, which can accommodate models in many empirical papers in the literature.

We also examined the computational performance of various estimators for estimating static discrete-choice games. Using two static-game models as organizing examples, we conducted Monte Carlo experiments to evaluate the finite-sample properties of the maximum-likelihood (ML), the two-step pseudo maximum-likelihood (2S-PML), the two-step least squares (2S-LS), and the NPL estimators. We examined different data generating processes, in which best-response stable and/or best-response unstable equilibria were used. Conditional on the same observables, we also varied the numbers of repeated observations in the data sets. These Monte Carlo exercises considered here are reasonably general and the static game model resembles to those studied in the literature.

Our Monte Carlo results demonstrate that, in most cases, the ML estimator performs the best. For all the estimators, the biases in the estimators decrease as the numbers of repeated observations increase in the data. The ML estimator produces average estimates that are within one standard deviation of the true parameter values, even when only five or ten repeated observations per market exist in the data.

The performance of the two-step estimators depends on the type of equilibria used in the data generating process. If the data are generated by best-response stable equilibria, then the 2S-PML estimator produces accurate parameter estimates with moderate numbers of repeated observations, says twenty five or more per market. When the data consist of best-response stable and best-response unstable equilibria, the two-step estimators have high finite-sample biases in almost all cases; the biases decrease considerably with large numbers (100 or 250) of repeated observations. In the experiments, in most cases, the NPL estimator is more biased than the ML estimator. When the data consist of best-response stable and unstable equilibria, the NPL algorithm fails to converge for all data sets in the experiment using the first model; the NPL algorithm converges in around ten data sets (out of one hundred) but produces highly biased parameter estimates in those converged runs in the experiment using the second model. Even when the data are generated by only best-response stable equilibria given there are multiple equilibria in some markets at true parameter values, the NPL algorithm, surprisingly, often fails to converge with small numbers of repeated observations per market; for example, with five repeated observations per market in the data, the NPL algorithm converges in less than five data sets using the first model and converges in sixty five data sets using the second model. The NPL algorithm converges more frequently with more repeated observations per market. This finding suggests that the NPL algorithm is not a reliable computational strategy unless the underlying equilibria in the data generating process are stable under best-response iterations and large numbers of repeated observations exist in the data.

The remainder of the paper is organized as follows: in Section 2, we describe a simple static discrete-choice game of incomplete information originally proposed by Rust (2008), while in Section 3, we discuss estimating this game using the method of maximum-likelihood under two computational strategies: the nested fixed-point (NFXP) algorithm and the constrained optimization approach. Using a numerical example, we illustrate that the likelihood function of NFXP is discontinuous. Thus, the outer-loop optimization problem in NFXP will be computationally intractable. we then present the constrained optimization formulation for the maximum-likelihood estimation and prove that the NFXP algorithm and the constrained optimization approach give the same objective (likelihood) value and same solution. We also discuss the two-step estimators as well as the NPL estimator. In Section 4, we present the results of the Monte Carlo experiments using the simple static game example, reporting the finite-sample performance of the ML estimator, the two-step estimators and the NPL estimator. In Section 5, we consider a more realistic and empirically relevant static game example developed by Ellickson and Misra (2011) based on the dynamic game model in Jia (2008) and present Monte Carlo results as well as post-estimation equilibrium analysis on We summarize the conclusions in Section 6.

2 Static discrete-choice games of incomplete information

The description of the example presented here follows closely the derivation of Rust (2008). For simplicity, we first describe the model for one market and then generalize the model to accommodate data observed in several differentiated markets. Similar but more general models of static discrete-choice games of incomplete information have been proposed and studied by Seim (2006), Sweeting (2009), Zhu and Singh (2009), Bajari et al. (2010), Misra and Ellickson (2008, 2011), Vitorino (2012) as well as Orhun (2013).

2.1 A simple static-game model with one market

Consider a two-player, discrete-choice game of incomplete information with *observed* as well as *unobserved* heterogeneity. Let the firms be labeled a and b, and let y_a and y_b denote the choices of firms a and b, respectively. For simplicity, we assume that each firm has two possible choices; see, for example, the entry-and-exit games in Berry (1992), and Seim (2006) as well as Ciliberto and Tamer (2009).

Let

$$y_a = \begin{cases} 1 & \text{if firm } a \text{ is active,} \\ 0 & \text{if firm } a \text{ is inactive;} \end{cases}$$

and define y_b similarly. The *ex post* utility functions of firms a and b are assumed to be

$$u_a(y_a, y_b, x_a, \epsilon_a) = \begin{cases} [\alpha + y_b(\beta - \alpha)]\, x_a + \varepsilon_{a1}, & \text{if } y_a = 1, \\ 0 + \varepsilon_{a0}, & \text{if } y_a = 0; \end{cases}$$

and

$$u_b(y_a, y_b, x_b, \epsilon_b) = \begin{cases} [\alpha + y_a(\beta - \alpha)]\, x_b + \varepsilon_{b1}, & \text{if } y_b = 1, \\ 0 + \varepsilon_{b0}, & \text{if } y_b = 0; \end{cases}$$

where a scalar x_a is the observed type and a (2×1) vector $\boldsymbol{\varepsilon_a} = (\varepsilon_{a0}, \varepsilon_{a1})$ is the unobserved type for firm a. For firm b, x_b and $\boldsymbol{\varepsilon_b}$ are defined similarly. The structural parameters (to be estimated) (α, β) measure the effect of the observed type x_a and x_b on firm a's and b's utility, respectively. Note that a firm's utility is a function of the joint decision of both firms, (y_a, y_b).

We assume that the observed types (x_a, x_b) are common knowledge among firms; that unobserved types $(\boldsymbol{\varepsilon_a}, \boldsymbol{\varepsilon_b})$ are private information; that $\boldsymbol{\varepsilon_a}$ and $\boldsymbol{\varepsilon_b}$ are independent; and that firm a knows the distribution of $\boldsymbol{\varepsilon_b}$ and firm b knows the distribution of $\boldsymbol{\varepsilon_a}$.

Because of the private information $\boldsymbol{\varepsilon_a}$, firm a's decision will be probabilistic from firm b's point of view. Let p_a denote firm b's belief of the probability that firm a will be active. Similarly, let p_b denote firm a's belief of the probability that firm b will be active. Given firm b's belief p_a, the expected utility of firm b for taking an action y_b is given by

$$\begin{aligned} U_b(y_b, x_b, \boldsymbol{\varepsilon_b}) &= p_a u_b(1, y_b, x_b, \boldsymbol{\varepsilon_b}) + (1 - p_a) u_b(0, y_b, x_b, \boldsymbol{\varepsilon_b}) \\ &= \begin{cases} p_a \beta x_b + (1 - p_a)\alpha x_b + \varepsilon_{b1}, & \text{if } y_b = 1, \\ \varepsilon_{b0}, & \text{if } y_b = 0. \end{cases} \end{aligned}$$

It follows that firm b will be active ($y_b = 1$) if and only if

$$U_b(1, x_b, \boldsymbol{\varepsilon_b}) > U_b(0, x_b, \boldsymbol{\varepsilon_b}).$$

We assume that each component in the error terms $\boldsymbol{\varepsilon_a}$ and $\boldsymbol{\varepsilon_b}$ has a Type 1, extreme-value distribution, so the probability density function is $f(\varepsilon) = \exp(\varepsilon)\exp[-\exp(\varepsilon)]$. Given firm b's belief p_a, the probability that firm b will be active is given by the standard binomial logit formula

$$\begin{aligned} p_b &= \Pr(y_b = 1) \\ &= \Pr\left[\boldsymbol{\varepsilon_b} | U_b(y_b = 1, x_b, \boldsymbol{\varepsilon_b}) > U_b(y_b = 0, x_b, \boldsymbol{\varepsilon_b})\right] \\ &= \frac{\exp[p_a \beta x_b + (1 - p_a)\alpha x_b]}{1 + \exp[p_a \beta x_b + (1 - p_a)\alpha x_b]} \\ &= \frac{1}{1 + \exp[-x_b\alpha + p_a x_b(\alpha - \beta)]} \\ &\equiv \Psi_b(p_a, p_b, x_b; \alpha, \beta). \end{aligned} \tag{1}$$

This formula can be thought of as a *best response function* for firm b given b's belief p_a. Similarly, the best response for firm a, given a's belief p_b, is

$$\begin{aligned} p_a &= \frac{\exp[p_b \beta x_a + (1 - p_b)\alpha x_a]}{1 + \exp[p_b \beta x_a + (1 - p_b)\alpha x_a]} \\ &= \frac{1}{1 + \exp[-x_a\alpha + p_b x_a(\alpha - \beta)]} \\ &\equiv \Psi_a(p_a, p_b, x_a; \alpha, \beta). \end{aligned} \tag{2}$$

A *Bayes–Nash equilibrium* is a pair of beliefs $\left(p_a^*, p_b^*\right)$ that are mutual best responses:

$$\begin{aligned} p_a^* &= \frac{1}{1 + \exp[-x_a\alpha + p_b^* x_a(\alpha - \beta)]} = \Psi_a(p_a^*, p_b^*, x_a; \alpha, \beta) \\ p_b^* &= \frac{1}{1 + \exp[-x_b\alpha + p_a^* x_b(\alpha - \beta)]} = \Psi_b(p_a^*, p_b^*, x_b; \alpha, \beta). \end{aligned} \tag{3}$$

To simplify the notation, let $\boldsymbol{\theta} = (\alpha, \beta)$, $\boldsymbol{x} = (x_a, x_b)$, $\boldsymbol{p} = (p_a, p_b)$ and $\boldsymbol{\Psi} = (\Psi_a, \Psi_b)$. We rewrite the Bayes–Nash (BN) equilibrium (3) as

$$\boldsymbol{p} = \boldsymbol{\Psi}(\boldsymbol{p}, \boldsymbol{x}; \boldsymbol{\theta}). \tag{4}$$

Given the parameters $\boldsymbol{\theta}$ and observed types $\boldsymbol{x}$, there are two unknowns in $\boldsymbol{p}$ in the two BN equilibrium equations defined in Eq. 3. Multiple solutions satisfying (3) can exist and, hence, multiple BN equilibria can exist. However, one can show that there are at most three equilibria in this model. The example below illustrates this case.

Example 1 Suppose $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0) = (5, -11)$ and firms' observed types $\boldsymbol{x} = (0.52, 0.22)$. Substituting these values into Eq. 3 and solving the following two equations for a BN equilibrium $\boldsymbol{p}^*$:

$$\begin{aligned} p_a &= \frac{1}{1 + \exp(-2.60 + 8.32 p_b)} \\ p_b &= \frac{1}{1 + \exp(-1.10 + 3.52 p_a)}. \end{aligned} \tag{5}$$

We used the constrained optimization solver, KNITRO, to solve this system of two equations and two unknowns. Since there could be multiple solutions, we tried one hundred different starting points and found the following three BN equilibria:

$$\begin{aligned} \boldsymbol{p}_1^* &= (0.030100, 0.729886), \\ \boldsymbol{p}_2^* &= (0.616162, 0.255615), \\ \boldsymbol{p}_3^* &= (0.773758, 0.164705). \end{aligned}$$

One can verify that the equilibrium $\boldsymbol{p}_2^*$ is not stable under best-response iterations, which means that best-response iterations will not converge to $\boldsymbol{p}_2^*$, even if the starting point is very close to that solution.[1] Note that best-response stable equilibrium is not among the common notions of stable equilibrium, such as strategic stability of Kohlberg and Mertens (1986), or Mertens (1989, 1991), or the evolutionarily stable strategy (ESS), studied in the game theory literature; see also Chapter 11 in Fudenberg and Tirole (1991).

2.2 A simple static-game model with multiple markets

We generalize the model described above to accommodate multiple differentiated markets in the observed data. Assume that there are M markets. The characteristics that differentiate these markets are the firms' observed types. Thus, two vectors of different observed types represent two different markets. We denote by $\boldsymbol{x}^m = (x_a^m, x_b^m)$ the firms' observed types in market m, for $m = 1, \ldots, M$. We assume firms have the same vector of structural parameters $\boldsymbol{\theta} = (\alpha, \beta)$ in all markets, but firms' decisions are independent across the markets.

For each market, there is a set of BN equilibrium equations, parameterized by the observed types. Let $\boldsymbol{p}^m = (p_a^m, p_b^m)$ denote a vector of a BN equilibrium in market m, which satisfies the following equation:

$$\boldsymbol{p}^m = \boldsymbol{\Psi}(\boldsymbol{p}^m, \boldsymbol{x}^m; \boldsymbol{\theta}), \qquad \text{for } m = 1, \ldots, M. \tag{6}$$

Let $\boldsymbol{P} = (\boldsymbol{p}^m)_{m=1}^M$ denote the collection of equilibrium probabilities and $\boldsymbol{X} = (\boldsymbol{x}^m)_{m=1}^M$ denote the collection of the firms' observed types for all markets,

[1] An equilibrium $\boldsymbol{p}$ is stable under best-response iteration if $\rho[\nabla_{\boldsymbol{p}} \boldsymbol{\Psi}(\boldsymbol{p}, \boldsymbol{x}; \boldsymbol{\theta})]$, the spectral radius of the Jacobian mapping $\nabla_{\boldsymbol{p}} \boldsymbol{\Psi}(\boldsymbol{p}, \boldsymbol{x}; \boldsymbol{\theta})$, is less than 1. It is easy to check that at $\boldsymbol{p}_2^*$, $\rho[\nabla_{\boldsymbol{p}} \boldsymbol{\Psi}(\boldsymbol{p}_2^*, \boldsymbol{x}; \boldsymbol{\theta}^0)] = 1.148$, while at $\boldsymbol{p}_1^*$ and $\boldsymbol{p}_3^*$, $\rho[\nabla_{\boldsymbol{p}} \boldsymbol{\Psi}(\boldsymbol{p}_1^*, \boldsymbol{x}; \boldsymbol{\theta}^0)] = 0.4099$ and $\rho[\nabla_{\boldsymbol{p}} \boldsymbol{\Psi}(\boldsymbol{p}_3^*, \boldsymbol{x}; \boldsymbol{\theta}^0)] = 0.8398$, respectively.

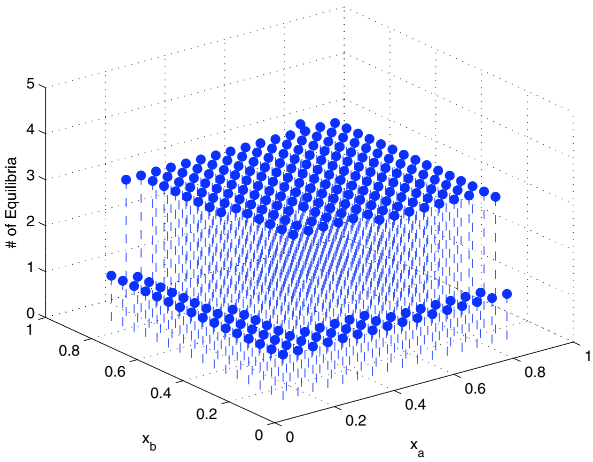


Fig. 1 Numbers of equilibria in different markets

respectively. With a slight abuse of notation, we simplify the BN equilibrium equations for all markets (6) as

$$\boldsymbol{P} = \boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}). \tag{7}$$

From Eq. 6, it is clear that markets with different observed types can have different BN equilibria. Also, the numbers of equilibria can be different for markets with different observed types, as illustrated in the following example.

Example 2 We selected the same parameter values $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0) = (5, -11)$ as in Example 1 and considered a case with 256 markets. Discretizing the interval [0.12, 0.87] into sixteen equally spaced grid points yields sixteen different observed types for each firm; jointly, there are $16 \times 16 = 256$ pairs of $\left(x_a^m, x_b^m\right)$, for $m = 1, \ldots, 256$. Each of these pairs defines a market. For each market m, we solved for the corresponding BN equilibrium (6) with 100 different starting values to find all the BN equilibria $\boldsymbol{p}^{m*} = \left(p_a^{m*}, p_b^{m*}\right)$ in that market.

In Fig. 1, we present the plot of the numbers of equilibria in each market. As one can see, there are three equilibria in most markets. For markets with low x_a and/or x_b, the equilibrium is, however, unique. A small change in the observed types can result in a relatively large change in the number of equilibria; for example, there are three equilibria in the market with observed types (0.17, 0.87), but there is only one equilibrium in the market with the observed types (0.12, 0.87).

3 Estimation

In this section, we describe the data generating process and discuss various estimators used to estimate static discrete-choice games of incomplete information.

3.1 Data generating process

Following a common assumption in the literature (see, for example, Aguirregabiria and Mira 2007, Bajari et al. 2007, Pakes et al. 2007 as well as Pesendorfer and Schmidt-Dengler 2008), we assume that in each market, only one equilibrium is played in the data. Since equilibrium solutions are different in different markets, as was demonstrated in Example 2, in the data we allow different equilibria to be played in different markets.

Assumption 2: In each market, the firms use the same equilibrium to play independently over the T periods. However, equilibria played across different markets are different.

Researchers observe firms' types $\boldsymbol{x}^m = \left(x_a^m, x_b^m\right)$ and decisions $\boldsymbol{y}^m = \left(y_a^{mt}, y_b^{mt}\right)_{t=1}^T$ in each market m over T periods in the data. Let $\boldsymbol{Z}^m$ denote the data observed in market m:

$$\boldsymbol{Z}^m = \left\{\boldsymbol{x}^m = \left(x_a^m, x_b^m\right), \ \boldsymbol{y}^m = \left(y_a^{mt}, y_b^{mt}\right)_{t=1}^T\right\}.$$

Data observed for all M markets over T periods are the collection of $\boldsymbol{Z}^m$ for all m, so

$$\boldsymbol{Z} = \{\boldsymbol{Z}^m, \ \text{for } m = 1, \ldots, M\}. \tag{8}$$

3.2 Maximum-likelihood estimation

We consider estimating the parameters of the discrete-choice game using the method of maximum-likelihood under Assumption 2. We first derive the logarithm of the likelihood function and the formulation of the maximum likelihood estimation problem. We then describe the NFXP algorithm for implementing the ML estimator and discuss its computational difficulties. To overcome the computational difficulties associated with the NFXP algorithm, we reformulate the ML estimation problem as a constrained optimization problem and show that the NFXP algorithm and the constrained optimization approach produce the same estimates.

3.2.1 Deriving the likelihood function for models with one market

We derive the logarithm of the likelihood function, first under a general equilibrium selection mechanism and then impose Assumption 2 that only one equilibrium is played in the data. To simplify the notation, we consider the case of one market and drop the superscript m for market index in the derivation below. We generalize the formulation for models with multiple markets in the next subsection.

Given the parameter vector $\boldsymbol{\theta}$ and firms' observed types $\boldsymbol{x}$, let $\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x}) = \{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k$ denote the set of BN equilibria, where $\bar{\boldsymbol{p}}_k(\boldsymbol{\theta}) = (\bar{p}_{ka}(\boldsymbol{\theta}), \bar{p}_{kb}(\boldsymbol{\theta}))$ is the k-th equilibrium that solves the BN equilibrium (4). Also, let $\boldsymbol{\lambda}(\boldsymbol{\theta}) = \{\lambda_k(\boldsymbol{\theta})\}_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}$ be the specified equilibrium selection mechanism, where $\lambda_k(\boldsymbol{\theta})$ is the probability that the k-th equilibrium is played in the data with $\sum_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|} \lambda_k(\boldsymbol{\theta}) = 1$. Following the derivation in Sweeting (2009) as well as de Paula (2013, Section 4), the logarithm of the likelihood function of observing the decisions $\boldsymbol{y} = \left(y_a^t, y_b^t\right)_{t=1}^T$ in the market at the parameters $\boldsymbol{\theta}$ is

$$\begin{aligned}&\mathbb{L}[\boldsymbol{\theta}, \boldsymbol{\lambda}(\boldsymbol{\theta}), \{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k]\\&=\sum_{t=1}^{T}\log\left\{\sum_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}\lambda_k(\boldsymbol{\theta})\left[\bar{p}_{ka}(\boldsymbol{\theta})^{y_a^t}\right]\left[(1-\bar{p}_{ka}(\boldsymbol{\theta}))^{1-y_a^t}\right]\left[\bar{p}_{kb}(\boldsymbol{\theta})^{y_b^t}\right]\left[(1-\bar{p}_{kb}(\boldsymbol{\theta}))^{1-y_b^t}\right]\right\}.\end{aligned} \tag{9}$$

The ML estimator $\left(\boldsymbol{\theta}^{MLE}, \boldsymbol{\lambda}^{MLE}(\boldsymbol{\theta}^{MLE})\right)$ then solves

$$\underset{\{\boldsymbol{\theta},\boldsymbol{\lambda}(\boldsymbol{\theta})\}}{\text{maximize}}\ \mathbb{L}[\boldsymbol{\theta}, \boldsymbol{\lambda}(\boldsymbol{\theta}), \{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k], \quad \text{s.t.} \sum_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}\lambda_k(\boldsymbol{\theta}) = 1. \tag{10}$$

Since the number of elements in the decision variables $\boldsymbol{\lambda}(\boldsymbol{\theta})$ varies, depending on the value of decision variables $\boldsymbol{\theta}$, the maximization problem formulated above is not computationally tractable. However, given Assumption 2 that only one equilibrium is played in the data, only one component in $\boldsymbol{\lambda}(\boldsymbol{\theta})$ as well as $\boldsymbol{\lambda}^{MLE}\left(\boldsymbol{\theta}^{MLE}\right)$ is one and all the other components are 0. Define

$$\begin{aligned}&L_i[(\boldsymbol{x}, \boldsymbol{y}); \bar{\boldsymbol{p}}_k(\boldsymbol{\theta}); \boldsymbol{\theta}]\\&=\sum_{t=1}^{T}\log\left\{\left[\bar{p}_{ka}(\boldsymbol{\theta})^{y_a^t}\right]\left[(1-\bar{p}_{ka}(\boldsymbol{\theta}))^{1-y_a^t}\right]\left[\bar{p}_{kb}(\boldsymbol{\theta})^{y_b^t}\right]\left[(1-\bar{p}_{kb}(\boldsymbol{\theta}))^{1-y_b^t}\right]\right\}\end{aligned} \tag{11}$$

Under Assumption 2, the maximization problem (10) is equivalent to

$$\begin{aligned}&\underset{\{\boldsymbol{\theta},\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\in\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})\}}{\text{maximize}}\ L_i[(\boldsymbol{x}, \boldsymbol{y}); \bar{\boldsymbol{p}}_k(\boldsymbol{\theta}); \boldsymbol{\theta}]\\&=\max_{\boldsymbol{\theta}}\left\{\max_{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\in\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})} L_i[(\boldsymbol{x}, \boldsymbol{y}); \bar{\boldsymbol{p}}_k(\boldsymbol{\theta}); \boldsymbol{\theta}]\right\}.\end{aligned}$$

Hence, the ML estimator for models with one market is defined as

$$\boldsymbol{\theta}^{MLE} = \underset{\boldsymbol{\theta}}{\text{argmax}}\left\{\max_{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\in\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})} L_i[(\boldsymbol{x}, \boldsymbol{y}); \bar{\boldsymbol{p}}_k(\boldsymbol{\theta}); \boldsymbol{\theta}]\right\}. \tag{12}$$

3.2.2 ML estimator for models with multiple markets

We present the formulation of the ML estimator for models with multiple markets. For ease of presentation, we provide a generalization of the ML estimator defined in Eqs. 11 and 12, and do not re-derive the likelihood function under a general equilibrium selection mechanism.

Given the parameter vector $\boldsymbol{\theta}$ and firms' observed types $\boldsymbol{x}^m$ for market m, let $\bar{\boldsymbol{p}}^m(\boldsymbol{\theta}) = (\bar{p}_a(\boldsymbol{x}^m; \boldsymbol{\theta}), \bar{p}_b(\boldsymbol{x}^m; \boldsymbol{\theta}))$ denote a BN equilibrium that solves (6). Given Assumption 2 and following Eq. 11, if the observed decisions $\boldsymbol{y}^m = \left(y_a^{mt}, y_b^{mt}\right)_{t=1}^T$ were generated by $\bar{\boldsymbol{p}}^m(\boldsymbol{\theta})$, then the logarithm of the likelihood of the observing data $\boldsymbol{Z}^m = (\boldsymbol{x}^m, \boldsymbol{y}^m)$ in market m at the parameters $\boldsymbol{\theta}$ is

$$\begin{aligned} L_i[\boldsymbol{Z}^m; \bar{\boldsymbol{p}}^m(\boldsymbol{\theta}); \boldsymbol{\theta}] \\ = \sum_{t=1}^T \left\{ y_a^{mt} \times \log[\bar{p}_a(\boldsymbol{x}^m; \boldsymbol{\theta})] + (1 - y_a^{mt}) \times \log[1 - \bar{p}_a(\boldsymbol{x}^m; \boldsymbol{\theta})] \right\} \\ + \sum_{t=1}^T \left\{ y_b^{mt} \times \log[\bar{p}_b(\boldsymbol{x}^m; \boldsymbol{\theta})] + (1 - y_b^{mt}) \times \log[1 - \bar{p}_b(\boldsymbol{x}^m; \boldsymbol{\theta})] \right\}, \end{aligned} \tag{13}$$

Let $\bar{\boldsymbol{P}}(\boldsymbol{\theta}) = (\bar{\boldsymbol{p}}^m(\boldsymbol{\theta}))_{m=1}^M$. Thus, the logarithm of the likelihood of observing the data $\boldsymbol{Z}$ for all markets at the parameters $\boldsymbol{\theta}$ is

$$L[\boldsymbol{\theta}, \bar{\boldsymbol{P}}(\boldsymbol{\theta})] = \sum_{m=1}^M L_i[\boldsymbol{Z}^m; \bar{\boldsymbol{p}}^m(\boldsymbol{\theta}), \boldsymbol{\theta}]. \tag{14}$$

Since multiple BN equilibria can exist for each $\boldsymbol{\theta}$, the ML estimator is defined as

$$\boldsymbol{\theta}^{MLE} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ \max_{\bar{\boldsymbol{P}}(\boldsymbol{\theta})} L[\boldsymbol{\theta}, \bar{\boldsymbol{P}}(\boldsymbol{\theta})] \right\}. \tag{15}$$

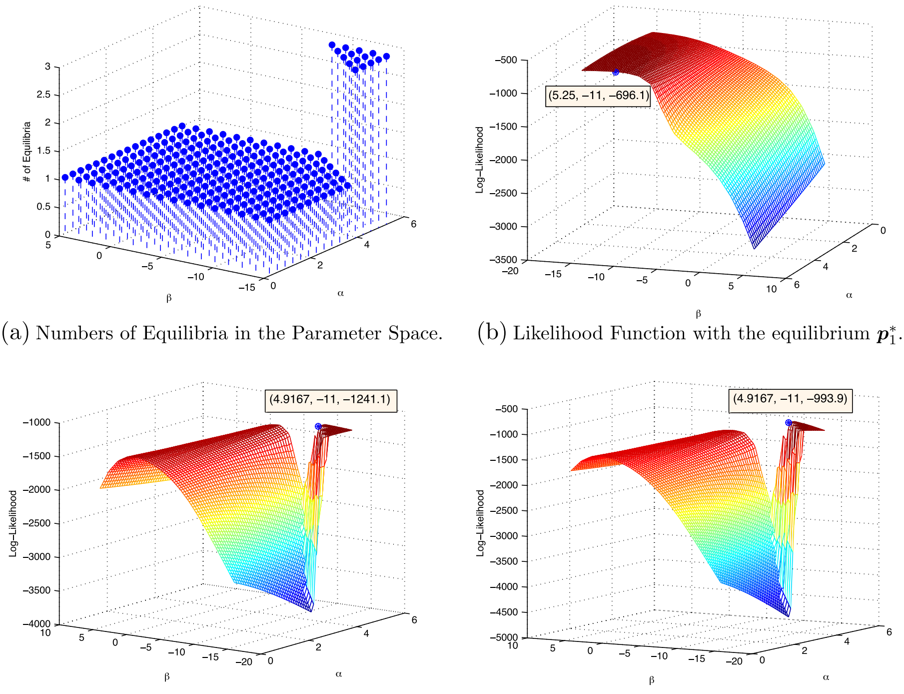
3.2.3 Solving the ML estimator using the NFXP approach

The NFXP algorithm of Rust (1987) has been proposed to compute the ML estimator $\boldsymbol{\theta}^{MLE}$ defined in Eq. 15. The implementation of the NFXP algorithm is described as follows: in the outer loop, search the structural parameter space over $\boldsymbol{\theta}$ to maximize the objective function $\left\{ \max_{\bar{\boldsymbol{P}}(\boldsymbol{\theta})} L[\boldsymbol{\theta}, \bar{\boldsymbol{P}}(\boldsymbol{\theta})] \right\}$; in the inner loop, for a given $\boldsymbol{\theta}$, find all the BN equilibria, evaluate the corresponding likelihood value at each BN equilibrium, and choose the equilibrium that yields the highest likelihood value. The algorithm then returns to the outer loop and repeats until it converges, or fails.

Two important computational difficulties arise when applying NFXP to solve the ML estimator (15): first, researchers must find all the equilibria for any given structural parameters $\boldsymbol{\theta}$; second, the objective function $\left\{ \max_{\bar{\boldsymbol{P}}(\boldsymbol{\theta})} L[\boldsymbol{\theta}, \bar{\boldsymbol{P}}(\boldsymbol{\theta})] \right\}$ can be a discontinuous function of $\boldsymbol{\theta}$. In the example below, we illustrate such a case.

Example 3 Consider the setting in Example 1, in which there are three BN equilibria given $\boldsymbol{\theta}^0 = (\alpha^0, \beta^0) = (5, -11)$ and firms' observed types $\boldsymbol{x} = (0.52, 0.22)$. We first assume that both firms use the equilibrium $\boldsymbol{p}_1^*$ to play the game 1,000 times and, hence, randomly generate 1,000 pairs of observed decisions $\boldsymbol{y_1} = \left(y_{a1}^t, y_{b1}^t\right)_{t=1}^{1000}$ independently using $\boldsymbol{p}_1^*$. Let $\boldsymbol{Z}_1$ denote this data set. Repeat the same procedure, but use $\boldsymbol{p}_2^*$ and $\boldsymbol{p}_2^*$ to generate data sets $\boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$, respectively. Fixing $\boldsymbol{x} = (0.52, 0.22)$, we then plotted the numbers of equilibria in the parameter space $\boldsymbol{\theta}$ and the corresponding logarithm of the likelihood function for each of the three data sets in Fig. 2. As one can see, the logarithm of the likelihood function for $\boldsymbol{Z}_1$ is continuous in $\boldsymbol{\theta}$, while the logarithms of the likelihood functions for $\boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are discontinuous. Recall that the equilibrium $\boldsymbol{p}_2^*$ is unstable under best-response iterations, while $\boldsymbol{p}_2^*$ is stable under best response. This example demonstrates that the discontinuity in the likelihood function does not depend on best-response stability of an equilibrium. Instead, the discontinuity arises from the change in the numbers of equilibria for different $\boldsymbol{\theta}$s.

In practice, almost all reliable and efficient numerical methods for solving optimization problems are based on variants of Newton's method, which requires the objective function and the constraints in the underlying problems to be differentiable. To date, the research in the optimization field on solving nonsmooth problems is, although increasing, still quite sparse. To the best of our knowledge, we are not aware of any reliable numerical methods or software that can solve problems with discontinuous functions in a robust manner.



(a) Numbers of Equilibria in the Parameter Space. (b) Likelihood Function with the equilibrium $\boldsymbol{p}_1^*$.

(c) Likelihood Function with the equilibrium $\boldsymbol{p}_2^*$. (d) Likelihood Function with the equilibrium $\boldsymbol{p}_3^*$.

Fig. 2 Numbers of equilibria and examples of the logarithms of likelihood functions. The *circle* in **b** – **d** indicates the maximizer obtained from grid search

3.2.4 ML estimation using the constrained optimization approach

We next describe the constrained optimization approach to estimating games of incomplete information under the method of maximum-likelihood.[2]

Let $\boldsymbol{p}^m = \left(p_a^m, p_b^m\right)$ denote any vector of probabilities in market m and $\boldsymbol{P} = (\boldsymbol{p}^m)_{m=1}^M$. Assuming the observed decisions $\boldsymbol{y}^m = \left(y_a^{mt}, y_b^{mt}\right)_{t=1}^T$ were generated by $\boldsymbol{p}^m$, we define the *augmented* logarithm of the likelihood function for observing $\boldsymbol{Z}^m$ in market m as

$$\begin{aligned}\mathcal{L}_i(\boldsymbol{Z}^m; \boldsymbol{p}^m, \boldsymbol{\theta}) &= \sum_{t=1}^{T}\left[y_a^{mt} \times \log(p_a^m) + (1 - y_a^{mt}) \times \log(1 - p_a^m)\right] \\ &\quad + \sum_{t=1}^{T}\left[y_b^{mt} \times \log(p_b^m) + (1 - y_b^{mt}) \times \log(1 - p_b^m)\right].\end{aligned} \tag{16}$$

The augmented logarithm of the likelihood of observing the full data $\boldsymbol{Z}$ for all markets is

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{P}) = \sum_{m=1}^{M} \mathcal{L}_i(\boldsymbol{Z}^m; \boldsymbol{p}^m, \boldsymbol{\theta}). \tag{17}$$

To ensure that, given $\boldsymbol{\theta}$, $\boldsymbol{P}$ is a collection of BN equilibria for all markets, we imposed the BN equilibrium (7) as constraints. The ML estimation problem is then

$$\begin{aligned}&\max_{(\boldsymbol{\theta}, \boldsymbol{P})} \quad \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{P}) \\ &\text{subject to} \quad \boldsymbol{P} = \Psi(\boldsymbol{P}, \boldsymbol{x}, \boldsymbol{\theta}).\end{aligned} \tag{18}$$

The decision variables in our formulation are $\boldsymbol{\theta}$ and $\boldsymbol{P}$. Note, too, that the structural parameters $\boldsymbol{\theta}$ do not directly enter either the augmented function (16) or the objective function (17). Instead of defining the likelihood function as a (potentially) discontinuous function of θ in Eq. 15, the objective function $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{P})$ is a smooth function of the equilibrium probabilities $\boldsymbol{P}$. Thus, the constrained optimization approach yields a smooth objective function as well as smooth constraints. Consequently, one can use state-of-the-art constrained optimization solvers to compute a solution to the ML estimation problem (18).

Following Proposition 1 in Su and Judd (2012), we state the equivalence in the likelihood value and solutions of the two optimization problems formulated in Eqs. 15 and 18.

Proposition 1 *Let $\bar{\boldsymbol{\theta}}$ be a solution of the ML estimation problem defined in* Eq. 15. *Denote $\bar{\boldsymbol{P}}^*(\boldsymbol{\theta}) = \underset{\bar{\boldsymbol{P}}(\boldsymbol{\theta})}{\operatorname{argmax}} L[\boldsymbol{\theta}, \bar{\boldsymbol{P}}(\boldsymbol{\theta})]$. Let $(\boldsymbol{\theta}^*, \boldsymbol{P}^*)$ be a solution of the constrained optimization problem* (18). *Then $L[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{P}}^*(\bar{\boldsymbol{\theta}})] = \mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{P}^*)$. If the model is identified, then $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^*$.*

[2]The derivation of the constrained optimization formulation under a general equilibrium selection mechanism is given in Appendix A.

Proof See Proposition 1 in Su and Judd (2012). □

Aitchison and Silvey (1958) as well as Silvey (1975) have shown that the ML estimator formulated in Eq. 18 is consistent as well as asymptotically normal. Similar results are also stated in Section 10.3 in Gourieroux and Monfort (1995). Below we state the asymptotic results of the ML estimator defined in Eq. 18 in the framework of Aitchison and Silvey (1958), assuming that the number of market M is fixed and number of periods T goes to infinity. We refer the readers to Aitchison and Silvey (1958) for detailed derivation and proof, and Silvey (1975) for a concise discussion.

Inference. Let $\bar{\boldsymbol{\vartheta}} = (\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{P}})$ be a solution of ML estimation problem (18) and $\boldsymbol{\vartheta}^0 = (\boldsymbol{\theta}^0, \boldsymbol{P}^0)$ be the true parameter vector of the data generating process, respectively. Denote the information matrix $\mathcal{I}_0 = -\mathbb{E}\left[\dfrac{\partial^2 \mathcal{L}(\boldsymbol{\vartheta}^0)}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}\right]$ and the constraint Jacobian $\boldsymbol{H}_0 = \dfrac{h(\boldsymbol{\vartheta}^0)}{\partial \boldsymbol{\vartheta}}$ at the true parameter vector $\boldsymbol{\vartheta}^0$. Let I be an identity matrix. Under assumptions and regularity conditions stated in Aitchison and Silvey (1958), the random variable $\sqrt{T}(\bar{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^0)$ is asymptotically normally distributed with

$$\sqrt{T}(\bar{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_0),$$

where $\boldsymbol{\Sigma}_0 = \mathcal{I}_0^{-1}\left[I - \boldsymbol{H}_0'\left(\boldsymbol{H}_0 \mathcal{I}_0^{-1} \boldsymbol{H}_0'\right)^{-1} \boldsymbol{H}_0 \mathcal{I}_0^{-1}\right]$.

3.3 Two-step estimators

Hotz and Miller (1993) pioneered using two-step estimators to estimate single-agent dynamic discrete-choice models. One attractive feature of two-step estimators is computational simplicity, at least when compared to the NFXP algorithm or the constrained optimization approach. In this section, we describe the 2S-PML estimator and the 2S-LS estimator for estimating the discrete-choice game.

For a given vector of probabilities $\widehat{\boldsymbol{P}} = (\hat{\boldsymbol{p}}^m)_{m=1}^M$, the pseudo likelihood function of observing the data $\boldsymbol{Z}^m$ in market m is defined as

$$\begin{aligned}
&L_i^{PML}(\boldsymbol{Z}^m, \hat{\boldsymbol{p}}^m; \boldsymbol{\theta}) \\
&\quad = \sum_{t=1}^{T} \left\{ y_a^{mt} \times \log[\Psi_a(\hat{p}_a^m, \hat{p}_b^m, \boldsymbol{x}^m, \boldsymbol{\theta})] + (1 - y_a^{mt}) \times \log[1 - \Psi_a(\hat{p}_a^m, \hat{p}_b^m, \boldsymbol{x}^m, \boldsymbol{\theta})] \right\} \\
&\qquad + \sum_{t=1}^{T} \left\{ y_b^{mt} \times \log[\Psi_b(\hat{p}_a^m, \hat{p}_b^m, \boldsymbol{x}^m, \boldsymbol{\theta})] + (1 - y_b^{mt}) \times \log[1 - \Psi_b(\hat{p}_a^m, \hat{p}_b^m, \boldsymbol{x}^m, \boldsymbol{\theta})] \right\}.
\end{aligned}$$

The pseudo likelihood of observing the full data $\boldsymbol{Z}$ for all markets is then

$$L^{PML}(\boldsymbol{\theta}, \widehat{\boldsymbol{P}}) = \sum_{m=1}^{M} L_i^{PML}(\boldsymbol{Z}^m, \hat{\boldsymbol{p}}^m; \boldsymbol{\theta}). \tag{19}$$

Notice that when defining the pseudo likelihood function (19), the component $\hat{\boldsymbol{p}}^m$ in $\widehat{\boldsymbol{P}}$ need not be a BN equilibrium. The pesudo likelihood function is well defined for any probabilities $\widehat{\boldsymbol{P}}$.

The 2S-PML estimator is as follows:

Step 1: Find $\widehat{\boldsymbol{P}}_0$, a consistent estimate of the true equilibrium probabilities $\boldsymbol{P}^0$.

Step 2: Fix $\widehat{\boldsymbol{P}}_0$. Solve the pseudo maximum-likelihood estimator:

$$\boldsymbol{\theta}^{2S-PML} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ L^{PML}(\boldsymbol{\theta}, \widehat{\boldsymbol{P}}_0). \tag{20}$$

One can also use the 2S-LS estimator of Pesendorfer and Schmidt-Dengler (2008). Instead of maximizing the pseudo likelihood function, one chooses structural parameters $\boldsymbol{\theta}$ to minimize the ℓ_2 norm of the errors in the BN equilibrium (7) in the 2S-LS estimator:[3]

Step 1: Find $\widehat{\boldsymbol{P}}_0$, a consistent estimate of the true equilibrium probabilities $\boldsymbol{P}^0$.

Step 2: Fix $\widehat{\boldsymbol{P}}_0$. Solve the least square problem:

$$\boldsymbol{\theta}^{2S-LS} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ \|\widehat{\boldsymbol{P}}_0 - \Psi(\widehat{\boldsymbol{P}}_0, \boldsymbol{x}; \boldsymbol{\theta})\|_2^2. \tag{21}$$

For two-step estimators, the optimization problem in the second step involves only structural parameters. Also, the BN equilibrium equations are not imposed in the second step and, hence, may not be satisfied. Because two-step estimators do not solve the BN equilibrium equations, they are computationally light. However, two-step estimators can perform poorly in finite samples when the first-step estimates are imprecise due to insufficient amounts of data, or when researchers do not choose suitable criterion functions in the second step; see the discussion in Pakes et al. (2007).

3.4 The nested pseudo-likelihood estimator

Recognizing the limitations of two-step estimators, Aguirregabiria and Mira (2007) have proposed the NPL estimator for estimating dynamic discrete games, aiming to reduce the finite-sample biases of the 2S-PML. They also proposed the NPL algorithm, a recursive iteration of the 2S-PML estimator, to compute a solution of the NPL estimator.

The NPL algorithm is described as follows: first, find an initial guess of equilibrium probabilities $\widetilde{\boldsymbol{P}}_0$. For $K \geq 1$, fix $\widetilde{\boldsymbol{P}}_{K-1}$ and solve the pseudo maximum-likelihood estimator for $\tilde{\boldsymbol{\theta}}_K$:

$$\tilde{\boldsymbol{\theta}}_K = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ L^{PML}(\boldsymbol{\theta}, \widetilde{\boldsymbol{P}}_{K-1}). \tag{22}$$

[3] Pesendorfer and Schmidt-Dengler (2008) proposed an asymptotic least-squares estimator and derived the asymptotically optimal weight matrix. Here, we do not derive the optimal weight matrix and use the identity matrix as the weight matrix in this example.

Given $\tilde{\boldsymbol{\theta}}_K$, obtain $\widetilde{\boldsymbol{P}}_K$ by one best-response iteration on the equilibrium (7):

$$\widetilde{\boldsymbol{P}}_K = \Psi(\widetilde{\boldsymbol{P}}_{K-1}, \boldsymbol{x}; \tilde{\boldsymbol{\theta}}_K). \tag{23}$$

Increase K by 1 and repeat the above procedure until convergence or if the maximum number of iterations $\bar{K}$ is reached, then declare a failed run and restart with a new initial guess.

When the NPL algorithm converges, a solution $(\boldsymbol{\theta}^{NPL}, \boldsymbol{P}^{NPL})$ satisfies the BN equilibrium (7). However, there are serious drawbacks with the NPL algorithm and, hence, the NPL estimator. First, since the NPL algorithm performs one best-response iteration (23) to update the equilibrium probabilities $\widetilde{\boldsymbol{P}}_K$ in the K^{th} recursive iteration, it will only converge to equilibria that are stable under best response. If the data were generated by equilibria that are unstable under best response iterations, then the NPL algorithm will either fail to converge or, worse, converge to wrong equilibria, which then leads to incorrect parameter estimates for the NPL estimator; see Pesendorfer and Schmidt-Dengler (2010) for such an example. Second, even when the data are generated by best-response stable equilibria, the NPL algorithm can generate cycling iterations and, consequently, fail to converge. Third, the NPL algorithm can take many recursive iterations before it converges and, computationally, it is not as efficient as the constrained optimization approach. We illustrate these points using the Monte Carlo experiments below.

4 Monte Carlo

We used Example 2 as the econometric model in the Monte Carlo experiments to study finite-sample performance of four estimators: the ML (with the constrained optimization approach), the 2S-PML, the 2S-LS, and the NPL estimators.[4] Recall that, in Example 2, the true parameter values are $\boldsymbol{\theta}^0 = (5, -11)$ and $M = 256$ differentiated markets exist.

4.1 Experiment specifications

In generating the data, we maintained the assumption that in each market only one equilibrium is played; however, different equilibria can be played in different markets. We describe the types of equilibria used in the data generating process below.

Scenario 1: The best-response stable equilibrium with lowest probability of being active for firm a is played in each market for markets with multiple equilibria.

Scenario 2: Randomly choose one of the best-response stable equilibria to be played in each market for markets with multiple equilibria.

[4] As noted by Pakes et al. (2007), the pseudo likelihood function is an inappropriate criterion function to use in the second step. The reason that we still use the 2S-PML estimator in our Monte Carlo experiments is to provide a direct comparison on the performance of the 2S-PML to that of the ML estimator and the NPL estimator.

Scenario 3: Randomly choose one equilibrium to be played in each market. The equilibrium chosen can be stable or unstable under best response iterations.

For each data set, we simulated the static game for T periods to generate T pairs of observed decisions, $\boldsymbol{y}^m = \left(y_a^{mt}, y_b^{mt}\right)_{t=1}^{T}$ in each market $m = 1, \ldots, 256$. Conditional on the observables $\boldsymbol{x}^m$, the number of periods T (ranging from 5 to 250) gives the number of repeated observations per market in the data. For each scenario, we constructed one hundred data sets for each T.

Since the firms' decisions are observed in every market, we used a frequency estimator to estimate the first-step equilibrium probabilities for two-step estimators. The estimated equilibrium probabilities $\widehat{\boldsymbol{P}}_{\text{freq}} = \left(\hat{\boldsymbol{p}}_{\text{freq}}^m\right)_{m=1}^{256}$ from the frequency estimator solve

$$\hat{p}_{a,\text{freq}}^m = \frac{1}{T}\sum_{t=1}^{T} y_a^{mt}, \quad \hat{p}_{b,\text{freq}}^m = \frac{1}{T}\sum_{t=1}^{T} y_b^{mt}, \text{ for } m = 1, \ldots, 256.$$

For the constrained optimization approach, there are $2 + 2 \times 256 = 514$ decision variables and 512 equality constraints (two equilibrium probabilities $\left(p_a^m, p_b^m\right)$ as well as two BN equilibrium equations in each market). To find the parameter estimates for each data set, we used one hundred different starting points for $\boldsymbol{\theta}$ and the frequency estimates $\widehat{\boldsymbol{P}}_{\text{freq}}$ as the starting value for $\boldsymbol{P}$. For the NPL estimator, we used $\widehat{\boldsymbol{P}}_{\text{freq}}$, $\frac{3\widehat{\boldsymbol{P}}_{\text{freq}}}{4}$, and $\frac{\widehat{\boldsymbol{P}}_{\text{freq}}}{2}$ as the three initial guesses of equilibrium probabilities for the NPL algorithm for each data set. In addition, we also used logit probabilities as an initial guess for the NPL algorithm. We choose `1.0e-6` as the convergence tolerance for the NPL algorithm and set the maximum number of NPL iterations to $\bar{K} = 1000$. If the difference of parameter values and equilibrium probabilities in successive iterates is less than the chosen tolerance, then we declare the NPL algorithm converges. If the number of NPL iterations reaches $\bar{K}$ before convergence, then we declared that the NPL algorithm failed to converge in that run. If the NPL algorithm failed to converge from all four starting values, we declared that the NPL algorithm failed to converge to a solution for the NPL estimator for that data set.

We coded the optimization problem for each estimator in AMPL and called the nonlinear optimization solver KNITRO to solve the problem. We use the default value `1.0-6` as optimality and feasibility tolerance. We chose AMPL as the programing platform because AMPL uses automatic differentiation to compute exact first-order and second-order derivatives efficiently and passes the derivative information together with sparsity structure of the constrained Jacobian and Hessian matrices to optimization solvers. The derivatives as well as the sparsity structure information are necessary for KNITRO (and other optimization solvers) to perform well, especially on large-scale problems. With this choice of software for numerical implementation, we hope to provide a fair comparison on the numerical performance of each estimator.

4.2 Monte Carlo results

The Monte Carlo results are reported in Tables 1, 2 and 3. Overall, our results indicate that the ML estimator performs the best and the NPL estimator the worst. For the two-step estimators, in general, the 2S-PML estimator performs better than the 2S-LS estimator.[5] For the NPL estimator, Aguirregabiria and Mira (2007) found that the NPL algorithm with logit estimates as an initial guess converged faster than that with frequency estimates. On the contrary, we find the NPL algorithm with frequency estimates as initial guesses performs better.[6] Except for $T = 5$ in Scenario 1, NPL with logit estimates either often fails to converge (in Scenario 1) or converges to wrong parameter estimates (in Scenario 2). Hence, we focus on the performance of NPL with frequency estimates as initial guesses in the discussion below.

For many experiments in Scenarios 1 and 2, the ML and the 2S-PML estimators do quite well at recovering the true parameter values. Even with small numbers of repeated observations ($T = 5$ or 10) per market, the mean of the ML estimator is within one standard deviation of the true parameters values; on the contrary, the mean of parameter estimates on α of two-step estimators are one standard devision away from the true value. The biases and the root mean square error (RMSE) in the ML estimator are significantly smaller than those of the two-step estimators when the numbers of repeated observations per market are small. The 2S-PML estimator produces accurate parameter estimates with twenty-five or more repeated observations per market in the data. Given sufficiently many repeated observations, the two-step estimators can provide good parameter estimates with little computational efforts. The NPL estimator, surprisingly, fails frequently for $T = 50$ or smaller; for example, the NPL algorithm converges in only two data sets for $T = 5$ and in 29 data sets for $T = 10$ in Scenario 1. In those failed runs, the NPL algorithm generates cycling iterations without any indication of achieving convergence. This finding suggests that the convergence of the NPL algorithm is not as robust as perceived in Aguirregabiria and Mira (2007), even when data are generated by best-response stable equilibria. In cases where the NPL algorithm does converge, the NPL estimator is more biased than the ML estimator or the two-step estimators. The NPL estimator requires large numbers of repeated observations ($T = 100$ for Scenario 1 and $T = 250$ for Scenario 2) to obtain estimates that are comparable to those of the ML and the two-step estimators.

In terms of computation time, the two-step estimators are fast, perhaps not a surprising result because the two-step estimators do not require solving a BN equilibrium. Both the constrained optimization approach and the NPL estimator impose the equilibrium constraints to be satisfied and, hence, require more computational time than the two-step estimators. As one can see from Tables 1 and 2, the constrained optimization approach requires only one to two seconds of computing time per

[5] One can improve the performance of the 2S-LS estimator by using the optimal weighting matrix, which has been suggested by Pesendorfer and Schmidt-Dengler (2008).

[6] In our experiment, the NPL algorithm usually converges to the same parameter estimates from starting values $\widehat{\boldsymbol{P}}_{\text{freq}}$, $\frac{3\widehat{\boldsymbol{P}}_{\text{freq}}}{4}$, and $\frac{\widehat{\boldsymbol{P}}_{\text{freq}}}{2}$.

Table 1 Scenario 1 – best-response stable equilibrium with lowest probabilities of entry for firm a used in DGP

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. NPL Iter.
		α	β				
	Truth	5	−11				
5	ML (Cons. Opt.)	5.234 (0.278)	−11.238 (0.506)	0.665	0.692	100	–
5	2S-PML	4.459 (0.276)	−10.646 (0.796)	1.058	0.040	100	–
5	2S-LS	4.514 (0.347)	−11.369 (1.100)	1.300	0.053	100	–
5	NPL (freq. prob.)	4.863 (0.241)	−10.019 (1.830)	1.639	36.051	2	987
5	NPL (logit prob.)	5.105 (0.193)	−10.173 (0.629)	1.057	28.477	39	762
10	ML (Cons. Opt.)	5.065 (0.143)	−11.111 (0.345)	0.393	0.441	100	–
10	2S-PML	4.787 (0.165)	−10.886 (0.529)	0.602	0.043	100	–
10	2S-LS	4.914 (0.238)	−11.473 (0.852)	1.002	0.055	100	–
10	NPL (freq. prob.)	5.054 (0.241)	−10.411 (1.830)	0.958	33.153	29	808
10	NPL (logit prob.)	5.096 (0.136)	-10.219 (0.482)	0.928	35.840	28	847
25	ML (Cons. Opt.)	5.018 (0.076)	−11.022 (0.181)	0.197	0.417	100	–
25	2S-PML	4.926 (0.114)	−11.040 (0.264)	0.298	0.058	100	–
25	2S-LS	5.014 (0.147)	−11.387 (0.479)	0.632	0.057	100	–
25	NPL (freq. prob.)	4.995 (0.081)	−10.607 (0.563)	0.688	29.122	71	543
25	NPL (logit prob.)	5.076 (0.096)	−10.164 (0.280)	0.888	50.674	28	856
50	ML (Cons. Opt.)	5.000 (0.061)	−11.000 (0.133)	0.146	0.398	100	–
50	2S-PML	4.956 (0.080)	−10.983 (0.198)	0.218	0.090	100	–
50	2S-LS	5.007 (0.109)	−11.119 (0.329)	0.365	0.056	100	–

Table 1 (continued)

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. NPL Iter.
		α	β				
	Truth	5	−11				
50	NPL (freq. prob.)	4.998 (0.070)	−10.665 (0.472)	0.581	32.133	86	409
50	NPL (logit prob.)	5.119 (0.093)	−10.226 (0.238)	0.821	80.510	16	913
100	ML (Cons. Opt.)	5.005 (0.046)	−10.996 (0.103)	0.112	0.858	100	–
100	2S-PML	4.985 (0.060)	−11.011 (0.164)	0.175	0.164	100	–
100	2S-LS	5.011 (0.077)	−11.090 (0.238)	0.265	0.056	100	–
100	NPL (freq. prob.)	5.005 (0.051)	−10.908 (0.283)	0.301	34.516	96	242
100	NPL (logit prob.)	5.061 (0.048)	−10.130 (0.249)	0.906	155.480	15	942
250	ML (Cons. Opt.)	5.000 (0.031)	−10.995 (0.062)	0.069	1.798	100	–
250	2S-PML	4.994 (0.037)	−11.002 (0.092)	0.099	0.379	100	–
250	2S-LS	5.005 (0.042)	−11.025 (0.152)	0.160	0.057	100	–
250	NPL (freq. prob.)	5.002 (0.051)	−10.955 (0.283)	0.198	57.083	100	174
250	NPL (logit prob.)	5.138 (0.022)	−10.276 (0.045)	0.739	383.240	3	985

Standard deviations are reported in parentheses. CPU time is the averaged time per run

starting point. Aguirregabiria and Mira (2007) suggested that the NPL estimator requires a relatively small additional computational cost over the 2S-PML estimator. In our experiments, however, the NPL algorithm requires on average more than 160 NPL iterations (i.e., iterating over the 2S-PML for more than 160 times) and around thirty seconds to converge per starting point for $T = 100$ or sixty seconds for $T = 250$. Hence, for Scenarios 1 and 2, the constrained optimization approach is about thirty times faster than the NPL algorithm in computing time for $T = 100$ and $T = 250$.

For Scenario 3, the ML estimator is the only estimator that recovers the true parameter values. Even though some equilibria used to generate data are unstable under best

Table 2 Scenario 2 – best-response stable equilibrium in each market used in DGP

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. PL Iter.
		α	β				
	Truth	5	−11				
5	ML	5.197	−11.189	0.588	0.803	100	–
	(Cons. Opt.)	(0.245)	(0.463)				
5	2S-PML	4.380	−10.427	1.132	0.040	100	–
		(0.263)	(0.711)				
5	2S-LS	4.395	−11.131	1.278	0.053	100	–
		(0.318)	(1.078)				
5	NPL	4.707	−8.534	2.574	34.847	4	975
	(freq. prob.)	(0.241)	(1.830)				
5	NPL	1.738	−3.318	8.346	28.712	3	988
	(logit prob.)	(0.026)	(0.046)				
10	ML	5.104	−11.038	0.354	0.472	100	–
	(Cons. Opt.)	(0.149)	(0.304)				
10	2S-PML	4.787	−10.831	0.615	0.043	100	–
		(0.181)	(0.523)				
10	2S-LS	4.893	−11.418	0.942	0.055	100	–
		(0.243)	(0.805)				
10	NPL	5.019	−9.732	1.534	28.135	46	682
	(freq. prob.)	(0.148)	(0.753)				
10	NPL	N/A	N/A	N/A	33.175	0	1000
	(logit prob.)	(N/A)	(N/A)				
25	ML	5.040	−10.992	0.214	0.245	100	–
	(Cons. Opt.)	(0.085)	(0.193)				
25	2S-PML	4.945	−10.943	0.335	0.059	100	–
		(0.113)	(0.307)				
25	2S-LS	5.022	−11.175	0.553	0.057	100	–
		(0.135)	(0.509)				
25	NPL	5.032	−10.087	1.229	25.661	75	469
	(freq. prob.)	(0.088)	(0.824)				
25	NPL	1.781	−3.364	8.287	46.136	1	1000
	(logit prob.)	(0.000)	(0.000)				
50	ML	5.009	−10.999	0.160	0.380	100	–
	(Cons. Opt.)	(0.060)	(0.149)				
50	2S-PML	4.967	−10.990	0.223	0.091	100	–
		(0.089)	(0.203)				
50	2S-LS	5.016	−11.106	0.374	0.058	100	–
		(0.111)	(0.343)				

Table 2 (continued)

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. NPL Iter.
		α	β				
	Truth	5	-11				
50	NPL (freq. prob.)	5.018 (0.071)	-10.243 (0.780)	1.087	30.148	86	384
50	NPL (logit prob.)	1.763 (0.000)	-3.386 (0.000)	8.274	69.820	1	998
100	ML (Cons. Opt.)	5.011 (0.046)	-10.982 (0.095)	0.107	0.821	100	–
100	2S-PML	4.995 (0.060)	-11.011 (0.164)	0.176	0.164	100	–
100	2S-LS	5.022 (0.077)	-11.090 (0.249)	0.275	0.059	100	–
100	NPL (freq. prob.)	5.024 (0.060)	-10.661 (0.650)	0.733	30.406	99	225
100	NPL (logit prob.)	1.775 (0.000)	-3.379 (0.000)	8.276	123.580	1	999
250	ML (Cons. Opt.)	5.003 (0.025)	-10.993 (0.057)	0.062	1.838	100	–
250	2S-PML	4.9957 (0.034)	-11.000 (0.103)	0.108	0.377	100	–
250	2S-LS	5.008 (0.040)	-11.025 (0.171)	0.176	0.060	100	–
250	NPL (freq. prob.)	5.010 (0.060)	-10.854 (0.650)	0.470	53.572	100	168
250	NPL (logit prob.)	1.774 (0.003)	-3.374 (0.000)	8.281	281.110	2	997

Standard deviations are reported in parentheses. CPU time is the averaged time per run

response, the mean of the ML estimator is within one standard derivation of the true values. This finding should not be surprising because constrained optimization algorithms do not rely on best-response iterations. Thus, the presence of best-response unstable equilibria in the data should not affect the performance of the ML estimator under the constrained optimization approach. For the two-step estimators, both the 2S-PML and the 2S-LS estimators reject the true parameter values $\boldsymbol{\theta}^0 = (5, -11)$. The mean of parameter estimates are at least one standard deviation away from the true values in all experiments; the biases are significantly higher than those of Scenarios 1 and 2. In general, the 2S-PML estimator performs better than the 2S-LS

Table 3 Scenario 3 – randomly chosen equilibrium in each market used in DGP

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. NPL Iter.
		α	β				
	Truth	5	-11				
5	ML	5.027	-10.743	0.661	1.346	100	–
	(Cons. Opt.)	(0.179)	(0.585)				
5	2S-PML	3.068	-7.279	4.228	0.043	100	–
		(0.208)	(0.512)				
5	2S-LS	2.918	-7.597	4.047	0.048	100	–
		(0.203)	(0.654)				
5	NPL	N/A	N/A	N/A	31.527	0	1000
	(freq. prob.)	(N/A)	(N/A)				
5	NPL	N/A	N/A	N/A	33.748	0	1000
	(logit prob.)	(N/A)	(N/A)				
10	ML	5.029	-10.816	0.394	0.641	100	–
	(Cons. Opt.)	(0.126)	(0.326)				
10	2S-PML	3.719	-8.535	2.812	0.042	100	–
		(0.165)	(0.403)				
10	2S-LS	3.459	-8.499	2.990	0.049	100	–
		(0.164)	(0.531)				
10	NPL	N/A	N/A	N/A	35.756	0	1000
	(freq. prob.)	(N/A)	(N/A)				
10	NPL	N/A	N/A	N/A	37.786	0	1000
	(logit prob.)	(N/A)	(N/A)				
25	ML	5.018	-10.964	0.189	0.512	100	–
	(Cons. Opt.)	(0.084)	(0.166)				
25	2S-PML	4.302	-9.663	1.537	0.060	100	–
		(0.122)	(0.268)				
25	2S-LS	3.959	-9.311	2.019	0.050	100	–
		(0.134)	(0.354)				
25	NPL	N/A	N/A	N/A	52.268	0	1000
	(freq. prob.)	(N/A)	(N/A)				
25	NPL	N/A	N/A	N/A	54.315	0	1000
	(logit prob.)	(N/A)	(N/A)				
50	ML	5.005	-11.007	0.150	0.669	100	–
	(Cons. Opt.)	(0.056)	(0.139)				
50	2S-PML	4.590	-10.280	0.865	0.093	100	–
		(0.099)	(0.230)				
50	2S-LS	4.279	-9.895	1.354	0.052	100	–
		(0.109)	(0.283)				

Table 3 (continued)

T	Estimator	Estimates		RMSE	CPU Time (sec.)	Num. of Data sets Conv.	Avg. NPL Iter.
		α	β				
	Truth	5	-11				
50	NPL (freq. prob.)	N/A (N/A)	N/A (N/A)	N/A	82.390	0	1000
50	NPL (logit prob.)	N/A (N/A)	N/A (N/A)	N/A	84.415	0	1000
100	ML (Cons. Opt.)	5.006 (0.045)	-10.997 (0.092)	0.102	1.252	100	–
100	2S-PML	4.773 (0.067)	-10.607 (0.165)	0.487	0.174	100	–
100	2S-LS	4.533 (0.084)	-10.285 (0.200)	0.881	0.053	100	–
100	NPL (freq. prob.)	N/A (N/A)	N/A (N/A)	N/A	150.220	0	1000
100	NPL (logit prob.)	N/A (N/A)	N/A (N/A)	N/A	152.560	0	1000
250	ML (Cons. Opt.)	5.000 (0.028)	-10.999 (0.057)	0.063	2.512	100	–
250	2S-PML	4.905 (0.043)	-10.828 (0.114)	0.231	0.410	100	–
250	2S-LS	4.905 (0.051)	-10.624 (0.157)	0.472	0.054	100	–
250	NPL (freq. prob.)	N/A (N/A)	N/A (N/A)	N/A	351.990	0	1000
250	NPL (logit prob.)	N/A (N/A)	N/A (N/A)	N/A	354.470	0	1000

Standard deviations are reported in parentheses. CPU time is the averaged time per run

estimator; however, 250 repeated observations are required for the 2S-PML estimator to produce reasonable parameter estimates. The NPL algorithm fails to converge and cannot compute a solution of the NPL estimator in all one hundred data sets in this scenario. Pesendorfer and Schmidt-Dengler (2010) have provided a counterexample that illustrates that the NPL algorithm produces wrong parameter values. By using the one best-response iteration update in Eq. (23), the NPL algorithm and, as a result, the NPL estimator implicitly requires that equilibria in the data are stable under best response. Obviously, this assumption is violated in the data generating process for Scenario 3.

5 Empirical application: discount retailers' entry decisions

5.1 The model

Based on the empirical application studied in Jia (2008), we consider a more realistic model of static game of incomplete information to examine the performance of the estimators discussed in previous sections. In this model, there are two discount retailers, Walmart and Kmart, making entry decisions in various markets, indexed by $m = 1, \ldots, M$. Following the specifications in Ellickson and Misra (2011) and Misra (2013), the entry probabilities of Walmart and Kmart in market m, denoted by p_W^m and p_K^m, respectively, are given as follows:

$$
\begin{aligned}
p_W^m &= \frac{\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_W \boldsymbol{x}_W^m - \delta p_K^m)}{1 + \exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_W \boldsymbol{x}_W^m - \delta p_K^m)} = \Psi_W(\boldsymbol{x}^m, \boldsymbol{x}_W^m, p_K^m; \boldsymbol{\theta}) \\
p_K^m &= \frac{\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_K \boldsymbol{x}_K^m - \delta p_W^m)}{1 + \exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_K \boldsymbol{x}_K^m - \delta p_W^m)} = \Psi_K(\boldsymbol{x}^m, \boldsymbol{x}_K^m, p_W^m; \boldsymbol{\theta}),
\end{aligned}
\tag{24}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}_W, \boldsymbol{\beta}_K, \delta)$ is the vector of structural parameters, and the data in each market m are county demographics $\boldsymbol{x}^m = \left(x_1^m, x_2^m, x_3^m\right)$, Walmart specifics $\boldsymbol{x}_W^m = \left(x_{W0}^m, x_{W1}^m, x_{W2}^m\right)$, and Kmart specifics $\boldsymbol{x}_K^m = \left(x_{K0}^m, x_{K1}^m\right)$. The descriptions of the data $\boldsymbol{x}^m, \boldsymbol{x}_W^m$, and $\boldsymbol{x}_K^m$ are given in the next subsection. Within structural parameters, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and δ are common (or identical) among the two firms, $\boldsymbol{\beta}_W = (\beta_{W0}, \beta_{W1}, \beta_{W2})$ are parameters associated with Walmart only, and $\boldsymbol{\beta}_K = (\beta_{K0}, \beta_{K1})$ are associated with Kmart only. The parameter δ captures the effect of the rival firm's entry decision on the focal firm's payoff, if the focal firm decides to enter the market. Following the notation used in Eq. (7) in Section 2, we represent the BN equilibrium equations for all markets as

$$
\boldsymbol{P} = \boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}),
$$

where $\boldsymbol{P} = \left(p_W^m, p_K^m\right)_{m=1}^M$, $\boldsymbol{\Psi} = (\Psi_W, \Psi_K)$ and $\boldsymbol{X} = \left(\boldsymbol{x}^m, \boldsymbol{x}_W^m, \boldsymbol{x}_K^m\right)_{m=1}^M$.

5.2 Data and experiment specifications

We used a subset of Jia's year 1997 data in our numerical experiments. Specifically, we used data from 200 markets, starting from county identifier 158 to 506.[7] In these 200 markets, 30.5 % (61 counties) had Walmart stores only, 4.5 % (9 counties) had Kmart stores only, and (13 % (26 counties) had both Walmart and Kmart stores, which are fairly representative of the same summary statistics in the whole data set.[8]

[7]These are Row 117 to 316 in Jia's data file, `XMat97.out`.

[8]In the whole date sets with 2065 counties, 33.6 % (694 counties) had Walmart stores only, 5 % (103 counties) had Kmart stores only, and 13.9 % (287 counties) had both Walmart and Kmart stores.

The descriptions of the data $\boldsymbol{x}^m$, $\boldsymbol{x}_W^m$, and $\boldsymbol{x}_K^m$ in each of the 200 markets are given below with the corresponding structural parameters in the parenthesis:

$$\boldsymbol{x}^m = \begin{cases} x_1^m : \text{log of county population} & (\alpha_1) \\ x_2^m : \text{log county retail sales per capita} & (\alpha_2) \\ x_3^m : \text{percentage of urban population} & (\alpha_3) \end{cases}$$

$$\boldsymbol{x}_W^m = \begin{cases} x_{W0}^m : \text{Walmart dummy} & (\beta_{W0}) \\ x_{W1}^m : \text{distance to Bentonville, AK} & (\beta_{W1}) \\ x_{W2}^m : \text{southern market dummy} & (\beta_{W2}) \end{cases}$$

$$\boldsymbol{x}_K^m = \begin{cases} x_{K0}^m : \text{Kmart dummy} & (\beta_{K0}) \\ x_{K1}^m : \text{midwest market dummy} & (\beta_{K1}). \end{cases}$$

We examined two sets of parameter values in our Monte Carlo experiments:

$$\begin{aligned} \boldsymbol{\alpha}^0 &= (1.75, 2.08, 1.50), \\ \boldsymbol{\beta}_W^0 &= (-16.95, -1.06, 0.88), \\ \boldsymbol{\beta}_K^0 &= (-24.26, 0.38), \\ \delta^0 &= 0.71 \text{ or } 6. \end{aligned} \tag{25}$$

In the first set, we chose $\delta^0 = 0.71$.[9] Given this set of true parameter values, one can verify that $\rho\left[\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}^0)\right]$, the spectral radius of the Jacobian mapping $\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}^0)$ is always less than 1, regardless of the equilibrium probabilities. This implies that a unique equilibrium exists in each of the 200 markets.

To create an example with multiple equilibria in some markets, the spectral radius $\rho\left[\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}^0)\right]$ needs to be greater than 1. This, in turn, requires that the true parameter value δ^0 needs to be greater than 4.[10] In the second set of parameter values, we chose true parameter value $\delta^0 = 6$ while keeping the other parameters at the same values given in Eq. 25. Indeed, with this choice of parameter values, we found three equilibria in 7 of the 200 markets. The county identifier of these 7 markets and the corresponding equilibrium probabilities are reported in Table 4.

We consider three different scenarios and describe the types of equilibrium used in the data generating process below. As before, the assumption that only one equilibrium is played in each market (Assumption 2) is maintained in generating the data.

Scenario 4: Parameter values given in Eq. 25 with $\delta^0 = 0.71$ are chosen as the truth. In this scenario, a unique equilibrium exists in each of the 200 markets.

Scenario 5: Parameter values given in Eq. 25 with $\delta^0 = 6$ are chosen as the truth. In each of the 7 markets with multiple equilibria, Equilibrium 1 (a best-response stable equilibrium) reported in Table 4 is played in the data.

[9] These numbers are close to those reported in Jia (2008), Ellickson and Misra (2011) and Misra (2013).

[10] We provide more detailed discussion on $\rho\left[\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta}^0)\right]$ and the condition on δ^0 in Appendix B.

Table 4 Markets with multiple equilibria with $\delta^0 = 6$

Market index	County identifier	Equilibrium probabilities (p_W, p_K)		
		Equilibrium 1	Equilibrium 2	Equilibrium 3
162	258	(0.0458, 0.8833)	(0.4378, 0.4187)	(0.8774, 0.0490)
164	261	(0.0927, 0.4662)	(0.1896, 0.3281)	(0.5482, 0.0537)
165	262	(0.0454, 0.9018)	(0.4749, 0.4110)	(0.8834, 0.0567)
171	273	(0.0358, 0.7788)	(0.3750, 0.3151)	(0.7509, 0.0460)
177	282	(0.0382, 0.6783)	(0.3539, 0.2407)	(0.6157, 0.0618)
181	288	(0.0405, 0.7092)	(0.3311, 0.2991)	(0.6926, 0.0465)
187	298	(0.0437, 0.8540)	(0.4011, 0.4066)	(0.8559, 0.0428)

Scenario 6: Parameter values given in Eq. 25 with $\delta^0 = 6$ are chosen as the truth. In each of the 7 markets with multiple equilibria, Equilibrium 2 (the best-response unstable equilibrium) reported in Table 4 is played in the data.

In simulating the data, we considered different number of repeated observations T, ranging from 5 to 50. For each scenario, we constructed one hundred data sets for each T. Since the 2S-LS estimator was outperformed by the 2S-PML estimator in the Monte Carlo experiments in Section 4, we excluded the 2S-LS estimator from this exercise. To find the parameter estimates for each data set, we used one hundred different starting points for the constrained optimization approach. For the 2S-PML estimator, we used frequency estimates of the equilibrium probabilities in the first step. For the NPL estimator, we used five different starting values of equilibrium probabilities and set the maximum number of NPL iterations to $\bar{K} = 1000$. We chose `1.0e-6` as the convergence tolerances for the NPL algorithm and the constrained optimization approach.

5.3 Estimation results

The Monte Carlo results, reported in Tables 5, 6 and 7, present qualitatively similar observations to those using the simple static example in Section 4. Overall, ML estimator still perform the best, producing estimates with mean that is within in one standard deviation of the true parameter values in all experiments. The 2S-PML estimator does quite well at recovering true values for most parameters, but has difficulties with the parameter δ, the coefficient that captures the strategic interactions among firms. In all experiments except for $T = 25$ and 50 in Scenario 4, the mean of the 2S-PML estimator on the parameter δ is one standard deviation away from the true value.

The performance of the NPL estimator varies across the three scenarios. In Scenario 4, the NPL estimator successfully converges in all one hundred data sets and produce accurate parameter estimates. However, the performance of the NPL estimator and the NPL algorithm worsens in Scenarios 5 and 6 with $\delta^0 = 6$. In Scenario 5, while the NPL estimator still does well at recovering true parameter values, the NPL

Table 5 Scenario 4 – Walmart-Kmart example with $\delta^0 = 0.71$ and unique equilibrium in each market

T	Estimator	Estimates									CPU Time (sec.)	Data sets Conv.	Avg. NPL Iter.
		α_1	α_2	α_3	β_{W0}	β_{W1}	β_{W2}	β_{K1}	β_{K2}	δ			
	Truth	1.75	2.08	1.50	−16.95	−1.06	0.88	−24.26	0.38	0.71			
5	ML	1.76	2.07	1.57	−16.94	−1.05	0.86	−24.22	0.35	0.73	0.96	100	–
		(0.21)	(0.22)	(0.38)	(2.26)	(0.17)	(0.25)	(2.01)	(0.23)	(0.58)			
5	2S-PML	1.63	1.93	1.50	−15.53	−1.05	0.91	−22.86	0.41	0.22	0.07	100	–
		(0.17)	(0.19)	(0.36)	(1.97)	(0.17)	(0.23)	(1.73)	(0.25)	(0.37)			
5	NPL	1.75	2.06	1.57	−16.88	−1.05	0.86	−24.16	0.35	0.71	3.48	100	62
		(0.20)	(0.21)	(0.39)	(2.22)	(0.17)	(0.25)	(1.96)	(0.24)	(0.55)			
10	ML	1.75	2.10	1.50	−17.19	−1.05	0.90	−24.40	0.35	0.70	0.82	100	–
		(0.13)	(0.17)	(0.31)	(1.66)	(0.11)	(0.17)	(1.54)	(0.18)	(0.37)			
10	2S-PML	1.66	1.99	1.45	−16.14	−1.04	0.94	−23.38	0.40	0.33	0.08	100	–
		(0.13)	(0.17)	(0.30)	(1.67)	(0.11)	(0.16)	(1.54)	(0.18)	(0.35)			
10	NPL	1.75	2.09	1.50	−17.17	−1.05	0.90	−24.38	0.35	0.69	7.98	100	95
		(0.14)	(0.17)	(0.31)	(1.64)	(0.11)	(0.17)	(1.53)	(0.18)	(0.37)			
25	ML	1.76	2.09	1.53	−17.10	−1.06	0.90	−24.37	0.35	0.74	0.78	100	–
		(0.09)	(0.09)	(0.24)	(0.89)	(0.08)	(0.12)	(0.85)	(0.11)	(0.28)			
25	2S-PML	1.69	2.02	1.50	−16.40	−1.05	0.92	−23.69	0.39	0.49	0.11	100	–
		(0.09)	(0.11)	(0.19)	(1.09)	(0.07)	(0.10)	(0.96)	(0.13)	(0.25)			
25	NPL	1.75	2.09	1.53	−17.10	−1.06	0.90	−24.36	0.35	0.73	7.19	100	77
		(0.09)	(0.10)	(0.19)	(1.09)	(0.07)	(0.10)	(0.94)	(0.13)	(0.24)			
50	ML	1.75	2.08	1.52	−17.03	−1.05	0.90	−24.30	0.37	0.70	1.03	100	–
		(0.06)	(0.08)	(0.14)	(0.80)	(0.05)	(0.07)	(0.68)	(0.08)	(0.19)			
50	2S-PML	1.72	2.04	1.50	−16.61	−1.05	0.91	−23.88	0.39	0.55	0.23	100	–
		(0.06)	(0.08)	(0.14)	(0.78)	(0.05)	(0.07)	(0.67)	(0.09)	(0.19)			
50	NPL	1.75	2.08	1.52	−17.04	−1.05	0.90	−24.30	0.37	0.70	9.03	100	59
		(0.06)	(0.08)	(0.14)	(0.81)	(0.05)	(0.07)	(0.69)	(0.08)	(0.19)			

Standard deviations are reported in parentheses. CPU time is the averaged time per run

Table 6 Scenario 5 – Walmart-Kmart example with $\delta^0 = 6$ and equilibrium 1 in DGP

T	Estimator	Estimates									CPU Time (sec.)	Data sets Conv.	Avg. NPL Iter.
		α_1	α_2	α_3	β_{W0}	β_{W1}	β_{W2}	β_{K1}	β_{K2}	δ			
	Truth	1.75	2.08	1.50	−16.95	−1.06	0.88	−24.26	0.38	6.00			
5	ML	1.73	2.05	1.52	−16.94	−0.98	0.69	−23.78	0.41	6.94	1.25	100	–
		(0.21)	(0.22)	(0.50)	(2.04)	(0.20)	(0.30)	(2.17)	(0.21)	(1.84)			
5	2S-PML	1.72	1.99	1.69	−16.32	−1.13	1.35	−24.09	0.64	4.57	0.07	100	–
		(0.23)	(0.22)	(0.58)	(2.21)	(0.19)	(0.30)	(1.95)	(0.34)	(0.66)			
5	NPL	1.76	2.09	1.61	−17.15	−1.07	0.97	−24.46	0.36	5.91	10.48	65	355
		(0.22)	(0.20)	(0.50)	(1.83)	(0.19)	(0.25)	(1.84)	(0.26)	(0.52)			
10	ML	1.75	2.10	1.54	−17.28	−1.03	0.85	−24.42	0.40	6.18	0.86	100	–
		(0.16)	(0.18)	(0.41)	(1.63)	(0.13)	(0.22)	(1.58)	(0.19)	(0.62)			
10	2S-PML	1.75	2.07	1.58	−17.03	−1.09	1.19	−24.53	0.48	5.14	0.08	100	–
		(0.16)	(0.19)	(0.47)	(1.73)	(0.13)	(0.22)	(1.58)	(0.19)	(0.62)			
10	NPL	1.75	2.12	1.54	−17.37	−1.05	0.96	−24.58	0.36	5.86	8.67	77	266
		(0.15)	(0.18)	(0.45)	(1.62)	(0.13)	(0.19)	(1.53)	(0.17)	(0.32)			
25	ML	1.76	2.07	1.55	−16.89	−1.06	0.88	−24.23	0.38	5.99	0.75	100	–
		(0.08)	(0.09)	(0.23)	(0.85)	(0.08)	(0.09)	(0.81)	(0.10)	(0.18)			
25	2S-PML	1.75	2.07	1.58	−16.92	−1.08	1.03	−24.35	0.41	5.57	0.12	100	–
		(0.10)	(0.12)	(0.29)	(1.08)	(0.08)	(0.14)	(1.01)	(0.17)	(0.25)			
25	NPL	1.76	2.08	1.55	−17.02	−1.06	0.93	−24.32	0.38	5.90	7.17	90	151
		(0.09)	(0.10)	(0.26)	(0.97)	(0.08)	(0.11)	(0.91)	(0.11)	(0.23)			
50	ML	1.76	2.09	1.53	−17.05	−1.06	0.89	−24.37	0.38	5.99	0.89	100	–
		(0.06)	(0.06)	(0.17)	(0.63)	(0.06)	(0.07)	(0.56)	(0.06)	(0.12)			
50	2S-PML	1.76	2.08	1.55	−17.03	−1.07	0.96	−24.41	0.41	5.77	0.23	100	–
		(0.07)	(0.08)	(0.22)	(0.73)	(0.06)	(0.11)	(0.69)	(0.11)	(0.16)			
50	NPL	1.75	2.09	1.54	−17.09	−1.06	0.91	−24.40	0.39	5.94	11.73	88	156
		(0.06)	(0.07)	(0.18)	(0.69)	(0.06)	(0.09)	(0.65)	(0.08)	(0.14)			

Standard deviations are reported in parentheses. CPU time is the averaged time per run

Table 7 Scenario 6 – Walmart-Kmart example with $\delta^0 = 6$ and equilibrium 2 in DGP

T	Estimator	Estimates									CPU Time (sec.)	Data sets Conv.	Avg. NPL Iter.
		α_1	α_2	α_3	β_{W0}	β_{W1}	β_{W2}	β_{K1}	β_{K2}	δ			
	Truth	1.75	2.08	1.50	−16.95	−1.06	0.88	−24.26	0.38	6.00			
5	ML	1.71	2.04	1.64	−16.73	−1.02	0.85	−23.78	0.36	6.17	1.50	100	–
		(0.22)	(0.22)	(0.51)	(2.04)	(0.19)	(0.35)	(1.97)	(0.25)	(1.03)			
5	2S-PML	1.66	1.92	1.69	−15.59	−1.12	1.34	−23.45	0.46	3.93	0.07	100	–
		(0.22)	(0.24)	(0.59)	(2.27)	(0.19)	(0.29)	(2.02)	(0.34)	(0.56)			
5	NPL	1.58	1.86	1.71	−20.46	−1.30	1.24	−22.47	0.41	4.93	21.79	7	745
		(0.43)	(0.66)	(0.26)	(9.08)	(0.57)	(0.42)	(6.01)	(0.24)	(0.48)			
10	ML	1.74	2.09	1.56	−17.15	−1.03	0.86	−24.27	0.39	6.04	1.17	100	–
		(0.16)	(0.18)	(0.41)	(1.63)	(0.13)	(0.22)	(1.58)	(0.19)	(0.62)			
10	2S-PML	1.72	2.00	1.57	−16.35	−1.09	1.20	−23.91	0.40	4.68	0.08	100	–
		(0.17)	(0.19)	(0.48)	(1.71)	(0.13)	(0.20)	(1.66)	(0.25)	(0.45)			
10	NPL	1.73	1.98	1.58	−16.31	−1.08	1.19	−23.64	0.25	4.77	24.70	6	747
		(0.20)	(0.09)	(0.40)	(1.55)	(0.16)	(0.12)	(1.05)	(0.14)	(0.18)			
25	ML	1.76	2.06	1.55	−16.88	−1.06	0.88	−24.18	0.39	5.96	0.88	100	–
		(0.09)	(0.09)	(0.24)	(0.89)	(0.08)	(0.12)	(0.85)	(0.11)	(0.28)			
25	2S-PML	1.74	2.04	1.56	−16.64	−1.08	1.04	−24.08	0.35	5.33	0.12	100	–
		(0.11)	(0.10)	(0.30)	(1.00)	(0.08)	(0.14)	(0.97)	(0.16)	(0.29)			
25	NPL	1.67	1.93	1.51	−18.83	−1.08	1.25	−23.13	0.37	7.28	32.84	13	708
		(0.22)	(0.45)	(0.18)	(8.01)	(0.12)	(0.47)	(3.93)	(0.32)	(8.75)			
50	ML	1.76	2.08	1.54	−17.00	−1.06	0.90	−24.31	0.38	5.98	0.98	100	–
		(0.06)	(0.07)	(0.17)	(0.70)	(0.07)	(0.08)	(0.60)	(0.06)	(0.16)			
50	2S-PML	1.75	2.07	1.54	−16.88	−1.07	0.98	−24.24	0.37	5.65	0.24	100	–
		(0.07)	(0.08)	(0.22)	(0.73)	(0.06)	(0.11)	(0.70)	(0.12)	(0.20)			
50	NPL	1.73	2.02	1.64	−16.40	−1.10	1.10	−23.95	0.30	4.85	47.47	11	634
		(0.06)	(0.07)	(0.17)	(0.59)	(0.05)	(0.07)	(0.64)	(0.07)	(0.07)			

Standard deviations are reported in parentheses. CPU time is the averaged time per run

algorithm fails to converge in 35 data sets for $T = 5$ and in 12 data sets for $T = 50$. In Scenario 6 where the best response unstable equilibria are played in the data, the NPL algorithm fails to converge frequently; for example, it converges in only 7 data sets for $T = 5$ and in 11 data sets for $T = 50$. While this finding is an improvement over the results in Scenario 3 where the NPL algorithm fails to converge in all one hundred data sets, the mean of the NPL estimator on the interaction coefficient δ from those converged data sets, except for $T = 25$, are two standard deviations or more away from the truth. Hence, the NPL estimator would reject the true parameter value $\delta^0 = 6$. Further examining the properties of the converged equilibrium by the NPL algorithm in each converged run, we find that the spectral radius $\rho\left[\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta})\right]$ evaluated at the converged equilibrium and the parameter estimates is strictly less than 1. This finding confirms the observation that implicitly, the NPL algorithm is searching for best response stable equilibrium in the iteration process. When best response unstable equilibria are played in the data, the NPL estimator cannot provide consistent estimates.

5.4 Post-estimation equilibrium analysis

In this subsection, we conduct post-estimation equilibrium analysis. To better understand the effects of using different estimators on the equilibrium structure (i.e., number of equilibria) and corresponding equilibrium probabilities, we use the mean of the three estimators from Scenario 6 and resolve the BN equilibrium model (6) for each of the 200 markets. In Table 8, we report only those markets in which the number of equilibria at the given parameter estimates differs from that at the true parameter values $\boldsymbol{\theta}^0$.

Our results indicate that using the mean of the ML estimator preserves the equilibrium structure, except in four markets for $T = 5$ and three markets for $T = 10$. On the contrary, using the mean of the 2S-PML and the NPL estimators tend to underpredidct the number of equilibria, especially when T is small. Recall that, at true parameter values θ^0 with $\delta^0 = 6$, there are 7 markets with three equilibria; see Table 4. Using the mean of the ML estimator, we find three equilibria in these 7 markets for all T. In addition, we also find three equilibria in another 4 markets (Market 137, 155, 174, and 185) for $T = 5$ and in another 3 markets (Market 137, 155 and 174) for $T = 10$. Using the mean of the 2S-PML and the NPL estimators for $T = 5$, we find very different equilibrium structure: a unique equilibrium exists in each of the 200 markets. As we increase T, the equilibrium structure implied by the mean of the 2S-PML and the NPL estimators become more similar to that at the true parameter values $\boldsymbol{\theta}^0$. For example, for $T = 50$, given the mean of the 2S-PML estimator, a unique equilibrium exists in one market (Market Index 164), and using the mean of the NPL estimator, a unique equilibrium exists in two markets (Market 164 and 177).

In Tables 9 and 10, we also report the corresponding equilibrium probabilities calculated using the mean of the three estimators for $T = 5$ and $T = 50$, respectively.[11] As shown in the tables, the equilibrium probabilities calculated using the mean of the

[11] Results for $T = 10$ and $T = 25$ provide similar observations to those for $T = 5$ and $T = 50$, and are omitted here.

Table 8 Post estimation analysis for scenario 6 – number of equilibria in each market at given parameter estimates

Market index	County identifier	Number of equilibria			
		$T=5$			
		$\boldsymbol{\theta}^0$	$\boldsymbol{\theta}^{MLE}$	$\boldsymbol{\theta}^{2S-PML}$	$\boldsymbol{\theta}^{NPL}$
137	184	1	3	1	1
155	241	1	3	1	1
162	258	3	3	1	1
164	261	3	3	1	1
165	262	3	3	1	1
171	273	3	3	1	1
174	277	1	3	1	1
177	282	3	3	1	1
181	288	3	3	1	1
185	293	1	3	1	1
187	298	3	3	1	1
		$T=10$			
		$\boldsymbol{\theta}^0$	$\boldsymbol{\theta}^{MLE}$	$\boldsymbol{\theta}^{2S-PML}$	$\boldsymbol{\theta}^{NPL}$
137	184	1	3	1	1
155	241	1	3	1	1
162	258	3	3	3	3
164	261	3	3	1	1
165	262	3	3	3	3
171	273	3	3	1	3
174	277	1	3	1	1
177	282	3	3	1	1
181	288	3	3	1	3
187	298	3	3	3	3
		$T=25$			
		$\boldsymbol{\theta}^0$	$\boldsymbol{\theta}^{MLE}$	$\boldsymbol{\theta}^{2S-PML}$	$\boldsymbol{\theta}^{NPL}$
123	164	1	1	1	3
134	179	1	1	1	3
162	258	3	3	3	1
164	261	3	3	1	1
165	262	3	3	3	1
171	273	3	3	3	1
177	282	3	3	1	1
181	288	3	3	3	1
187	298	3	3	3	1

Table 8 (continued)

Market index	County identifier	Number of equilibria			
		$T = 50$			
		$\boldsymbol{\theta}^0$	$\boldsymbol{\theta}^{MLE}$	$\boldsymbol{\theta}^{2S-PML}$	$\boldsymbol{\theta}^{NPL}$
162	258	3	3	3	3
164	261	3	3	1	1
165	262	3	3	3	3
171	273	3	3	3	3
177	282	3	3	3	1
181	288	3	3	3	3
187	298	3	3	3	3

Table 9 Post-estimation analysis for scenario 6 – equilibrium probabilities at given parameter estimates for $T = 5$

Market index	County identifier	Parameter values	Equilibrium probabilities (p_W, p_K)		
			Equilibrium 1	Equilibrium 2	Equilibrium 3
137	184	$\boldsymbol{\theta}^0$	(0.0248, 0.7693)	–	–
		$\boldsymbol{\theta}^{MLE}$	(0.0218, 0.7727)	(0.4725, 0.1739)	(0.6143, 0.0807)
		$\boldsymbol{\theta}^{2S-PML}$	(0.1663, 0.4619)	–	–
		$\boldsymbol{\theta}^{NPL}$	(0.0000, 0.6584)	–	–
155	241	$\boldsymbol{\theta}^0$	–	–	(0.4625, 0.0705)
		$\boldsymbol{\theta}^{MLE}$	(0.0968, 0.4156)	(0.1752, 0.3048)	(0.4937, 0.0579)
		$\boldsymbol{\theta}^{2S-PML}$	–	–	(0.2700, 0.1843)
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.4418)	–	–
162	258	$\boldsymbol{\theta}^0$	(0.0458, 0.8833)	(0.4378, 0.4187)	(0.8774, 0.0490)
		$\boldsymbol{\theta}^{MLE}$	(0.0393, 0.8893)	(0.4321, 0.4157)	(0.8843, 0.0418)
		$\boldsymbol{\theta}^{2S-PML}$	–	(0.5617, 0.3465)	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.8433)	–	–
164	261	$\boldsymbol{\theta}^0$	(0.0927, 0.4662)	(0.1896, 0.3281)	(0.5482, 0.0537)
		$\boldsymbol{\theta}^{MLE}$	(0.0683, 0.5190)	(0.2101, 0.3103)	(0.5788, 0.0442)
		$\boldsymbol{\theta}^{2S-PML}$	–	–	(0.3264, 0.1955)
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.5058)	–	–
165	262	$\boldsymbol{\theta}^0$	(0.0454, 0.9018)	(0.4749, 0.4110)	(0.8834, 0.0567)
		$\boldsymbol{\theta}^{MLE}$	(0.0393, 0.9067)	(0.4665, 0.4103)	(0.8909, 0.0482)
		$\boldsymbol{\theta}^{2S-PML}$	(0.3289, 0.6156)	–	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.8656)	–	–
171	273	$\boldsymbol{\theta}^0$	(0.0358, 0.7788)	(0.3750, 0.3151)	(0.7509, 0.0460)
		$\boldsymbol{\theta}^{MLE}$	(0.0308, 0.7812)	(0.3649, 0.3123)	(0.7564, 0.0390)
		$\boldsymbol{\theta}^{2S-PML}$	–	(0.3199, 0.3693)	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.6950)	–	–

Table 9 (continued)

Market index	County identifier	Parameter values	Equilibrium probabilities (p_W, p_K)		
			Equilibrium 1	Equilibrium 2	Equilibrium 3
174	277	$\boldsymbol{\theta}^0$	–	–	(0.6032, 0.0404)
		$\boldsymbol{\theta}^{MLE}$	(0.0969, 0.4761)	(0.1511, 0.3940)	(0.6206, 0.0346)
		$\boldsymbol{\theta}^{2S-PML}$	–	–	(0.3720, 0.1611)
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.4969)	–	–
177	282	$\boldsymbol{\theta}^0$	(0.0382, 0.6783)	(0.3539, 0.2407)	(0.6157, 0.0618)
		$\boldsymbol{\theta}^{MLE}$	(0.0330, 0.6881)	(0.3421, 0.2466)	(0.6350, 0.0510)
		$\boldsymbol{\theta}^{2S-PML}$	–	(0.2562, 0.3267)	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.6030)	–	–
181	288	$\boldsymbol{\theta}^0$	(0.0405, 0.7092)	(0.3311, 0.2991)	(0.6926, 0.0465)
		$\boldsymbol{\theta}^{MLE}$	(0.0348, 0.7183)	(0.3252, 0.2980)	(0.7040, 0.0394)
		$\boldsymbol{\theta}^{2S-PML}$	–	(0.3279, 0.2902)	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.6347)	–	–
185	293	$\boldsymbol{\theta}^0$	(0.0722, 0.4518)	–	–
		$\boldsymbol{\theta}^{MLE}$	(0.0608, 0.4766)	(0.3236, 0.1524)	(0.3740, 0.1164)
		$\boldsymbol{\theta}^{2S-PML}$	(0.2153, 0.2216)	–	–
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.4420)	–	–
187	298	$\boldsymbol{\theta}^0$	(0.0437, 0.8540)	(0.4011, 0.4066)	(0.8559, 0.0428)
		$\boldsymbol{\theta}^{MLE}$	(0.0375, 0.8626)	(0.3984, 0.4038)	(0.8645, 0.0368)
		$\boldsymbol{\theta}^{2S-PML}$	–	–	(0.5729, 0.2807)
		$\boldsymbol{\theta}^{NPL}$	(0.0001, 0.8102)	–	–

ML estimator match those at the true parameter values well, especially for $T = 50$. Although for $T = 5$, using the mean of ML estimator results in overpredicting the number of equilibria in 4 markets (Market 137, 155, 174, and 185), one of the three ML equilibria in each of those 4 markets is close to the true equilibrium at the true parameter values $\boldsymbol{\theta}^0$; see Table 9. For the 2S-PML estimator, the equilibrium probabilities calculated using the mean of the 2S-PML estimator for $T = 5$ are off from the true equilibrium probabilities by at least 10 % in most markets, except for Market 171 and 181. When we use the mean of the 2S-PML estimator for $T = 50$, the implied equilibrium probabilities are fairly accurate, with the difference being less than 2 % from the true equilibrium in most cases; see Table 10.

Using the mean the NPL estimator for $T = 5$, we find only one type of equilibrium in the markets: Walmart enters the market with very low probability (0.01 % or less) and Kmart enters with high probability (between 44 % in Market 155 and 86 % in Market 165); see Table 9. Clearly, this policy is very different from the equilibrium probabilities at the true parameter values. For $T = 50$, using the mean

Table 10 Post-estimation analysis for scenario 6 – equilibrium probabilities at given parameter estimates for $T = 50$

Market index	County identifier	Parameter values	Equilibrium probabilities (p_W, p_K)		
			Equilibrium 1	Equilibrium 2	Equilibrium 3
162	258	$\boldsymbol{\theta}^0$	(0.0458, 0.8833)	(0.4378, 0.4187)	(0.8774, 0.0490)
		$\boldsymbol{\theta}^{MLE}$	(0.0466, 0.8836)	(0.4392, 0.4199)	(0.8776, 0.0499)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0620, 0.8628)	(0.4415, 0.4239)	(0.8570, 0.0656)
		$\boldsymbol{\theta}^{NPL}$	(0.1423, 0.7862)	(0.4361, 0.4690)	(0.7991, 0.1318)
164	261	$\boldsymbol{\theta}^0$	(0.0927, 0.4662)	(0.1896, 0.3281)	(0.5482, 0.0537)
		$\boldsymbol{\theta}^{MLE}$	(0.0928, 0.4687)	(0.1911, 0.3287)	(0.5501, 0.0541)
		$\boldsymbol{\theta}^{2S-PML}$	–	–	(0.5021, 0.0751)
		$\boldsymbol{\theta}^{NPL}$	–	–	(0.4317, 0.1296)
165	262	$\boldsymbol{\theta}^0$	(0.0454, 0.9018)	(0.4749, 0.4110)	(0.8834, 0.0567)
		$\boldsymbol{\theta}^{MLE}$	(0.0463, 0.9021)	(0.4766, 0.4123)	(0.8836, 0.0578)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0608, 0.8846)	(0.4841, 0.4118)	(0.8622, 0.0763)
		$\boldsymbol{\theta}^{NPL}$	(0.1330, 0.8237)	(0.5078, 0.4310)	(0.7951, 0.1581)
171	273	$\boldsymbol{\theta}^0$	(0.0358, 0.7788)	(0.3750, 0.3151)	(0.7509, 0.0460)
		$\boldsymbol{\theta}^{MLE}$	(0.0363, 0.7775)	(0.3754, 0.3147)	(0.7493, 0.0467)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0501, 0.7451)	(0.3808, 0.3107)	(0.7116, 0.0649)
		$\boldsymbol{\theta}^{NPL}$	(0.1234, 0.6321)	(0.3894, 0.3209)	(0.5975, 0.1468)
177	282	$\boldsymbol{\theta}^0$	(0.0382, 0.6783)	(0.3539, 0.2407)	(0.6157, 0.0618)
		$\boldsymbol{\theta}^{MLE}$	(0.0387, 0.6773)	(0.3550, 0.2402)	(0.6139, 0.0629)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0545, 0.6355)	(0.3771, 0.2196)	(0.5482, 0.0966)
		$\boldsymbol{\theta}^{NPL}$	(0.1391, 0.4982)	–	–
181	288	$\boldsymbol{\theta}^0$	(0.0405, 0.7092)	(0.3311, 0.2991)	(0.6926, 0.0465)
		$\boldsymbol{\theta}^{MLE}$	(0.0412, 0.7074)	(0.3313, 0.2987)	(0.6904, 0.0473)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0585, 0.6654)	(0.3320, 0.2975)	(0.6472, 0.0665)
		$\boldsymbol{\theta}^{NPL}$	(0.1790, 0.4817)	(0.2826, 0.3599)	(0.5231, 0.1489)
187	298	$\boldsymbol{\theta}^0$	(0.0437, 0.8540)	(0.4011, 0.4066)	(0.8559, 0.0428)
		$\boldsymbol{\theta}^{MLE}$	(0.0444, 0.8542)	(0.4024, 0.4077)	(0.8561, 0.0436)
		$\boldsymbol{\theta}^{2S-PML}$	(0.0601, 0.8292)	(0.4011, 0.4138)	(0.8339, 0.0576)
		$\boldsymbol{\theta}^{NPL}$	(0.1466, 0.7326)	(0.3746, 0.4753)	(0.7752, 0.1148)

of the NPL estimates tends to overestimate (underestimate) the entry probability of Walmart (Kmart) in Equilibrium 1 but underestimate (overestimate) the entry probability of Walmart (Kmart) in Equilibrium 3; see Table 10. In some cases, the difference between the equilibrium probabilities at the mean of the NPL estimator and at the true equilibrium probabilities is around 10 % or larger; see, for example, Market 171 or 181 in Table 10.

6 Conclusion

We have proposed a constrained optimization formulation of the ML estimation problem for static games of incomplete information and conducted Monte Carlo experiments to examine the finite-sample performance of the ML estimator, the 2S-PML and 2S-LS estimators, and the NPL estimator. Our Monte Carlo results demonstrate that the finite-sample performance of the ML estimator is superior to those of the two-step estimators as well as the NPL estimator, particularly when small numbers of repeated observations exist in the data. The 2S-PML and 2S-LS estimators perform well when there are sufficiently many repeated observations to estimate equilibrium probabilities accurately in the first step; however, they can be biased when only small numbers of repeated observations are available or when the data are generated by best-response unstable equilibria. In practice, if there are sufficient amounts of data available to researchers to accurately estimate choice probabilities in the first stage, then 2S-PML and 2S-LS estimators provide a viable and computationally efficient way to obtain parameter estimates. The NPL estimator frequently fails to converge when only small numbers of repeated observations are available or when best-response unstable equilibria are played in the data. In the latter case, the NPL estimator produces highly biased estimates when the NPL algorithm converges. Given that one cannot test whether data were generated by best-response stable or unstable equilibria prior to estimating the model, we believe the ML estimator implemented by the constrained optimization approach is the better choice, provided the size of the estimation problem to be solved is computationally manageable.

In the post-estimation equilibrium analysis, we also find that with small number of repeated observations, the 2S-PML and the NPL estimators can predict fairly different equilibrium structure and equilibrium policies when the data are generated by best-response unstable equilibria, while the ML estimator preserve the equilibrium structure and equilibrium probabilities quite well.

While we do not have a general characterization theorem to indicate which estimator will work better under certain situations or which numerical algorithm will converge faster in practice on specific models, our findings from Monte Carlo experiments are still important and valuable. Researchers need to be aware of the trade-off between computational costs and the accuracy of parameter estimates as well implied equilibrium probabilities at the parameter estimates in the post-estimation equilibrium analysis. To avoid unintended consequences, researchers also need to understand the implications of choosing specific computational procedures to implement specific estimation methods.

Acknowledgments I thank the editor Greg M. Allenby and two anonymous referees for insightful comments and helpful suggestions. I have benefited greatly from discussions with Panle Jia Barwick, Kenneth L. Judd, Sanjog Misra, Ariel Pakes, Peter E. Rossi, and John Rust, and comments from seminar participants at the 2010 QME Conference, Boston College, and Harvard University. I am very grateful to John Rust for developing and suggesting the first static-game example used in this paper and to Harry J. Paarsch for carefully reading this paper. I acknowledge the financial support from the IBM Faculty Research Fund at The University of Chicago Booth School of Business. All errors are mine.

Appendix A: The constrained optimization formulation for the ML estimator under general equilibrium selection mechanism

In this appendix, we derive the constrained optimization formulation of the maximum likelihood estimation problem under a general equilibrium selection mechanism. To simplify the notation, we only consider the case of one market and drop the superscript m for market index in the notation used in Section 3.

Given the parameter vector $\boldsymbol{\theta}$ and firms' observed types $\boldsymbol{x}$, let $\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x}) = \{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k$ denote the set of BN equilibria, where $\bar{\boldsymbol{p}}_k(\boldsymbol{\theta}) = (\bar{p}_{ka}(\boldsymbol{\theta}), \bar{p}_{kb}(\boldsymbol{\theta}))$ is the k-th equilibrium that solves the BN equilibrium (4). For the constrained optimization approach, given $\boldsymbol{\theta}$, an equilibrium selection mechanism $\boldsymbol{\lambda} = \{\lambda_k\}_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}$ and any given set of probabilities $\{\boldsymbol{p}_k = (p_{ka}, p_{kb})\}_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}$ (which do not need to be the equilibrium probabilities that solve the BN equilibrium equation), we define the augmented logarithm of the likelihood function of observing the decisions $\boldsymbol{y} = (y_a^t, y_b^t)_{t=1}^T$ as

$$\mathfrak{L}\left[\boldsymbol{\theta}, \boldsymbol{\lambda}, \{\boldsymbol{p}_k\}_k\right] = \sum_{t=1}^{T} \log \left\{ \sum_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|} \lambda_k \left[(p_{ka})^{y_a^t}\right]\left[(1-p_{ka})^{1-y_a^t}\right]\left[(p_{kb})^{y_b^t}\right]\left[(1-p_{kb})^{1-y_b^t}\right] \right\}. \tag{26}$$

The constrained optimization formulation of the ML estimator is then defined as:

$$\begin{aligned} &\underset{\{\boldsymbol{\theta},\boldsymbol{\lambda},\{\boldsymbol{p}_k\}_k\}}{\text{maximize}} && \mathfrak{L}\left[\boldsymbol{\theta}, \boldsymbol{\lambda}, \{\boldsymbol{p}_k\}_k\right] \\ &\text{s.t.} && \sum_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|} \lambda_k = 1, \\ & && \boldsymbol{p}_k = \boldsymbol{\Psi}(\boldsymbol{p}_k, \boldsymbol{x}; \boldsymbol{\theta}), \quad k = 1, \ldots, |\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x})|. \end{aligned} \tag{27}$$

Notice the constrained optimization problem formulated above is not computable because the number of constraints, $|\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x})|$, in Eq. 27 depends on the values of decision variables $\boldsymbol{\theta}$. Nonetheless, we can characterize the equivalence in objective value of the two problems (10) and (27) in the following proposition.

Proposition 2 *Let* $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}))$ *be a solution of the ML estimation problem defined in* Eq. 10 *with corresponding equilibrium probabilities* $\{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_{k=1}^{|\mathcal{E}(\bar{\boldsymbol{\theta}},\boldsymbol{x})|}$. *Let* $(\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \{\boldsymbol{p}_k^*\}_{k=1}^{|\mathcal{E}(\theta^*,\boldsymbol{x})|})$ *be a solution of the constrained optimization problem* (27). *Then* $\mathbb{L}[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_k] = \mathfrak{L}\left[\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \{\boldsymbol{p}_k^*\}_k\right]$. *If the model is identified, then* $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^*$.

Proof By definition, $\mathbb{L}[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_k] \geq \mathbb{L}[\boldsymbol{\theta}, \boldsymbol{\lambda}(\boldsymbol{\theta}), \{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k]$ for any given $\boldsymbol{\theta}$, $\boldsymbol{\lambda}(\boldsymbol{\theta})$ and $\{\bar{\boldsymbol{p}}_k(\boldsymbol{\theta})\}_k$ that satisfy the BN equilibrium equation $\boldsymbol{p} = \Psi(\boldsymbol{p}, \boldsymbol{x}, \boldsymbol{\theta})$. Since the pair $(\boldsymbol{\theta}^*, \{\boldsymbol{p}_k^*\}_{k=1}^{|\mathcal{E}(\theta^*,\boldsymbol{x})|})$ satisfies the BN equation as the constraints in Eq. 27, it follows that

$$\mathbb{L}[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_k]] \geq \mathbb{L}\left[\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \{\boldsymbol{p}_k^*\}_k\right] = \mathfrak{L}\left[\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \{\boldsymbol{p}_k^*\}_k\right].$$

Conversely, since the tuple $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_{k=1}^{|\mathcal{E}(\bar{\boldsymbol{\theta}},\boldsymbol{x})|})$ satisfies the constraints in Eq. 27, we have

$$\mathfrak{L}\left[\boldsymbol{\theta}^*, \boldsymbol{\lambda}^*, \{\boldsymbol{p}_k^*\}_k\right] \geq \mathfrak{L}[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_k] = \mathbb{L}[\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\lambda}}(\bar{\boldsymbol{\theta}}), \{\bar{\boldsymbol{p}}_k(\bar{\boldsymbol{\theta}})\}_k].$$

If the model is identified, the solution is unique, so $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^*$. □

Assuming only one equilibrium is played in the data, only one component in $\boldsymbol{\lambda}$ is 1 and all the others are 0. Then the constrained optimization problem (27) is equivalent to

$$\begin{aligned} &\underset{\{\boldsymbol{\theta},\{\boldsymbol{p}_k\}_k\}}{\text{maximize}} \quad \max_{k=1,\ldots,|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|} \left[\sum_{t=1}^{T} \log\left\{(p_{ka})^{y_a^t}(1-p_{ka})^{1-y_a^t}(p_{kb})^{y_b^t}(1-p_{kb})^{1-y_b^t}\right\}\right] \\ &\text{s.t.} \qquad \boldsymbol{p}_k = \boldsymbol{\Psi}(\boldsymbol{p}_k, \boldsymbol{x}; \boldsymbol{\theta}), \quad k = 1, \ldots, |\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x})|. \end{aligned} \tag{28}$$

Notice that the constrained optimization problem formulated in Eq. 28 is also not computable because its size (the number of constraints, $|\mathcal{E}(\boldsymbol{\theta}, \boldsymbol{x})|$ and decision variables $\{\boldsymbol{p}_k\}_k$) depend on the values of decision variables $\boldsymbol{\theta}$; it also involves finding all the equilibrium solutions $\{\boldsymbol{p}_k\}_{k=1}^{|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}$ at any given $\boldsymbol{\theta}$. We present a computationally tractable reformulation of the problem (28) as follows:

$$\begin{aligned} &\underset{\{\boldsymbol{\theta},\boldsymbol{p}\}}{\text{maximize}} \sum_{t=1}^{T} \log\left\{(p_a)^{y_a^t}(1-p_a)^{1-y_a^t}(p_b)^{y_b^t}(1-p_b)^{1-y_b^t}\right\} \\ &\text{s.t.} \qquad \boldsymbol{p} = \boldsymbol{\Psi}(\boldsymbol{p}, \boldsymbol{x}; \boldsymbol{\theta}). \end{aligned} \tag{29}$$

Notice that the size (number of decision variables and constraints) of the problem (29) is fixed and does not depend on the value of $\boldsymbol{\theta}$. Moreover, we do not need to find all equilibria at any given $\boldsymbol{\theta}$; only one equilibrium $\boldsymbol{p}$ is needed since only one equilibrium constraint is imposed. The equivalence in the objective value and solutions between the problems (28) and (29) is stated below. We omit the proof since it follows similar arguments used in proving Proposition 2.

Proposition 3 *Let* $\left(\boldsymbol{\theta}^*, \{\boldsymbol{p}_k^*\}_{k=1}^{|\mathcal{E}(\boldsymbol{\theta}^*,\boldsymbol{x})|}\right)$ *denote the solution to the optimization problem* (28) *with* $\boldsymbol{p}_{\bar{k}}^* \in \{\boldsymbol{p}_k^*\}_k$ *being the equilibrium that maximizes the objective function, i.e.,*

$$\bar{k} = \underset{k=1,\ldots,|\mathcal{E}(\boldsymbol{\theta},\boldsymbol{x})|}{\text{argmax}} \left[\sum_{t=1}^{T} \log\left\{(p_{ka}^*)^{y_a^t}\left(1-p_{ka}^*\right)^{1-y_a^t}(p_{kb}^*)^{y_b^t}\left(1-p_{kb}^*\right)^{1-y_b^t}\right\}\right].$$

Then, the vector $\left(\boldsymbol{\theta}^*, \boldsymbol{p}_{\bar{k}}^*\right)$ *is also an optimal solution to the optimization problem* (29). *Conversely, let* $\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{p}}\right)$ *denote a solution of the problem* (29). *Then the pair* $\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{p}}\right)$ *together with all the other equilibrium probabilities* $\{\boldsymbol{p}_k\}_k \setminus \hat{\boldsymbol{p}}$ *at* $\hat{\boldsymbol{\theta}}$ *is an solution to the problem* (28).

Finally, the formulation in Eq. 29 *can be generalized to the constrained optimization problem for the ML estimator defined in* Eq. 18 *for the case of observing data in multiple markets.*

Appendix B: Creating examples with multiple equilibria

In this appendix, we derive the formula for calculating the spectral radius $\rho\,[\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P},\boldsymbol{X};\boldsymbol{\theta})]$ for the model in Section 5. We first derive the formula for calculating the spectral radius for one market and then generalize the formula to the case with multiple markets. Recall that in Section 5, for $m=1,\ldots,M$, we have

$$p_W^m = \frac{\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_W\boldsymbol{x}_W^m - \delta p_K^m)}{1+\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_W\boldsymbol{x}_W^m - \delta p_K^m)} = \Psi_W(\boldsymbol{x}^m,\boldsymbol{x}_W^m,p_K^m;\boldsymbol{\theta})$$

$$p_K^m = \frac{\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_K\boldsymbol{x}_K^m - \delta p_W^m)}{1+\exp(\boldsymbol{\alpha}'\boldsymbol{x}^m + \boldsymbol{\beta}'_K\boldsymbol{x}_K^m - \delta p_W^m)} = \Psi_K(\boldsymbol{x}^m,\boldsymbol{x}_K^m,p_W^m;\boldsymbol{\theta}),$$

We represent the BN equilibrium equations above for market m as

$$\boldsymbol{p}^m = \boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta}),$$

where $\boldsymbol{p}^m=(p_W^m,p_K^m)$, $\boldsymbol{\Psi}=(\Psi_W,\Psi_K)$, and $\boldsymbol{X}^m=(\boldsymbol{x}^m,\boldsymbol{x}_W^m,\boldsymbol{x}_K^m)$. The Jacobian of the mapping $\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta})$ with respect to $\boldsymbol{p}^m$ is

$$\nabla_{\boldsymbol{p}^m}\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial\Psi_W}{\partial p_W^m} & \frac{\partial\Psi_W}{\partial p_K^m} \\ \frac{\partial\Psi_K}{\partial p_W^m} & \frac{\partial\Psi_K}{\partial p_K^m} \end{bmatrix} = \begin{bmatrix} 0 & p_W^m(1-p_W^m)\delta \\ p_K^m(1-p_K^m)\delta & 0 \end{bmatrix}.$$

The eigenvalues λ^m of $\nabla_{\boldsymbol{p}^m}\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta})$ satisfy the following conditions:

$$\det\left(\nabla_{\boldsymbol{p}^m}\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta})-\lambda^m I\right) = (\lambda^m)^2 - p_W^m(1-p_W^m)p_K^m(1-p_K^m)\delta^2 = 0,$$

where I is the identity matrix. Hence, we have

$$\lambda^m = \pm\delta\sqrt{p_W^m(1-p_W^m)p_K^m(1-p_K^m)}.$$

The spectral radius of the Jacobian mapping $\nabla_{\boldsymbol{p}^m}\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta})$ for market m is then

$$\rho[\nabla_{\boldsymbol{p}^m}\boldsymbol{\Psi}(\boldsymbol{p}^m,\boldsymbol{X}^m;\boldsymbol{\theta})] = \max|\lambda^m| = |\delta|\sqrt{p_W^m(1-p_W^m)p_K^m(1-p_K^m)}.$$

Next we generalize the derivation above to the case with multiple markets. Recall we represent the BN equilibrium equations for all markets $m=1,\ldots,M$ as

$$\boldsymbol{P} = \boldsymbol{\Psi}(\boldsymbol{P},\boldsymbol{X};\boldsymbol{\theta}),$$

where $\boldsymbol{P}=(p_W^m,p_K^m)_{m=1}^M$, $\boldsymbol{\Psi}=(\Psi_W,\Psi_K)$ and $\boldsymbol{X}=(\boldsymbol{x}^m,\boldsymbol{x}_W^m,\boldsymbol{x}_K^m)_{m=1}^M$. Observe that the Jacobian matrix $\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P},\boldsymbol{X};\boldsymbol{\theta})$ has the block-diagonal structure

$$\nabla_{\boldsymbol{P}}\boldsymbol{\Psi}(\boldsymbol{P},\boldsymbol{X};\boldsymbol{\theta}) = \begin{bmatrix} \nabla_{\boldsymbol{p}^1}\boldsymbol{\Psi}(\boldsymbol{p}^1,\boldsymbol{X}^1;\boldsymbol{\theta}) & 0 & \cdots & 0 \\ 0 & \nabla_{\boldsymbol{p}^2}\boldsymbol{\Psi}(\boldsymbol{p}^2,\boldsymbol{X}^2;\boldsymbol{\theta}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \nabla_{\boldsymbol{p}^M}\boldsymbol{\Psi}(\boldsymbol{p}^M,\boldsymbol{X}^M;\boldsymbol{\theta}) \end{bmatrix}.$$

Consequently, the spectral radius of the Jacobian mapping $\nabla_{\boldsymbol{P}} \boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta})$ is given as

$$\rho[\nabla_{\boldsymbol{P}} \boldsymbol{\Psi}(\boldsymbol{P}, \boldsymbol{X}; \boldsymbol{\theta})] = \max_{m=1,\ldots,M} |\lambda^m| = \max_{m=1,\ldots,M} \left\{ |\delta| \sqrt{p_W^m (1 - p_W^m) p_K^m (1 - p_K^m)} \right\}.$$

To create an example with multiple equilibria in some markets, we need the spectral radius $\rho[\nabla_{\boldsymbol{P}} \boldsymbol{\Psi}(\boldsymbol{P}^*, \boldsymbol{X}; \boldsymbol{\theta}^0)]$ evaluated at structural parameters $\boldsymbol{\theta}^0$ and the corresponding equilibrium P to be greater than 1, which implies $|\delta^0| \sqrt{p_W^m (1 - p_W^m) p_K^m (1 - p_K^m)} > 1$ for some m. This, in turns, requires that

$$|\delta^0| > 4,$$

since $\sqrt{p_W^m (1 - p_W^m) p_K^m (1 - p_K^m)} < 1/4$. As a conservative choice, we choose $\delta^0 = 6$ in the second set of true parameter values in Eq. 25.

References

Arcidiacono, P., & Miller, R.A. (2011). CCP estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, *79*, 1823–1868.

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, *75*, 1–53.

Aitchison, J., & Silvey, S.D. (1958). Maximum likelihood estimation of parameters subject to restraints. *Annals of Mathematical Statistics*, *29*, 813–828.

Bajari, P., Benkard, C.L., Levin, J. (2007). Estimating dynamic games of imperfect competition. *Econometrica*, *75*, 1331–1370.

Bajari, P., Hong, H., Krainer, J., Nekipelov, D. (2010). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics*, *28*, 469–482.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, *60*, 889–917.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, *77*, 1791–1828.

de Paula, A. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Econonomics*, *5*, 107–131.

Ellickson, P., & Misra, S. (2008). Supermarket pricing strategies. *Marketing Science*, *27*, 811–828.

Ellickson, P., & Misra, S. (2011). Estimating discrete games. *Marketing Science*, *30*, 997–1010.

Fudenberg, D., & Tirole, J. (1991). In *Game theory*. Cambridge: MIT Press.

Gourieroux, C., & Monfort, A. (1995). In *Statistics and econometric models* Vol. 1. Cambridge: Cambridge University Press.

Hotz, J., & Miller, R.A. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, *60*, 497–529.

Jia, P. (2008). What happens when Wal-Mart comes to town: an empirical analysis of the discount retail industry. *Econometrica*, *76*, 1263–1316.

Judd, K., Renner, P., Schmedders, K. (2012). Finding all pure strategy equilibria in dynamic and static games with continuous strategies. *Quantitative Economics*, *3*, 289–331.

Kohlberg, E., & Mertens, J.-F. (1986). On the strategic stability of equilibria. *Econometrica*, *54*, 1003–1038.

Mertens, J.-F. (1989). Stable equilibria–a reformulation: Part I. Definition and basic properties. *Mathematics of Operations Research*, *14*, 575–625.

Mertens, J.-F. (1991). Stable equilibria–a reformulation: Part II. Discussion of the definition, and further results. *Mathematics of Operations Research*, *16*, 694–753.

Misra, S. (2013). Markov chain Monte Carlo for incomplete information discrete games. *Quantitative Marketing and Economics*, *11*, 117–153.

Orhun, A.Y. (2013). Spatial differentiation in the supermarket industry: the role of common information. *Quantitative Marketing and Economics*, *11*, 3–37.

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

Pakes, A., Porter, J., Ho, K., Ishii, J. (2011). In *Moment Inequalities and Their Applications*. Working paper: Harvard University.

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

Pesendorfer, M., & Schmidt-Dengler, P. (2010). Sequential estimation of dynamic discrete games: a comment. *Econometrica*, *78*, 833–842.

Rust, J. (1987). Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica*, *55*, 999–1033.

Rust, J. (2008). Private communication.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, *37*, 619–640.

Su, C.-L., & Judd, K.L. (2012). Constrained optimization approaches to estimation of structural models. *Econometrica*, *80*, 2213–2230.

Silvey, S.D. (1975). In *Statistical inference*. London: Chapman & Hall.

Sweeting, A. (2009). The strategic timing of radio commercials: an empirical analysis using multiple equilibria. *RAND Journal of Economics*, *40*, 710–742.

Vitorino, M.A. (2012). Empirical entry games with complementarities: an application to the shopping center industry. *Journal of Marketing Research*, *49*, 175–191.

Zhu, T., & Singh, V. (2009). Spatial competition with endogenous location choices: an application to discount retailing. *Quantitative Marketing and Economics*, *7*, 1–35.