Quant Mark Econ (2013) 11:39–81
DOI 10.1007/s11129-012-9127-6

Bayesian estimation of discrete games of complete information

Sridhar Narayanan

Received: 30 May 2011 / Accepted: 6 September 2012 / Published online: 10 October 2012

Abstract Estimation of discrete games of complete information, which have been applied to a variety of contexts such as market entry, technology adoption and peer effects, is challenging due to the presence of multiple equilibria. In this paper, we take a Bayesian MCMC approach to this problem, specifying a prior over multiple equilibrium selection mechanisms reflecting the analysts uncertainty over them. We develop a sampler, using the reversible jump algorithm to generate draws from the posterior distribution of parameters across these equilibrium selection rules. The algorithm is flexible in that it can be used both in situations where the equilibrium selection rule is identified and when it is not, and accommodates heterogeneity in equilibrium selection. We explore the methodology using both simulated data and two empirical applications, one in the context of joint consumption, using a dataset of casino visit decisions by married couples, and the second in the context of market entry by competing chains in the retail stationery market. We demonstrate the importance of accounting for multiple equilibrium selection rules in these applications and show that taking an empirical approach to the issue, such as the one we have demonstrated, can be useful.

Keywords Discrete games · Multiple equilibria · Bayesian estimation · Markov Chain Monte Carlo methods · Reversible jump algorithm

JEL Classification C11 · C31 · C35 · M31

S. Narayanan (✉)
Graduate School of Business, Stanford University, Stanford, CA 94305, USA
e-mail: sridhar.narayanan@stanford.edu

1 Introduction

In a number of situations, agents' payoffs from their actions are not independent of actions of other agents. For instance, the profits that a firm would earn from entering a market would depend on the entry decisions of its competitors, or the benefits to adopting a technology or a new product with network effects depend on adoption decisions of others. Discrete games have been applied to several such contexts where decisions are discrete or can be discretized. Discrete games of complete information were first applied by Bresnahan and Reiss (1990) and Berry (1992) in market entry contexts. The actions of agents are modeled as the Nash equilibrium outcomes of a static game where all agents are fully informed about the payoffs of all other agents. Subsequently, such games have been applied to product quality choice (Mazzeo 2002), pricing strategy choice (Zhu et al. 2009) and joint consumption (Hartmann 2010) amongst others. A specific issue with such games is the absence of a one-to-one mapping between parameters and outcomes—a problem of multiplicity of equilibria. The extant literature has proposed ways to deal with multiplicity, ranging from the modeling of outcomes that are unique even in the presence of multiplicity (Bresnahan and Reiss 1990; Berry 1992), specification of a sequential rather than simultaneous move game (Berry 1992), assuming an ad-hoc equilibrium selection rule (Hartmann 2010), randomizing the equilibria (Soetevent and Kooreman 2007), obtaining the identified set estimates rather than unidentified point estimates of parameters (Ciliberto and Tamer 2009) and empirically estimating the equilibrium selection rule (Bajari et al. 2010). In this paper, we propose an alternative hierarchical Bayesian approach to deal with this issue of multiplicity, in which the uncertainty of the analyst over the multiple equilibria is accounted for and posterior parameter estimates obtained that reflect this uncertainty.

The problem of multiple equilibria was explored in detail by Bresnahan and Reiss (1990, 1991) and Berry (1992) in their seminal studies of market entry. The main issue with estimation in the presence of multiple equilibria is that for a given set of parameters, observable covariates and unobservable variables, there is the potential for more than one equilibrium outcome. This causes the likelihood to be ill-defined because the probabilities of the potential outcomes for a particular observation add up to a value greater than 1. Such an econometric model is incomplete (Tamer 2003) and cannot be directly estimated using a maximum likelihood procedure without simplifying assumptions. The literature on discrete games has made various simplifying assumptions, but these assumptions either reduce the scope of the problems that can be studied (e.g. the study of unique number of entrants rather than their identities in Bresnahan and Reiss 1990) or can lead to inconsistency of estimates if the assumption is invalid (e.g. the assumed sequence of entry in Berry 1992, or the assumption of equilibrium selection in Hartmann 2010). Alternatively, the literature has taken a partial identification approach to the problem, identifying bounds for parameters rather than point estimates (see for instance

Beresteanu et al. 2011; Ciliberto and Tamer 2009; Pakes et al. 2006). While this approach has appeal since it does not rely on equilibrium selection assumptions, an issue often faced by the empirical researcher in adopting these approaches is the difficulty in conducting counterfactual simulations, where the researcher needs to take a stand on equilibrium selection in order to simulate outcomes. A third approach to the problem of multiplicity has been to take an empirical route, estimating the equilibrium selection rule (Bajari et al. 2010, see also Bjorn and Vuong 1984) in the case of complete information games. However, this approach either requires the researcher to observe a set of variables that affect one player's payoffs, but can be excluded from the payoff functions of other players, or relies on an identification at infinity argument, where there are observations with sufficiently large values of covariates. In practice, it may be difficult to find such excluded variables in some contexts, or observations that meet the requirements for identification at infinity. Empirical approaches to the issue of multiplicity have also been adopted in the case of incomplete information games (see for instance Su and Judd forthcoming; Misra 2012) but are not discussed here in detail since our focus is on complete information games. Further, for detailed reviews of the empirical literature on discrete games, the reader is referred to Berry and Tamer (2007) and Ellickson and Misra (2011).

We develop a hierarchical Bayesian approach to estimate parameters of the payoff functions of the players in the presence of multiple equilibria. Essentially, this approach takes a Bayesian model selection route to addressing the problem of multiplicity, by conditioning the model on equilibrium selection. Conditional on equilibrium selection, the likelihood is well-defined, and hence the parameters of this conditional model can be estimated. Thus, the problem can be considered to be one where there are multiple (conditional on equilibrium selection) models, over which the analyst has uncertainty. We develop a sampling algorithm for this problem, using a reversible jump MCMC algorithm to sample the parameters over the multiple models corresponding to the multiple equilibria jointly with indicators for the models themselves. This procedure generates draws from the posterior distribution of parameters across the multiple equilibria, reflecting our uncertainty about equilibrium selection. Thus, the proposed method provides a practical tool to deal with multiple equilibria in the estimation of discrete games.

We demonstrate our proposed algorithm first using simulated data. We show that it is able to recover parameters of a discrete game with more than two players with a high degree of accuracy. We further demonstrate through these simulations the downsides of making ad-hoc equilibrium selection mechanisms. We then apply the methodology to a social interaction game involving casino visits by married couples. We estimate the parameters of the model, and demonstrate how the parameter estimates as well as estimates of counterfactual simulations are biased when the presence of multiple equilibria is ignored or dealt with using an ad-hoc equilibrium selection rule. We finally present an application to a competitive market entry game in the retail stationery market. Our estimates show that the commonly assumed equilibrium selection rule has low posterior probability in this market, raising questions about the assumption for other markets as well.

The rest of the paper is organized as follows. We first set up an illustrative model to demonstrate the issue of multiplicity for two different contexts—one where players' payoffs are positively affected by the presence of other players, and another where they are negatively affected. We briefly discuss existing approaches to estimating discrete games of complete information, including approaches that have been used to deal with multiplicity of equilibria. Next, we explain in detail our proposed Bayesian method to estimate the parameters of the conditional model and model indicators using a reversible jump MCMC sampler. We discuss conceptual issues related to the methodology, as well as issues related to implementation. We next demonstrate our methodology using simulated data, and then present our empirical application. We discuss the results of the empirical analysis, pointing in particular to the consequences of assumptions on equilibrium selection. We finally conclude.

2 Multiple equilibria in discrete games

We first set up an illustrative model for a market entry game, modifying the discussion for a social interactions game subsequently. Assume that there are a total of $i = 1, \ldots, I$ players in the game, and $t = 1 \ldots T$ outcomes observed for these players. In a market entry context, the players are firms considering entry in the T markets, and the outcomes represent the market structure in each market, i.e. which of the firms choose to enter and which do not. Let the payoff functions for firm i in market t be the following if it chooses to enter the market.

$$\pi_{it} = \alpha_i + X_{it}\beta - \gamma \cdot 1\left(\sum_{j\neq i} y_{jt} > 0\right) - \gamma\delta \cdot 1\left(\sum_{j\neq i} y_{jt} > 1\right) - \gamma\delta^2 \cdot 1\left(\sum_{j\neq i} y_{jt} > 2\right) - \ldots + \varepsilon_{it} \quad (1)$$

Here, α_i is a player specific intercept and represents the intrinsic profitability of the firm. X_{it} is a vector of exogenous covariates, representing factors like the population size and other demographic factors of market t, and β is a vector of coefficients for these covariates. After this are a set of terms that represent the effect of the presence of competitors in market t. y_{it} represents the decision of firm i to enter market t and takes the value 1 if the firm chooses to enter and 0 otherwise. The first of the competitive effect terms thus captures the impact on firm i's profits if it has one competitor in the market, and subsequent terms capture the effect of additional competitors. The parameter γ and δ

are assumed to be positive, and thus the competitive effect is assumed to be negative. Finally, there is an additively separable error which is unobservable to the econometrician, but is observed by all the firms. This may represent unobserved fixed costs of entering the market, which is firm specific. The common knowledge of the errors amongst all players of the game makes this a complete information game. The specification laid out above is similar to the one in Hartmann (2010) and captures the competitive effect in a parsimonious way.[1]

Each firm considers its own entry decision, taking into account the optimal decisions of its competitors. It chooses to enter the market if it expects positive profits, and chooses to stay out of the market otherwise. The outcome is thus a Nash equilibrium of this game, with an equilibrium in pure strategies guaranteed with the restrictions on parameters that we have imposed (non negative γ and δ).

As has been pointed out in the literature, such a game has multiple equilibria. To illustrate this point, we consider the simplest case of a two player game without covariates. The profit functions of the two firms are then given by

$$\pi_{it} = \alpha_i - \gamma \cdot y_{-i,t} + \varepsilon_{it}, \quad i \in \{1, 2\} \tag{2}$$

where $y_{-i,t}$ is the entry decision of the competing firm.

Let us first consider the case where neither firm enters market t. Thus, both y_{1t} and y_{2t} are 0 in this case. The profits for the firms in this situation are given by

$$\begin{aligned} \pi_{1t} &= \alpha_1 + \varepsilon_{1t} < 0 \\ \pi_{2t} &= \alpha_2 + \varepsilon_{1t} < 0 \end{aligned} \tag{3}$$

Thus, for neither firm to enter

$$\begin{aligned} \varepsilon_{1t} &< -\alpha_1 \\ \varepsilon_{2t} &< -\alpha_2 \end{aligned} \tag{4}$$

The equilibrium conditions for the other three outcomes, i.e. firm 1 entering alone, firm 2 entering alone, and both firms entering the market can be similarly evaluated. The full set of equilibrium conditions are depicted graphically in Fig. 1. Consider panel A of this figure. There are nine regions depicted in this panel, which we shall refer to as cells. These cells are labeled I through

[1] It is necessary to give structure to the competitive effect, particularly for small number of players since there are limited degrees of freedom available. For instance, in a two player game, one can estimate at most three parameters in addition to the coefficients for exogenous covariates—for instance two firm-specific intercepts and a competitive effect parameter. In a 3-player game, one can estimate at most seven parameters, and hence we would not be able to estimate the most general competitive effect even in this case, where there were for instance a different effect of each firm on each of its competitors, in addition to firm-specific intercepts. As the number of players increase, the degrees of freedom increase exponentially, giving more flexibility in specifying the competitive effects. The specific model shown here is for illustrative purposes—the methodology can be extended to other models of discrete games of complete information.

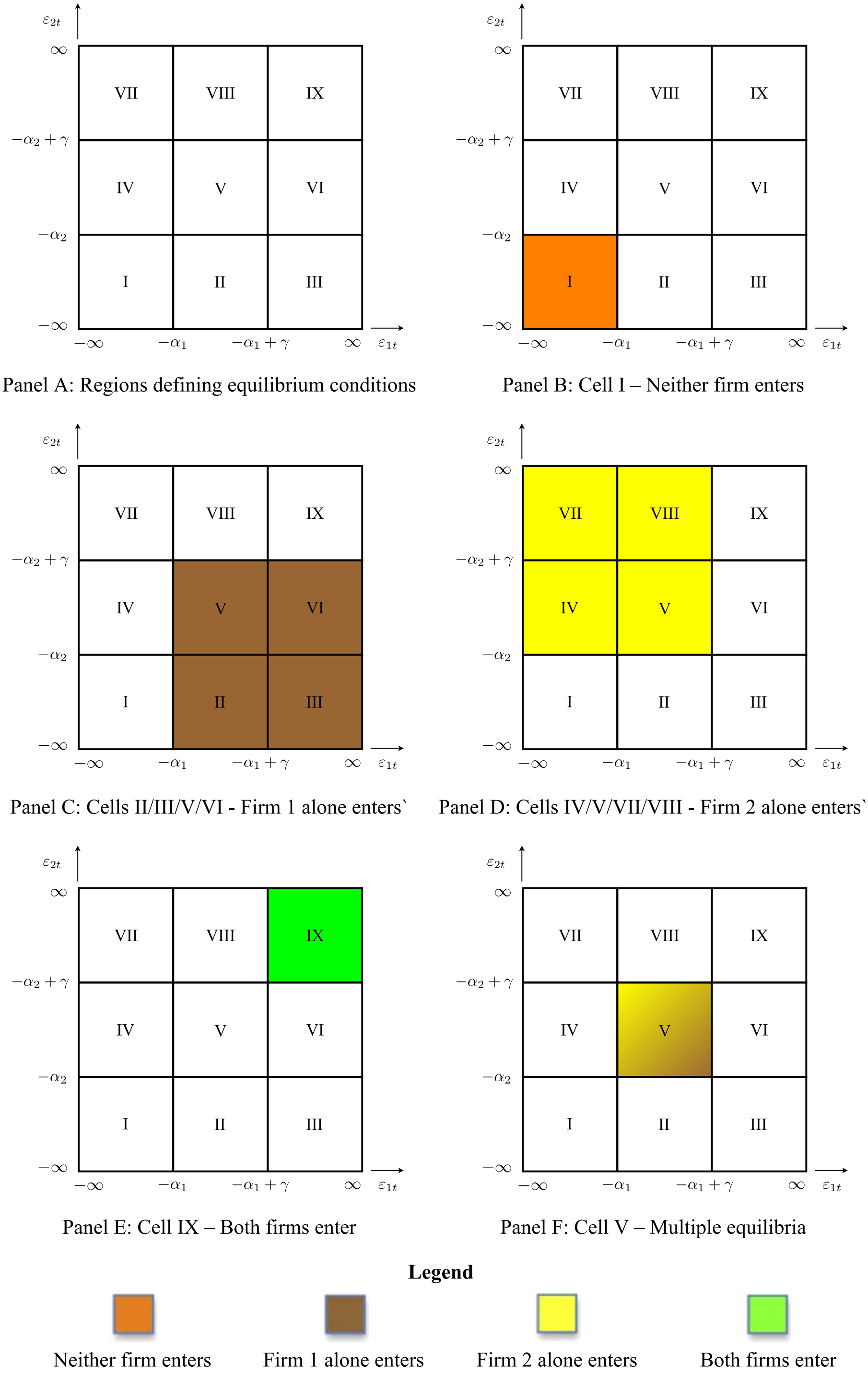


Fig. 1 Equilibrium conditions—discrete game of market entry

IX. These cells are defined by the equilibrium conditions for the different outcomes. Consider panel B of the figure. Cell I, which corresponds to neither firm entering (i.e. satisfying the equilibrium condition in Eq. 4) is shaded. Similarly panels C, D and E are graphical representations of the equilibrium conditions for firm 1 entering alone, firm 2 entering alone and both firms entering respectively. Note from panels C and D that they share a cell—cell V, i.e. both equilibria are satisfied in this cell. This is the region of multiple equilibria, highlighted in panel F of the figure. This cell is consistent with two equilibria—either firm 1 alone enters, or firm 2 alone enters.

Discrete games can also be set up to model social interactions, where players benefit from joint consumption of a product. A slightly modified version of the model used for analyzing market entry could be used for this context, although the equilibrium conditions, and the nature of multiplicity of equilibria in this case are distinctly different. Let the payoffs of the two players be given by

$$\pi_{it} = \alpha_i + X_{it}\beta + \gamma \cdot 1\left(\sum_{j\neq i} y_{jt} > 0\right) + \gamma\delta \cdot 1\left(\sum_{j\neq i} y_{jt} > 1\right) + \gamma\delta^2 \cdot 1\left(\sum_{j\neq i} y_{jt} > 2\right) + \ldots + \varepsilon_{it} \tag{5}$$

Note that the main difference between the payoffs in this social interaction game and those in the market entry game in Eq. 1 is the sign of the interaction terms. In this case, consumption by another player increases the payoffs from consumption. Thus, for $\gamma > 0$ and $\delta > 0$, we are guaranteed the existence of an equilibrium in pure strategies for any realization of the errors ε_{it} and any values of the covariates and the parameters.

The equilibrium conditions for a two-player game with social interactions can be derived in a similar fashion as for the market entry game. Assuming there are no other covariates in the model, the payoffs are given by

$$\pi_{it} = \alpha_i + \gamma \cdot y_{-i,t} + \varepsilon_{it} \tag{6}$$

where $y_{-i,t}$ is the consumption decision of the other player.

Figure 2 depicts the equilibrium conditions for the various outcomes. Panel B of this figure shows us the conditions for the equilibrium where neither player consumes—this corresponds to cells I, II, IV and V. Panels C and D respectively show the conditions for player 1 consuming alone (Cell III) and player 2 consuming alone (Cell VII) respectively. Panel E shows the conditions for joint consumption by both players (Cells V, VI, VIII and IX). As can be seen in Panel F, multiplicity of equilibria arises from the fact that Cell V is consistent with two equilibria—neither player consuming and both of them consuming. Thus, the change in sign of the interaction effect in the payoff function has a non-trivial impact on the nature of the multiplicity of equilibria. In particular, the overlap now is between the equilibrium where neither player consumes and the one where both players consume.

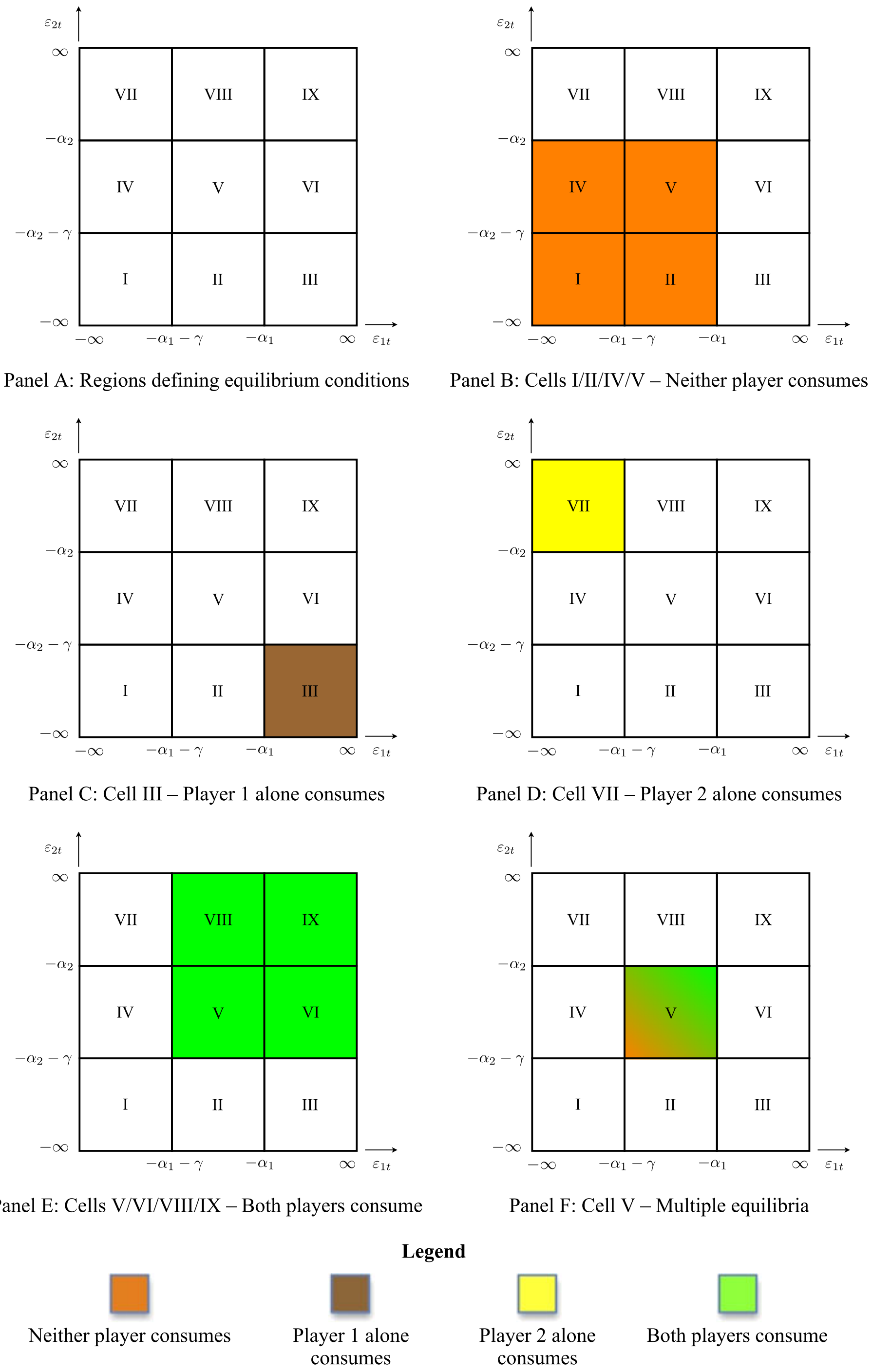


Fig. 2 Equilibrium conditions—discrete game of social interactions

With greater than two players in the game, the issue of multiplicity remains and the number of cells with multiple equilibria increases. While it is somewhat tedious to work out the equilibrium conditions for a large number of players by hand, it is straightforward to enumerate the equilibria using a computer. This is because irrespective of the values of the parameters, the cells that correspond to more than one equilibrium are the same for a given number of players and a given type of game (i.e. market entry or social interactions). The addition of covariates to the model also does not fundamentally change the equilibrium conditions and the cells corresponding to the different equilibria. The covariates merely shift the boundaries of the cells corresponding to the different equilibria.

Covariates can help in identification of equilibrium selection rules in the presence of exclusion restrictions. Specifically, when there are covariates that enter the payoff function of one player but not that of the other players, a shift in the covariates would alter the outcomes differently for different equilibrium selection rules. This can be utilized by the empirical researcher to identify the equilibrium selection rule simultaneously with the parameters of the payoff functions. Note that in the absence of exclusion restrictions, a shift in the outcomes with a shift in the covariates can be rationalized by different parameters for each equilibrium selection rule, and hence the parameters and equilibrium selection are not simultaneously identified. We refer the reader to Bajari et al. (2010) for a detailed discussion of this issue.

3 Estimating discrete games of complete information

3.1 Overview

Our proposed Bayesian approach to estimating discrete games of complete information starts by making distributional assumptions on the errors and specifying prior distributions for the parameters. Due to multiple equilibria, the likelihood is not well defined and hence a standard Markov Chain procedure such as a Metropolis Hastings algorithm is not feasible. However, conditional on a selected equilibrium, the likelihood is well-defined. The main idea behind our proposed methodology is to treat each equilibrium selection rule as a model, and to set up an estimation procedure that augments the parameter state space with a model space. Thus, the estimation routine navigates both parameter and model (equivalently equilibrium selection rule) spaces, and generates joint posterior draws of parameters and model indicators. When the equilibrium is identified, for instance through the presence of excluded variables, the estimation routine generates posterior estimates of model parameters as well as model indicators. When the equilibrium is not identified, the posterior estimates of the model indicators would be the same as the prior indicators (up to a simulation error), but the procedure is able to uncover the posterior distribution of parameters that spans the various models and thus reflects the analyst's (prior) uncertainty about the equilibria.

Our procedure is closest to that of Bajari et al. (2010) but with some significant differences. Like them, we take an empirical approach to the problem, considering a finite subset of possible equilibrium selection rules, and letting the data tell us about the likelihood that any of these rules are actually used in the data generating process. The identification of equilibrium selection rules in this case is made possible by the presence of suitable exclusion restrictions. In this sense, our method can be considered a Bayesian analog of their approach, where we specify a prior probability distribution over the set of equilibrium selection rules, and obtain posterior estimates of the probability distribution. Where our approach differs from their approach is that we allow for the situation where there isn't enough information in the data to identify the equilibrium selection rule, i.e. where we do not have suitable exclusion restrictions. In this case, our approach allows us to obtain posterior estimates of the other parameters that reflect the analyst's prior uncertainty over the set of equilibrium selection rules. Furthermore, our hierarchical Bayes approach allows us to specify heterogeneity in equilibrium selection, with different sets of players choosing different equilibria. The methodologies that assume a single equilibrium selection rule (such as in Berry 1992 and Hartmann 2010, for instance) are also nested in our approach, since they are akin to placing a prior with a probability of one on a single equilibrium selection rule. Our approach generalizes this to a case where a more general prior is placed over a set of equilibrium selection rules, potentially allowing for heterogeneity in these rules, and identifying the equilibrium selection rules when the data allow for that.

We first discuss a Bayesian approach to estimating a discrete game, with the use of an ad-hoc equilibrium selection rule as has been often used in the literature. We then build on this to develop a methodology for dealing with multiplicity of equilibria.

3.1.1 Bayesian estimation with an ad-hoc equilibrium selection rule

To illustrate the proposed methodology, we first start with the case of an entry game with two players, and subsequently extend it to cases with a greater number of players and to games of social interaction. Let the payoff functions of the two players be given by

$$\pi_{it} = \alpha_i + X_{it}\beta - \gamma \cdot y_{-i,t} + \varepsilon_{it} \tag{7}$$

where the notation is the same as before. π_{it} represents firm i's profits in market t. α_i represents a firm specific intercept. The vector X_{it} represents a set of exogenous covariates, which potentially includes a set of excluded variables. If it is an excluded variable, it takes the value 0 for all observations for the player whose payoff it is excluded from. The coefficient for the covariates is β, which includes the coefficients for the excluded variables. The interaction between firms is captured by the coefficient γ. Finally, ε_{it} is assumed to be an error that is uncorrelated across markets.

The conditions for entry are therefore

$$y_{it} = \begin{cases} 1 & \text{if} \quad \pi_{it} > 0 \\ 0 & \text{if} \quad \pi_{it} < 0 \end{cases} \tag{8}$$

Conditional on an equilibrium being selected, the model is a bivariate probit model. The equilibrium conditions are similar to those described earlier, except that covariates are included and the equilibrium selection rule is imposed. Note from Fig. 1 that there are two possible equilibria for a region of the error space (shown as Cell V in the figure)—one where firm 1 enters the market, and another where firm 2 enters the market. Take the case of the first of these two equilibria being selected, i.e. firm 1 enters but firm 2 does not when errors are in Cell V. This could result, for instance, from the true data generating process being one with sequential entry, with firm 1 making entry decisions before firm 2. In such a case, the regions of the error space corresponding to the four possible outcomes are below

1. $(y_{1t}, y_{2t}) = (0, 0) \Rightarrow (\varepsilon_{1t}, \varepsilon_{2t}) \in$ Cell I
2. $(y_{1t}, y_{2t}) = (0, 1) \Rightarrow (\varepsilon_{1t}, \varepsilon_{2t}) \in$ Cells IV/VII/VIII
3. $(y_{1t}, y_{2t}) = (1, 0) \Rightarrow (\varepsilon_{1t}, \varepsilon_{2t}) \in$ Cells II/III/V/VI
4. $(y_{1t}, y_{2t}) = (1, 1) \Rightarrow (\varepsilon_{1t}, \varepsilon_{2t}) \in$ Cell IX

Note here that the boundaries of the cells would be modified from that in Fig. 1 due to the presence of covariates in the model. Thus α_i would be substituted by $\alpha_i + X_{it}\beta$ in all the cell boundaries on both axes. For a given guess of parameters, the likelihood for a given observation is the sum of the probabilities of the errors being in the cells corresponding to that outcome. With a suitable distributional assumption on the errors, the likelihood can be evaluated. We assume that the unobservables are normally distributed with mean 0 and unit variance, i.e.

$$\varepsilon_{it} \sim N(0, 1) \tag{9}$$

In order to specify the likelihood for a particular observation, we need to find the probability of the unobservables being in one of the cells corresponding to the outcome for that observation. This probability is the sum of the probabilities for the unobservables to lie in each of the cells for that outcome. The probability for the unobservables to be in any given cell j for the tth observation is given by

$$P_{jt} = \int_{A_{jt}} f(\varepsilon_t)\, d\varepsilon_t \tag{10}$$

where ε_t is the vector of unobservables, $f(\varepsilon_t)$ is the distribution of ε_t and A_{jt} denotes the region of the error space for the jth cell (i.e. the boundaries of the cell, which vary with observation because of variation in the covariates X_{it}).

Since we have assumed ε_t to be distributed normally, this reduces to the following expression in our two-player example

$$\begin{aligned} P_{jt} &= \int_{l_{j1t}}^{u_{j1t}} \int_{l_{j2t}}^{u_{j2t}} \phi\left(\varepsilon_{1t}\right) \phi\left(\varepsilon_{2t}\right) d\varepsilon_{1t} d\varepsilon_{2t} \\ &= \left[\Phi\left(u_{j1t}\right) - \Phi\left(l_{j1t}\right)\right]\left[\Phi\left(u_{j2t}\right) - \Phi\left(l_{j2t}\right)\right] \end{aligned} \tag{11}$$

where l_{jit} and u_{jit} represent the lower and upper boundaries of the jth cell for ε_{it}, $\phi(\cdot)$ represents the normal density function and $\Phi(\cdot)$ represents the normal distribution function. The boundaries of the cells have been defined earlier. For instance, for cell I, $l_{I1t} = l_{I2t} = -\infty$, $u_{I1t} = -\alpha_1 - X_{1t}\beta$ and $u_{I2t} = -\alpha_2 - X_{2t}\beta$. The boundaries of the other cells are similarly obtained.

The likelihood of the tth observation is then

$$\begin{aligned} \mathcal{L}_t\left(\theta; X_t, y_t\right) = 1\left(y_{1t}=0, y_{2t}=0\right) & \\ \times \left(P_{I,t}\right) + 1\left(y_{1t}=0, y_{2t}=1\right)&\left(P_{IV,t} + P_{VII,t} + P_{VIII,t}\right) \quad (12) \\ +1\left(y_{1t}=1, y_{2t}=0\right)&\left(P_{II,t} + P_{III,t} + P_{V,t} + P_{VI,t}\right) \\ +1\left(y_{1t}=1, y_{2t}=1\right)&\left(P_{IX,t}\right) \quad (13) \end{aligned}$$

where $1(\cdot)$ is an indicator function, taking the value 1 if the conditions in it are true and 0 otherwise, y_t is the vector of decisions, X_t is the matrix formed by stacking all covariate vectors (X_{it}), and θ is the vector of all parameters.

The likelihood of the data is then

$$\mathcal{L}\left(\theta; X, y\right) = \prod_t \mathcal{L}_t\left(\theta; X_t, y_t\right) \tag{14}$$

Here, y is the stacked vector of outcomes, and X is the stacked matrix of covariates across all observations.

To complete the model, we need to specify prior distributions for the parameters—α_1, α_2, β and γ. Since γ is constrained to be positive, we reparametrize it as $\exp(\tilde{\gamma})$. Thus,

$$\theta = \begin{pmatrix} \alpha_1 & \alpha_2 & \beta & \tilde{\gamma} \end{pmatrix} \tag{15}$$

Let the prior distribution for θ be

$$\theta \sim N\left(\mu, \Sigma\right) \tag{16}$$

The posterior distribution of parameters is then given by

$$f\left(\theta | X, y, \mu, \Sigma\right) \propto \mathcal{L}\left(\theta; X, y\right) f\left(\theta | \mu, \Sigma\right) \tag{17}$$

The parameters can be estimated in an MCMC procedure that utilizes the Metropolis-Hastings algorithm (Chib and Greenberg 1995). This involves generating a sequence of draws from a candidate density and then modifying it through a rejection step to achieve the detailed balance condition. This sequence of draws converges to the posterior distribution of parameters, allowing us to do inference or conduct counterfactual experiments.

We have discussed the estimation of a market entry discrete game with two players, with the ad-hoc equilibrium selection rule of firm 1 entering in the case of multiple equilibria (where the two equilibria are of either firm 1 entering the market alone, or firm 2 entering it alone). But the methodology can be easily extended to more players. For instance, in a three player entry game, there are a total of six cells that are consistent with two equilibria each, and two cells that are consistent with three equilibria each. Conditional on X and ε (i.e. within a cell), there are up to three equilibria. However, there are a total 576 ($= 2^6 \cdot 3^2$) possible combinations of equilibria if we did not condition on X and ε. Conditional on assuming an equilibrium selection rule, which assigns cells with multiple consistent outcomes to one of them, there is a unique assignment of each cell to an outcome, and hence the likelihood for such a game is well-specified. Given the likelihood, a Metropolis Hastings algorithm for simulation from the posterior distributions of parameters is relatively straightforward to implement.

The methodology can also be extended to other equilibrium selection mechanisms, that do not assign entire cells to a particular outcome. For instance, the assumption in Berry (1992) is that in cases of multiple equilibria, the firms enter in the order of profitability. This can be justified on the basis of an underlying assumption of sequential entry, where the firm with the greatest expected profits in a market enters first, followed by the next most profitable firm and so on. Thus, the equilibrium selection mechanism selects the equilibrium that maximizes industry profits market by market.

Implementing the equilibrium selection rule described above requires a modification of the procedure we have described earlier where the equilibrium selection rule allocated entire cells in the region of multiplicity to particular outcomes. If we assumed that the firms entered in decreasing order of profitability, cells corresponding to the region of uncertainty would need to be partly allocated to outcomes. To see this, let us first discuss this in the context of the two-player market entry game described earlier and depicted in Fig. 1. The region for multiple equilibria is in Cell V. The two outcomes that this cell is consistent with involve either of the firms entering the market alone. Consider first the outcome of firm 1 entering, and consider for notational simplicity the case of no covariates. The equilibrium selection rule is that the more profitable firm enters the market in the case of multiple equilibria. Hence, it must be the case that the profits of firm 1 entering alone must be greater than the profits of firm 2 entering, given a realization of the errors. Thus,

$$\pi_1 (y_1 = 1, y_2 = 0) > \pi_2 (y_1 = 0, y_2 = 1) \tag{18}$$

This implies that

$$\alpha_1 + \varepsilon_1 > \alpha_2 + \varepsilon_2 \tag{19}$$

Thus, the condition for firm 1 being the more profitable firm and hence the outcome of $(y_1 = 1, y_2 = 0)$ being picked over $(y_1 = 0, y_2 = 1)$ is

$$\varepsilon_2 < -\alpha_2 + \alpha_1 + \varepsilon_1 \tag{20}$$

This condition describes a straight line with intercept $(-\alpha_2 + \alpha_1)$ and a slope equal to 1, i.e it is a 45 degree line intersecting the ε_2-axis at $(-\alpha_2 + \alpha_1)$. It is easy to see that it this line bisects the cell in two, running from the lower left corner to the upper right corner. In region under the line in Cell V, firm 1 enters the market and in the region above the line, firm 2 enters. This is shown graphically in Fig. 3.

To estimate the model with this equilibrium selection rule, we need to ensure that the likelihoods corresponding to each outcome are correctly evaluated, with the probability of the outcome of firm 1 or firm 2 entering the market alone including the appropriate part of Cell V.

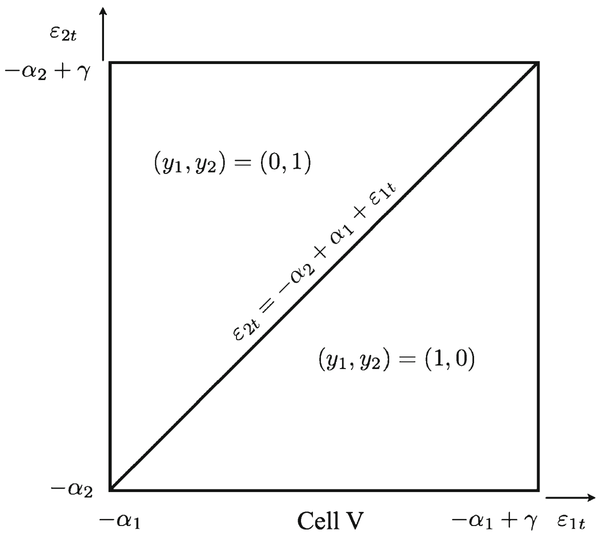
This procedure can easily be extended to games involving more than two players. It is easy to verify for a three-player game of market entry, for instance, that with an equilibrium selection rule that has the firms entering a market in decreasing order of profitability, each cell with multiple equilibria would similarly be equally divided between the different outcomes that are consistent with the cell. As in the two-player example illustrated above, the likelihood for each observation could be constructed through assignment of the cells partially to multiple outcomes, instead of being assigned exclusively to one outcome.

To summarize, for any game with multiple equilibria,the likelihood is well-specified and a Metropolis Hastings procedure can be set up to sample from the posterior distribution of parameters as long as an equilibrium selection rule is assumed, i.e. all the cells (or parts thereof) are uniquely assigned to specific outcomes.

3.1.2 Bayesian estimation with multiple equilibrium selection rules

In the previous sub-section, we have discussed a Bayesian estimators of the parameters of a discrete game with an ad-hoc equilibrium selection assumption.

Fig. 3 Equilibrium conditions—sequential entry in order of profitability



We now extend it to the case where the analyst is unwilling to make an assumption of a particular equilibrium selection mechanism, but instead has some prior beliefs over a finite set of potential equilibrium selection mechanisms. For instance, in the case of the two player game, where there is one region (Cell V in Fig. 1) of multiplicity, the analyst may believe that there are three potential equilibrium selection rules—one that allocates the entire cell to the outcome of firm 1 entering, another that allocates it entirely to firm 2 entering, and a third that users the equilibrium selection mechanism in Fig. 3 and consequently splits the cell between the two outcomes. The analyst believes that all observations in the data generating process have outcomes generated using one of these equilibrium selection rules. Consider each of these equilibrium selection rules to be a model. Thus, there are multiple models—one corresponding to each equilibrium selection rule, each of which can be estimated using a Metropolis Hastings algorithm. However, such a procedure may be cumbersome as the number of equilibrium selection rules increase. Further, when the equilibrium selection rule is identified, through the presence of suitable exclusion restrictions, it is harder to conduct inference about the posteriors on the models themselves. Finally, this approach would not allow us to incorporate heterogeneity in equilibrium selection. We propose a simulator using the reversible jump algorithm, which allows us to simulate posterior distributions of parameters across multiple models (equilibrium selection rules) and directly provides us posterior estimates on the equilibrium selection rules themselves. Further, it lends itself to allowing for heterogeneity in equilibrium selection.

We first introduce the basics of the reversible jump algorithm, since it is new to the Marketing and Economics literatures, and then discuss its application to our problem of estimation of discrete games. A novel aspect of our use of the reversible jump sampler is that we allow for heterogeneity in models across units, which to our knowledge is new to the reversible jump literature.

Reversible jump algorithm The reversible jump algorithm (Green 1995; Green and Hastie 2009) is a Markov Chain Monte Carlo algorithm that generates draws from a stationary distribution for an across-model state space which is potentially trans-dimensional. For instance, in modeling count data, there could be two potential models—a Poisson model, which is a single-parameter model and a Negative Binomial model, which has two parameters. The reversible jump algorithm is an extension of the Metropolis Hastings algorithm, which simultaneously draws a model indicator as well as parameters. Like the Metropolis Hastings algorithm, it does this by generating draws from a candidate density, which is modified through a rejection step to satisfy the detailed balance condition.

Formally, let there be a countable set $\mathcal{K}$ of models, with the k being a model indicator and θ_k being the parameter vector of dimension n_k corresponding to the kth model. Let the data be represented by D. Let $\mathcal{L}\,(D|k, \theta_k)$ be the likelihood of the data corresponding to the kth model and $p\,(\theta_k|k)$ be the density of the prior for the parameters of this model. Also, let there be a prior across models, with its density specified as $p(k)$. Then the joint posterior of the model and parameters is given by

$$\pi(k, \theta_k | D) \propto \mathcal{L}(D | k, \theta_k)\, p(\theta_k | k)\, p(k) \tag{21}$$

Let t denote a move type, involving a forward move from (k, θ_k) to $\left(k' \theta'_{k'}\right)$ and with a corresponding reverse move t' and let there be a countable set $\mathcal{T}$ of such moves. Let $\theta'_{k'}$ have dimension $n_{k'}$. Let $m_t(k, \theta_k)$ indicate the probability of this move, which could be dependent on both the current model and parameter values of the current state. The reverse probability would thus be $m_{t'}\left(k', \theta'_{k'}\right)$. For the forward move, let us draw a vector of random numbers u with dimension r_t from a known distribution $f_t(u)$, and let the vector u' be drawn for the reverse move. Let the dimensions of u' be r'_t, which satisfies the condition $n_k + r_t = n_{k'} + r'_t$and the distribution it is drawn from be $f'_t\left(u'\right)$. Let the function that defines the transformation from (θ_k, u) to $\left(\theta'_{k'}, u'\right)$ be denoted by $h_t(\theta_k, u)$ and the inverse function by $h'_t\left(\theta'_k, u'\right)$, such that

$$\begin{aligned} \left(\theta'_{k'}, u'\right) &= h_t(\theta_k, u) \\ (\theta_k, u) &= h'_t\left(\theta'_{k'}, u'\right) \end{aligned} \tag{22}$$

Finally, let the probability that the move is accepted be denoted by $\alpha_t((k, \theta_k),$ $(k', \theta'_k))$ and the reverse probability be $\alpha'_t\left(\left(k', \theta'_k\right), (k, \theta_k)\right)$. To generate draws from the target density, in this case the joint posterior distribution of parameters and model indicators, it is sufficient to construct a Markov Chain whose transition kernel satisfies the Detailed Balance condition. It is straightforward to show that this is the case when

$$\begin{aligned} &\int\int_{(k,\theta_k)\in A, \left(k',\theta'_{k'}\right)\in B} \pi(k, \theta_k | D)\, m_t(k, \theta_k)\, f_t(u)\, \alpha_t\left((k, \theta_k), \left(k', \theta'_{k'}\right)\right) d(k, \theta_k)\, du \\ &\quad = \int\int_{\left(k',\theta'_{k'}\right)\in B, (k,\theta_k)\in A} \pi\left(k', \theta'_{k'} | D\right) m'_t\left(k', \theta'_{k'}\right) f'_t(u) \alpha'_t\left(\left(k', \theta'_{k'}\right), (k, \theta_k)\right) d\left(k', \theta'_{k'}\right) du' \end{aligned} \tag{23}$$

This condition says that it is equally likely for the chain to move from set A to B as it is to move from B to A. This will be satisfied if the move of type t is accepted with probability

$$\alpha_t\left((k, \theta_k), \left(k', \theta'_{k'}\right)\right) = \min\left[1, \frac{\pi\left(k', \theta'_{k'} | D\right) m'_t\left(k', \theta'_{k'}\right) f'_t\left(u'\right)}{\pi(k, \theta_k | D)\, m_t(k, \theta_k)\, f_t(u)} \left| \frac{\partial h_t(\theta_k, u)}{\partial(\theta_k, u)} \right| \right] \tag{24}$$

The last term, is the Jacobian for the transformation from (θ_k, u) to $\left(\theta'_{k'}, u'\right)$.

To summarize the algorithm, we propose a candidate move from (k, θ_k) to $\left(k', \theta'_k\right)$ with probability $m_t(k, \theta_k)$ by first generating draws for u from the density $f_t(u)$ and then transforming (θ_k, u) to $\left(\theta'_{k'}, u'\right)$ using the function $h_t(\cdot)$. We then accept this move with probability $\alpha_t\left((k, \theta_k), \left(k', \theta'_{k'}\right)\right)$. Repeated draws from this chain converge to the posterior distribution $\pi(k, \theta_k | D)$.

Reversible jump algorithm applied to discrete games In the case of discrete games, the multiple models correspond to the multiple equilibrium selection rules. Henceforth, we will use the term model and equilibrium selection rule interchangeably. Thus, in a two-player entry game for instance, we may have two models, with a region of the error space (Cell V in Fig. 1) being consistent with one of two potential outcomes in the two respective rules. A given set of parameters that rationalizes observed outcomes for one model would not rationalize the observed outcomes for the second model. Thus, while the dimensions of the parameters do not differ for the different models, the parameter spaces are different. As we have discussed earlier, the reversible jump sampler is an MCMC sampler that traverses a broader parameter space that is the union of parameter spaces across multiple models. We thus employ this sampler for our context even though ours is not a trans-dimensional problem for which reversible jump samplers have been typically employed.

Consider a move type t, which involves a candidate move from model k to model k' as the forward move and with a corresponding reverse move t' from k' to k. With the assignment of priors to the parameters conditional on model $p\left(\theta_k|k\right)$ and model indicator $p\left(k\right)$, the posterior is given by Eq. 21. We assume a multinomial prior on the model indicator. Thus,

$$k \sim \text{Multinomial}\left(p\right) \tag{25}$$

where p is a vector of prior probabilities associated with each of the $\mathcal{K}$ models.

Let the probability of move t be independent of the current state of the parameter θ_k. i.e. $m_t\left(k, \theta_k\right) = m_t\left(k\right)$. Since the move type indicates the models corresponding to the current state and the candidate move, this can be more simply denoted by $m_{k,k'}$. The probability of the reverse move is similarly $m_{k',k}$. We make the further assumption that

$$m_{k,k'} = p_{k'} \tag{26}$$

where $p_{k'} = p\left(k'\right)$ is the element of the prior p corresponding to model k', In other words, we generate a candidate move to model k' with its prior probability and independent of the current model. This is not necessary for our algorithm (for instance, one could use a uniform probability of moves across models) but as we will see, this is a natural choice for the jump probability and also simplifies the expression for the acceptance ratio.

Let the random vector u have the same dimension as the parameter vector, and let this be a draw from a mean-zero normal distribution with a variance-covariance matrix Σ_u. We define the variable transformation function $h_t\left(\theta_k, u\right)$ as follows

$$\begin{aligned} \theta'_{k'} &= g_t\left(\theta_k\right) + u \\ u' &= u \end{aligned} \tag{27}$$

The function $g_t\left(\theta_k\right)$ transforms the parameters from the parameter space of model k to that of model k' by matching a set of moments. The set of moments to be matched are specific to the empirical application, but could include the predicted shares of the full set of outcomes at various values of covariates. In our empirical applications, we match these moments for the mean levels of covariates and for values that are one standard deviation apart on either side of the mean. The number of moments to be matched needs to be at least as large as the dimension of θ_k. We conduct this moment matching through an optimization step, which minimizes the mean squared difference between the sample analogs of these moments for the two models.

The reverse function $h'_t\left(\theta'_{k'}, u'\right)$ is the inverse of this transformation and is given by

$$\begin{aligned} \theta_k &= g'_t\left(\theta'_k - u'\right) \\ u &= u' \end{aligned} \tag{28}$$

where $g'_t\left(\theta'_{k'}\right)$ is the inverse of $g_t\left(\theta_k\right)$.

The Jacobian is thus given by

$$\left|\frac{\partial h_t\left(\theta_k, u\right)}{\partial\left(\theta_k, u\right)}\right| = \left|\begin{array}{cc} \dfrac{\partial\left[g_t\left(\theta_k\right)+u\right]}{\partial\theta_k} & \dfrac{\partial\left[g_t\left(\theta_k\right)+u\right]}{\partial u} \\ \dfrac{\partial u}{\partial\theta_k} & \dfrac{\partial u}{\partial u} \end{array}\right| = \left|\begin{array}{cc} \dfrac{\partial g_t\left(\theta_k\right)}{\partial\theta_k} & I_{n_k} \\ 0 & I_{n_k} \end{array}\right| = \left|\frac{\partial g_t\left(\theta_k\right)}{\partial\theta_k}\right| \tag{29}$$

This Jacobian does not have an analytical solution, and is hence evaluated numerically in our empirical applications.[2]

The candidate move described above is like that in a random walk Metropolis Hastings sampler, except that the candidate density is centered at a transformed value of the current state instead of the current state itself. When the candidate move is to the same model as in the current state, the algorithm reduces to a standard random walk Metropolis Hastings algorithm. To see this, note that for $k' = k$, the function $g_t\left(\theta_k\right)$ that equalizes the moments for the current and candidate models is simply

$$g_t\left(\theta_k\right) = \theta_k$$

The Jacobian in this case equals 1. The probabilities $m_t\left(k, \theta_k\right)$ and $m'_t\left(k', \theta'_{k'}\right)$ are equal, as are the densities $f_t(u)$ and $f'_t\left(u'\right)$ and hence the acceptance probability in Eq. 24 reduces to the standard expression for a random walk Metropolis Hastings algorithm.

[2]A necessary condition for the methodology to work is that the transformation function $g_t(\theta_k)$ is differentiable, allowing for the evaluation of the Jacobian. This is achieved by the absence of any discontinuities in this function. An informal way for the analyst to verify that this is the case is to evaluate the Jacobian for the reverse transformation and check if this is the inverse of the Jacobian for the forward transformation. Discontinuities in the transformation function would lead to the reverse Jacobian deviating from the inverse of the forward Jacobian. In our simulations as well as empirical applications, we did not encounter any instances where this happened.

Returning to the general case of a candidate move from model k to model k', the acceptance probability in Eq. 24 can be written as

$$\alpha_t\big((k,\theta_k),\big(k',\theta_k'\big)\big) = \min\left[1, \frac{\mathcal{L}\big(D|k',\theta_{k'}'\big)\, p\left(\theta_{k'}'|k'\right) p\big(k'\big)\, p(k)\, f_t'\left(u'\right)}{\mathcal{L}(D|k,\theta_k)\, p(\theta_k|k)\, p\,(k)\, p(k')\, f_t\,(u)} \left|\frac{\partial g_t\,(\theta_k)}{\partial\,(\theta_k)}\right|\right] \tag{30}$$

where we have written out the expressions for $\pi\,(k,\theta_k|D)$ and $\pi\,\left(k',\theta_{k'}'|D\right)$, substituted in the values of $m_t\,(k,\theta_k)$ and $m_t\,\left(k',\theta_{k'}'\right)$ using Eq. 26 and that for the Jacobian from Eq. 29. Noting that $f_t\,(u) = f_t'\,\left(u'\right)$ and canceling terms, we obtain the acceptance probability for across-model moves as

$$\alpha_t\left((k,\theta_k),\left(k',\theta_{k'}'\right)\right) = \min\left[1, \frac{\mathcal{L}\left(D|k',\theta_{k'}'\right) p\left(\theta_{k'}'|k'\right)}{\mathcal{L}\,(D|k,\theta_k)\; p\,(\theta_k|k)}\left|\frac{\partial g_t\,(\theta_k)}{\partial\,(\theta_k)}\right|\right] \tag{31}$$

For within model moves, this acceptance probability is the standard random walk Metropolis Hastings ratio

$$\alpha_t\left((k,\theta_k),\left(k,\theta_k'\right)\right) = \min\left[1, \frac{\mathcal{L}\left(D|k,\theta_k'\right) p\left(\theta_k'|k\right)}{\mathcal{L}\,(D|k,\theta_k)\; p\,(\theta_k|k)}\right] \tag{32}$$

Every move need not be an across-model move. In general, it is efficient to have several sweeps of draws, with an across-model move being potentially attempted at the beginning of every sweep, with several within-model moves within each sweep. A summary of the algorithm and its operational aspects, such as tuning and assessment of convergence are in the Appendix.

Identification of equilibrium selection mechanisms and the reversible jump algorithm So far, we have laid out the method by which we can estimate discrete games with multiple equilibrium selection rules using a reversible jump algorithm. However, the methodology does not assume anything about the identification of equilibrium selection rules. When the equilibrium selection rules (models in our terminology) are not identified, the posterior densities conditional on equilibrium selection would be the same at the mode for each models. On average, the algorithm would draw each of the models with its prior probability. Thus, when the model is unidentified, the posterior probabilities for the model indicators would essentially be the same as the prior probabilities. The posterior estimates of the parameters conditional on the model would not be affected by the fact that the posterior probabilities are unidentified. However, the marginal posterior distributions of the parameters would reflect the prior probabilities of the models.[3] When the model is

[3]The marginal posteriors are analogous to the set identified estimates in the moment inequality approaches, and gives the analyst an idea about how uncertainty about equilibrium selection manifests itself in uncertainty for parameters. However, the marginal posteriors cannot be directly used for predictive purposes, for which the analyst would need to use the posterior conditional on model.

identified, the posterior densities of the parameters would be different at the modes of the different models. The models with higher modes would on average be more likely be drawn, since the acceptance probabilities of these models would be higher. Thus, the probabilities of drawing each model would depend on their relative levels of the posterior densities, and would differ from the prior probabilities. The stronger the identification of the model, the weaker would be the influence of the prior on the posterior probabilities. Thus, this approach nests both cases—when the equilibrium is unidentified and when it is identified. When there are no excluded variables available, the empirical researcher still benefits by obtaining parameters estimates that reflect her own uncertainty about the true equilibrium. When excluded variables are available, the researcher can let the data inform her about the true equilibrium.

We have discussed identification of equilibrium selection rules earlier and noted that we can only have a finite set of equilibrium selection rules, with the set being smaller than the number of moments available for identification. For instance, a three-player game of market entry has 576 ways in which cells with multiple equilibria can be allocated entirely to specific outcomes and it is unlikely that we would have the number of moments required to identify these 576 equilibrium selection rules. Instead, we may be interested in identifying the probabilities of a smaller set of equilibria being played. For instance, we may consider two sets of equilibrium selection rules—one that assumes that firms enter in decreasing order of profitability (Berry 1992) and another set consisting of all equilibrium selection rules that allocate cells entirely to specific outcomes. In a social entry context, we might be interested in an rule that selects the Pareto dominant equilibrium (Hartmann 2010), with a set of other equilibrium selection rules grouped together. In each of these two instances, we could classify the two sets of equilibrium selection rues as two models, with the specific equilibrium selection rule within each set randomized at each draw.[4]

Accommodating heterogeneity in equilibrium selection So far, our assumption about the data generating process has been that there is an equilibrium selection rule that applies to all observations. However, in several contexts, one might be interested in modeling heterogeneity in equilibrium selection, where different units (e.g. sets of individuals, firms etc. for which we have repeated observations) use different equilibrium selection rules. Consider, for

[4]Such a randomization has also been used in Soetevent and Kooreman (2007), but in their case across all equilibria rather than a subset of them. Randomization in this case is akin to placing a uniform prior over the equilibrium selection rule within each set. While it is not necessary for the purpose of inference, particularly of the model indicator itself, to keep track of each specific equilibrium selection rule, it is necessary to retain the joint draws of specific equilibrium selection rule and parameters for the sake of counterfactual simulations since parameters are consistent with a specific equilibrium selection rule.

instance, the context of joint consumption of a good by friends (Hartmann 2010), modeled as the outcome of a discrete game of complete information. In such games, one of the equilibria is always Pareto dominant. Instead of making the typical ad-hoc assumption that this equilibrium is selected, one could estimate the probability of its selection when one has suitable excluded variables. However, a valid question of interest might be that different sets of friends do not have the same equilibrium selection rule. Since inference and evaluation of counterfactual outcomes depend on equilibrium selection, it is likely important in such situations to model heterogeneity in equilibrium selection.

Accommodating heterogeneity in equilibrium selection is feasible using the reversible jump algorithm we have laid out earlier in this section. Consider N units (e.g. sets of friends in the joint consumption example) for each of which we observe data $\mathcal{D}_i$, where i indexes the unit. Let k_i be an indicator of the model for the ith unit and θ_{ik_i} represent the parameter vector for the ith unit and the ith model. Then the unit level equivalent of the posterior density in Eq. 21 is

$$\pi\left(k, \theta_k | D\right) \propto \mathcal{L}\left(D_i | k_i, \theta_{ik_i}\right) p\left(\theta_{ik_i} | k_i\right) p\left(k_i\right) \tag{33}$$

It might be infeasible to specify priors for the parameter vector for each unit separately and hence it might be useful to cast this model in a hierarchical framework, for instance by assuming that

$$\theta_{i,k} \sim \mathrm{N}\left(q_i \lambda_k, V_{\theta,k}\right)$$

where q_i is a vector of observed covariates, λ_k is a population-level parameter vector specific to the model and $V_{\theta,k}$ is a population-level variance-covariance matrix. The model would be completed by specifying priors on these population-level parameters. With an assumption of conditional independence of the unit-level parameters θ_{ik_i}, we could set up an overall Gibbs sampler, with the population-level parameters being drawn as straightforward Gibbs steps using the full-conditional distributions of these parameters, and the unit-level parameters being drawn through a reversible jump step. Thus, if we assume a conditionally conjugate prior of the form given below

$$\lambda_k \sim \mathrm{N}\left(\bar{\lambda}, V_{\theta,k} \otimes A^{-1}\right) \forall k \in \mathcal{K} \tag{34}$$

$$V_{\theta,k} \sim \text{Inverse Wishart}\left(\nu, V\right) \forall k \in \mathcal{K} \tag{35}$$

we could derive the full conditional densities of the population-level parameters as follows

$$\operatorname{vec}\left(\lambda_k\right) | \cdot \sim \mathrm{N}\left(\operatorname{vec}\left[\left(Q_k' Q_k + A\right)^{-1}\left(Q_k' \Theta_k + a\bar{\lambda}\right)\right], V_{\theta,k} \otimes \left(Q_k' Q_k + A\right)^{-1}\right)$$

$$V_{\theta,k} \sim \text{Inverse Wishart}\left(\nu + n_k, V + S_k\right)$$

where Q_k is the matrix formed by stacking up all the q_i vectors corresponding to model k, Θ_k is similarly the stacked matrix of $\theta_{i,k}$ vectors for the model, n_k is the number of such models and S_k is the sum of squared residuals matrix. Thus, once we have draws of unit-level $\theta_{i,k}$ parameters, the population-level parameters can be drawn model by model as simple Gibbs steps. The hierarchical setup is similar to that for standard hierarchical Probit models (see for instance Rossi et al. 1996), and the set of draws at the population level are very similar for those models, except that we have separate population-level parameters for each model. The algorithm would involve a series of reversible jump steps, one for each unit, within a larger Gibbs sampler. We would draw a model indicator for each individual and could conduct inference on the posterior probabilities of equilibrium selection for each unit. One last point to make in this section is that even when the equilibrium selection rules are not identified, for instance, when there are no exclusion restrictions available, our experimentation with this method using Monte Carlo simulations show that it is still useful to cast the unit level model in a hierarchical framework, specifically when the payoff functions include covariates. The reason is that across model moves depend on efficient moment matching between models to ensure that the sampler accepts candidate draws at a sufficient rate. Since the moment are matched for specific values of covariates (e.g. mean, mean plus one standard deviation, etc.), the posterior densities for the actual observations in the data depend on how close the covariates in the observations are to those used in the moment matching conditions. The covariates are typically closer within units, and there could be typically large dispersions away from these values when we pool observations across units, causing high rejection rates. Thus, specifying heterogeneity in equilibrium selection has practical benefits even in the absence of exclusion restrictions. Note that the parameters of the payoff functions and any heterogeneity in those are still identified, and specifying heterogeneity in equilibrium selection does not affect inference for these parameters.

4 Monte Carlo simulations

4.1 Two player game of social interaction

4.1.1 When the equilibrium selection rule is not identified

We first simulate a social interaction game with two players, with players deciding whether to consume the good jointly with their partner, alone or not at all. We assume that there are no exclusion restrictions available. We observe 1,000 such decisions by the individuals, with four potential outcomes in each observation. There are three parameters in the model, the intercepts for the two players (α_1 and α_2), and a social interaction effect of the consumption effect of the other player on the focal player's payoffs (γ). We draw the errors (ε_1 and ε_2) from independent standard normal distributions. We then simulate the outcomes by checking which cell the error draw falls in. In case the error falls in Cell V in Fig. 2, we assume that the Pareto dominant equilibrium is played, where the outcome of both players consuming Pareto dominates the outcome of neither consuming. We estimate the model using our algorithm, assuming a uniform prior over the multiple equilibria, and using diffuse priors for the other parameters.

We report the estimates in Table 1, in which we report the conditional posterior estimates of parameters, conditional on the model, and also the marginal estimates. Note that there are two models in this case, one corresponding to the Pareto dominant equilibrium being selected (Equilibrium 1 in the Table) and the other corresponding to the Pareto dominated equilibrium being selected (Equilibrium 2 in the Table). The estimates conditional on the true model (i.e. Equilibrium 1) are quite close to the true parameter values. But the estimates conditional on the incorrect model are far away from the true parameter values for the two intercept parameters α_1 and α_2. We also depict the posterior densities for these parameters in Fig. 4. The panels on the left show the marginal densities, while the panels on the right show the conditional densities. The true parameter values are depicted by vertical lines. The conditional densities for α_1 and α_2 once again show that the conditional densities are relatively far apart, with the densities conditional on the selection of Pareto dominant equilibrium being close to the true parameter values, while those for the Pareto dominated equilibrium do not enclose the true values. Moreover, the marginal densities for these parameters show two modes, reflecting the two equilibria. Also, the algorithm draws the Pareto dominant and Pareto dominated models in roughly the same proportion as the prior probabilities for the models (which was 0.5 for each of the models).

These estimates demonstrate two important things. First, they show the effect of equilibrium selection on the parameter estimates. The parameter estimates conditional on the incorrect model being selected are far away from the truth, and in the case of these simulations, do not even enclose the truth within the 95 % posterior credible interval. Specifically, assuming the Pareto dominated model, which allocates the entire region of multiplicity (Cell V in Fig. 2) to the outcome of neither player consuming, leads to an overestimation of the intercepts for the two players. This makes sense, given that with a smaller region allocated for the outcome of both players consuming, the observed outcomes can only be rationalized by increasing the intercepts. While we don't show the maximum likelihood estimates for these models, the estimates assuming the incorrect equilibrium show a similar pattern, with the truth not lying within the 95 % confidence interval for these parameters. Second, these estimates show that our methodology is working, recovering the

Table 1 Monte Carlo simulation—2 player game of social interaction

Parameter	True value	Posterior estimates (mean/std. dev.)		
		Marginal	True model	Model 2
α_1	−0.2	−0.1346 (0.0790)	−0.2040 (0.0492)	−0.0755 (0.0434)
α_2	−0.4	−0.3274 (0.0823)	−0.3989 (0.0549)	−0.2665 (0.0438)
$\log(\gamma)$	−0.5	−0.4994 (0.0848)	−0.4873 (0.0862)	−0.5098 (0.0822)
Probability ($k = i$)			0.52	0.48

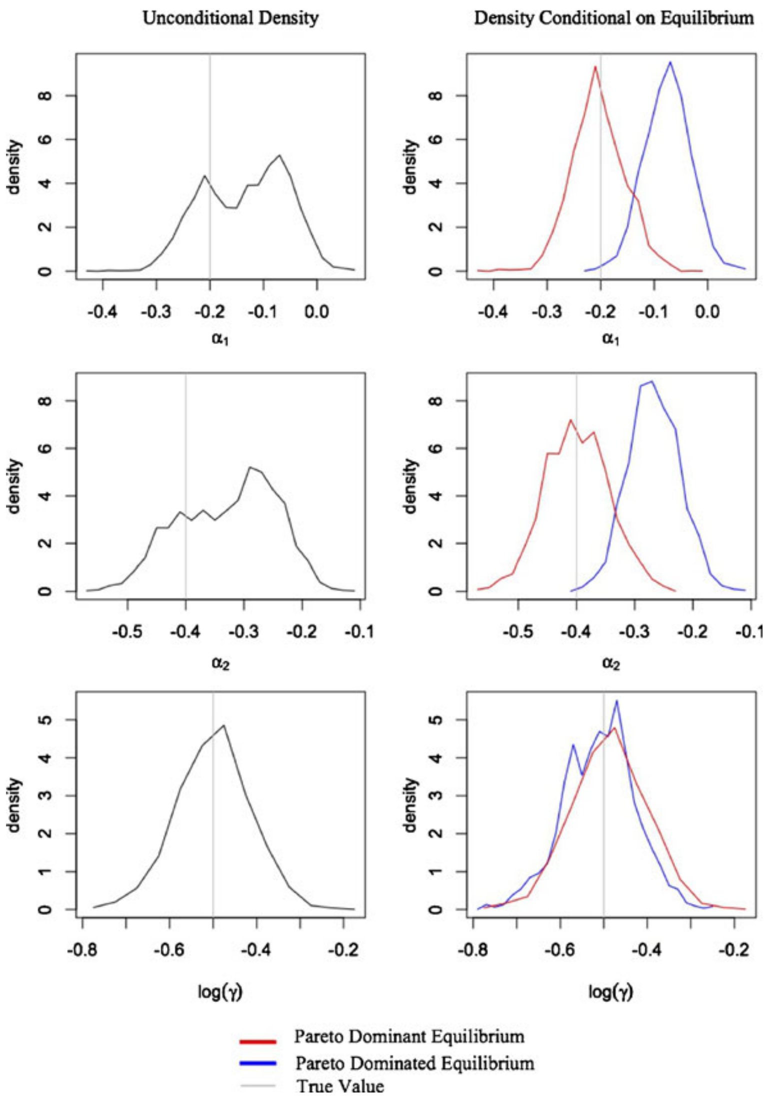


Fig. 4 Posterior densities: two player game of social interaction

truth conditional on the true equilibrium selection, and recovering a potentially multi-modal posterior distribution for parameters. Thus, these estimates demonstrates how our methodology complements existing methodologies.

4.1.2 When equilibrium selection is identified

In this simulation, we add excluded covariates to the data generating process, which identifies the equilibrium selection rule as discussed earlier. We simulate one covariate for each of the two players, assuming that it is uniformly distributed between −1 and 1. We assume values for the coefficients for these excluded covariate, and keep everything else the same as the previous simulation, including the assumption that the Pareto dominant equilibrium is selected. We again estimate the model using our reversible jump algorithm, assuming an equal prior probability for both the Pareto dominant and Pareto dominated equilibria, and diffuse priors for the other parameters.

The results for this simulation are reported in Table 2. The results look similar to those for the previous simulation. Once again, we find that the intercepts for the two players depend critically on the assumption on equilibrium selection. The conditional estimates for the true model for all the parameters enclose the true value within the 95 % credible interval, but not for the incorrect model. We also find that the estimates of the coefficients for the covariates, and of the social interaction parameter, are similar for both models. The interesting result is that the correct model, i.e. with the selection of the Pareto dominant equilibrium, is drawn more than twice as often as the Pareto dominated model. This demonstrates the identification of the equilibrium selection rule in the presence of excluded covariates.

Figure 5 shows the posterior densities of three parameters—the two intercepts and the social interaction parameter (the coefficients for the two excluded covariates are not shown, but have similar patterns as the social interaction parameter). The marginal density is a mixture of the densities for the two models, but with higher contribution from the Pareto dominant model as it is drawn more often than the other model. The key point to note is that the marginal density in this case is closer to that for the true model and does not have multiple modes, though it has broader shoulders than the density for either model. As the posterior probability for the a particular model increases, this marginal density should move closer and closer to the conditional density for the that model. Thus, with better identification, we should get tighter estimates of the parameters.

We have shown through these simulations that the algorithm is able to do a reasonably good job of recovering the parameters, and reflecting the uncertainty over equilibrium selection. It spans both extremes, when the model is unidentified, and when it is known with certainty (which would be equivalent to setting a prior probability of 1 on that model) and the continuum of situations in between, where the model is not known with certainty, but can be identified using the data.

Table 2 Monte Carlo simulation—2 player game of social interaction—with covariates

Parameter	True value	Posterior estimates (mean/std. dev.)		
		Marginal	True model	Other models
α_1	−0.2	−0.1273 (0.0630)	−0.1569 (0.0491)	−0.0685 (0.0433)
α_2	−0.4	−0.3424 (0.0676)	−0.3720 (0.0453)	−0.2837 (0.0510)
$\log(\gamma)$	−0.5	−0.6515 (0.0968)	−0.6509 (0.0970)	−0.6526 (0.0963)
β_1	0.2	0.2176 (0.0710)	0.2173 (0.0710)	0.2181 (0.0711)
β_2	0.2	0.1873 (0.0723)	0.1870 (0.0711)	0.1880 (0.0747)
Probability ($k = i$)			0.68	0.32

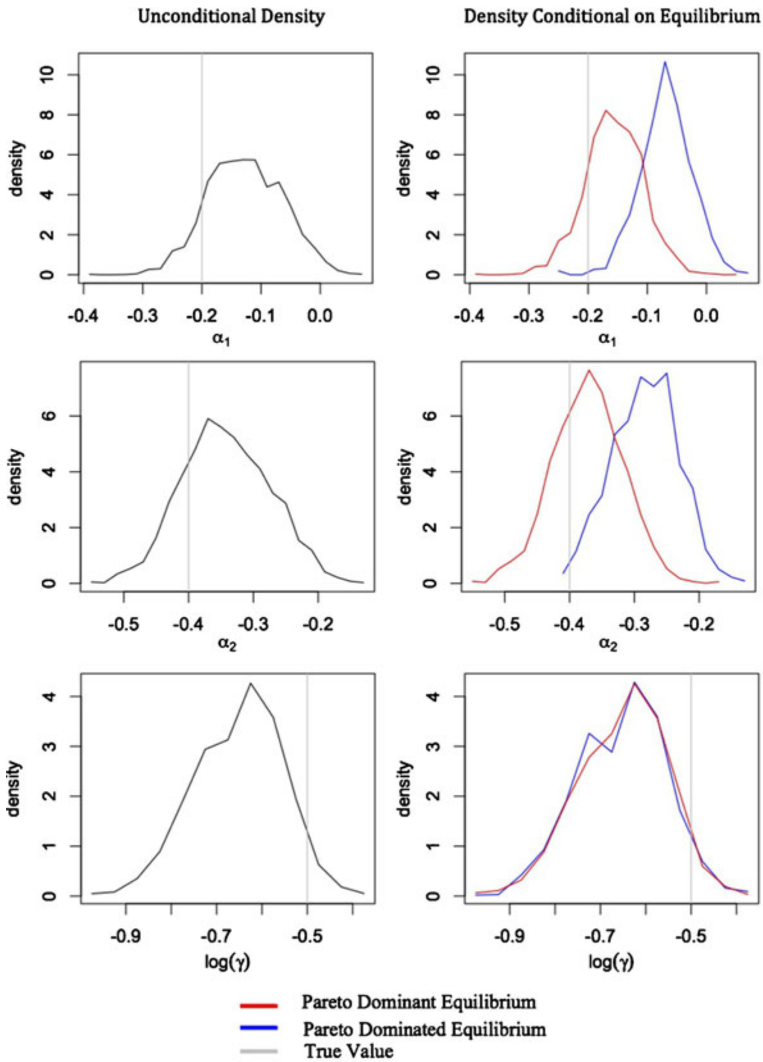


Fig. 5 Posterior densities: two player game of social interaction—with covariates

4.2 Three-firm competitive game

We next conduct a simulation for a competitive game, with three players and with payoff functions (reduced form profit functions) as defined in Eq. 1. We also add excluded variables in this simulation, thus the equilibrium selection rule is identified (up to the number of moments available). Similar to the Monte Carlo simulations in the case of the 2-player social interaction game, we first assume parameter values. The parameters in this simulation include the intercepts—α_1, α_2 and α_3, the two competitive effect parameters—γ and

δ, and the coefficients for the three excluded covariates θ_1, θ_2 and θ_3. We assume an equilibrium selection rule that picks the equilibrium that maximizes industry profits. The underlying assumption of such an equilibrium is that firms enter in decreasing order of profits. We then draw the errors for 1,000 observations from independent standard normal distributions and generate the outcomes, applying the equilibrium selection rule when the errors are in a cell with multiple equilibria. Once again, estimation was carried out using our reversible jump approach, specifying diffuse priors for the parameters of the profit functions. We specify that the true model, i.e. with the equilibrium selection favoring higher industry profits, has a 50 % prior probability, and the set of all other equilibrium selection mechanisms that allocate entire cells to outcomes has a 50 % prior probability. Note that this second set consists of 576 equilibrium selection rules, with each being a combination of allocation of cells to particular equilibria. There are six cells with two equilibria, and two other cells with three equilibria, giving us 576 such combinations.

The results of this simulation are in Table 3. There are several interesting results in this table. First, unlike in the case of the social interaction game, the parameter estimates for the true model and for the set of other models are quite close. This reflects the fact that in this case, the true model involves parts of the cells with multiple equilibria being allocated to different outcomes, rather than entire cells. Since the set of other models have all possible combinations of entire cells being allocated to outcomes, the parameter estimates represent a mixture of these combinations. However, the true model is also like a mixture of these other models. Thus, the estimates of the true model and the set of other models are close. Note however, that the standard deviations of the estimates for the set of other models is higher than that for the true model in the case of every parameter with the exception of the α parameter, which is estimated with a very high variance in both cases, reflecting the fact that these data perhaps do not identify the parameter very well.

Figure 6 shows the posterior densities for the three intercepts for this simulation. We see that the modes of the conditional densities, conditional on the true model and the set of other models are very close. However, estimates for

Table 3 Monte Carlo simulation—3 player competitive game

Parameter	True value	Posterior estimates (mean/std. dev.)		
		Marginal	True model	Other models
α_1	−0.2	−0.2434 (0.0742)	−0.2396 (0.0664)	−0.2498 (0.0855)
α_2	−0.4	−0.3758 (0.0770)	−0.3761 (0.0652)	−0.3755 (0.0935)
α_3	−0.6	−0.5963 (0.0850)	−0.5966 (0.0767)	−0.5958 (0.0973)
$\log(\gamma)$	−0.5	−0.4136 (0.1117)	−0.4110 (0.1081)	−0.4180 (0.1174)
$\log\left(\frac{\delta}{1-\delta}\right)$	−0.6	−3.4221 (10.2048)	−3.6594 (10.4287)	−3.0214 (9.8036)
θ_1	0.2	0.2884 (0.1049)	0.2887 (0.1014)	0.2877 (0.1105)
θ_2	0.2	0.3943 (0.1140)	0.3995 (0.1095)	0.3854 (0.1208)
θ_3	0.2	0.1680 (0.1205)	0.1684 (0.1143)	0.1672 (0.1302)
Probability ($k = i$)			0.56	0.44

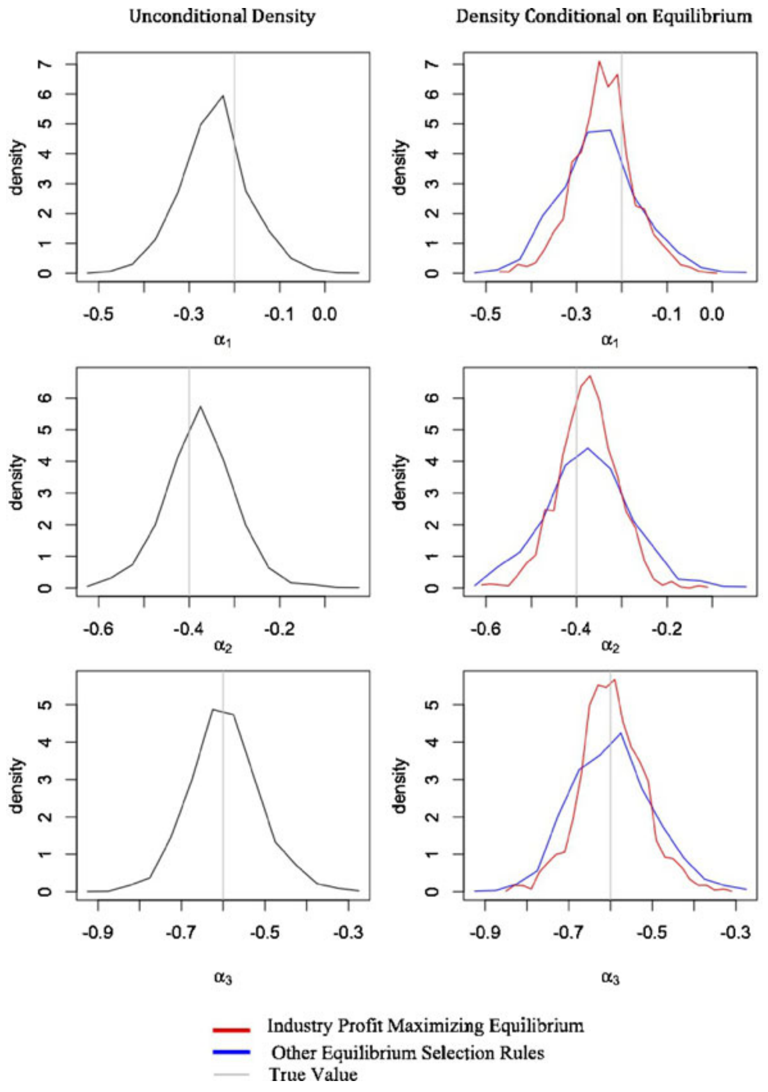


Fig. 6 Posterior densities: three player competitive game—with covariates

the true model have higher peaks and those for the other models have broader shoulders. Of course, if the true model were one of the other equilibrium selection rules, the posterior estimates of parameters might not be close, like in the case of the simulations for the social interaction games. To demonstrate this, we re-estimate the simulation, testing between two specific equilibrium selection rules with entire cells allocated to outcomes and report the posterior estimates in Table 4. We see that posterior densities for the two models don't have the same degree of overlap as in the previous simulation, and that the true model is drawn significantly more frequently than the alternative model.

Table 4 Testing between two specific equilibrium selection rules—3 player competitive game

Parameter	True value	Posterior estimates (mean/std. dev.)	
		Model 1	Model 2
α_1	−0.2	−0.2215 (0.0677)	−0.2624 (0.0607)
α_2	−0.4	−0.4132 (0.0614)	−0.3449 (0.0666)
α_3	−0.6	−0.6418 (0.0629)	−0.6802 (0.0696)
$\log(\gamma)$	−0.5	−0.4516 (0.0774)	−0.4616 (0.0938)
$\log\left(\frac{\delta}{1-\delta}\right)$	−0.6	−3.0801 (10.5871)	−4.4857 (9.7847)
θ_1	0.2	0.3302 (0.1236)	0.2366 (0.0801)
θ_2	0.2	0.3669 (0.1242)	0.3698 (0.1088)
θ_3	0.2	0.1404 (0.0985)	0.1602 (0.1097)
Probability $(k = i)$		0.61	0.39

5 Empirical applications

5.1 Casino visits by married couples

5.1.1 Context and data description

This empirical application considers joint decision making by couples in the context of visiting casinos. The decision of whether the husband visits the casino on his own, the wife visit alone, neither of them visit or both visit can be modeled as the outcome of a discrete game played between them, with the visit by the partner assumed to positively impact the utility of each person. The goal of such an exercise might be to measure this partner influence, which could be used by the firm to target promotions in a way that fully accounts for such social effects. Such contexts have been studied in the literature in the past (Hartmann 2010; Narayan et al. 2006). In particular, Hartmann (2010) models the decision by friends of playing golf separately or with their friends as a discrete game.[5]

Our data consists of visit decisions by consumers to a casino in Las Vegas over an eight quarter period, captured using information on loyalty cards. The loyalty cards did not cost anything to the consumers and hence an overwhelming majority of regular consumers had loyalty cards. We classify a pair of consumers as husband and wife if they are adults over 25 years of age, share the same last name, the same address, have different genders, and have ages that are not more than ten years apart. The focus on heterosexual couples is because it is more likely that they share the same last name. The filtering on age difference is meant to ensure that we don't include cases of a parent living with their child and the minimum age limit attempts to eliminate cases of siblings living with one or more parents. We also exclude from the sample

[5]We abstract away from issues such as whether assumptions like a non-cooperative game, and of a positive utility from participation by partners should be relaxed, since our focus is on the estimation methodology.

couples for whom the very first observation in the data is more than six months after the first period in the data, to try and eliminate left-censoring of the data. Finally, we exclude couples who had multiple visits within any of the quarters, given that our analysis is at the quarterly level, to ensure that we don't have cases where, for instance, the husband visited on one trip and the wife visited on another. This set of filtering rules is unlikely to be fully accurate, since there could be cases of siblings living in the same address which might be picked up as couples, and of couples that don't share the same last name and would be missed in our analysis. Hence, we do not claim these results as representative of all couples who are consumers of the casino, but of those we have selected in our sample. We thus arrive at a sample of 3,046 couples in the data, for each of whom we observe visits to the casino, either separately or together. We convert this information into a vector of quarterly visit decisions for each of the eight quarters in our data. Each element of the vector can take one of four values—neither husband nor wife visiting the casino, both visiting, or either of them visiting separately.

The summary statistics for these data are in Table 5. A majority of observations are those where neither partner visits the casino. The second highest frequency of observations is for both partners visiting the casino together. We do not observe other covariates for these couples, and hence the equilibria are not identified in this case, either through exclusion restrictions or through identification at infinity arguments.

5.1.2 Estimation

We assume payoff functions similar to that in Eq. 5. Thus, the payoffs from visiting the casino for couple i at quarter t are given by

$$\begin{aligned}\pi_{iht} &= \alpha_{ih} + \gamma_i 1\left(y_{iwt} = 1\right) + \varepsilon_{iht} \\ \pi_{iwt} &= \alpha_{iw} + \gamma_i 1\left(y_{iht} = 1\right) + \varepsilon_{iwt} \end{aligned} \tag{36}$$

Here, the h and w subscript index husband and wife respectively, and y_{iht} and y_{iwt} respectively denote the decision of the husband and the wife to visit the casino. The ε_{iht} and ε_{iwt} represent unobservable (to the analyst) factors that affect the payoffs, that both husband and wife know perfectly, and are assumed to be identical and independently distributed as standard normal distributions. Note that one modification in this setup compared to that in Eq. 5 is that couples are allowed to be heterogeneous in their parameters. Given the moments we have (four decisions and hence three probabilities), we can identify three payoff parameters in total. Hence, we restrict the interaction

Table 5 Couples' casino visit decisions—summary statistics

Decisions	% of observations
Neither husband nor wife visits	91.1
Only husband visits	2.3
Only wife visits	1.1
Both husband and wife visit	5.5

effect γ_i to be the same for husbands and wives, though it is allowed to be heterogeneous across couples. We reparametrize this parameter as $\exp(\tilde{\gamma}_i)$ to ensure that it is positive.

Let k index the model, i.e. the equilibrium selection rule. In this case, there is one cell with two equilibria, and hence we consider two models. One, which is the Pareto dominant model, allocates the entire Cell V to the equilibrium of both husband and wife visiting, whereas the other—the Pareto dominated model - allocates the entire cell to the outcome of neither visiting. Let $k = 1$ index the Pareto dominant model and $k = 2$ index the Pareto dominated model.

We cast the model in a hierarchical context, by assuming that the couple-level parameters are distributed as follows, with these priors being conditioned on the model.

$$\theta_i | k_i = \begin{pmatrix} \alpha_{ih} & \alpha_{iw} & \tilde{\gamma}_i \end{pmatrix} \sim N\left(\theta_{k_i}, V_{\theta k_i}\right) \tag{37}$$

To complete the model, we specify conditionally conjugate priors on these prior parameters

$$\begin{aligned} \theta_k &\sim \mathrm{N}(\mu, \Sigma) \quad \forall k \in \{1, 2\} \\ V_\theta &\sim \text{Inv. Wishart}(\nu, V) \quad \forall k \in \{1, 2\} \end{aligned} \tag{38}$$

and prior probabilities for the models themselves, which we set to be uniform. Thus,

$$p_1 = p_2 = 0.5 \tag{39}$$

We use the reversible jump algorithm described in Section 3.1.2 to draw the θ_i parameters and the model indicators k_i.

5.1.3 Results

There are three main parameters of interest in this case, the intercept for the husband, the intercept for the wife and the social interaction parameter. These parameters are assumed to be heterogeneous across the sample. Instead of reporting the large number of individual-level parameters, we focus on population level means for the three parameters denoted by the vector $\mu \equiv \begin{pmatrix} \alpha_h & \alpha_w & \tilde{\gamma} \end{pmatrix}$. The parameter estimates are reported in Table 6. As expected when the models are not identified, the two models are drawn as per the prior probability, with the posterior probabilities being 0.47 and 0.53 for the two

Table 6 Population-level parameter estimates—casino visits by couples

Parameter	Posterior estimates (mean/std. dev.)		
	Marginal	Pareto dominant	Pareto dominated
α_h	−1.9066 (0.4285)	−2.1178 (0.4688)	−1.6954 (0.2414)
α_w	−2.3509 (0.5094)	−2.4199 (0.4713)	−2.2820 (0.5361)
$\tilde{\gamma}$	0.6528 (0.5158)	0.3209 (0.2330)	0.9848 (0.5074)
Probability ($k = i$)		0.47	0.53

models. We see that the parameter estimates conditional on the two models are quite different from each other. Specifically, the Pareto dominant model predicts a lower value for the social interaction parameter, and lower values for the two intercepts. Note that the logarithm of the social interaction parameter is reported in the table. When this is exponentiated, the ratio of mean to standard deviation is higher for the Pareto dominant model than for the Pareto dominated model, indicating that there is smaller variance relative to the mean in the case of the Pareto dominant model.

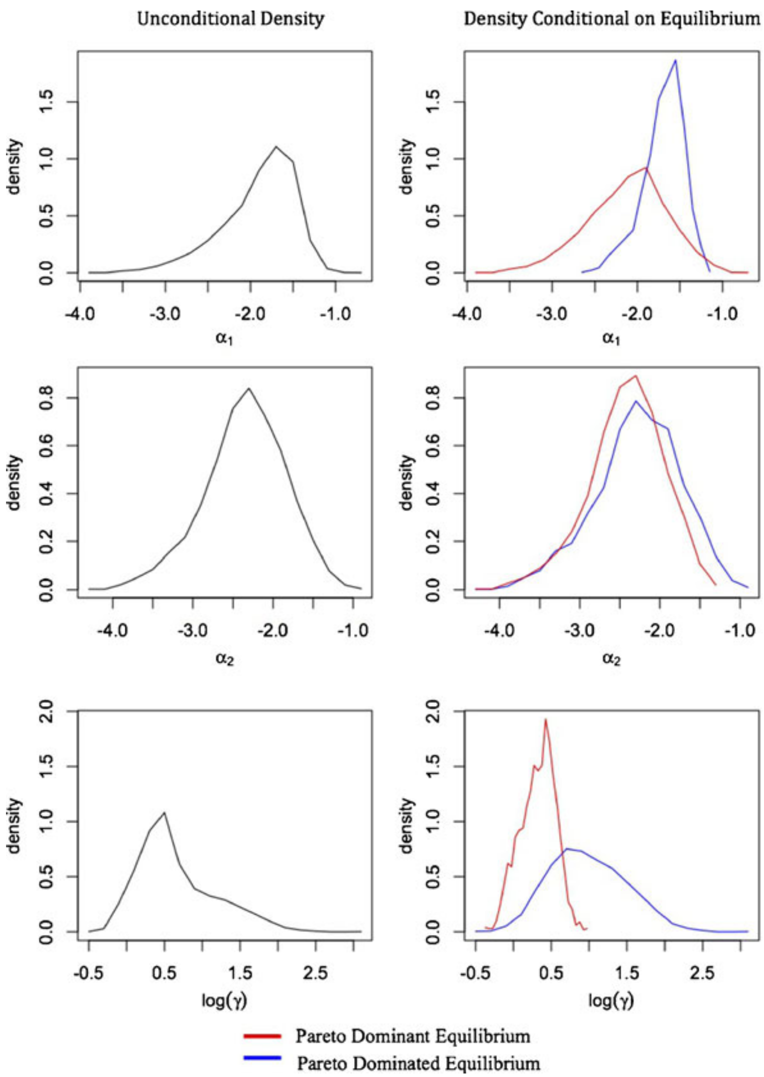


Fig. 7 Posterior densities—population-level parameters—casino visits by couples

Figure 7 shows the posterior densities for the parameters. The main points from these density plots is that the marginal densities are not bimodal, they have much broader shoulders than the conditional densities, particularly for the husbands' intercept and the social interaction parameter. Since we don't have exclusion restrictions in our data to identify the models, the best the analyst can do is to let posterior estimates reflect her own prior uncertainty over models. If she can be sure that the Pareto dominant equilibrium is always selected, she could set the prior probability for that model to be 1. In that case, the estimated posterior densities would look like the conditional densities for the Pareto dominant models in the figure (shown in red). However, if the analyst does not have that certainty, the posterior estimates would deviate from these estimates to the extent that the posterior estimates of the Pareto dominated model differ from those for the Pareto dominant model.

Next, we consider a set of counterfactual simulations of the kind that the casino might be interested in conducting to illustrate the implications of our results. For instance, the casino might be interested in conducting a promotional campaign, where coupons are sent to consumers to encourage them to visit the casino. In the presence of social interactions however, the social interaction could make it beneficial for the firm to target the promotion to only one of the two partners when sending promotional offers to couples. For a sufficiently high social interaction effect, it would induce the other partner to also visit the casino, without the casino offering them anything. Thus, measuring the social multiplier effect might be very useful for the firm. Since our data do not contain information on promotional offers sent to consumers, this variable is not in our model. We therefore consider a hypothetical promotional offer that changes the consumer's probability of visiting the casino by a certain percentage.

Consider a promotional offer that is sent to all husbands, which raises their likelihood of visiting the casino, but is not transferable to the wife. Since we don't observe promotions in our data, we simulate this by changing the intercepts of all husbands by an amount that is sufficient to increase the average husband's probability of visiting the casino alone by 10 %. We simulate the outcomes using the estimated parameters, but with the modified intercepts for husbands. We compare these outcomes with the simulated outcomes using the estimated parameters, including the estimated intercepts for husbands. We find that this does not just increase the probability that husbands alone visit the casino, but also that of couples visiting. As husbands' probability of visiting the casino increases, that for wives increases indirectly through the social interaction mechanism, since they derive incremental utility of visiting the casino with their partners. We conduct a similar simulation for a promotional offer sent to wives. We report the incremental increase in the number of couples visiting the casino for both of these promotions in Table 7. In each case, we report three different average effects, one for each model and an unconditional measure. The main point to note from this table is that there is a big difference in the estimates of the social multiplier effect depending on the model. It is much lower for the Pareto dominant model than for the Pareto dominated model, both when husbands and wives get the promotional offer.

Table 7 Effect of promotion to one partner on joint visits by couples

Parameter	% increase in the number of couples visiting	
	Promotion to husbands	Promotion to wives
Unconditional	13.8071	3.1854
Conditional on Pareto dominant	3.3603	1.2351
Conditional on Pareto dominated	19.3572	5.3370

5.2 Location decisions by stores in the stationery retail industry

5.2.1 Context and data description

In this application, we consider geographical location decisions by firms in the stationery retail industry. There are three major chains in this industry—Staples, Office Depot and Office Max, with market shares and number of stores in that order. Store location decisions in this industry have been a matter of interest in the recent past (Manuszak and Moul 2008), particularly due to the proposed merger between Staples and Office Depot. This merger was opposed by the Federal Trade Commission (FTC), and the firms eventually decided not to pursue the merger. The reason for the opposition by the FTC was that average prices in markets where both stores were present were lower than those in markets where only Staples was present. Thus, it was felt that the merger would end up raising prices, presumably because the merged firm would re-optimize its location decisions and close down one of the stores in many of the markets where both stores were present. An understanding of how firms make location decision in this industry requires an understanding of their profits from entry into markets, and the competitive interactions between them, and discrete games of market entry offer a natural way to do so.

The data consist of location decisions for the three major firms in the market, downloaded from the firms' websites using an automated script. We geocode the street address for each store's location into latitudes and longitudes, and use census information to determine which Core Based Statistical Area (CBSA) the store is located in. A CBSA consist of a "core area consisting of a substantial population nucleus, together with adjacent communities that have a high degree of economic and social integration with that core' (Census 2011). It combines together cities and suburbs, with most people commuting within a CBSA rather than across CBSAs. We aggregate the store location information at the CBSA level, and for the purpose of this application, consider whether the store for a chain is present in the CBSA or not. The choice of CBSA as a market is to try and minimize the possibility that stores in a market draw customers from nearby markets. A CBSA is large enough to ensure that such across-market customer movements are relatively small, but small enough to ensure that we have a sufficient number of markets and sufficient variation in the data for reasonable empirical analysis. There were totally 933 CBSAs in our data, and thus 933 markets in which we observe whether each of the three chains was present or not.

We also collected demographic information for these markets from the US census data for 2001.While there are a number of variables in these data, we use the population and per capita income in each market in our empirical analysis, since our preliminary analysis showed that these variables were the most significant. The exclusion of the other variables (such as employment profile, racial profile of the households, income distributions etc.) did not change the results very much. In addition, we include three dummy variables for the census region the market is located in—Northeast, Midwest, South or West—and a dummy variable for whether the CBSA was classified as a metropolitan area or a micropolitan area. A CBSA is classified as a metropolitan area if it contains at least one urban area with a population of 50,000 people or more, and as a micropolitan area if it contains at least one urban area with a population between 10,000 and 50,000 but none with a population exceeding 50,000 people. In place of census region, we also tried using dummies for census divisions, which are nine in number and constitute smaller geographical regions than census regions. We find that they do not add very much to the analysis (the likelihoods are not very much higher than those when we use dummies for census regions) but add significantly to the number of parameters to be estimated. We also included a variable indicating the distance of the market from the centroid of the US, to account for the different attractiveness of the coastal regions of the US vs. the regions in the heart of the country.

Finally, we included a set of excluded variables to help identify the equilibrium. The excluded variables we use are the distances from the corporate headquarters of the firms, a variable that has been used before in the literature (Jia 2008; Holmes 2011; Zhu et al. 2009). The rationale for its inclusion in the profit function is that firms' costs of managing stores that are far away from the corporate headquarters are higher than those for stores that are nearby. Thus, the distance of the market from corporate headquarters would play a role in the firm's market entry decision. However, after controlling for market characteristics (including locational advantages of some parts of the country, accounted for using regional dummies), the distance of a market from a competitor's corporate headquarters should not affect a firm's profitability

Table 8 Market structure of the stationery market

Variable	Number of markets (CBSAs)
Total markets	933
Staples present	411
Office depot present	307
Office max present	288
Staples only	190
Office depot only	61
Office max only	71
Staples & office depot	79
Staples & office max	50
Office depot & office max	75
All three chains	92
None	315

Table 9 Summary statistics—demographic variables

Variable	Median	Mean	Std. dev.
Population	71,918	241,746	623,039
Per capita income	24,755	25,391	4,834
Distance from staples HQ (miles)	911	1,028	608
Distance from office depot HQ (miles)	938	1,052	567
Distance from office max HQ (miles)	548	626	438
Distance from center of the US (miles)	621	673	358

in a particular market and consequently its decision to locate in that market. Thus, these would serve as excluded variables, directly affecting only the firms' own decisions, but not those of competitors.

We next provide some summary statistics for the data. As mentioned earlier, there were a total of 933 markets in the data. First, we give a sense of how the stores are distributed across the markets in Table 8. Staples is present in the largest number of markets, followed by Office Depot and Office Max in that order. None of the three chains are present in just over a third of the markets, and all three are present in 92 markets, presumably the largest and most important markets. Next, we look at the demographic variables in our data (Table 9). The distribution of population is skewed, reflecting the small number of markets with large populations, and a relatively large number with small populations. Staples has its corporate headquarters in Framingham, Massacheusetts, Office Depot in Boca Raton, Florida and Office Max in Naperville, Illinois, and we report the summary statistics for the distances from these places to the centroids of each market. We also report the distance of the market from the geographical center of the continental United States. Not reported in the table is the fact that there are 360 metropolitan markets, and 573 micropolitan markets. The largest numbers of markets are in the South at 392, followed by the Midwest, with 278, the West with 169 and Northeast, with 94.

Table 10 Reduced form probit regressions—stationery chain locations

Covariate	Dependent variable: market presence for		
	Staples	Office depot	Office max
Intercept	−0.2887 (0.3817)	−2.3217 (0.4969)	−0.8052 (0.4165)
Staples present	–	−1.1893 (0.1835)	−1.2658 (0.1509)
Office depot present	−1.1816 (0.1784)	–	−0.8796 (0.1808)
Office max present	−1.3211 (0.1525)	−1.0867 (0.1947)	–
Distance from staples HQ	−0.6607 (0.1770)	0.1594 (0.1902)	−0.2082 (0.1671)
Distance from office depot HQ	−0.1908 (0.1453)	−0.3973 (0.1961)	−0.1948 (0.1481)
Distance from office max HQ	−0.1370 (0.1289)	−0.2923 (0.1726)	−0.3472 (0.1411)
Distance from center of the US	−0.1305 (0.1103)	−0.4656 (0.1560)	0.0881 (0.1308)
Population	0.8582 (0.0878)	1.2174 (0.1313)	1.0059 (0.0998)
Income	1.6562 (0.3671)	2.1094 (0.4594)	1.1783 (0.3886)
Dummies	Metro, region	Metro, region	Metro, region

Before we move on to the estimation and results for this application, we discuss a set of reduced form probit regressions for the location decisions of the firms. We estimate one regression each for the three firms. The covariates include the demographic variables, region and metro dummies, and distances from each firm's corporate headquarters. We also include the presence of the competing stores as covariates. The purpose of these regressions, besides checking if the estimates have the expected signs, is to verify that the excluded variables, i.e. the distances of the markets from the corporate headquarters, affect the firm's own location decisions, but not those of its competitors. We have taken logs of appropriately scaled distances, populations and incomes to account for concavity in these effects. These results are reported in Table 10. We find expected signs for the parameters. First, the effect of the presence of competitors is negative, i.e. a firm is less likely to locate in a market if its competitor is also present there, everything else remaining equal. The distances from the corporate headquarters negatively affects the respective firm's location decisions, but not of its competitors. Population and income have the expected positive signs. While they are not reported here, the coefficients for the dummy variables have the expected signs. These regression give us confidence to proceed with the estimation of the discrete game, using distances from corporate headquarters as excluded variables.

5.2.2 Estimation

We use the payoffs in Eq. 1 for the three firms' profit functions

$$\pi_{it} = \alpha_i + X_{it}\beta - \gamma \cdot 1\left(\sum_{j\neq i} y_{jt} > 0\right) - \gamma\delta \cdot 1\left(\sum_{j\neq i} y_{jt} > 1\right) + \varepsilon_{it} \underset{i.i.d.}{\sim} \mathrm{N}(0,1) \tag{40}$$

Here $i \in \{\text{Staples, OfficeDepot, OfficeMax}\}$. An underlying assumption in this payoff function is that the competition effect is the same across firms, and depends only on the number of competing firms irrespective of their identity. Since there are a total of eight possible outcomes in a 3-player entry game, there are seven moments available to identify the intercepts and competitive interaction parameters. Here, we have five parameters, and in principle could relax the model to estimate two other terms. Since the purpose of this application is to illustrate the estimation approach, we stay with this simpler formulation. In this application, we have exclusion restrictions (which constitute a part of X_{it}) to identify the equilibrium. With three excluded variables, we don't have enough moments to identify more than three equilibrium selection rules in this model. Like in the case of the Monte Carlo simulation reported in Section 4.2, we test between the sequential entry rule and a set of 576 other equilibrium selection rules, each of which allocates entire cells to outcomes. We assume a prior probability of 50 % each for these two sets of equilibria.

Finally, we assume diffuse conditionally conjugate priors for the parameters. More details are available from the author upon request.

5.2.3 Results

The main parameters of interest include the model indicator (k), which informs us about the posterior probability for the sequential entry equilibrium assumption, the intercepts for the three firms (α_{Staples}, $\alpha_{\text{OfficeDepot}}$ and $\alpha_{\text{OfficeMax}}$), the two competitive interaction parameters (γ and δ) and the vector of coefficients for covariates (β). The parameter estimates for the two models and the unconditional parameter estimates are summarized in Table 11. A notable result is that the posterior probability of the sequential entry equilibrium is 0.39, much lower than the prior of 0.5. While not reported here, this pattern was replicated for other priors (such as 0.7 prior probability for that equilibrium). This is a notable result since it is common in the literature to assume this equilibrium. The parameter estimates show a similar pattern as in the Monte Carlo simulations—the parameters for the sequential entry model are close to the average for the set of other equilibrium selection rules, with lower standard deviations. The parameters are largely along expected lines. The intercepts are the highest for Staples, followed by that for Office Depot and Office Max in that order. This reflects the overall sizes of the chains. The competition effect is estimated relatively tightly, while the parameter that tells us how much the competition effect is for a second competitor is close to 1 ($\frac{\exp(3.5736)}{1+\exp(3,5736)} = 0.9727$) suggesting that the competition effect is close to being linear in the number of competitors.

There is not very much difference between the parameter estimates for the sequential entry equilibrium selection rule and for the set of other rules.

Table 11 Parameter estimates—stationery retailers market location application

Parameter	Posterior estimates (mean/std. dev.)		
	Marginal	Sequential entry	Other equilibria
α_{Staples}	−0.4829 (0.3417)	−0.4773 (0.3178)	−0.4867 (0.3567)
$\alpha_{\text{Officedepot}}$	−0.7857 (0.3410)	−0.7970 (0.3183)	−0.7783 (0.3550)
$\alpha_{\text{Officemax}}$	−0.8671 (0.3286)	−0.8702 (0.3162)	−0.8650 (0.3366)
$\log(\gamma)$	−0.4908 (0.1673	−0.4918 (0.1638)	−0.4902 (0.1724
$\log\left(\frac{\delta}{1-\delta}\right)$	3.5736 (3.3626)	3.3084 (3.1105)	3.7501 (3.6940)
$\beta_{\text{Distance from staples HQ}}$	−0.5606 (0.1811	−0.5599 (0.1665)	−0.5610 (0.2011)
$\beta_{\text{Distance from office depot HQ}}$	−0.3889 (0.1893)	−0.4176 (0.1850)	−0.3697 (0.1920)
$\beta_{\text{Distance from office max HQ}}$	−0.0815 (0.1400)	−0.0881 (0.1322)	−0.0770 (0.1500)
$\beta_{\text{Distance from US center}}$	−0.0770 (0.1097)	−0.0726 (0.1092)	−0.0799 (0.1100)
$\beta_{\text{Population}}$	0.7393 (0.1157)	0.7396 (0.1148)	0.7391 (0.1171)
β_{Income}	0.9369 (0.3309)	0.9658 (0.3175)	0.9176 (0.3382)
$\beta_{\text{Metropolitan}}$	0.8437 (0.1820)	0.8405 (0.1743)	0.8458 (0.1928
β_{Midwest}	−0.6239 (0.2210)	−0.6488 (0.1966)	−0.6073 (0.2343)
$\beta_{\text{Northeast}}$	−1.0130 (0.2523)	−1.0292 (0.2336)	−1.0021 (0.2635)
β_{South}	−0.6480 (0.2006)	−0.6671 (0.1960)	−0.6353 (0.2027)
Probability$(k = i)$		0.39	0.61

Hence, we don't explore these in further detail. However, our results do point to the fact that researchers need to consider equilibria other than the standard sequential entry equilibrium. In this application, we don't consider other specific equilibria, but this deserves further investigation.

6 Conclusion

In this paper, we consider the issue of multiplicity of equilibria in the case of discrete games of complete information, and demonstrate a Bayesian approach to deal with this. An issue that has been difficult to address satisfactorily ever since the literature on discrete games originated is that there are multiple equilibria in the case of these games.. One approach that the literature has taken is to avoid the problem altogether by focusing on outcomes that do not suffer from multiplicity. However, this restricts the nature of problems that can be analyzed to ones that can be modeled as competitive games with homogenous players. A second approach has been to make an ad-hoc equilibrium selection assumption. While many of these assumptions are reasonable, they are often not testable and don't have theoretical basis either. A third, more recent approach in the literature, has been to take an empirical approach to equilibrium selection, allowing the data to tell the analyst about the likelihood of a specific equilibrium being selected. Another approach has been to estimate bounds for parameters when they are not point-identified due to multiplicity of equilibria. Our approach complements these approaches, allowing for situations when equilibria are identified and when they are not. When the equilibrium selection rule is identified, our approach provides joint posterior density estimates for equilibrium selection rule and parameters. When it is not, it gives us marginal posterior estimates that are potentially multi-modal, to reflect the multiple equilibria.

Specifically, we develop an MCMC approach to estimating the equilibrium selection rules and parameters jointly, based on a reversible jump sampler. The reversible jump sampler is a useful algorithm to draw parameter estimates in cross-dimensional spaces. We adapt this approach to a context that is not cross-dimensional, but is across non-comparable parameter spaces nevertheless. We demonstrate how to practically implement this algorithm to the context of discrete games. We further extend the approach to allow for heterogeneity in equilibrium (model) selection. We show how to cast our model in a hierarchical framework, allowing for estimation of individual-level parameters and posterior equilibrium probabilities. This is a significant novel contribution of this approach to the literature on discrete games. While we do not explicitly demonstrate this, this hierarchical framework lends itself to the study of relationships between payoff parameters and equilibrium selection.

Acknowledgements I thank Wes Hartmann, Jean-Francois Houde and Paul Grieco for useful discussions, and the participants at the Marketing Science Conference 2010, Quantitative Marketing and Economics Conference 2011, and seminars at Goethe University, University of California at Davis and University of Texas at Dallas for useful comments. All errors are my own.

Appendix: Details of the algorithm

The algorithm for our sampler has the following steps:

1. The current draw is (k, θ_k).
2. Generate a candidate draw of model k' from the multinomial distribution Multinomial (p) where p is also the vector of prior probabilities for the various models.
3. Using a set of moment matching conditions, generate $g(\theta_k)$. This is done by minimizing the distance between the moments for the state (k, θ_k) k and $\left(k', g(\theta_k)\right)$.
4. Numerically evaluate $\frac{\partial g(\theta_k)}{\partial \theta_k}$.
5. Generate a draw for u from the normal distribution $\mathrm{N}(0, \Sigma_u)$
6. Set the candidate parameter vector $\theta'_{k'} = g(\theta_k) + u$
7. Calculate the acceptance probability $\alpha_t\left((k, \theta_k), \left(k', \theta'_{k'}\right)\right)$.
8. Generate $\mu \sim$ Uniform $(0, 1)$
9. If $\mu < \alpha_t\left((k, \theta_k), \left(k', \theta'_{k'}\right)\right)$ accept the candidate draw $\left(k, \theta'_{k'}\right)$, else set the new draw to be equal to the previous one (k, θ_k).

Tuning the reversible jump algorithm

One of the important choices to be made in practically implementing a Metropolis Hastings algorithm is the candidate density. In the case of a random walk algorithm with a normal candidate, this choice boils down to the choice of the variance covariance matrix for the candidate move. In principle, this choice does not matter as long as the density fulfills certain regularity conditions, but in practice this is a crucial choice. Intuitively, if the candidate moves to a very far away location from the current draw, it might be in a region where the target density is very low, and hence the move would get rejected. If the step size is too small, such that the candidate draw has a very similar target density as the current draw, the probability of acceptance becomes very high. In both cases, information is lost due to high autocorrelation of the draws. This means that it would take a very large number of draws to dissipate initial conditions, and to traverse the regions where the target density is non-negligible. Thus, a significant part of practical implementation of a Metropolis Hastings algorithm involves tuning the algorithm.

The issue of tuning assumes even greater importance in the case of the reversible jump algorithm. This is because the algorithm traverses parameter spaces across models, which are not entirely comparable. At an extreme, consider the case where $g(\theta_k) = \theta_k$ for all moves, including across model moves. In principle, such a sampler would work, but the rejection rate of candidate draws might be unacceptable high and may render the algorithm practically useless. This is because the candidate draws generating by taking a deviation from the current draw may have very low target density for the candidate model, making it almost certain that the candidate draw would be rejected. The algorithm would not make across-model moves in such a case. The moment-matching step in our algorithm reduces the chance of a prohibitively high rejection rate, by ensuring that the target densities for the current draw and the transformed draw (i.e. $g(\theta_k)$ are at least at similar levels. But tuning is still a non-trivial part of making the reversible jump algorithm work. One challenge is that a small step size for one model might be a large one for another. Hence, it would be attractive to come up with an adaptive version of the algorithm that has some degree of 'automatic' tuning as has been attempted for the standard Metropolis Hastings algorithm (Roberts and Rosenthal 2001). These adaptive algorithms adjust the variance of the candidate draw to keep the acceptance rate close to a pre-specified target level. The problem with this approach is that there is typically no clear criterion (i.e. no generalizable choice of target acceptance rate) that can be used for automating the tuning, particularly in the case of the reversible jump algorithm. Thus, one would have to use relatively arbitrary criterion (such as an arbitrarily chosen 'optimal' acceptance rate) to adaptively adjust the tuning. An alternative approach in the case of the Metropolis Hastings algorithm is to use a proposal that utilizes information about the gradient of the target density to generate efficient candidate draws in situations where rejection rates of the standard random walk algorithm are prohibitively high. This underlies the so-called *Langevin* algorithm (Roberts and Tweedie 1996; Roberts and Rosenthal 2006), which relies on generating draws that satisfies two conditions. First, it ensures that the acceptance rates are equal to 1 when the candidate draw is equal to the current draw. Second, it ensures that the gradient of the acceptance rate with respect to the candidate is equal to zero when the candidate is the current draw. A similar *Langevin-like* approach was introduced by Brooks et al. (2003) for the reversible jump algorithm, by proposing a draw that ensures that the acceptance rate is equal to 1 when candidate is a transformed version of the current draw, and by equating the gradient of the acceptance rate to zero with respect to a transformed version of the current draw to 0, at the current draw. One of the costs of this approach is that it can be computationally expensive, particularly in the context where the transformation function $g(\theta_k)$ cannot be computed analytically.

We have tried both approaches in our Monte Carlo simulations and found in our context that an automatic tuning approach worked well, although it comes at the cost of using an arbitrary criterion such as the one in Roberts and Rosenthal (2001). We find that the sampler mixes quite well in this case, with an acceptance rate ranging from about 10 to 25 %; We use this procedure in our empirical applications.

Assessing convergence

Assessing convergence of any MCMC sampler can be tricky and often depends on intuitive methods such as plotting the chains of the parameter draws and looking at measures such as autocorrelations. The issue of assessing convergence becomes much more complicated in the case of the reversible jump algorithm due to the across-model parameter space it samples from, and due to the difficulties in assessing convergence of the model indicator itself. One way to assess convergence of parameters is to see if there is good mixing of the parameters within each model. Depending on the dimensions of the parameter vector and the number of models involved, this can be an onerous task to do manually. An alternative is to compute a summary statistic or moment for each set of parameters that is invariant across models and assess convergence of that measure both within and across models (Brooks and Giudici 2001). Assessing convergence of the model indicators can also be tricky since it is a discrete variable. One indicator of convergence would be that successive blocks of draws have very similar proportions of indicators for the various models or to use more formal tests of convergence such as the hypothesis tests suggested by Brooks et al. (2003).

References

Bajari, P., Hong, H., Ryan, S.P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica, 78*(5), 1529–1568.

Beresteanu, A., Molchanov, I., Molinari, F. (2011). Sharp identification regions in models with convex moment predictions. *Econometrica, 79*(6), 1785–1821.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica, 60*(4), 889–918.

Berry, S., & Tamer, E. (2007). Identification in models of oligopoly entry. In R. Blundell, W. Newey, T. Persson (Eds.), *Advances in economics and econometrics, ninth world congress* (Vol. 2). Cambridge University Press.

Bjorn, P.A., & Vuong, Q.H. (1984, unpublished). Simultaneous equation models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation. California Institute of Technology. http://economics.mit.edu/files/4432.

Bresnahan, T., & Reiss, P. (1990). Entry in monopoly markets. *Review of Economic Studies, 57*(4), 531–553.

Bresnahan, T., & Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy, 99*(5), 977–1009.

Brooks, S.P., & Giudici, P. (2001). Markov chain Monte Carlo convergence assessment via two-way ANOVA. *Journal of Computational and Graphical Statistics, 9*(2), 266–285.

Brooks, S.P., Giudici, P., Philippe, A. (2003). Nonparametric convergence assessment for MCMC model selection. *Journal of Computational and Graphical Statistics, 12*(1), 1–22.

Brooks, S.P., Giudici, P., Roberts, G.O. (2003). Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions. *Journal of the Royal Statistical Society, Series B, 85*(1), 3–55.

Census, U.S. (2011). Metropolitan and Micropolitan statistical areas. http://www.census.gov/population/www/metroareas/aboutmetro.html. Accessed 1 May 2011.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis hastings algorithm. *American Statistician, 49*(4), 327–335.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica, 77*(6), 1791–1828.

Ellickson, P.B., & Misra, S. (2011). Estimating discrete games. *Marketing Science, 30*(6), 997–1010.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika, 82*(4), 711–732.

Green, P.J., & Hastie, D.I. (2009). Reversible jump MCMC. Working Paper, University of London.

Hartmann, W.R. (2010). Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science, 29*(4), 585–601.

Holmes, T.J. (2011). The diffusion of wal-mart and economies of density. *Econometrica, 79*(1), 253–302.

Jia, P. (2008). What happens when walmart comes to town: an empirical analysis of the discount retailing industry. *Econometrica, 76*(6), 1263–1316.

Manuszak, M.D., & Moul, C.C. (2008). Prices and endogenous market structure in office supply superstores. *Journal of Industrial Economics, 56*(1), 94–112.

Mazzeo, M.J. (2002). Product choice and oligopoly market structure. *RAND Journal of Economics, 33*(2), 221–242.

Misra, S. (2012). Markov chain Monte Carlo for incomplete information discrete games. *Quantitative Marketing and Economics*. doi:10.1007/s11129-012-9128-5.

Narayan, V., Yang, S., Assael, H. (2006). Estimating the interdependence of television program viewership between spouses: a Bayesian simultaneous equation model. *Marketing Science, 25*(4), 336–249.

Pakes, A., Porter, J., Ho, K., Ishii, J. (2006). Moment inequalities and their application. Working Paper, Harvard University.

Roberts, G.O., & Rosenthal, J.S. (2001). Optimal scaling for various metropolis hastings algorithms. *Statistical Science, 16*(4), 351–367.

Roberts, G.O., & Rosenthal, J.S. (2006). Examples of adaptive MCMC. Working Paper, University of Toronto.

Roberts, G.O., & Tweedie, R.L. (1996). Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli, 2*(4), 341–363.

Rossi, P.E., McCulloch, R.E., Allenby, G.M. (1996). The value of purchase history date in target marketing. *Marketing Science, 15*(4), 321–340.

Soetevent, A.R., & Kooreman, P. (2007). A discrete choice model with social interactions with an application to high school teen behavior. *Journal of Applied Econometrics, 22*(3), 599–624.

Su, C.-L., & Judd, K.L. (forthcoming). Constrained optimization approaches to estimation of structural models. *Econometrica*. doi:10.2139/ssrn.1085394.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies, 70*(1), 147–165.

Zhu, T., Singh, V., Manuszak, M.D. (2009). Market structure and competition in the retail discount industry. *Journal of Marketing Research, 46*(4), 453–466.