Quant Mark Econ (2013) 11:117–153
DOI 10.1007/s11129-012-9128-5

Markov chain Monte Carlo for incomplete information discrete games

Sanjog Misra

Received: 3 March 2011 / Accepted: 6 September 2012 / Published online: 13 October 2012

Abstract This paper outlines a Bayesian approach to estimating discrete games of incomplete information. The MCMC routine proposed features two changes to the traditional Metropolis–Hastings algorithm to facilitate the estimation of games. First, we propose a new approach to sample equilibrium probabilities using a probabilistic equilibrium selection rule that allows for the evaluation of the parameter posterior. Second, we propose a differential evolution based MCMC sampler which is capable of handling the unwieldy posterior that only has support on the equilibrium manifold. We also present two applications to demonstrate the feasibility of our proposed methodology.

Keywords Discrete games · MCMC · Nested fixed point · Differential evolution

JEL Classification C11 · C7 · M3 · L2

1 Introduction

Firms and consumers make choices which are influenced by the choices made by others. In particular, the payoffs from discrete choices that such economic agents make often depend on the choices of other agents. For example, profits from entry depend on if competitors enter, a given supermarket's pricing strategy choice might depend on what its competition does. Individual choices

S. Misra (✉)
UCLA Anderson School of Management, 110 Westwood Plaza, B.409, Los Angeles, CA 90095-1481, USA
e-mail: Sanjog.misra@anderson.ucla.edu

are no different and one can think of a plethora of examples where peer effects have a significant impact on our choices.[1]

The estimation of discrete choice models with such strategic and social interactions is non-trivial. The key issue, is that the correspondence between parameters and the likelihood is multivalued. In other words, for any particular guess of the parameters there might exist multiple values of the choice probabilities. This gives rise to the need for a mechanism that prescribes which values are relevant for the construction of the likelihood. The researcher then has to define a rule that picks from this set of probabilities to construct the sample likelihood. This in itself would not be problematic, but the current knowledge base does not have methods available that can fully characterize this equilibrium set, at least not in a cost effective manner. This has turned researchers away from estimation methods that involve solving for fixed points since if we do not have a guarantee that we can find all equilibria any specification of the equilibrium selection rule will by construction be incomplete. The models we consider in this paper are described by discrete choice likelihoods where choice probabilities are defined as the outcome of such fixed point computations. Usually we will know that there an equilibrium exists but in many cases there many be more than one equilibria. As we described above this multiplicity problem complicates the issue of taking the model to data.

There are a few approaches that one might take to resolving this problem. The first is a Nested Fixed Point approach originally proposed by Rust (1987) in the for single agent dynamic discrete choice models. In the context of games, the idea is to construct an inner loop that searches for all possible equilibria for each parameter guess and then uses a pre-determined equilibrium selection rule to construct the likelihood. The outer loop searches over the parameter space to maximize the likelihood. While the nested fixed point approach works quite well in single agent settings (because of a unique fixed point), it is often infeasible in multi-player settings where multiplicity may arise. This is because there exist, to the best of the author's knowledge, no established analytical or computational methods available that are able to find all equilibria and guarantee the exact characterization of the equilibrium manifold. There have been a few papers that use the nested fixed point approach (see e.g. Seim 2006; Zhu and Singh 2009) with either the implicit or explicit assumption that there exists only one equilibria.

An alternative, and quite popular, set of methods are based on the idea of employing a pseudo-likelihood. This idea has been popularized by the pioneering work of Aguirregabiria and Mira (2007) and Bajari et al. (2007) in the context of dynamic games. These authors build on the estimators first developed by Hotz and Miller (1993) and quantify player beliefs by estimating reduced form representation of the equilibrium probabilities (also known as

[1]For a detailed review of such models the reader is refered to Ellickson and Misra (2011) and Draganska et al. (2008). In this paper we focus our attention to games of incomplete information. For a Bayesian treatment of games of complete information see Narayanan (2012).

the conditional choice probabilities or *CCPs*) directly. By recognizing that in equilibrium the probabilities will only be a function of observable state variables the researcher can in a first stage estimate these probabilities using flexible methods. In a subsequent stage these estimated probabilities are taken as given and 'plugged-in' into the structural model to estimate the relevant parameters. Implementations for static games have been proposed by Aguirregabiria (2004) and Bajari et al. (2007) and have been applied by Ellickson and Misra (2008) among others. The pseudo-likelihood approach gives consistent estimates of the structural parameters without having to solve the game, search for the various equilibria or implement a selection rule. Unfortunately, if the researcher wishes to engage in any policy experiments she is left with no option but to revisit the problem of solving the game. Further, these pseudo-likelihood approaches require large amounts of data and admitting unobservables into the framework is difficult.

Finally, there is the recent approach advocated by Judd and Su (2006) that uses a direct and exhaustive search over the parameters and equilibria space. The authors term their approach Mathematical Programming with Economic Constraints (MPEC) and argue for a constrained optimization approach to estimation. The recipe is fairly straightforward and entails maximizing the likelihood subject to the equilibrium constraints. In essence, the equilibrium probabilities are treated as parameters to be estimated along with the structural parameters. The implicit selection rule employed picks from the set of possible equilibria that one which is most likely (in terms of the likelihood).

In this paper we propose an alternative Bayesian methodology that samples the parameters and equilibria jointly from the appropriate posterior. Our approach combines elements from nested fixed point and stochastic search approaches to develop a new methodology. We propose a conditional prior over the unobserved equilibrium probabilities which facilitates sampling. This then allows us to construct an MCMC routine that stochastically explores the posterior while restraining its attention only to the set of points in the posterior that are consistent with the equilibria of the game. Our proposed algorithm uses a Differential Evolution based proposal step that explores the posterior surface efficiently without becoming 'stuck' in any given area. The method is particularly appealing because it solves the game as defined by the researcher at each iteration. In other words the likelihood is precisely defined, requires no approximations or first step estimators. Further, because of the fact that we are using a Bayesian approach, all inference is exact. Consequently, no expensive post-estimation bootstraps or asymptotic arguments are needed to compute 'confidence' intervals. The MCMC routine, provides the researcher with the posterior density of the parameters and chosen equilibria. Finally, the approach also facilitates naturally the implementation of counterfactuals and policy experiments. Given a particular set of parameters a modified sampler can be run that only samples the equilibrium probabilities.

The rest of this paper is organized as follows: We first discuss the context of our paper by showcasing the multiplicity problem. In the next section we present our proposed algorithm and provide a discussion of how and why

it works with examples. Following that, use our sampler in the context of two "real" research problems and conclude with a discussion of directions for future research.

2 The setup: Static discrete games

In what follows, players are indexed by i and take actions $a_i \in \{0, 1, ..., K\}$ with $a = (a_1, ..., a_n)$ denoting the collection of actions across all n players. The state vector for player i is denoted s_i and $s = (s_1, ..., s_n)$ denotes the set of states across players. We will assume that the utility player i derives from choosing a given action $a_i (= k)$ is additively separable in the private information component $\varepsilon_i (k)$ and can be written

$$u(\mathbf{a}, \mathbf{s}, \varepsilon_i) = \tilde{v}_k(\mathbf{a}, s_i; \theta) + \varepsilon_i(k). \tag{1}$$

If we assume that $F(\varepsilon)$ is the Extreme Value density and is i.i.d. across players and actions the probability of choosing action k can be denoted by the set of fixed point equations,

$$\mathbf{p}_i^e = \Gamma_i\left(a_i = k | \mathbf{p}^e, s_i; \theta\right) \equiv \frac{\exp\left(v_k\left(\mathbf{p}^e, s_i; \theta\right)\right)}{\sum_{k' \in K} \exp\left(v_{k'}\left(\mathbf{p}^e, s_i; \theta\right)\right)} \tag{2}$$

where $v_k\left(\mathbf{p}^e, s_i; \theta\right) = E_{\varepsilon(\mathbf{a}_{-i})}\left[\tilde{v}_k(\mathbf{a}, s_i; \theta) | \mathbf{s}\right]$. Any vector of choice probabilities across players $\mathbf{p}^e \in \mathbb{P}$ is called an equilibrium if, for a given state vector $s \in \mathbb{S}$ and for parameter vector $\theta \in \Theta$, it satisfies the above fixed point equations for all players $i = 1..N$.[2] We then say that $\left\{\mathbf{p}^e, \theta, s\right\} \in \mathbb{P} \times \Theta \times \mathbb{S}$ is an *equilibrium point* and set of equilibrium points denoted by $\mathcal{P}^e \subseteq \mathbb{P} \times \Theta \times \mathbb{S}$ is termed the *equilibrium manifold* (Balasko 1988).

As with most static games of incomplete information, we make a number of standard assumptions. First, we assume that the ε are private information and i.i.d. across players and actions. In this paper we also assume both the ε and the actions of competitors $(\mathbf{a}_{-i})$ enter a player's payoff function (u_i) in an additive-separable manner. The assumption of i.i.d. errors is a fundamental aspect of the informational assumption of private information. The assumption that the ε are independent across actions is for convenience and a more general setup could allow for correlations. Further, as mentioned before, for purposes of convenience we will assume that the ε follow an Extreme Value distribution. Finally, players are assumed to move simultaneously given their beliefs. Given these assumptions on the primitives of the game each set of equilibrium probabilities characterize a Bayes–Nash equilibrium of the game.

[2] The probability space $\mathbb{P}$ is the product of N $K-$unit cubes and is defined as $\mathbb{P} = \prod_{i=1}^{N} \mathbb{P}^K$, where

$$\mathbb{P}^K = \left\{(p_1, ..., p_K) \in \mathbb{R}^K | \Sigma_{k=1}^K p_k = 1 \text{ and } p_k \geq 0 \text{ for all } k\right\}$$

$\mathbb{S}$ and Θ are the state and parameter spaces respectively.

2.1 An illustrative example

To illustrate the context and the nuances of the issues we deal with we start with a simple example. Consider a stylized coordination game with two players ($i = A$ or B) each of whom can pick one of two possible actions ($k = 1$ or 2). Let the payoffs be defined as

$$\begin{aligned} \tilde{u}_{A1} &= \theta_0 + \theta_1 s_A + \delta Y_{B1} + \varepsilon_{A1} \\ \tilde{u}_{A2} &= 0 + \varepsilon_{A2} \end{aligned} \tag{3}$$

Recall that the ε are private information so players make have to make decisions based on expected payoffs

$$\begin{aligned} u_{A1} &= E_{\varepsilon_B}(\tilde{u}_{A1}) = \theta_0 + \theta_1 s_A + \delta P_{B1} + \varepsilon_{A1}. \\ u_{A2} &= E_{\varepsilon_B}(\tilde{u}_{A2}) = 0 + \varepsilon_{A2} \end{aligned} \tag{4}$$

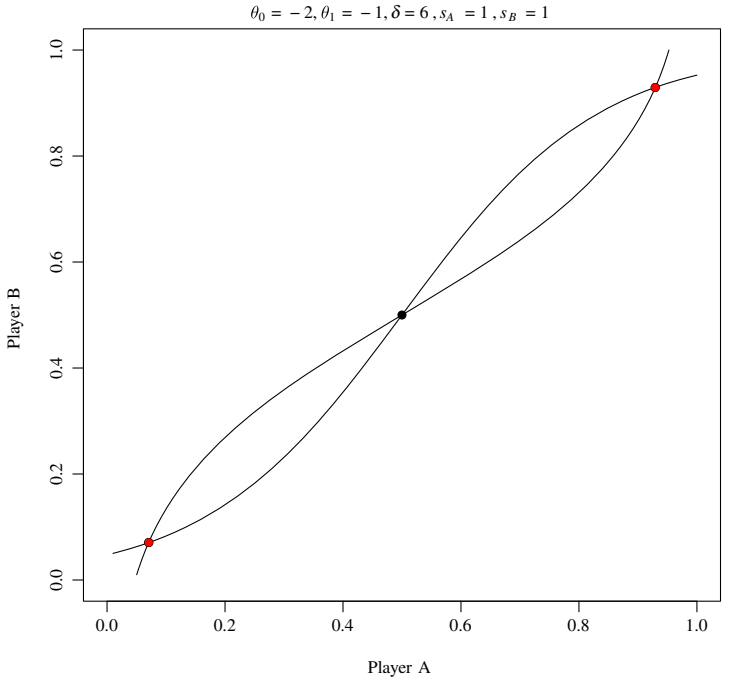
Expected payoffs for player B can be constructed analogously. Assuming Extreme Value (Type I) error gives us the familiar Logit form and equilibrium probabilities are then described by,

$$\begin{aligned} P^*_{A1} &= \frac{\exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{B1}\right)}{1 + \exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{B1}\right)} \\ P^*_{B1} &= \frac{\exp\left(\theta_0 + \theta_1 s_B + \delta P^*_{A1}\right)}{1 + \exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{A1}\right)}. \end{aligned} \tag{5}$$

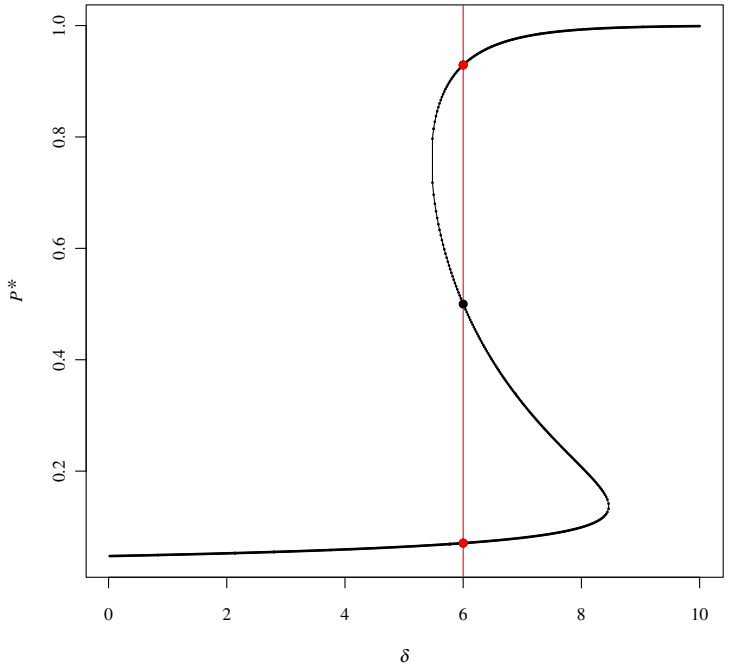
With $P^*_{A2} = 1 - P^*_{A1}$ and $P^*_{B2} = 1 - P^*_{B1}$. Since both sides of the above equations are confined to lie between 0 and 1, we know from Brouwer's fixed point theorem that an equilibrium exists. That is, the equations will have at least one point of intersection. However, there is no guarantee that this will be the only point of intersection and consequently we may be faced with the possibility of multiple equilibria. For example assume that $s_A = s_B = 1, \theta_0 = -2, \theta_1 = -1$ and $\delta = 6$. The reaction curves of the two players are shown in panel (a) of Fig. 1. These particular parameters generate three possible equilibria,

$$\begin{aligned} e_1 &: \left\{P^*_{A1} = 0.9293,\ P^*_{B1} = 0.9293\right\} \\ e_2 &: \left\{P^*_{A1} = 0.5000,\ P^*_{B1} = 0.5000\right\} \\ e_3 &: \left\{P^*_{A1} = 0.0707,\ P^*_{B1} = 0.0707\right\} \end{aligned} \tag{6}$$

There are depicted by the points of intersection of the reaction curves. Two of these equilibria (e_1, e_3) are stable (red dots) while e_2 is not. That is, a best response approach starting even in close proximity to e_2 would move away from it. In general, multiplicity can occur in a number of situations and for a variety of parameters. The second panel (Fig. 1b) plots for the set of equilibria for games where the strategic interaction parameter $\delta \in [2, 10]$. The three equilibria corresponding to our example lie on the vertical line with $\delta = 6$. In this setting, for the assumed values of s, and θ, multiple equilibria



(a) Reaction Functions



(b) The Equilibrium Manifolds

Fig. 1 Equilibria for a simple discrete game

exist for the range $\delta \in [5.4641, 8.4644]$. Outside this range there is a unique equilibrium in the game. Keep in mind that this plot simply depicts a slice of the equilibrium manifold since we are only allowing δ to vary. The nature of the reaction functions and the resultant set of equilibria will change as the value of the parameters and state variables change.

Figures 2 and 3 show some example of reaction curves and equilibria for the game described above but with different values for s, θ and δ. The slopes of the reaction functions are governed by the strategic interaction parameter δ. The first set of plots (Fig. 2) have positive strategic interaction parameters

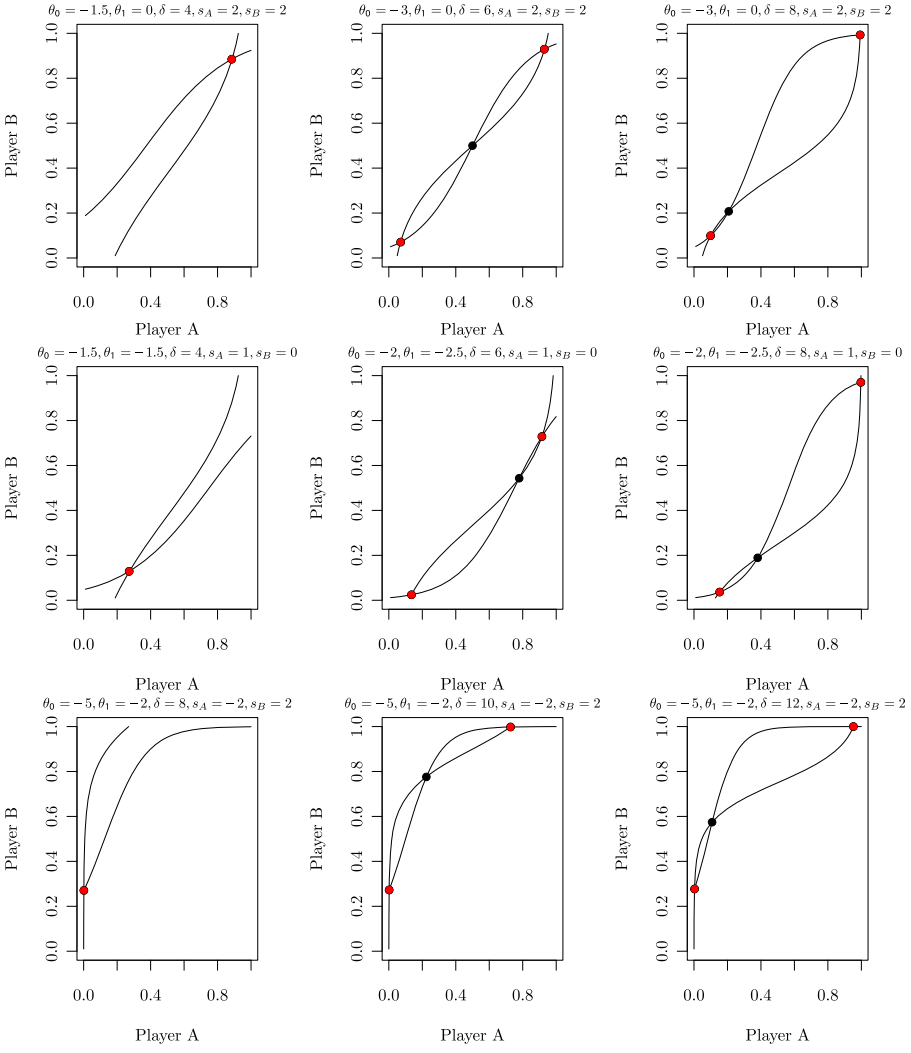


Fig. 2 Example of games with $\delta > 0$

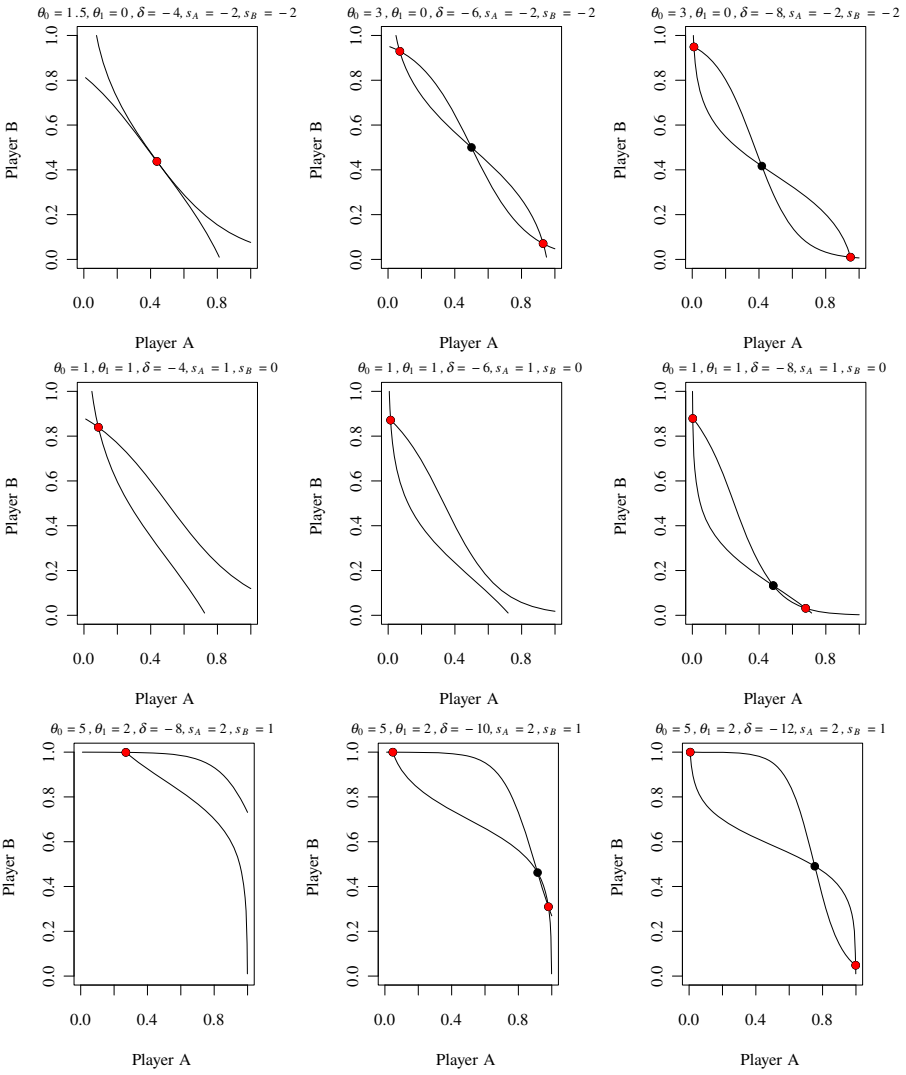


Fig. 3 Example of games with $\delta < 0$

(i.e. $\delta > 0$) while the second set (Fig. 3) have the opposite ($\delta < 0$). Further, in most cases δ increases in absolute value from left to right.

The prevalence of multiplicity and the corresponding manifold it generates creates problems for estimation. The researcher does not know the following:

(i) The true value of the parameters,
(ii) If there are multiple equilibria at the true parameter values and
(iii) If there happen to be multiple equilibria which particular subset of equilibria were "played" to generate the observed choice data.

A complete econometric model requires the explicit addressing of *all* three of these issues. Most empirical work (see e.g. Bajari et al. 2010; Sweeting 2009) assumes (as we do) that there is a single equilibrium played in any given market. This helps alleviate problems related to the third point, albeit not completely. It simply assumes away the possibility that there might be any sort of mixing or randomization over equilibria within a market. Addressing second issue in a comprehensive manner involves the complete characterization of the equilibrium manifold. This is a difficult task and while there are ongoing attempts we know of no established approach that accomplishes this task. In our estimation approach we will work around this problem by exploring the equilibrium manifold using MCMC methods. This approach will ultimately allow us to obtain estimates of the true parameters as well.

3 MCMC for discrete games

3.1 Constructing the posterior

In order to set up the posterior we need to introduce notation related to data. We will assume that the researcher has data across M independent markets. Define $y_m = \{y_{mi}(a_i)\}_{i \in N_m}$ as a vector of indicators of actions chosen by the N_m players in market m. Further, let the set $\mathcal{P}_m^e(\theta)$ contain all equilibria supported by θ and $\mathbf{s}_m$ with elements $\mathbf{p}_m^e$. Further, let the function $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$ denote the probability that equilibrium vector $\mathbf{p}_m^e$ was chosen from $\mathcal{P}_m^e(\theta)$ in market m. By construction each $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) \geq 0$ and $\sum_{\mathbf{p}_m^e \in \mathcal{P}_m^e} \pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) = 1$. The likelihood contribution from market m can then be written as,

$$\mathcal{L}_m\left(\mathbf{y}_m|\theta, \vartheta, \mathbf{s}_m\right) = \sum_{\mathbf{p}_m^e \in \mathcal{P}_m^e} \pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) \left[\prod_{i \in N_m} \prod_{a_i \in K} p_{mi}^e(a_i)^{y_{mi}(a_i)}\right]. \tag{7}$$

Adding priors $\pi_\theta(\theta)$ and $\pi_\vartheta(\vartheta)$ allows us to write down the posterior for θ and ϑ

$$\pi\left(\theta, \vartheta|\mathbf{y}, \mathbf{s}\right) \propto \prod_m \mathcal{L}_m\left(\mathbf{y}_m|\theta, \vartheta, \mathbf{s}_m\right) \pi_\theta(\theta) \pi_\vartheta(\vartheta). \tag{8}$$

Alternatively, we could work with the *joint* posterior of θ, ϑ and $\mathbf{P}^e = \left\{\mathbf{p}_m^e\right\}_{m=1}^M$,

$$\pi\left(\theta, \vartheta, \mathbf{P}^e|\mathbf{y}, \mathbf{s}\right) \propto \prod_{m \in M} \left[\prod_{i \in N_m} \prod_{a_i \in K} p_{mi}^e(a_i)^{y_{mi}(a_i)}\right] \pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) \pi_\theta(\theta) \pi_\vartheta(\vartheta) \tag{9}$$

Note that while the equilibrium set $\mathcal{P}_m^e(\theta)$ is a deterministic function of the parameters and state variables the chosen equilibrium $\left(\mathbf{p}_m^e\right)$ is not. Rather, it

is chosen from the set based on some mechanism. Since this mechanism is unobserved, the researcher needs to specify a distribution that stochastically picks an equilibrium from $\mathcal{P}_m^e(\theta)$ and captures the associated uncertainty. Consequently, we can interpret $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$ as a "prior" (or a mixing distribution) instead of as an equilibrium selection rule. In the Bayesian paradigm these distinctions are in any case irrelevant and result in the same estimator. The semantic change does, however, facilitate the construction of the MCMC routine in what follows. We note here that the prior is over which equilibrium gets played not how it was arrived at since that is completely determined by the structure of the game. The reader might find it odd that the "prior" conditions on other parameters (θ). This is perfectly fine from a theoretical perspective since we can always define

$$\pi\left(\mathbf{p}_m^e, \vartheta, \theta\right) = \pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)\pi_\theta(\theta)\,\pi_\vartheta(\vartheta)\,. \tag{10}$$

as a joint prior. Also note that the fact that the equilibria are a function of the state variables and consequently enter the prior is also not a problem as long as s is strictly exogenous. Note that for now we are working with the assumption that the researcher has the ability to specify $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$ precisely, which as we have already noted may be infeasible. We will discuss relaxing this requirement later.

3.1.1 MCMC along the equilibrium manifold

If we assume that $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e\right)$ can be explicitly evaluated then the construction of an MCMC routine that samples from the appropriate posterior is straightforward. The key here is to recognize that the equilibrium probabilities are conditioned on θ and ϑ and we cannot separate them. The idea, rather, is to sample θ, ϑ and $\mathbf{p}^e$ in tandem. One example of such an algorithm is presented below.

Algorithm 1 outlines a basic Metropolis–Hastings routine that samples parameters and equilibrium probabilities together. There are three key steps in the routine. First, we draw a candidate parameter vectors θ' and ϑ' from some proposal densities $q_\theta\left(\theta|\theta^{(t)}\right)$ and $q_\vartheta\left(\vartheta|\vartheta^{(t)}\right)$. In the next step, we condition on these values and draw a candidate equilibrium vector from the proposal density $q_p\left(\mathbf{p}_m^e|\mathbf{p}_m^{(t)}\right)$ in each market. Finally, we accept these candidates with probability α_{M-H}.

There are two major problems with this routine: First, it requires us to write down an analytical expression for the complete posterior, which may be unavailable. In other words, unless the researcher has the ability to characterize the equilibrium manifold Algorithm 1 is infeasible. The second problem is that, even if we could evaluate the posterior the MCMC routine described above would mix very slowly. To see this note that unless the proposal $q_p\left(\mathbf{p}_m^e|\mathbf{p}_m^{(t)}\right)$ places mass only on the equilibrium manifold the posterior will be zero and the chain will not move. We thus need a proposal that enforces the structure of the game. One possibility is to set $q_p = \pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e\right)$ which would generate equilibria that are consistent with the game and priors about the manner in

Algorithm 1 Basic Metropolis–Hastings

At iteration $t+1$

Parameter Step:

Draw θ' from the candidate density $q_\theta\left(\theta|\theta^{(t)}\right)$

Draw ϑ' from the candidate density $q_\vartheta\left(\vartheta|\vartheta^{(t)}\right)$

Manifold Step

For each market m draw $\mathbf{p}'_m$ from some $q_p\left(\mathbf{p}^e_m|\mathbf{p}^{(t)}_m\right)$

Metropolis–Hastings Accept-Reject Step:

Jointly accept the vector $\left\{\theta', \vartheta', \mathbf{P}' = \left\{\mathbf{p}'_m\right\}_{m=1}^M\right\}$ with probability,

$$\alpha_{M-H} = \min\left\{\frac{\pi\left(\theta',\vartheta',\mathbf{P}'|\mathbf{y},\mathbf{s}\right)}{\pi\left(\theta^{(t)},\vartheta^{(t)},\mathbf{P}^{(t)}|\mathbf{y},\mathbf{s}\right)}\frac{q_\theta\left(\theta^{(t)}|\theta'\right)q_\vartheta\left(\vartheta|\vartheta^{(t)}\right)\prod_m q_p\left(\mathbf{p}^{(t)}_m|\mathbf{p}'_m\right)}{q_\theta\left(\theta'|\theta^t\right)q_\vartheta\left(\vartheta'|\vartheta^{(t)}\right)\prod_m q_p\left(\mathbf{p}'_m|\mathbf{p}^{(t)}_m\right)}, 1\right\}.$$

and update $\left\{\theta^{(t+1)}, \vartheta^{(t+1)}, \mathbf{P}^{(t+1)}\right\}$ accordingly.

which equilibria are selected. This modified MCMC algorithm is presented as Algorithm 2.

In Algorithm 2, each p^e_{mi} is an equilibrium probability drawn from the set $\mathcal{P}^e_m(\theta')$ with density proportional to $\pi_p\left(\mathbf{p}^e_m; \vartheta', \mathcal{P}^e_m(\theta')\right)$. In other words the MCMC chain only moves along the equilibrium manifold. Note that the modified Metropolis–Hastings acceptance probability $\tilde{\alpha}_{M-H}$ no longer

Algorithm 2 Modified Metropolis–Hastings

At iteration $t+1$

Parameter Step:

Draw θ' from the candidate density $q_\theta\left(\theta|\theta^{(t)}\right)$

Draw ϑ' from the candidate density $q_\vartheta\left(\vartheta|\vartheta^{(t)}\right)$

Manifold Step

For each market m draw $\mathbf{p}'_m$ from $\pi_p\left(\mathbf{p}^e_m; \vartheta', \mathcal{P}^e_m\left(\theta'\right)\right)$

Metropolis–Hastings Accept-Reject Step:

Jointly accept the vector $\left\{\theta', \vartheta', \mathbf{P}' = \left\{\mathbf{p}'_m\right\}_{m=1}^M\right\}$ with probability,

$$\tilde{\alpha}_{M-H} = \min\left\{\frac{\left[\prod_{m\in M, i\in N_m, a_i\in K}\left[p'_{mi}\right]^{y_{mi}(a_i)}\right]\pi_\theta\left(\theta'\right)\pi_\vartheta\left(\vartheta'\right)}{\left[\prod_{m\in M, i\in N_m, a_i\in K}\left[p^{(t)}_{mi}\right]^{y_{mi}(a_i)}\right]\pi_\theta\left(\theta^{(t)}\right)\pi_\vartheta\left(\vartheta^{(t)}\right)}\frac{q_\theta\left(\theta^{(t)}|\theta'\right)q_\vartheta\left(\vartheta|\vartheta^{(t)}\right)}{q_\theta\left(\theta'|\theta^t\right)q_\vartheta\left(\vartheta'|\vartheta^{(t)}\right)}, 1\right\}.$$

and update $\left\{\theta^{(t+1)}, \vartheta^{(t+1)}, \mathbf{P}^{(t+1)}\right\}$ accordingly.

contains $\pi_p\left(\mathbf{p}_m^e; \vartheta', \mathcal{P}_m^e(\theta)\right)$ since the proposal density cancels with terms in the posterior. In fact, all components in $\tilde{\alpha}_{M-H}$ are known to the researcher. What remains, now, is to construct a routine that allows us to draw from $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$ and specify proposal densities q_θ and q_ϑ.

3.2 Priors over equilibria

3.2.1 Specifying $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$

Recall that the task at hand is the construction of a prior that reasonably reflects the a priori beliefs of the researcher about the equilibrium that would be picked. We first note that the prior for $\mathbf{p}_m^e$ in most cases will be of the following form,

$$\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) \propto \pi\left(\mathbf{p}_m^e | \vartheta, \tilde{\mathcal{P}}_m\right) \pi\left(\tilde{\mathcal{P}}_m^e | \mathcal{P}_m^e(\theta)\right) \tag{11}$$

where $\tilde{\mathcal{P}}_m^e \subseteq \mathcal{P}_m^e(\theta)$ and $\mathcal{P}_m^e(\theta)$ is a set of all possible equilibrium points supported by (θ, s_m). There are various reasons why this particular specification will be needed. In almost any empirical application the researcher will need to place restrictions on the type of equilibria being considered. For example, one may place positive mass only on those equilibria that are consistent with some other pre-specified rule such as best-response stability (e.g. Sweeting 2009).

The function $\pi\left(\mathbf{p}_m^e | \vartheta, \tilde{\mathcal{P}}_m\right)$ is also specified by the researcher and assigns probability to each equilibrium in the admissible set $\tilde{\mathcal{P}}_m^e$. While there are many specifications one could choose, we need one that facilitates sampling. We follow Sweeting (2009) who adopts a Multinomial Logit specification and prescribe the following form

$$\pi\left(\mathbf{p}_m^e | \vartheta, \tilde{\mathcal{P}}_m\right) = \frac{\exp\left(\vartheta' \Psi\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)}{\sum_{\tilde{\mathbf{p}}_m \in \tilde{\mathcal{P}}_m^e} \exp\left(\vartheta' \Psi\left(\tilde{\mathbf{p}}_m, \mathbf{s}\right)\right)}. \tag{12}$$

In the above specification Ψ are a set of functions chosen by the researcher and ϑ are parameters to be estimated. For example, in an entry game we could specify $\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right) = \sum_{i \in N_m} \left(\mathbf{p}_{im}^e\right)^2$, which would pick equilibria with higher Herfindahl scores. Alternatively, we could use $\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right) = \sum_{i \in N_m} \mathbf{p}_{im}^e$, which weighs those equilibria where more players are likely to enter more heavily. Sweeting (2009) makes Ψ a function of market characteristics which is another alternative. Clearly, if $\tilde{\mathcal{P}}_m^e$ is a singleton then obviously $\pi\left(\mathbf{p}_m^e | \vartheta, \tilde{\mathcal{P}}_m\right) = 1$.

To be clear, we are not specifying a rule for how the equilibrium was generated. That aspect is completely governed by the primitive assumptions placed on the structure of the game. Unlike complete information games we cannot use alternative structures as equilibrium selection rules. For example, a Berry (1992) type rule that imposes sequence of moves based on some criterion (e.g. firms with higher profits move first) will generate equilibria that are inconsistent with those in a simultaneous move game. This is because unlike complete information games the equilibria is depicted by probabilities not actions. In a sense we are simply imposing a prior on which of the equilibria

generated by the game was played in the data generating process. The rule should be seen as agnostic about the process by which the equilibrium was selected and serves only as an approximation. As Sweeting (2009) points out this specification can be thought of as a randomization device or as the proportion of markets playing the particular equilibrium profile. From the point of view of the researcher, it simply reflects the prior belief that a particular equilibrium is played in the data.

Evaluating $\pi\left(\mathbf{p}_m^e|\vartheta, \tilde{\mathcal{P}}_m\right)$ explicitly is possible only if we could enumerate $\tilde{\mathcal{P}}_m$. To our knowledge there is no currently established algorithm that accomplishes this objective.[3] Having said that, there are indeed applications where this might be feasible and if so the likelihood can be directly evaluated. For example, Sweeting (2009) starts his search from the extreme vertices of the $N \times K$-unit cube and uses a best response iteration to find all equilibria. He shows that this approach finds all the equilibria in his game. Our objective is to propose a method that works in the more general case there such enumeration may not be possible. Below we show that we can sample from Eq. 11 even though we cannot evaluate it.

3.2.2 Sampling from $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$

To facilitate our MCMC routine we need to be able to draw equilibria from our choice of $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$. To do this we specify $\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right)$ as follows,

$$\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) = \frac{\exp\left(\vartheta'\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right) + \ln \mathcal{D}\left(\mathbf{p}_m^e\right)\right)}{\sum_{\tilde{\mathbf{p}}_m \in \tilde{\mathcal{P}}_m^e} \exp\left(\vartheta'\Psi\left(\tilde{\mathbf{p}}_m, \mathbf{s}\right) + \ln \mathcal{D}\left(\mathbf{p}_m\right)\right)}. \tag{13}$$

where $\mathcal{D}\left(\mathbf{p}_m^e\right)$ is the domain of attraction of the fixed point $\mathbf{p}_m^e$. Note that by definition each $\mathcal{D}\left(\mathbf{p}_m^e\right) \geq 0$ since every stable fixed point must have a domain of attraction and, $\bigcup_{\mathbf{p}_m^e \in \tilde{\mathcal{P}}_m^e} \mathcal{D}\left(\mathbf{p}_m^e\right) = \mathbb{P}$ since $\mathcal{D}\left(\mathbf{p}_m^e\right)$ are non-overlapping and every point in $\mathbb{P}$ must be in the domain of attraction of some fixed point. Our assumption implicitly imposes a particular specification on $\pi\left(\tilde{\mathcal{P}}_m^e|\mathcal{P}_m^e(\theta)\right)$ where we place mass only on best-response stable equilibria.[4] Second our specification is chosen to facilitate sampling from the equilibrium manifold without fully characterizing it. We document this sampling approach in Algorithm 3.

In the above routine, $\mathbf{I}_{\vartheta \leq 0}$ is an indicator vector with ones where the subscripted condition is satisfied and zero otherwise, $\circ$ denotes the Hadamard (element-by-element) product and $\underline{\Psi}$ and $\overline{\Psi}$ denote the basis vectors with the minimal and maximal values of the elements. Finally, $\mathbf{\Gamma}$ is the best response

[3] There are ongoing attempts to find all equilibria using Homotopy methods (Bajari et al. 2011) that appear promising albeit computationally expensive.

[4] This is not required. We could just as easily alter the definition of $\mathcal{D}\left(p_m^e\right)$ to be the Newtonian basin of attraction and use Newton's root solving method to generate candidate equilibria. As long as every equilibria can be generated with positive probability our MCMC routine will converge.

Algorithm 3 Sampling from the prior

Draw ϱ_m^0 from $\mathbf{U}_{N_m}[0,1]$
Iterate Best Responses(until convergence)
$$\varrho_m^{\kappa+1} = \mathbf{\Gamma}\left(\varrho_m^{\kappa}; s, \theta\right)$$
Take converged values $\tilde{\mathbf{p}}_m^e = \varrho_m^{\kappa^*}$ as a draw from $\mathcal{D}\left(\mathbf{p}_m^e\right)$
Define $\bar{V}_m = \exp\left(\vartheta'\left(\mathbf{I}_{\vartheta\leq 0} \circ \underline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right) + \mathbf{I}_{\vartheta>0} \circ \overline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)\right)$
Draw $v = U\left[0, \bar{V}_m\right]$
Accept $\tilde{\mathbf{p}}_m^e$ if $\vartheta'\Psi\left(\tilde{\mathbf{p}}_m^e, \mathbf{s}\right) \geq \ln v$
else restart procedure.

operator defined earlier (i.e. the collection of choice probabilities denoted by Eq. 2).[5]

An explanation of the algorithm is in order. First we note that

$$\pi_p\left(\mathbf{p}_m^e; \vartheta, \mathcal{P}_m^e(\theta)\right) \propto \exp\left(\vartheta'\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right)\right) \times \mathcal{D}\left(\mathbf{p}_m^e\right) \tag{14}$$

So we simply need to draw equilibria proportional to the weights described by the RHS of the above equation. The algorithm essentially proceeds as follows: First it samples a candidate equilibrium from $\tilde{\mathcal{P}}_m(\theta)$ with probabilities proportional to $\mathcal{D}\left(\mathbf{p}_m^e\right)$. By definition, starting from a random point on the $N_m \times K$-unit cube and iterating using a best-response type iteration will generate a candidate equilibria with probability proportional to its domain of attraction. Once a candidate is available the algorithm uses an accept-reject rule to obtain a draw.

To implement the accept-reject step we have to construct an envelope probability for $\exp\left(\vartheta'\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)$. The trick is to realize that given the parameters and the data the maximum possible value of $\exp\left(\vartheta'\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)$ can be computed quite easily. For example if $\Psi\left(\mathbf{p}_m^e, \mathbf{s}\right) = \sum_{i\in N_m} \mathbf{p}_{im}^e$ then we know that $\overline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right) = N_m$ and $\underline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right) = 0$. Now if $\vartheta > 0$ then $\bar{V}_m = \exp(\vartheta N_m)$ and alternatively if $\vartheta \leq 0$ then $\bar{V}_m = \exp(0) = 1$. The rule

$$\bar{V}_m = \exp\left(\vartheta'\left(\mathbf{I}_{\vartheta\leq 0} \circ \underline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right) + \mathbf{I}_{\vartheta>0} \circ \overline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)\right) \tag{15}$$

simply implements this more generally. Once we have $\bar{V}_m$ a simple accept-reject rule obtains the draw. The rationale behind the last bit is that rejection sampling from a multinomial distribution does not require the probabilities but only the relative weights. As long as we can envelope these weights then by

[5] One may opt to use other more efficient iterative methods to speed up convergence here. In our applications we use Mann iterations which seem to work well. Regular Picard iterations also gave essentially identical results but in some cases took a long time to converge. The reader is referred to the excellent book by Berinde (2007) for a review of Mann, Picard and other fixed point iteration methods and their properties.

drawing a given index (in our case an equilibrium) and then using the accept-reject rule as prescribed provides accurate draws.[6]

A special case of our specification is when $\vartheta = 0$. In this case the algorithm integrates over the equilibria according to their domains of attractions. In general, identifying parameters related to the equilibrium selection is a difficult task and using this simplified specification might be a useful strategy. We will discuss this in a later section.

3.2.3 Informativeness of the chosen prior

While our proposed prior might seem somewhat arbitrary, it is not completely unreasonable. Clearly we would prefer a uniform prior over equilibria since it is as close to uninformative as we could be. However, such a prior is infeasible in general since it requires knowledge of the number of equilibria or requires the sampler to find all equilibria at each iteration. To ascertain the degree to which our choice of prior influences our inferences we present a simple but stylized example. Consider a case with asymmetric firms $\theta = (-3, 1, 6)$ and $x = (-1.75, 2)$ with payoffs and game-structure described in Section 2.1. In this game there are two stable equilibria $\{(0.0687, 0.3571), (0.7442, 0.9697)\}$. We will assume that $\vartheta = 0$ so that the domains of attraction effectively impose a "prior" with weights (0.5835, 0.4165) on the two equilibria. We assume that we have a single market. The table below depicts the posterior weights on each equilibria conditional on the single observation.

Observation	Posterior Weights	
	Equilibrium #1 (0.0687,0.3571)	Equilibrium #2 (0.7442,0.9697)
(0, 0)	0.8158	0.1842
(0, 1)	0.1216	0.8784
(1, 0)	0.7327	0.2673
(1, 1)	0.0452	0.9546
Prior	0.5835	0.4165

Clearly, even with a single observation, the data on actions chosen are informative about the equilibrium picked. Given a choice the algorithm will tend to pick those equilibria that are most consistent with the data observed. So, if we see $a = (0, 0)$ the posterior weights tend to weigh quite heavily on equilibrium #1 and similarly $a = (1, 1)$ tends to favor equilibrium #2. Consider

[6]Suppose we wanted to draw from a multinomial with probability $\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$. All we would need are weights $\left(\text{say } w = (50, 25, 25)\right)$ that are proportional to π and the knowledge that the maximum weight is bounded above by some value (say $\bar{V} = 100$). Then by drawing an index uniformly (j) and accepting the draw only if $w_j > U\left[0, \bar{V}\right]$ we will get draws exactly proportional to the requisite probabilities. In our case the indices are the equilibria that are generated non-uniformly (based on their domain of attraction). An R version of this algorithm for multinomial draws is available from the author upon request.

a more extreme case where our prior happens to place a measly 10 % weight on the high equilibrium and the data happen to be $a = (1, 1)$. Even in this extreme case the posterior probability of the high equilibrium being chosen turns out to be around 0.77. Further reducing the prior weight to a mere 5 % weight brings the posterior down to 0.61 but here again the larger weight is placed on the equilibrium #2. The main takeaway from this exercise it that the posterior weights tend to be informed primarily by the data. Of course, if the prior has the additional flexibility of free parameters (ϑ) then the impact of the imposed part of the prior is further reduced.

3.3 A general differential evolution MCMC routine

To complete our algorithm we need to adopt particular specifications for q_θ and q_ϑ. Typical choices include Random–Walk specifications or Independence samplers with normal errors. In the context of discrete games and the nature of the posterior these approaches will be inefficient since the posterior is only supported on the equilibrium manifold. In a multi-parameter setting the posterior is bound to be quite unwieldy with numerous modes and ridges, sharp turns, falls and peaks. It is well known that standard Metropolis–Hastings routines have a tendency to get "stuck" and find it hard to traverse "bumpy" posteriors efficiently. Thus, even armed with the sampling scheme described earlier exploring the posterior in real world problems with many dimensions is difficult. To deal with this we propose a perturbed Differential-Evolution MCMC sampler. This algorithm is described below.

There are two key improvements that this algorithm offer over the one proposed earlier. First, this algorithm is a population based MCMC and runs D chains in parallel. This allows for a better exploration of the posterior. Second, each chain d borrows information from other chains as the routine progresses. This interpolation happens in the parameter step where we have replaced the usual multivariate normal proposal q_θ with a Differential Evolution (DE) proposal. Differential Evolution (Storn and Price 1997) is a population based algorithm where randomly chosen elements of the population "mate" to create "offspring" that are superior. Our approach follows Braak (2006) and perturbs the original deterministic DE proposal with a symmetric shock (υ) to ensure that the draws have unbounded support. In particular, our implementation assumes that the υ are distributed standard normal. The parameters φ_1 and φ_2 are both positive and are held fixed during the runs.[7]

The DE step ensures that the routine explores the posterior more efficiently since the routine balances exploration and exploitation of the parameter space using information across chains. In particular, the proposal is self tuning and

[7] In our implementation we follow suggestions of Storn and Price (1997) and Braak (2006) to set $\varphi_1 = 2.38/\sqrt{2D}$ and $\varphi_1 = 10^{-4}$.

Algorithm 4 Differential evolution MCMC for discrete games

At iteration $t+1$

For each chain d

Parameter Step:

Draw θ'_d and ϑ'_d using a differential evolution step

$$\theta'_d = \theta_d^{(t)} + \varphi_1 \cdot \left[\left(\theta_{d'}^{(t)} - \theta_{d''}^{(t)}\right)\right] + \varphi_2 \upsilon_\theta$$

$$\vartheta'_d = \vartheta_d^{(t)} + \varphi_1 \cdot \left[\left(\vartheta_{d'}^{(t)} - \vartheta_{d''}^{(t)}\right)\right] + \varphi_2 \upsilon_\vartheta \quad \text{where } d \neq d' \neq d''$$

Manifold Step

For each market m

M0: Draw ϱ_m^0 from $\mathbf{U}_{N_m}[0,1]$

Iterate Best Responses (until convergence)

$$\boldsymbol{\varrho}_m^{\kappa+1} = \boldsymbol{\Gamma}\left(\boldsymbol{\varrho}_m^{\kappa}; s, \theta'\right)$$

Take converged values $\tilde{\mathbf{p}}_m^e = \boldsymbol{\varrho}_m^{\kappa^*}$ as a draw from $\mathcal{D}\left(\mathbf{p}_m^e\right)$

Define $\bar{V}_m = \exp\left(\vartheta'\left(\mathbf{I}_{\vartheta \leq 0} \circ \underline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right) + \mathbf{I}_{\vartheta > 0} \circ \overline{\Psi}\left(\mathbf{p}_m^e, \mathbf{s}\right)\right)\right)$

Draw $\upsilon = U\left[0, \bar{V}_m\right]$

Accept $\tilde{\mathbf{p}}_m^e$ if $\vartheta' \Psi\left(\tilde{\mathbf{p}}_m^e, \mathbf{s}\right) \geq \ln \upsilon$

else go to **M0**.

Metropolis–Hastings Accept-Reject Step:

Jointly accept the vector $\left\{\theta', \vartheta', \mathbf{P}' = \left\{\mathbf{p}'_m\right\}_{m=1}^M\right\}$ with probability,

$$\tilde{\alpha}_{M-H} = \min\left\{\frac{\left[\prod_{m \in M, i \in N_m, a_i \in K}\left[p'_{mi}\right]^{y_{mi}(a_i)}\right]\pi_\theta\left(\theta'\right)\pi_\vartheta\left(\vartheta'\right)}{\left[\prod_{m \in M, i \in N_m, a_i \in K}\left[p_{mi}^{(t)}\right]^{y_{mi}(a_i)}\right]\pi_\theta\left(\theta^{(t)}\right)\pi_\vartheta\left(\vartheta^{(t)}\right)}, 1\right\}.$$

and update $\left\{\theta^{(t+1)}, \vartheta^{(t+1)}, \mathbf{P}^{(t+1)}\right\}$ accordingly.

possesses a kind of "divergence property" (Storn and Price 1997). When the posterior is flat the distances between the chains ($[\theta_{d'} - \theta_{d''}]$) increase rapidly so as to explore the surface more efficiently. As we get to areas where the posterior has significant and dense mass the distances shrink the steps become shorter and the chains tend to coalesce. This is similar to other modified MCMC methods such as the Metropolis adjusted Langevin algorithm (MALA) of Roberts and Tweedie (1996) in which proposals are generated based on $\mu' = \mu^{(t)} + \frac{\sigma^2}{2}\nabla \ln \pi\left(\mu^{(t)}\right) + \sigma \upsilon$ and are a function of the gradient of the posterior evaluated at the current state. The idea is the step-size in the algorithm changes with $\nabla \ln \pi\left(\mu^{(t)}\right)$ by growing when it is large and shrinking when it is small. When the mode of the posterior is reached the gradient is zero. The DE proposals mimic this behavior, require no explicit calculation of the gradient yet provide the similar benefits. In the next section we discuss the properties of our proposed algorithm.

3.4 Properties of the algorithm

Our proposed algorithm is non standard and it is important that we document its properties. We show below that our basic scheme results in a reversible chain that satisfies detailed balance and results in draws from the true joint posterior. We then discuss the properties of the differential evolution MCMC routine and use results in Braak (2006) to describe why each of the multiple chains has the joint posterior as its unique invariant density.

3.4.1 Properties of the basic MCMC scheme (Algorithm 1)

In what follows we are deliberately more general than needed and will work with Algorithm 1 (Algorithm 2 and each chain of the DE-MCMC routine are special cases.)

To begin, let $q\left(\theta', \vartheta', \mathbf{P}'|\theta^{(t)}, \vartheta^{(t)}\right) = q\left(\mathbf{P}'|\theta', \vartheta'\right) q\left(\theta'|\theta^{(t)}\right) q\left(\vartheta'|\vartheta^{(t)}\right)$ and $\Upsilon^{(t)} = \left\{\theta^{(t)}, \vartheta^{(t)}, \mathbf{P}^{(t)}\right\}$ then from earlier we have,

$$\begin{aligned}\alpha\left(\Upsilon^{(t)}, \Upsilon'\right) &= \min\left(1, \frac{\pi\left(\theta', \vartheta', \mathbf{P}'|\mathbf{y}, \mathbf{s}\right) q\left(\theta^{(t)}, \vartheta^{(t)}, \mathbf{P}^{(\mathbf{t})}|\theta', \vartheta'\right)}{\pi\left(\theta^{(t)}, \vartheta^{(t)}, \mathbf{P}^{(t)}|\mathbf{y}, \mathbf{s}\right) q\left(\theta', \vartheta', \mathbf{P}'|\theta^{(t)}, \vartheta^{(t)}\right)}\right) \\ &= \min\left(1, \frac{\pi\left(\Upsilon'|\mathbf{y}, \mathbf{s}\right) q\left(\Upsilon^{(t)}|\theta', \vartheta'\right)}{\pi\left(\Upsilon^{(t)}|\mathbf{y}, \mathbf{s}\right) q\left(\Upsilon|\theta^{(t)}, \vartheta^{(t)}\right)}\right). \end{aligned} \tag{16}$$

Now $\Upsilon^{(t)}$ evolves as follows

$$\Upsilon^{(t+1)} = \begin{cases} \Upsilon' & \text{with probability } \alpha\left(\Upsilon^{(t)}, \Upsilon'\right) \\ \Upsilon^{(t)} & \text{with probability } 1 - \alpha\left(\Upsilon^{(t)}, \Upsilon'\right) \end{cases} \tag{17}$$

The Transition kernel is then defined as

$$\begin{aligned}\mathcal{T}\left(\Upsilon^{(t)}, \Upsilon^{(t+1)}\right) &= \rho\left(\Upsilon^{(t)}, \Upsilon^{(t+1)}\right) q\left(\Upsilon^{(t+1)}|\Upsilon^{(t)}\right) \\ &\quad + \left(1 - \int\int \sum_{P'_e \in \mathcal{P}_{\theta'}} \rho\left(\Upsilon^{(t)}, \Upsilon'\right) q\left(\theta', \vartheta', \mathbf{P}'|\theta^{(t)}, \vartheta^{(t)}\right) d\theta' d\vartheta'\right) \\ &\quad \times \delta_{\Upsilon^{(t)}}\left(\Upsilon^{(t+1)}\right) \end{aligned} \tag{18}$$

where δ is Dirac's delta. This is the typical MH transition kernel augmented with the discrete distribution pertaining to the equilibria. Now, let $\pi^*\left(\Upsilon|\mathbf{y}, \mathbf{s}\right)$ denote the stationary distribution of the posterior. For detailed balance we need to show that

$$\pi^*\left(\Upsilon|\mathbf{y}, \mathbf{s}\right) \mathcal{T}\left(\Upsilon, \Upsilon'\right) = \pi^*\left(\Upsilon'|\mathbf{y}, \mathbf{s}\right) \mathcal{T}\left(\Upsilon', \Upsilon\right). \tag{19}$$

Examining $\pi^*(\Upsilon)\mathcal{T}\left(\Upsilon, \Upsilon'\right)$ we see that it is a sum for two terms. If each satisfies detailed balance then the sum does as well. We show these in turn First note that

$$\begin{aligned}\pi^*(\Upsilon|\mathbf{y},\mathbf{s})\rho(\Upsilon,\Upsilon')q(\Upsilon'|\Upsilon) &= \pi^*\left(\Upsilon|\mathbf{y},\mathbf{s}\right)\rho(\Upsilon,\Upsilon')q\left(\mathbf{P}'|\theta',\vartheta'\right)q\left(\theta'|\theta\right)q\left(\vartheta'|\vartheta\right)\\ &= \min\begin{pmatrix}\pi^*\left(\Upsilon'|\mathbf{y},\mathbf{s}\right)q\left(\mathbf{P}'|\theta',\vartheta'\right)q\left(\theta'|\theta\right)q\left(\vartheta'|\vartheta\right),\\ \pi^*\left(\Upsilon|\mathbf{y},\mathbf{s}\right)q\left(\mathbf{P}|\theta,\vartheta\right)q\left(\theta|\theta'\right)q\left(\vartheta|\vartheta'\right)\end{pmatrix}\\ &= \pi^*(\Upsilon'|\mathbf{y},\mathbf{s})\rho(\Upsilon',\Upsilon)q\left(\mathbf{P}|\theta,\vartheta\right)q\left(\theta|\theta'\right)q\left(\vartheta|\vartheta'\right).\end{aligned} \tag{20}$$

Similarly for the second part of the sum we have

$$\begin{aligned}&\pi^*\left(\Upsilon|\mathbf{y},\mathbf{s}\right)\left(1-\int\int\sum_{P'_e\in\mathcal{P}_{\theta'}}\rho\left(\Upsilon,\Upsilon'\right)q\left(\theta',\vartheta',\mathbf{P}'|\theta,\vartheta\right)d\theta' d\vartheta'\right)\delta_{\Upsilon}\left(\Upsilon'\right)\\ &= \pi^*\left(\Upsilon'|\mathbf{y},\mathbf{s}\right)\left(1-\int\int\sum_{P_e\in\mathcal{P}_{\theta}}\rho\left(\Upsilon',\Upsilon\right)q\left(\theta,\vartheta,\mathbf{P}|\theta',\vartheta'\right)d\theta d\vartheta\right)\delta_{\Upsilon'}\left(\Upsilon\right)\end{aligned} \tag{21}$$

since for any $\Upsilon \neq \Upsilon'$ we have $\delta_{\Upsilon}\left(\Upsilon'\right) = \delta_{\Upsilon}\left(\Upsilon'\right) = 0$. The equality is satisfied trivially for $\Upsilon = \Upsilon'$. This establishes detailed balance in the general case. Since we have assumed that $q\left(\mathbf{P}'|\theta',\vartheta'\right) = \pi\left(\mathbf{P}|\theta',\vartheta'\right)$ and symmetric proposals q_θ and q_ϑ we get some simplifications in the acceptance ratio but all the above detailed balance equations still continue to hold. Given detailed balance, we know that the invariant density of the chain will be $\pi^*\left(\Upsilon|\mathbf{y},\mathbf{s}\right)$ and that the chain is reversible (Robert and Casella Theorem 6.46, p. 230).

To prove that the generated Markov chain is ergodic it is only necessary that we show that it is irreducible and aperiodic. To prove irreducibility we need

$$q\left(\Upsilon'|\Upsilon\right) > 0 \text{ for all } \left\{\Upsilon,\Upsilon'\right\} \in \text{supp}\left(\pi\right). \tag{22}$$

That is, we need to show that the candidate generating function allows the every state to be reached with positive probability from any other state. Since the posterior only has support on the equilibrium manifold as defined by the stable equilibria of the game, this condition implies that q should allow the state to move to any point on the equilibrium manifold (conditional on the value of $\mathbf{s}$) from any other point on the manifold. Since q_θ and q_ϑ have support on the real line and we start the algorithm which defines $q\left(\mathbf{P}|\theta,\vartheta\right)$ from a uniform draw on the standard simplex this guarantees the irreducibility of the chain. To see this note that if we repeatedly drew equilibria with different starting points we would eventually find all stable equilibria. In other words since q_θ and q_ϑ can move to any point $\{\theta_0,\vartheta_0\}$ in the joint parameter space and we can reach any equilibrium supported by $\{\theta_0,\vartheta_0\}$ with positive probability all state communicate. This is a critical aspect of the algorithm.

If for example, we drew the starting points uniformly from the vertices of the simplex there would be no such guarantee since they may not collectively cover the domains of attraction of all equilibria. Finally, the chain is aperiodic since MH acceptance ratio ensures that the probability of remaining at any given state is strictly positive. Taken together, these conditions satisfy all conditions (see Robert and Casella, pp. 272–274) for ergodicity. Intuitively, the algorithm works because we are moving along the equilibrium manifold in a manner that allows us to retrace our steps and move to any point on the manifold in a single step.

3.4.2 Properties of the DE-MCMC algorithm

The Differential evolution based routine described in Algorithm 4 possesses all the properties of a valid MCMC chain. The intuition behind this is straightforward. The DE proposals are essentially a random walk and the manifold step can be seen as part of the evaluation of the posterior. When coupled with the Metropolis-Hastings rule standard convergence properties apply. In particular, each component DE-MCMC chain has the posterior distribution of $\mu = \{\theta, \vartheta\}$ as its unique stationary distribution. We discuss this claim below.

The DE proposal step is symmetric since the probability of any move from μ_d^t to μ_d^{t+1} is equal to the reverse step. In other words,

$$\begin{aligned} \mu_d^{(t)} &= \mu_d^{(t+1)} - \varphi_1 \cdot \left[\mu_{d'}^{(t)} - \mu_{d''}^{(t)}\right] - \varphi_2 \upsilon \\ &= \mu_d^{(t+1)} + \varphi_1 \cdot \left[\mu_{d''}^{(t)} - \mu_{d'}^{(t)}\right] + \varphi_2 \upsilon \end{aligned} \tag{23}$$

This follows from the fact that the pair $\left(\mu_{d'}^t, \mu_{d''}^t\right)$ is equally likely as $\left(\mu_{d''}^t, \mu_{d'}^t\right)$ and by the assumption that the density of υ is symmetric. Given reversibility, if $\mu_d^{(t+1)}$ is drawn from the posterior then detailed balance follows immediately from the use of the Metropolis–Hastings acceptance ratio. The conditional (on the other chains) stationary distribution of a given chain d then must be the posterior distribution of μ. Further, since the conditional stationary distribution does not depend on the state of the other chains the stationary distribution of the chain must also be the posterior of μ. In order for the stationary distribution to be unique we need to show that the MCMC chain is aperiodic and irreducible. Since, the proposal density q_μ induced by the DE step for a particular chain d is essentially a random walk aperiodicity is almost surely guaranteed. Finally, the unbounded support of υ and that υ is a proper density implies that all $q_\mu > 0$. It follows then that every set within the support of μ that has positive measure can be reached in a single step, ensuring that the chain is irreducible. Since the chain is aperiodic, irreducible and satisfies detailed balance with respect π^* it must have π^* as its unique stationary distribution.

Coupled with the discussion earlier we can claim that the draws obtained from our proposed sampler will be from the joint posterior of interest. In what

follows we illustrate the implementation of our algorithm with example and real applications.

3.5 Illustrating the algorithm

3.5.1 Example 1: How the algorithm works

To illustrate the working of the algorithm we use the simple static discrete game with two players and two actions described in Section 2.1. The probability of player i choosing action $a_i = 1$ is denoted by,

$$P_i = \Psi\left(a_i = 1|\theta\right) \equiv \frac{\exp\left(\theta_0 + \delta P_{-i}\right)}{1 + \exp\left(\theta_0 + \delta P_{-i}\right)}. \tag{24}$$

For the purposes of our example we will let $\theta_0 = -3$ and $\delta = 6$. Note that this is exactly equivalent to the setting in the example discussed in Section 2.1. As before, for these parameter values there exist three possible symmetric equilibria at $\{0.0707, 0.5, 0.9293\}$. Of these the equilibrium at 0.5 is unstable. Next, we generate 10,000 data points from the high equilibrium $[P_1 = P_2 = 0.9293]$. The data generated is presented in the table below.

	$a_1 = 0$	$a_1 = 1$
$a_2 = 0$	62	689
$a_2 = 1$	668	8581

For now assume that $\theta_0 = -3$ is known and the researcher only wishes to estimate δ (and of course infer $\mathbf{p}^e$). Note that because the game is symmetric and there exist no covariates, the likelihood depends on a single equilibrium probability. Based on this, we can map out the likelihood which is supported on the equilibrium manifold. Figure 4 depicts the likelihood with the thickness of each dot representing the height of the posterior.

Note that there is no mass in any area but on the ribbon that defines the manifold. The aim of the algorithm described earlier is to trace out this manifold accepting points on it that stochastically enhance the posterior. To understand how the approach works consider panel (a) in Fig. 5. We start with a randomly chosen starting point in the parameter-probability space denoted by the grey diamond. The manifold step then uses best-response iterates to find the closest fixed point located on the manifold (denoted by the hollow circle). In this case, obviously, the draw is accepted since there is positive mass at this point. The next iteration then begins with a new candidate for the parameter and a new starting point for ϱ_0 (grey solid circle). Again the Manifold step iterates to find the closest fixed point which is also accepted. This goes on with some points being accepted and others rejected (dark cross) until the algorithm is interrupted by the researcher. The results from one run of this algorithm are presented in panel (b) of Fig. 2.

The red (gray) dots indicate accepted draws while the black dots are rejected draws. The plot depicts a fairly dense posterior for θ_1 with a mean of 6.01. The

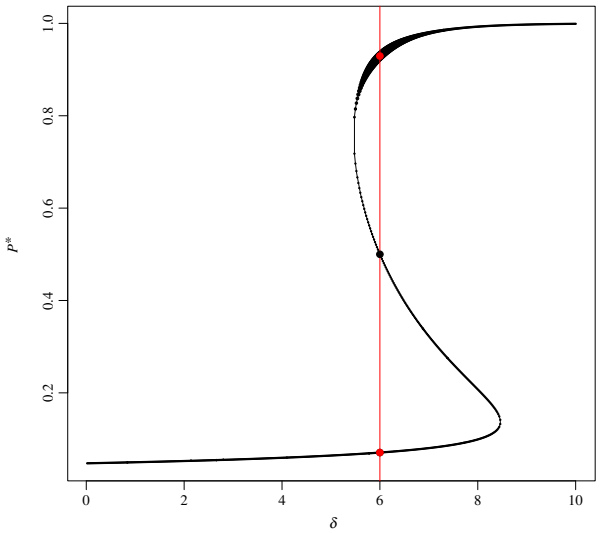


Fig. 4 The equilibrium manifold

MCMC routine also finds the appropriate equilibria with a high degree of precision with a mean of 0.93. The dashed lines offer a 95 % highest posterior density region. While the example described above is elementary, it serves to illustrate the basic ideas behind our routine.

3.5.2 Example 2: A more involved example

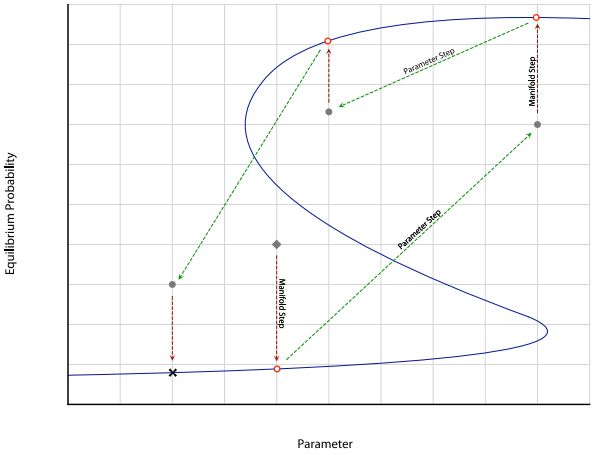
We now examine a more realistic example modeled after Sweeting (2009). In this model players play a coordination game such that payoffs are

$$\begin{aligned} u_{A1} &= \theta_0 + \theta_1 s_A + \delta P_{B1} + \varepsilon_{A1} \\ u_{A2} &= \delta P_{B2} + \varepsilon_{A2} \end{aligned} \tag{25}$$

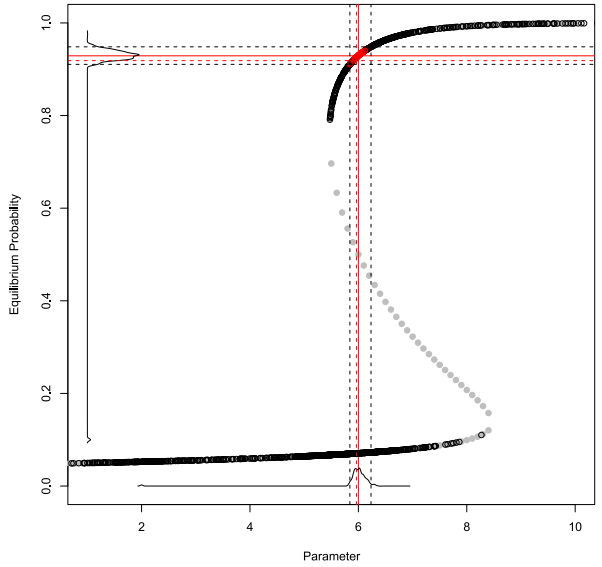
with player B's payoffs defined analogously. The equilibrium probabilities are then implicitly defined by

$$\begin{aligned} P^*_{A1} &= \frac{\exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{B1}\right)}{\exp\left(\delta\left(1 - P^*_{B1}\right)\right) + \exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{B1}\right)} \\ P^*_{B1} &= \frac{\exp\left(\theta_0 + \theta_1 s_B + \delta P^*_{A1}\right)}{\exp\left(\delta\left(1 - P^*_{A1}\right)\right) + \exp\left(\theta_0 + \theta_1 s_A + \delta P^*_{A1}\right)}. \end{aligned} \tag{26}$$

We will assume that $\theta_0 = -0.25, \theta_1 = 1$ and $\delta = 3$. These parameters will often generate multiple equilibria with at most two stable equilibria. In the



(a) MCMC Steps



(b) MCMC Draws

Fig. 5 MCMC over the equilibrium manifold

case when this happens we will assume that equilibria are picked based on the following rule

$$\pi\left(\mathbf{p}_m^e|\vartheta,\tilde{\mathcal{P}}_m\right)=\frac{\exp\left(\lambda\left(P_{A1}^{*(1)}+P_{B1}^{*(1)}\right)\right)}{\exp\left(\lambda\left(P_{A1}^{*(1)}+P_{B1}^{*(1)}\right)\right)+\exp\left(\lambda\left(P_{A1}^{*(2)}+P_{B1}^{*(2)}\right)\right)} \tag{27}$$

with $\lambda=2$. Thus, equilibria that favor both players picking action 1 are more likely to be accepted. Note that the data generating mechanism finds all stable

equilibria using Sweeting's approach of starting from the vertices of the unit cube and then chooses one based on the rule described above. That is, the domain of attraction weights are not used in the DGP. We generate data for $M = 5{,}000$ markets with s_{Am} and s_{Bm} drawn from standard densities. The generated equilibria are depicted in Fig. 6. Of these markets 397 markets had multiple (exactly 2) equilibria while the rest had a single stable equilibria. As the plots show, in the case of multiple equilibria the odds of picking the equilibrium with larger probabilities are substantially higher. The choices generated from these equilibria are described below.

	$B1$	$B2$
$A1$	2078	269
$A2$	238	2415

We used the Differential Evolution MCMC routine with $D = 15$ chains, each with randomly chosen starting values. We ran the sampler for 15,000 iterations and retained the last 5,000 for inference. The sampler converged quite quickly and draws from the converged chains seemed to locate all

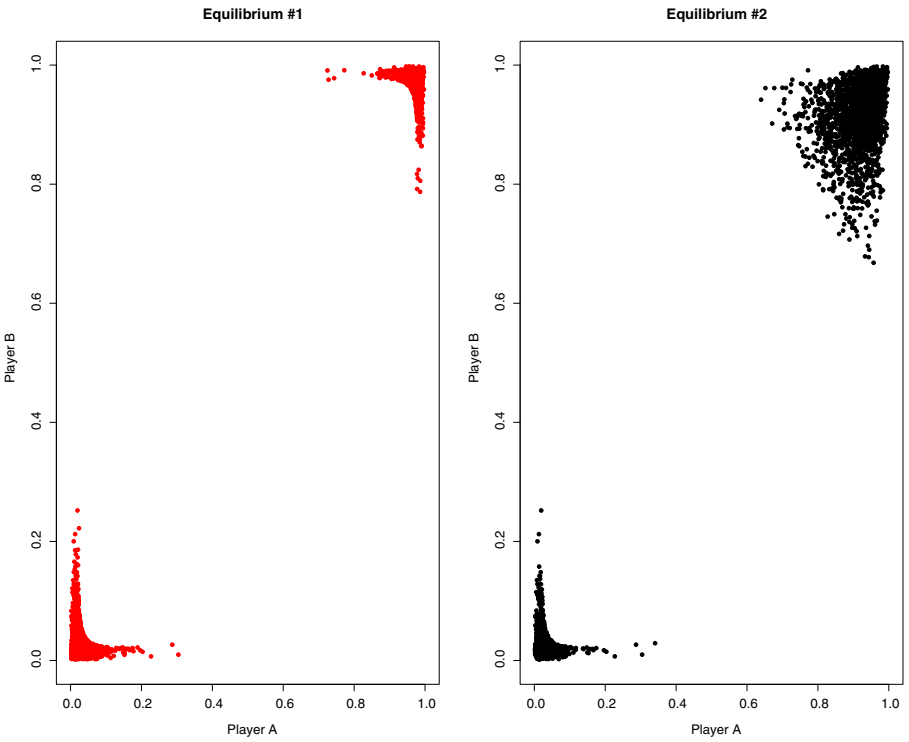


Fig. 6 Equilibria in simulated data (Example 2)

Table 1 MCMC results for Example 2

Parameter	True values	95 %-HPD
θ_0	-0.25	$(-0.2653, -0.2305)$
θ_1	1	(0.9744, 1.0653)
δ	3	(2.9613, 3.0385)
λ	2	(1.6891, 2.2906)

parameters quite accurately. The 95 % Highest Posterior Density (HPD) regions are reported in Table 1 and each contains the relevant true value. The posterior densities are presented in Fig. 7.

Overall, we believe our sampler performed quite well with limited data. The results on recovering λ are especially encouraging but should be taken with a word of caution. Our particular example generated a large number of markets with multiple equilibria and the variation in the covariates was limited. Effectively the data offers a reasonable number of replications of the game. This is critical to identifying λ. In real applications where we have only cross-sectional data such identification may be infeasible. We discuss this in the next section.

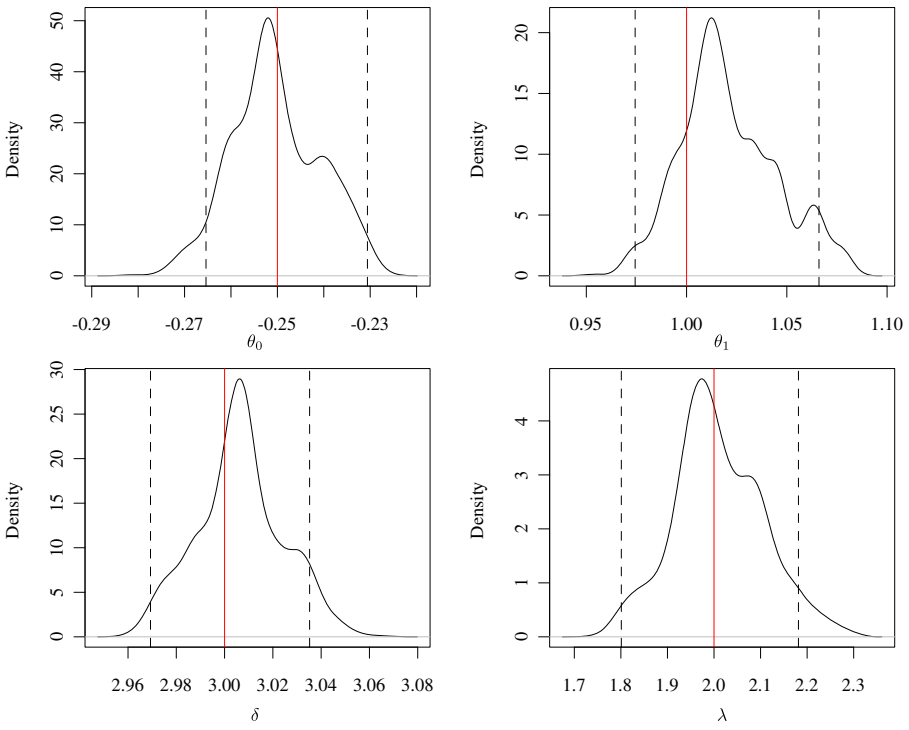


Fig. 7 Posterior densities (Example 2)

3.6 Identification

We do not aim to provide a formal discussion of identification here since these have been discussed before. We refer the reader to Sweeting (2009), Bajari et al. (2010) and Pesendorfer and Schmidt-Dengler (2008) for excellent discussions on the topic. In this section we will summarize the identification results from the literature and informally discuss the assumptions and data requirements for the identification of the models discussed in this paper. We will first discuss parameter identification assuming that the equilibrium probabilities (Conditional Choice Probabilities or CCPs) are known. We will then discuss the identification of the CCPs themselves.

3.6.1 Parameter identification given equilibrium probabilities

Assuming that the CCP's are known identification of the parameters involves showing that these probabilities can be inverted to identify the parameter. There are two cases worth discussing here (1) When players are identical and (2) When players are heterogeneous.

If players are identical and at least two equilibria are played in the data then the parameters of the model are identified. To see this consider example 1. In this setup if only a single equilibrium is played in the data then θ_0 and δ cannot be separately identified since there will be an infinite number of curves that can pass though the equilibrium point. In this case one would need to normalize θ_0 or, as in our example, assume it is known, in order to identify δ. If on the other hand two equilibria are played in the data then we can pinpoint the (logit) curve that passes through these points and identify both θ_0 and δ. Sweeting (2009) discusses this identification in more formal detail.

When players exhibit observed heterogeneity, as in our applications, identification is achieved on account of excluded variation. This idea is discussed in some detail in Bajari et al. (2010) as well as in other papers. The basic idea is that if there exists at least one covariate that impacts the payoffs of each player but that does not directly influence those of other players (except via their actions) then the parameters are identified. In this case the CCPs of other players exhibit variation that is independent of the state variables of any given player. This variation helps nail down the strategic parameter and the rest of the parameters are identified from the variation in the remaining state variables. In empirical applications using the methods presented here we strongly encourage the use of such exclusions.

3.6.2 Identification of equilibrium probabilities

Under the assumption that a single equilibrium is played in the entire data the identification of the equilibrium probabilities (CCPs) is trivial. Since we observe the state variables and the actions the reduced form probabilities are estimable. With enough (infinite) data these probabilities can be identified exactly. The more interesting case is when we assume that a single equilibrium

is played in each market. In this case again, as long as we have panel data with enough repetitions the CCPs are identified. With cross-sectional data the identification argument relies inability of a common model to adequately rationalize the variation in the choices. Sweeting (2009) shows that the identification of the equilibrium probabilities and associated equilibrium selection parameters (in our case ϑ) with symmetric players boils down to examining the conditions for the identification of a finite mixture model with cross-sectional data. The case where the players (and equilibria) are not symmetric on account of observed characteristics is less transparent but identification is still achieved as long as there is sufficient variation in types relative to the number of equilibria. Our own experience (with a large number of Monte-Carlo experiments) is that as long as enough markets exhibit multiplicity and the excluded state variables vary sufficiently the equilibrium probability and the selection parameters seem identified. The fact that unlike complete information games the number of equilibria do not increase significantly with the number of players or actions (Bajari et al. 2011).

Finally, we would like to discuss the issue of heterogeneity and common unobservables. There is clearly a confound between identifying heterogeneity and multiple equilibria with cross-sectional data. Since the variation required to obtain the CCPs with multiple equilibria is essentially the same that would identify heterogeneity identification is questionable. Again if we had panel data with enough replications of the game, identification would not be a concern. Alternatively, if we gave up on identifying ϑ (as we do in our discount store application) then the same variation can be used to parametrically identify unobserved heterogeneity across markets.

4 Implementation

4.1 Application: Supermarket pricing strategy

To test the feasibility of our proposed MCMC approach we apply our algorithm to a reasonably sized, real data based research question. The application we consider is the decision of supermarkets to choose everyday low pricing (EDLP) strategies.[8] We focus our attention on 852 markets where there are only two stores (see Fig. 8). The model can be simply summarized by the following probability of player i choosing EDLP in market m,

$$P_{im}(\text{EDLP}) = \frac{\exp\left(X'\theta + \delta_m P_{-im}(\text{EDLP})\right)}{\exp\left(X'\theta + \delta_m P_{-im}(\text{EDLP})\right) + \exp\left(\delta_m P_{-im}(\text{PROMO})\right)} \quad (28)$$

[8]For more details on the data, the definition of markets etc. the reader is refered to Ellickson and Misra (2008).

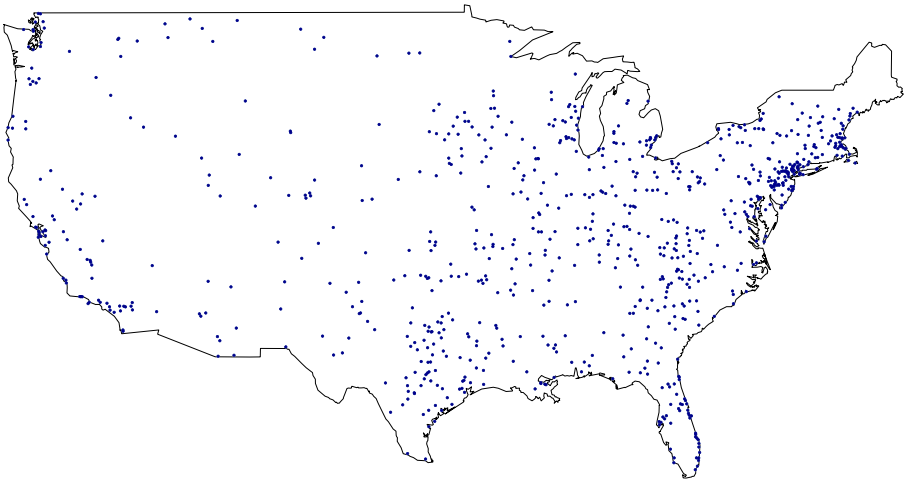


Fig. 8 2-strore markets in the US

where

$$X = \left\{ \begin{array}{c} \text{Store Size} \\ \text{Proportion of Minorities in Population} \\ \text{Vertical Integration Dummy,} \end{array} \right\}$$

We extend Ellickson and Misra (2008) by allowing key parameter of interest δ, which measures the impact competitive strategy choices have on the focal player's payoffs, to vary across markets. In particular, we let

$$\delta_m = \delta_0 + \delta_1 \text{Size of Market } \left(in \text{ Sq.Miles}\right) \tag{29}$$

Finally, as in the earlier example we used a prior specification where

$$\pi\left(\mathbf{p}_m^e | \vartheta, \tilde{\mathcal{P}}_m\right) = \frac{\exp\left(\lambda\left(P_{1m}^{(1)}(\text{EDLP}) + P_{2m}^{(1)}(\text{EDLP})\right)\right)}{\exp\left(\lambda\left(P_{1m}^{(1)}(\text{EDLP}) + P_{2m}^{(1)}(\text{EDLP})\right)\right) + \exp\left(\lambda\left(P_{1m}^{(2)}(\text{EDLP}) + P_{2m}^{(2)}(\text{EDLP})\right)\right)} \tag{30}$$

We have restricted our attention to a small set of variables to model this choice. We note that the store size and vertical integration dummies act as exclusion variables that identify the model. Ellickson and Misra (2008) find that $\delta > 0$, albeit with a larger dataset and more variables. The 2-player markets are fairly well spread out over the continental US and are plotted in Fig. 7.

Table 2 Pricing strategy game—2-step and MCMC results

	2-STEP		MCMC	
Variable	Estimate	Std. error	Post. mean	95 %HPD
Intercept	−1.3338	0.4659	−1.5321	(−1.96, −1.04)
Store size	0.6085	0.3358	0.6496	(0.13, 1.30)
Vert. integration	0.7653	0.1612	0.7785	(0.51, 1.05)
Proportion minority	0.9402	0.3108	1.1512	(0.62, 1.71)
δ_0	0.7983	0.2137	0.5516	(0.03, 1.32)
δ_1(size of market)	0.0493	0.0028	0.0464	(−0.01, 0.09)
λ			−1.0911	(−93.11, 98.48)

4.2 Details

We set up our MCMC routine with $D = 20$. That is, we use 20 chains that will communicate with each other. The MCMC chain was run for 15,000 iterations and the last 5,000 were retained for inference. The initial values for the routine were chosen at random rather than possibly more informative values.[9]

We allowed for a maximum of one thousand best response iterates in the manifold step although no more than 200 were ever used. The chain converged quite quickly for the scale of the problem and in no case did the best-response iteration fail to converge. The acceptance rate was about 14 %[10] on average with chain specific acceptance rates ranging from 6.7 to 63.3 %. The chains did have some autocorrelation and we thinned the chain by subsampling. A casual observation of trace plots showed that the pooled chain has converged.[11]

4.3 Results

Parameter estimates are presented in Table 2. For comparison purposes we also estimated the model using a 2-step approach and find that both methods recover qualitatively similar parameters. The differences may be on account of a number of differences between the estimators and we will no speculate here.

While this is a fairly limited analysis of the problem, the results accord well with those obtained by Ellickson and Misra (2008) who used pseudo-likelihood methods. The strategic effect, while still positive, is lower in our results. The range δ evaluated at the posterior means of δ_0 and δ_1 across markets was (0.7984, 3.08). At the converged draws multiplicity existed only in a few markets (between 4 and 11) and it is not altogether surprising that the λ parameter is effectively not identified and reflects only the information in the prior. Changing the equilibrium selection rule did not alter the qualitative findings and we do not report those results here.

[9]For example, a simple two step estimator might offer reasonable starting values.

[10]As a frame of reference, a traditional MH routine resulted in a <2 % acceptance rate after 200,000 iterations.

[11]Standard convergence diagnostices such as *R*-statistic of Gelman and Rubin (1992) also confirm this. Trace plots and convergence diagnostics are available from the authors upon request.

4.3.1 Equilibrium probabilities

As a by product of the estimation routine the MCMC also provides a natural way to construct posterior estimates of the equilibrium probabilities. Since the equilibrium probabilities are a function of the state variables and the parameters, the draws of the parameters coupled with the manifold step allows us to find the density of equilibrium probabilities in any market. In Fig. 9 we present examples of the different equilibria estimated in our application. We note that most of the markets were of the type depicted in the top left panel (Market #11).

In the figure each line corresponds to the best response curve conditioned on the state variables and a draw from the posterior of θ. These MCMC draws allow us to construct confidence sets for equilibrium probabilities For example, the top left hand panel shows a market where the equilibrium probabilities for both players range within [0.13, 0.25] with the mean around {0.174, 0.189}. The equilibrium is not symmetric as player two always has a higher probability.

In some cases since the state variables are identical, the equilibrium turns out to be symmetric and lies on the 45° line. In other cases, the equilibrium probabilities are asymmetric. Further we also see that in some cases the

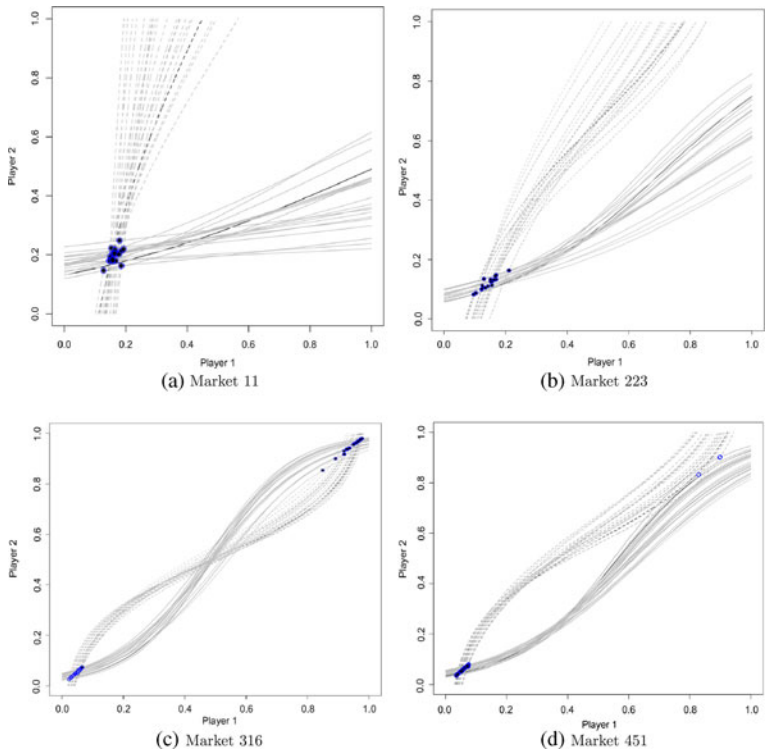


Fig. 9 Posterior equilibrium probabilities for four markets

probabilities are a more dispersed than in others. Again, this happens because the state variables in some markets differ significantly across players. We note here that, the draws for the parameters are held the same across markets. The differences arise only on account of the state variables and the strategic interaction parameter. The fact that our algorithm traces multiple equilibria is evident in the two plots with multiple intersections of the reaction functions. In one case (Market #316) the algorithm sampled both equilibria relatively frequently but only accepted the lower equilibrium once in the set of draws we considered. This market had both players adopting EDLP when the state variables could have easily justified PROMO. The strategic interaction parameter for this market was among the highest and this essentially generated multiple equilibria.

Overall, the DE-MCMC algorithm works fairly efficiently and without much supervision. As we mentioned earlier for the 2-store markets the chain converged in less than three hours. The routine can also be used as an equilibrium sampling routines based on parameters estimated from 2-step/NPL approaches. There are a number of other extensions worth considering, we discuss some of these in what follows.

5 Heterogeneity and common unobservables

Unlike Pseudo-likelihood approaches the MCMC methods used in this paper do not have to construct beliefs since the routine relies on a nested fixed point approach. As such, incorporating additional unobservable (to the researcher) components is not a problem since we do not have to worry about the consistency of the first step estimators. For example, market heterogeneity $\left(\text{say } \theta_m\right)$ that accounts for common unobservables can also be added without major modifications.

The basic idea would be to construct a hierarchical specification,

$$\theta_m = \bar{\theta} + \sigma_\theta^m \eta_m \tag{31}$$

and then draw θ_m for each market (conditioned on a candidate $\bar{\theta}$) and search the equilibrium manifold to construct the posterior. Note that care must be taken to ensure that $\bar{\theta}$ is updated only if the θ_m it spawns moves the MCMC chain. Even with this, the increase in complexity is limited since the manifold search section remains the same. Player heterogeneity (θ_i) can be handled similarly but requires that we observe the same player operating in multiple markets (e.g. same chains across markets), i.e. that we have panel data.

5.1 Application: Discount store entry

To illustrate how unobservables can be accommodated into the proposed MCMC approach we use a "real world" example: an entry game between WalMart and Kmart discount stores. We will assume that there are only two

players[12] Kmart and WalMart who compete across a number of small local markets. We will further assume that their entry patterns can be modeled as a static discrete game.

The dataset we use is drawn from Jia (2008) and was constructed by the author for her analysis of the discount retail industry. Our model is a simplified version of her framework, and assumes that the two chains make independent entry decisions across a collection of local markets. As in Jia (2008) we only use markets in which each firm operates at most one store. A local market is defined to be a county which gives us 2,065 relatively small and isolated markets which are assumed to be independent replications of the discrete game where the two firms decide whether to "enter" or not. In keeping with the assumptions outlined earlier the game will be assumed to be one of incomplete information where firms make simultaneous moves.

5.1.1 Specification

We keep the specification of the latent profits simple and include key covariates identified by Jia (2008) in the model. The probability of any firm $i \in \{WalMart, Kmart\}$ entering market m is denoted by,

$$p_{im}(\text{ENTRY}|\Theta, \xi_m, \mathbf{s}_m) = \frac{1}{1 + \exp\left(X'\alpha + Z_i'\beta + \delta p_{-im}(\text{ENTRY}|\Theta, \xi_m, \mathbf{s}_m) + \xi_m\right)} \tag{32}$$

In the above $\Theta = \{\alpha, \beta, \delta\}$ and the observable components X and Z_i contain common and player specific covariates respectively. These are describe below.

$$X = \left\{ \begin{array}{r} \text{Population} \\ \text{Retail Sales per capita} \\ \text{Urban} \end{array} \right\}$$

$$Z_W = \left\{ \begin{array}{r} \text{WalMart Intercept} \\ \text{Distance to Bentonville, AK} \\ \text{Southern Market Dummy} \end{array} \right\}$$

$$Z_K = \left\{ \begin{array}{r} \text{Kmart Intercept} \\ \text{MidWest Market Dummy} \end{array} \right\}$$

In addition, the model the proposed model contains a market level random effect ξ_m which is assumed to be Normally distributed with mean zero and variance σ_ξ^2. The role of the common unobservable is important in uncovering the true strategic effect since ignoring such market level effects could exaggerate the strategic interaction effect. At the same time, incorporating such unobservables into discrete games is non-trivial since we have to solve for the equilibrium probabilities. Using the notation defined earlier, with the action

[12] During the time period of this data Target was not a major player.

space $a = \{\text{ENTER}, \text{NOT ENTER}\}$ we can write the likelihood for a given market as,

$$\mathcal{L}_m\left(\mathbf{y}_m|\theta, \mathbf{s}_m\right) = \int \sum_{\mathbf{p}_m^e \in \mathcal{P}_m^e(\theta,\mathbf{s})} \pi_p\left(\mathbf{p}_m^e|\Theta, \xi_m, \mathbf{s}_m\right) \times \prod_{i\in N_m}\prod_{a_i\in K}\left[p_{im}^e\left(a_i|\Theta, \xi_m, \mathbf{s}_m\right)\right]^{y_{im}(a_i)} d\mathcal{F}_N\left(\xi_m|\sigma_\xi\right) \quad (33)$$

where the equilibrium probabilities are defined by the fixed point of

$$p_i^e\left(\text{ENTRY}|\Theta, \xi_m, \mathbf{s}_m\right) = \frac{1}{1+\exp\left(X'\alpha + Z_i'\beta + \delta p_{-im}^e\left(\text{ENTRY}|\Theta, \xi_m, \mathbf{s}_m\right) + \xi_m\right)}. \quad (34)$$

In this application we do not specify an equilibrium selection rule and simply draw equilibria proportional to their domains of attraction. As we discussed earlier, with cross-sectional data identifying both heterogeneity and the selection rule parameters is untenable. What further exacerbates the problem is that in this application, there doesn't seem to be significant evidence of multiplicity which render any identification of equilibrium selection moot.

5.1.2 Implementation details

A quick look at the likelihood makes it clear why the estimating discrete games in the presence of unobservables is computationally demanding. The key problem arises on account of the common unobservables which need to be integrated out. This is non-trivial since the equilibrium probabilities are conditioned on the value of ξ_m. For example, a Monte-Carlo approach would involve drawing ξ_m for each market and solving for the fixed point for each draw. At first glance this problem prohibit the use of the MCMC approach since we have to jointly accept (or reject) the parameters and equilibrium probabilities. This would entail jointly accepting/rejecting hundreds if not thousands of parameters. Fortunately, we can block parameters into two sets and run the modified routine for each set. That is, we construct two blocks θ and $\{\xi_m\}$ and apply the proposed DE-MCMC routine to each block sequentially by conditioning on the other block.

To be more specific, given draws of the parameter vector θ we sample $\{\xi, \mathbf{P}\}$ jointly using the DE-MCMC routine. We then take the draws of ξ as given and sample $\{\theta, \mathbf{P}\}$. This increases the computational burden since the algorithm requires us to find the equilibrium manifold twice for each market (at each draw) once for the common parameters and once for the market specific effects. This increase however turns out not to be prohibitive from an implementation point of view. In the current application $M = 2{,}065$ and the algorithm converges in less than a couple of hours. Note also, that this blocking strategy does not alter the properties of the MCMC chain since we are always sampling equilibria every time a new parameter set is proposed. More

generally, any Gibbs-type algorithm for each individual parameter would also be acceptable (albeit computationally expensive) as long as "new" candidate equilibria were also being generated at the same time. Essentially, all such algorithms that explore the manifold by sampling parameters and equilibria together will work.

We used $D = 50$ communicating chains for this application, ran the MCMC chain was run for 25,000 iterations and retained the last 5,000 for inference. We allowed for a maximum of one thousand best-response iterates in the manifold step although no more than 200 were ever used. The chain converged quite quickly (less than 2 h) and in no case did the best-response iteration fail to converge. The converged chain had an acceptance rate of about 21 % for the base parameters and ξ_m was accepted on average in 45 % of the markets at any given draw. As with the pricing case study some autocorrelation was found and we thinned the chain by subsampling. For the sake of brevity and to avoid repetition trace plots and other convergence details are not presented here but are available from the author upon request.

5.1.3 Results

The results from the MCMC run are presented in Table 3. For the sake of comparison we have also included results from a MCMC run of a baseline model that ignores market level heterogeneity. The direction and magnitude of the effects in the *base model* are comparable to those presented in Jia (2008) and in Ellickson and Misra (2011) who use the same data. The comparison to Ellickson and Misra (2011) is noteworthy since they use the same specification as us but different estimation approaches and obtain essentially similar results. Our numerical results are slightly different since the posterior means does not coincide with parameters describing the Likelihood mode.

The estimates provide significant evidence for market level heterogeneity. The posterior mean for the standard deviation parameter (σ_ξ) of the market

Table 3 Table 2: Discount store entry game—MCMC results

	Base model	Model with common market unobservables	
Variable	Posterior mean	Posterior mean	95 % HPD
Common effects			
Population	1.92	2.29	(0.72, 4.30)
Retail sales per capita	1.66	1.51	(0.18, 3.25)
Urban	1.84	1.43	(0.81, 1.95)
δ	0.87	0.29	(0.01, 2.07)
σ_ξ	–	0.90	(0.78, 1.03)
Wal-Mart specific effects			
Intercept (Wal-Mart)	−14.04	−13.20	(−14.71, −11.94)
Distance to Bentonville, AK	−0.97	−0.65	(−1.15, −0.11)
South	0.54	0.11	(−1.65, 1.55)
Kmart specific effects			
Intercept (Kmart)	−21.65	−18.53	(−21.47, −15.39)
MidWest	0.23	−0.24	(−1.71, 1.43)

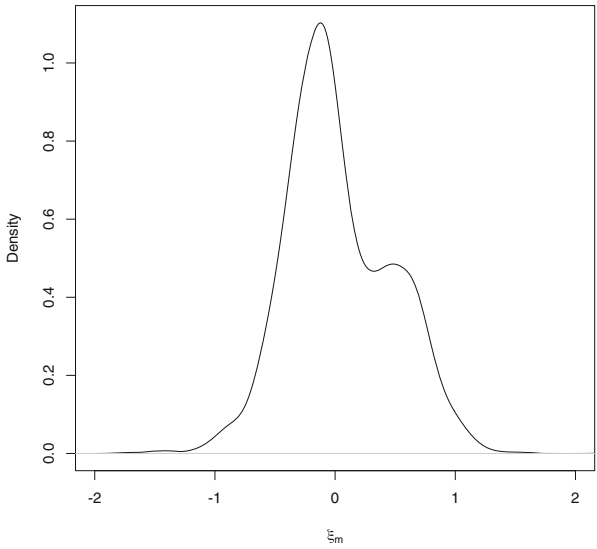


Fig. 10 Density plot of market level effects (posterior means)

effect density was found to be 0.90 with a 95 % HPD (Highest Posterior Density) region of (0.78, 1.03). Figure 10 shows the density plot of the posterior means of of the heterogeneity elements (ξ_m) across the 2,065 markets in the data. A marginal likelihood comparison across the models also provides "strong evidence" for heterogeneity. Apart from the heterogeneity constructs exhibiting strong statistical significance they also afford some substantially meaningful insights. The inclusion of market effects eliminates the role played by region (MidWest and South) dummies. This is not altogether surprising since local market effects are a more powerful explanatory mechanism for local decisions. A more important finding is that the inclusion of this source of heterogeneity significantly reduces (but does not eliminate) the strategic interaction effect. This is similar to the findings in Orhun (2012).

Since this exercise was simply to document the feasibility of the method we will not belabor a more detailed discussion of the results but will conclude this discussion by stating our belief that the MCMC procedures outlined in this paper can be used for most discrete games of incomplete information games without much computational cost.

6 Concluding remarks

While this paper has focused on static discrete games, the methods presented here are fairly general and could be applied to other contexts as well with appropriate modifications.

In a number of cases the researcher may have access to additional data that arises as a consequence of the decisions made by the players. For example, in Ellickson and Misra (2008) the authors have access to revenue data that is conditioned on the choice of a supermarkets pricing strategy choice. The

authors present a semiparametric approach to correcting for selectivity in the revenue regression in the context of a first stage discrete game among many players. While their approach is agnostic about the distribution of the structural errors a more parametric approach could be used to construct an appropriate joint likelihood for the observed data and then used in our proposed MCMC scheme.

Our attention in this paper has been limited to *simultaneous* move games of incomplete information. The method could easily be modified to apply to games with sequential moves or games with a sequence of simultaneously moving sets of players (say incumbents then entrants). Such a specification was used by Jia (2008) in her analysis of discount store entry. Of course, a pure sequential move game offers the additional simplification of uniqueness which reduces the computational burden significantly. There may be contexts, however, where the order of the moves is unknown. In such a case the methods presented in this paper could easily be modified to estimate the model by drawing a random permutation for the order of play at each iteration. An application of this type and the conditions for the identification of the order of play are presented in Einav (2010) in the context of movie release timing.

The implementation of Bayesian methods (in particular MCMC methods) to the estimation of dynamic discrete choice models is a new and exciting area of research. To our knowledge there is no published work on the use of these methods in the context of dynamic games. The problem with adopting a MCMC approach stems from the fact that a dynamic game involves the solution to a coupled fixed point problem. That is, the researcher has to solve for the fixed point in value function space as well as for the equilibrium in the players action space. Even with the efficiency "tricks" presented here this poses a computationally challenging problem.

Having said that, there may be particular cases where a MCMC approach may be viable. Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2011) propose alternative probability based characterizations of optimal policy function in the context of single and multi-agent decision makers. In some very special cases (e.g. terminal states, finite dependence, discrete state spaces) rather startling simplifications emerge which could be exploited to obtain computational gains. One example is the work by Ellickson et al. (2011) who show how a dynamic game of pricing strategy choice can be estimated using simple methods by exploiting the terminal nature of the exit decision. Combining these clever recharacterizations of the value function space with the methods proposed in this paper might offer a promising avenue for future research.

In summary, this paper outlines a simple approach to sampling parameters from a likelihood supported on the equilibrium manifold of a discrete game. The application demonstrates the viability and validity of the approach in a simple setting by applying the proposed methods to two research datasets. The basic ideas are general, and can be applied to other settings fairly straightforwardly. We hope that this research spurs interest in the estimation of discrete games using MCMC methods.

References

Arcidiacono, P., & Miller, R. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica, 7*(6), 1823–1868.

Aguirregabiria, V. (2004). Pseudo maximum likelihood estimation of structural models involving fixed-point problems. *Economics Letters, 84*(3), 335–340.

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

Balasko, Y. (1988). *Foundations of the theory of general equilibrium*. London: Academic Press.

Bajari, P., Benkard, C.L., Levin J. (2007). Estimating dynamic models of imperfect competition. *Econometrica, 75*(5), 1331–1370.

Bajari, P., Hong, H., Krainer, J., Nekipolov, D. (2010). Estimating static models of strategic interactions. *Journal of Business and Economic Statistics, 28*(4), 469–482.

Bajari, P., Hong, H., Krainer, J., Nekipolov, D. (2011). *Computing equilibria in static games of incomplete information using the all-solution homotopy*. Working paper.

Berinde, V. (2007). Iterative approximation of fixed points. In *Lectures notes in mathematics*. Berlin: Springer.

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

Braak, C. (2006). Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing, 16*(3), 239–249

Draganska, M., Misra, S., Aguirregabiria, V., Bajari, P., Einav, L., Ellickson, P., Horsky, D., Narayanan, S., Orhun, Y., Reiss, R., Seim, K., Singh, V., Thomadsen, R., Zhu, T. (2008). Discrete choice models of firms' strategic decisions. *Marketing Letters, 19*(3–4), 399–416.

Einav, L. (2010). Not All rivals look alike: estimating an equilibrium model of the release date timing game. *Economic Inquiry, 48*(2), 369–390.

Ellickson, P., & Misra, S. (2008). Supermarket pricing strategies. *Marketing Science, 27*(5), 811–828.

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

Ellickson P., Misra, S., Nair, H. (2011). Repositioning dynamics and pricing strategy. *Journal of Marketing Research* (forthcoming).

Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–511.

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

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

Judd, K.L., & Su, C.L. (2006). *A new optimization approach to maximum likelihood estimation of structural models*. Working paper, Hoover Institution.

Narayanan, S. (2012). Bayesian estimation of discrete games of complete information. *Quantitative Marketing and Economics*. doi:10.1007/s11129-012-9127-6.

Orhun, A. (2012). Spatial differentiation in the supermarket industry: the role of common information. *Quantitative Marketing and Economics*. doi:10.1007/s11129-012-9123-x.

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

Robert, G.O., & Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika, 83*(1), 95–110.

Rust, J. (1987). Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica 55*(5), 999–1033.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics, 37*(3), 619–640.

Storn, R., & Price, K. (1997). Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11*, 341–359.

Sweeting, A. (2009). The strategic timing of radio commercials: an empirical analysis using multiple equilibria. *RAND Journal of Economics, 40*(4), 710–742.

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