

NBER WORKING PAPER SERIES

IDENTIFICATION AND EFFICIENT SEMIPARAMETRIC ESTIMATION OF A  
DYNAMIC DISCRETE GAME

Patrick Bajari  
Victor Chernozhukov  
Han Hong  
Denis Nekipelov

Working Paper 21125  
<http://www.nber.org/papers/w21125>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
April 2015

The authors acknowledge generous research supports from the NSF and able research assistance from Timothy Armstrong and Boyu Wang. We would like to thank our colleagues and numerous conference and seminar participants for helpful comments. The usual disclaimer applies. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Patrick Bajari, Victor Chernozhukov, Han Hong, and Denis Nekipelov. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game  
Patrick Bajari, Victor Chernozhukov, Han Hong, and Denis Nekipelov  
NBER Working Paper No. 21125  
April 2015  
JEL No. C01,C14,C57,C7,C73,L0

**ABSTRACT**

In this paper, we study the identification and estimation of a dynamic discrete game allowing for discrete or continuous state variables. We first provide a general nonparametric identification result under the imposition of an exclusion restriction on agent payoffs. Next we analyze large sample statistical properties of nonparametric and semiparametric estimators for the econometric dynamic game model. We also show how to achieve semiparametric efficiency of dynamic discrete choice models using a sieve based conditional moment framework. Numerical simulations are used to demonstrate the finite sample properties of the dynamic game estimators. An empirical application to the dynamic demand of the potato chip market shows that this technique can provide a useful tool to distinguish long term demand from short term demand by heterogeneous consumers.

Patrick Bajari  
University of Washington  
331 Savery Hall  
UW Economics Box 353330  
Seattle, Washington 98195-3330  
and NBER  
Bajari@uw.edu

Victor Chernozhukov  
Department of Economics  
MIT  
Cambridge, MA 02142  
vchern@mit.edu

Han Hong  
Stanford University  
Landau Economics Building  
579 Serra Mall  
Stanford, CA 94305  
doubleh@stanford.edu

Denis Nekipelov  
Department of Economics  
University of Virginia  
254 Monroe Hall  
Charlottesville, VA 22904-4182  
dn4w@virginia.edu

# 1 Introduction

In this paper, we study the identification and efficient sieve estimation of a dynamic discrete game. We provide a general nonparametric identification result under the imposition of an exclusion restrictions on agents payoffs and analyze large sample statistical properties of nonparametric and semiparametric estimators for the econometric dynamic game model. We also show how to achieve semiparametric efficiency of dynamic discrete choice models using a sieve based conditional moment framework of Ai and Chen (2003). Numerical simulations are used to demonstrate the finite sample properties of the dynamic game estimators.

A dynamic discrete game is a generalization of a dynamic discrete choice model as in Rust (1987), Hotz and Miller (1993). As in these earlier papers, agents in the model are assumed to solve a dynamic programming problem. Payoffs in each time period depend on the agent's actions, the state variables and random preference shocks. Given current choices, the state variables evolve according to a law of motion which can depend on an agent's actions. A dynamic game generalizes this single agent model to allow the payoffs of one agent to depend on the actions of other agents. Dynamic game models are applicable in many areas such as industrial organization dynamic oligopoly with collusions, e.g. Fershtman and Pakes (2009). Recently, a number of papers have proposed methods to estimate dynamic games including Aguirregabiria and Mira (2002, 2007), Berry, Pakes, and Ostrovsky (2003), Pesendorfer and Schmidt-Dengler (2010), Bajari, Benkard, and Levin (2007) and Jenkins, Liu, McFadden, and Matzkin (2004).

Our identification framework for dynamic discrete games builds on an extensive recent literature including Aguirregabiria and Mira (2007), Berry, Pakes, and Ostrovsky (2003), Pesendorfer and Schmidt-Dengler (2010). In particular, the important contribution by Magnac and Thesmar (2002) developed a recoverability technique for a single agent dynamic model that also provides a much credited building block to identify closely related dynamic discrete games. Our framework builds directly on this profound literature and makes use of a recoverability method that follows directly from the single agent model analyzed in Magnac and Thesmar (2002). We generalize in that we allow the state variables to be either discrete or continuous. This is attractive for empirical work since in many applications state

variables are naturally modeled as continuous. The model is nonparametrically identified if the researcher is willing to make exclusion restrictions, that is, not all state variables can enter the payoffs of all agents. Such restrictions are commonly imposed in empirical research. For example, cost and demand shifters for one firm are frequently excluded from the payoffs of other firms. These include the distance to its distribution center in retail store entry models (Jia (2008) and Holmes (2011)), and the distance of a firm to the project location in highway procurements (Somaini (2011)).

Second, we analyze semiparametric and non-parametric estimation procedures. In a semiparametric model, imposing parametric functional form on the static payoff also allows for possible cross equation restrictions between players that can differentiate the current setup from most existing models. We begin with the analysis of a semiparametric setup, where we only use non-parametric identification assumptions and parameterize the identifiable payoffs of the players without additional restrictions on the state transition law. We find the semiparametric efficiency bound for the payoff parameters, which is the minimum variance of the parameter estimates without parametric assumptions regarding the state transition. Moreover, we show that obtaining the semiparametrically efficient estimates does not require solving for equilibria of the game and computing the corresponding likelihood function. We demonstrate that the treatment of the player's decision problem as a moment equation, generated by her first-order condition allows use to estimate the payoff parameters in one step. We also show that the estimation procedure that allows one to achieve the semiparametric efficiency bound belongs to our class of one-step estimation methods. This is a new approach to the analysis of dynamic games and it generalizes the existing two-step estimation techniques such as those proposed by Aguirregabiria and Mira (2007), Berry, Pakes, and Ostrovsky (2003) and Pesendorfer and Schmidt-Dengler (2010).

An additional advantage of our approach is that it does not rely on the discreteness of the state space which is in particular achieved by using an estimation approach that does not need preliminary estimation of the continuation values of the players. In applied work, many researchers choose to discretize a continuous state variable. Increasing the number of grid points in a two step estimator reduces the bias of the first stage. However, this comes at the cost of increasing the variance of the first stage estimates. In fact, when there are  $d = 4$

or more continuous state variables, it can be shown that it is not possible to obtain through discretization  $\sqrt{T}$  consistent and asymptotically normal parameter estimates in the second stage, where  $T$  is the sample size.<sup>2</sup> Therefore, discretizing the state space does not provide a solution to continuous state variables. The estimation approach of Bajari, Benkard, and Levin (2007) allows for continuous state variables. However, it requires a parametric first stage and the resulting estimates will be biased if the first stage is misspecified.

Third, we find that the reduction of the estimation procedure to one stage allows us to estimate payoffs of players fully non-parametrically. The structure of the non-parametric estimator is based on the player's first-order condition similarly to the semiparametric case. The estimates of the payoff function have a slower than parametric convergence rate. This rate depends on the smoothness of the distribution of the state transition as well as on the support condition on the policy functions of the players. We analyze the non-parametric estimator from the perspective of the mean-square optimality and offer a choice of trimming for the sieve representation of the payoff functions as well as the value functions that provides the procedure with the minimum mean squared error while converging at an optimal non-parametric rate. For the non-parametric estimator we develop a unified large sample theory that nests both continuous and discrete state variables as special cases.

Furthermore, the one step semiparametric conditional moment estimator proposed in this paper does not require one to determine numerically the value functions as they are nonparametrically estimated within the conditional moment formulation. This constitutes a considerable advantage over an identification based estimator and other existing multi-step estimation procedure for dynamic discrete choice models.

Section 2 discusses identification in a static discrete game model. Section 3 extends the identification analysis to a dynamic game. Section 3.7 develops nonparametric and semiparametric estimation methods which follow the lines of the identification conditions to construct estimates for the payoffs based on the non-parametric estimates of the conditional choice probabilities. Section 4 demonstrates how the multi-stage estimation strategy can be improved by representing the decision problem of a player in a dynamic game as a

---

<sup>2</sup>To see this follow the arguments in Newey and McFadden (1994). A requirement for the second stage parameter estimates to be  $\sqrt{T}$  consistent is that both the bias  $\sqrt{T}h^2$  and the variance  $1/(Th^d)$  should converge to 0 as  $T \rightarrow \infty$ . This is not possible when  $d > 4$ .

conditional moment equation. Moreover, it demonstrates how to obtain the estimates for the payoff parameters with the minimum variance over the class of models without parametric assumptions regarding the choice probabilities. It also derives the asymptotic distribution for the estimates of the payoff parameters. Section 4.3 briefly describes nonparametric estimation and section 5 demonstrates the finite sample properties of the estimators in a monte carlo simulation. Section 7 concludes.

## 2 Nonparametric identification of static games

We begin by describing the model for the case of static games. This serves two purposes. First, this will allow us to discuss our modeling assumptions in a simpler setting. Second, we prove the identification for the static model. This will highlight some key ideas in our identification of the full dynamic model and also will be used as a step in the identification of the more general dynamic model.

In the model, there are a finite number of players  $i = 1, \dots, n$ . Each player simultaneously chooses an action  $a_i \in \{0, 1, \dots, K\}$  out of a finite set. We assume that the set of actions are identical across players. This assumption is for notational convenience only and could easily be relaxed. Let  $A = \{0, 1, \dots, K\}^n$  denote the set of possible actions for all players and  $a = (a_1, \dots, a_n)$  denote a generic element of  $A$ . Also, let  $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$  denote a vector of strategies for all players excluding  $i$ . The vector  $s_i \in S_i$  denotes the state variable for player  $i$ . The set  $S_i$  can be discrete, continuous or both. Also, define  $S = \prod_i S_i$  and let  $s = (s_1, \dots, s_n) \in S$  denote a vector of state variables for all  $n$  players. We assume that  $s$  is common knowledge to all players in the game and is observable to the econometrician.

For each agent, there are  $K + 1$  private shocks  $\epsilon_i(a_i)$  indexed by the actions  $a_i$ . Let  $\epsilon_i = (\epsilon_i(0), \dots, \epsilon_i(K))$  have a density  $f(\epsilon_i)$  and assume that the shocks  $\epsilon_i$  are i.i.d across agents and actions  $a_i$ . We shall assume that  $\epsilon_i(a_i)$  is distributed extreme value.

**Assumption 1** *The error terms  $\epsilon_i(a_i)$  are distributed i.i.d. across actions and agents. Furthermore, the error term has an extreme value distribution with density*

$$f(\epsilon_i) = \exp(-(\epsilon_i + \bar{\gamma})) \exp(-\exp(-(\epsilon_i + \bar{\gamma}))).$$

In the above  $\bar{\gamma}$  is the Euler's constant,  $\bar{\gamma} \approx 0.577$ . This location shift ensures that the error terms have mean zero.

We could easily weaken this assumption. However, it is commonly used in the applied literature and will allow us to write a number of formulas in closed form which will simplify our study of identification. The vNM utility function for player  $i$  is:

$$u_i(a, s, \epsilon_i) = \Pi_i(a_i, a_{-i}, s) + \epsilon_i(a_i).$$

In the above,  $\Pi_i(a_i, a_{-i}, s)$  is a scalar which depends on  $i$ 's own actions, the actions of all other agents  $a_{-i}$  and the entire vector of state variables  $s$ . We assume that the iid preference shocks  $\epsilon_i(a_i)$  are private information for player  $i$ . The assumption that the error term  $\epsilon_i(a_i)$  is private information is not universal in the literature. For example, Bresnahan and Reiss (1991) assume that the error terms are common knowledge. However, this model requires quite different econometric methods which account for the presence of multiple equilibrium and the possibility of mixed strategies.

A strategy for agent  $i$  is a function  $a_i = \delta_i(s, \epsilon_i)$  which maps the state  $s$  and agent  $i$ 's private information  $\epsilon_i$  to an action  $a_i$ . Note that agent  $i$ 's strategy does not depend on  $\epsilon_{-i}$  since this is assumed to be private information to the other agents in the game. Define

$$\sigma_i(a_i = k|s) = \int 1 \{ \delta_i(s, \epsilon_i) = k \} f(\epsilon_i) d\epsilon_i.$$

This is the probability that agent  $i$  will play strategy  $k$  after we margin out  $\epsilon_i$ .

In equilibrium, player  $i$ 's belief is that  $j$  will play strategy  $k$  with probability  $\sigma_j(a_j = k|s)$ . Therefore,  $i$ 's expected utility from choosing the strategy  $a_i$  is  $\sum_{a_{-i}} \Pi_i(a_i, a_{-i}, s) \sigma_{-i}(a_{-i}|s) + \epsilon_i(a_i)$ .

Moving forward, it will be useful to define the *choice specific value function* as

$$\Pi_i(a_i, s) = \sum_{a_{-i}} \Pi_i(a_i, a_{-i}, s) \sigma_{-i}(a_{-i}|s). \tag{1}$$

Note that we can write the expected utility from choosing  $a_i$  as

$$\Pi_i(a_i, s) + \epsilon_i(a_i). \tag{2}$$

Recall that the error terms are distributed extreme value. Standard results about the logit model plus the definition of the choice specific value function imply that

$$\sigma_i(a_i|s) = \frac{\exp(\Pi_i(a_i, s))}{\sum_{a'_i \in A_i} \exp(\Pi_i(a'_i, s))}.$$

**Definition 1** Fix the state  $s$ . A Bayes-Nash equilibrium is a collection of probabilities,  $\sigma_i^*(a_i = k|s)$  for  $i = 1, \dots, n$  and  $k = 0, \dots, K$  such that for all  $i$  and all  $k$

$$\begin{aligned} \sigma_i^*(a_i|s) &= \frac{\exp(\Pi_i(a_i, s))}{\sum_{a'_i \in A_i} \exp(\Pi_i(a'_i, s))} \quad \text{and} \\ \Pi_i(a_i, s) &= \sum_{a_{-i}} \Pi_i(a_i, a_{-i}, s) \sigma_{-i}^*(a_{-i}|s). \end{aligned}$$

An equilibrium requires the actions of all players to be a best response to the actions of all other players. Moving forward, it is convenient to define an equilibrium in terms of  $\sigma_i(a_i|s)$  instead of  $\delta_i(s, \epsilon_i)$ .

## 2.1 Identification of the static model

An important question is whether it is possible for us to identify the parameters of our model. One approach to identification is to impose parametric restrictions on  $\Pi_i(a, s)$ . In what follows, we allow  $\Pi_i(a, s)$  to be a general function of  $s$  and do not specify the payoffs  $\Pi_i(a_i, s)$  parametrically. We identify  $\Pi_i(a_i, a_{-i}, s_i)$  nonparametrically by imposing exclusion restrictions on this function.

Our proof of identification is constructive. Assuming that the population probabilities  $\sigma_i(a_i = k|s)$  for all  $k, i$  and  $s$  are known, we reverse engineer the  $\Pi_i(a_i, a_{-i}, s_i)$  that rationalize the data. Simple algebra implies that

$$\sigma_i(a_i = k|s) = \frac{\exp(\Pi_i(a_i, s))}{\sum_{a'_i \in A_i} \exp(\Pi_i(a'_i, s))} \quad (3)$$

$$\log(\sigma_i(a_i = k|s)) - \log(\sigma_i(a_i = 0|s)) = \Pi_i(a_i = k, s) - \Pi_i(a_i = 0, s) \quad (4)$$

Equation (4) is the well known Hotz-Miller inversion. This equation implies that it is possible to learn the choice specific payoff functions,  $\Pi_i(a_i = k, s)$  up to a first difference from knowledge of the choice probabilities  $\sigma_i(a_i = k|s)$ . Since these choice-specific payoff



functions can only be learned up to a first difference, we need to impose the normalization that an “outside good” action always yields zero utility:

$$\Pi_i(a_i = 0, a_{-i}, s) = 0. \quad (5)$$

**Assumption 2** For all  $i$ , all  $a_{-i}$  and all  $s$ ,  $\Pi_i(a_i = 0, a_{-i}, s) = 0$ .

Having identified the choice specific payoff functions  $\Pi_i(a, s)$ , we next turn to the problem of identifying primitive mean utilities  $\Pi_i(a_i, a_{-i}, s)$ . The definition of the choice specific payoff function implies that these two objects are related by the following equation:

$$\Pi_i(a_i, s) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s) \Pi_i(a_i, a_{-i}, s), \quad \forall i = 1, \dots, n, a_i = 1, \dots, K. \quad (6)$$

Given  $s$ , this is a system of  $n \times K$  equations, because there are  $n$  agents and for each agent, there are  $K + 1$  choices. Then  $\Pi_i(a_i, a_{-i}, s)$  are  $n \times K \times (K + 1)^{n-1}$  free parameters in equation (6). Recall that for each agent, we have normalized the utility for the action  $a_i = 0$  to zero regardless of the actions of the other players. Therefore, for each agent  $i$ , there are  $K \times (K + 1)^{n-1}$  free parameters corresponding to the  $K$  actions available to  $i$  which yield nonzero utility and the  $(K + 1)^{n-1}$  actions of the other agents. Clearly,  $n \times K \times (K + 1)^{n-1} > n \times K$ , which implies that the model is underidentified.

In order to identify the model, we will impose exclusion restrictions on  $i$ 's payoffs. Partition  $s = (s_i, s_{-i})$ , and assume that

$$\Pi_i(a_i, a_{-i}, s) = \Pi_i(a_i, a_{-i}, s_i) \quad (7)$$

depends only on the subvector  $s_i$ . In other words, we are excluding some component of  $s$  from  $i$ 's payoffs. Such assumptions are commonly used in applied work. For example, many oligopoly models predict that after we control for  $-i$ 's strategies,  $i$ 's profits are not influenced by certain cost or demand shifters for  $-i$ .

If we impose these exclusion restrictions, we can rewrite (6) as

$$\Pi_i(a_i, s_{-i}, s_i) = \sum_{a_{-i}} \sigma_{-i}(a_{-i}|s_{-i}, s_i) \Pi_i(a_i, a_{-i}, s_i), \quad (8)$$

If there are  $(K + 1)^{n-1}$  points in the support of the conditional distribution of  $s_{-i}$  given  $s_i$ , we will have more equations than unknowns.

**Theorem 1** *Suppose that Assumptions 1 and 2 hold. Also suppose that for each  $s_i$ , there exist  $(K+1)^{n-1}$  points in the support of the conditional distribution of  $s_{-i}$  given  $s_i$ . Assume that the  $(K+1)^{n-1}$  equations defined by (8) are linearly independent almost surely. Then the latent utilities  $\Pi_i(a_i, s_{-i}, s_i)$  are identified for almost every  $s_i$  and  $s_{-i}$ .*

We can alternatively state a rank condition, similar to the linear least squares regression model, that is sufficient for identification. This rank condition requires that given each  $s_i$ , the second moment matrix of the “regressors”  $\sigma_{-i}(a_{-i}|s_{-i}, s_i)$ ,

$$E\sigma_{-i}(a_{-i}|s_{-i}, s_i)\sigma_{-i}(a_{-i}|s_{-i}, s_i)' \tag{9}$$

is nonsingular. Intuitively, we interpret  $\Pi_i(a_i, s_{-i}, s_i)$  as the dependent variable in an ols regression and  $\sigma_{-i}(a_{-i}|s_{-i}, s_i)$  as a regressor.

The rank condition in Theorem 1 is stated in terms of the equilibrium choice probabilities, which implicitly depend on the primitive parameters in the instantaneous payoffs  $\Pi_i(a_i, a_{-i}, s)$ . For a given set of primitive utility and transition density parameters, whether the induced equilibrium choice probabilities satisfy the rank condition may not be easy to verify. In the two by two case if the discount rate is close to zero, the rank condition holds as long as the instantaneous payoffs are different between the two players. Pesendorfer and Schmidt-Dengler (2010) gave elegant conditions for the discrete state space model in more details. Verifying the rank condition for continuous state space models can be more difficult. Theorem 1 does have the advantage that it can be verifiable using observed data. Furthermore, while whether the equilibrium choice probabilities induced by an arbitrarily given set of model primitive parameters satisfy the rank condition is an open question, it always holds for the primitive model utility functions recovered by the nonparametric identification procedure. By construction, the nonparametrically identified primitive utility functions satisfy the Bellman equation and the fixed point conditions for the Markov perfect equilibrium, and induce equilibrium choice probabilities that satisfy the rank condition as long as it is satisfied in the data.

We also note the identification arguments above can be extended where there are some state variables  $s_0$  common to all payoffs. While this might slightly weaken the source of the exclusion restriction, it can be important in practice. In order for the identification

arguments to go through, the discount rate  $\beta$  is assumed to be known. This follows from known nonidentification results in one-player dynamic optimization problems from e.g. Rust (1987) and in dynamic auctions from Jofre-Bonet and Pesendorfer (2003).

### 3 Nonparametric identification of dynamic games

#### 3.1 Dynamic game of incomplete information

In this section, we extend our model to allow for non-trivial dynamics. Our model is similar to the framework proposed by Aguirregabiria and Mira (2007), Berry, Pakes, and Ostrovsky (2003), and Pesendorfer and Schmidt-Dengler (2010) and Pesendorfer, Schmidt-Dengler, and Street (2008). Period returns are defined using a static logit model. The current actions  $a$  and state influence the future evolution of the state variable. We shall restrict attention to Markov perfect equilibrium. The methods that we propose here could be applied to other dynamic games, such as a finite horizon where payoffs and the law of motion are time dependent. These extensions require considerable additional notational complexity.

#### 3.2 The Environment

##### 3.2.1 Payoffs

In the model, there are  $t = 1, \dots, \infty$  time periods. At each time  $t$ , we let  $a_{it} \in \{0, 1, \dots, K\}$  denote the choice for agent  $i$ . We shall assume that the choice set is identical for all agents and does not depend on the state variable. Both assumptions could be dropped at the cost of notational complexity. Let  $s_{i,t} \in S_i$  denote the state variable for agent  $i$  at time  $t$ . As in the previous section,  $S_i$  is a collection of real valued vectors and the state can either be continuous or discrete.

Let  $\varepsilon_{it} = (\varepsilon_{it}(0), \dots, \varepsilon_{it}(K))$  denote a vector of iid shocks to agent  $i$ 's payoffs at time  $t$ . As in the previous section, we shall assume that the error terms are distributed extreme value. Player  $i$ 's period utility function is

$$u_i(a_{it}, a_{-it}, s_t, \varepsilon_{it}) = \Pi_i(a_{it}, a_{-it}, s_t) + \varepsilon_{it}(a_{it}).$$

As in the previous section, we shall develop the model assuming that  $\Pi_i(a_{it}, a_{-it}, s_t)$  is a general function of the state variables rather than a member of a particular parametric family. Let  $\sigma_i(a_i|s)$  denote the probability that  $i$  plays  $a_i$  given that the state is  $s$ . As in the previous section, we define  $\Pi_i(a_{it}, s)$  as  $\Pi_i(a_i, s) = \sum_{a_{-i}} \Pi_i(a_i, a_{-i}, s) \sigma_{-i}(a_{-i}|s)$ .

### 3.3 Value Functions

In the model, the evolution of the state variable depends on the current state and the actions of all players. We assume that the state variable evolves according to a first order Markov process  $\bar{g}(s'|s, a_i, a_{-i})$ . As before,  $s$  is perfectly observed by the agent and the econometrician. Player  $i$  maximizes expected discounted utility using a discount factor  $\beta$ .

Let  $W_i(s, \epsilon_i; \sigma_{-i})$  be player  $i$ 's value function given  $s$  and  $\epsilon_i$ . The value function holds fixed the strategies of the other agents  $\sigma_{-i}$ . The value function then satisfies the following recursive relationship:

$$W_i(s, \epsilon_i; \sigma_{-i}) = \max_{a_i \in A_i} \left\{ \Pi_i(a_i, s) + \epsilon_i(a_i) \right. \\ \left. + \beta \int \sum_{a_{-i}} W_i(s', \epsilon'_i; \sigma_{-i}) \bar{g}(s'|s, a_i, a_{-i}) \sigma_{-i}(a_{-i}|s) f(\epsilon'_i) d\epsilon'_i ds' \right\}. \quad (10)$$

At each state, agents choose an action  $a_i \in A_i$  to maximize expected discounted utility. The term  $\Pi_i(a_i, s) + \epsilon_i(a_i)$  is the current period return from choosing  $a_i$ . The second term captures  $i$ 's utility from future time periods. In our model, agents choose their actions simultaneously. Therefore, agent  $i$ 's beliefs about the evolution of the state given his current information will be  $\sum_{a_{-i}} \bar{g}(s'|s, a_i, a_{-i}) \sigma_{-i}(a_{-i}|s)$ . This integrates out agent  $i$ 's uncertainty about the actions of  $-i$ . The agent also needs to take into account expectations about next periods preference shocks,  $\epsilon'_i$ , by integrating out their distribution using the density  $f(\epsilon'_i)$ .

**Definition 2** *A Markov perfect equilibrium is a collection of policy functions,  $\delta_i(s, \epsilon_i)$  and corresponding conditional choice probabilities  $\sigma_i(a_i|s)$  such that for all  $i$ , all  $s$  and all  $\epsilon_i$ ,  $\delta_i(s, \epsilon_i)$  maximizes the value function  $W_i(s, \epsilon_i; \sigma_{-i})$  defined in (10)*

In a Markov perfect equilibrium, an agent's strategy  $\delta_i(s, \epsilon_i)$  is restricted to be a function of the state  $(s, \epsilon_i)$ . This solution concept restricts equilibrium behavior by not allowing

for time dependent punishment strategies, such as trigger strategies or tit-for-tat which do not depend on payoff relevant state variables. While the Markov perfect equilibrium assumption restricts behavior considerably, it has the advantage that equilibrium behavior can be expressed using familiar techniques from dynamic programming. Since the focus of this paper is on nonparametric identification and estimation, existence of equilibrium will be taken as given in the following analysis.

### 3.4 Nonparametric identification

Next, we turn to the problem of identification of the model. The strategy for identifying the model will be similar to the static model. We begin with some preliminaries by first defining the choice specific value function and deriving some key equations that must hold in our dynamic model.

The starting point of our analysis is to define the choice specific value function

$$V_i(a_i, s) = \Pi_i(a_i, s) + \beta \int \sum_{a_{-i}} W_i(s', \epsilon'_i; \sigma) \bar{g}(s' | s, a_i, a_{-i}) \sigma_{-i}(a_{-i} | s) f(\epsilon'_i) d\epsilon'_i ds'. \quad (11)$$

Similar to (1), the choice specific value function is the expected utility from choosing the action  $a_i$ , excluding the current period error term  $\epsilon_i(a_i)$ . As in the static setting, the term  $\Pi_i(a_i, s)$  integrates out player  $i$ 's expectations about the actions of the other players. In a dynamic setting, however, we have to include the utility from future time periods. We do this by integrating out the value function  $W_i(s', \epsilon'_i; \sigma)$  with respect to next periods private information,  $\epsilon'_i$ , and state  $s'$ . In words, we can interpret the choice specific value function as the returns, excluding  $\epsilon_i(a_i)$ , from choosing  $a_i$  today and then reverting to the solution to the dynamic programming problem (10) in all future time periods. Next, we define the ex ante value function, or social surplus function, as

$$V_i(s) = \int W_i(s, \epsilon_i; \sigma) f(\epsilon_i) d\epsilon_i \quad (12)$$

The ex ante value function is the expected value of  $W_i$  tomorrow given that the state today is  $s$ . In order to compute this expectation, we integrate over the distribution of  $\epsilon_i$  given that the current state is  $s$ .

Using equations (11) and (12), the ex ante and choice specific value functions are related to each other through the following equation

$$V_i(a_i, s) = \Pi_i(a_i, s) + \beta E [V_i(s')|s, a_i]. \quad (13)$$

In the dynamic model, if the state is equal to  $s$ , the ex ante value function is related to the choice specific value function by:

$$V_i(s) = E_{\epsilon_i} \max_{a_i} [V_i(a_i, s) + \epsilon_i(a_i)]. \quad (14)$$

That is, the utility maximizing action maximizes the sum of the choice specific value function plus the private information  $\epsilon_i(a_i)$ . As in the static model, the equilibrium probabilities and the choice specific value functions are related through the following equation

$$\sigma_i(a_i|s) = \frac{\exp(V_i(a_i, s))}{\sum_{a'_i} \exp(V_i(a'_i, s))}. \quad (15)$$

### 3.5 Constructive Proof of Identification

As in the static model, we prove the identification of our model constructively. Our strategy is to assume that the econometrician has knowledge of the population choice probabilities  $\sigma_i(a_i|s)$ . We then show that it is possible to uniquely recover  $\Pi_i(a_i, a_{-i}, s)$  after making appropriate normalizations and checking a rank condition.

As in the static model, we begin by taking the log of both sides of (15). Straightforward algebra implies that

$$\log(\sigma_i(a_i = k|s)) - \log(\sigma_i(a_i = 0|s)) = V_i(a_i = k, s) - V_i(a_i = 0, s) \quad (16)$$

This equation demonstrates that it is possible to recover the choice specific value functions up to a first difference, if we know the population choice probabilities.

Next, it follows from (14) and the properties of the extreme value distribution that:

$$\begin{aligned} V_i(s) &= E_{\epsilon_i} \max_{a_i} V_i(a_i, s) + \epsilon_i(a_i) = \log \sum_{k=0}^K \exp(V_i(k, s)) \\ &= \log \sum_{k=0}^K \exp(V_i(k, s) - V_i(0, s)) + V_i(0, s). \end{aligned} \quad (17)$$

The second equality follows from a property of the multinomial logit specification (derived in e.g. Anderson, DePalma, and Thisse (1992)). Using (16), (17) can also be written more concisely as

$$V_i(s) = -\log \sigma_i(a_i = 0|s) + V_i(0, s). \quad (18)$$

In particular, (18) shows that  $V_i(s)$  is known as soon as  $V_i(0, s)$  is.

We now combine (17) with equation (13) to yield:

$$\begin{aligned} V_i(0, s) &= \Pi_i(a_i = 0, s) + \beta E[V_i(s')|s, a_i = 0] \\ &= \Pi_i(a_i = 0, s) + \beta E[-\log \sigma_i(a_i = 0|s') + V_i(0, s')|s, a_i = 0] \\ &= \Pi_i(a_i = 0, s) + \beta E[-\log \sigma_i(a_i = 0|s') |s, a_i = 0] \\ &\quad + \beta E[V_i(0, s')|s, a_i = 0]. \end{aligned} \quad (19)$$

Next, suppose that we are willing to make the “outside good” assumption as in equation (5). Then equation (16) implies that:

$$\begin{aligned} V_i(0, s) &= \beta E[-\log \sigma_i(a_i = 0|s') + V_i(0, s')|s, a_i = 0] \\ &= \beta E[-\log \sigma_i(a_i = 0|s') |s, a_i = 0] \\ &\quad + \beta E[V_i(0, s')|s, a_i = 0]. \end{aligned}$$

Since the population probabilities  $\sigma_i(a_i = k|s)$  are assumed to be known for the purposes of our identification argument, the term

$$\beta E[-\log \sigma_i(a_i = 0|s') |s, a_i = 0]$$

can be treated as a known constant. Then, equation (19) is a functional equation involving the unknown function  $V_i(0, s)$ . Blackwell’s sufficient conditions (e.g. Theorem 3.3 in Stokey, Lucas, and Prescott (1989)) imply that for fixed  $\sigma_i(a_i|s)$ , (19) is a contraction mapping and therefore there is a unique solution for  $V_i(0, s)$ , which can be computed e.g. using methods discussed in section 5.2. As a result, we have shown that  $V_i(0, s)$  is identified. Moreover,  $V_i(k, s)$  is identified for all  $k$  by substituting  $V_i(0, s)$  into (16). Finally, we note that the ex ante value functions can be identified by (17) or (18) given that we have identified the  $V_i(k, s)$ .

Next, note that (13) implies that

$$\Pi_i(a_i = k, s) = V_i(a_i = k, s) - \beta E [V_i(s') | s, a_i = k]. \quad (20)$$

Our identification arguments imply that both terms on the right hand side of (20) are known. This implies that  $\Pi_i(a_i = k, s)$  is identified. Alternatively, using (13), (16) and (18) one obtains

$$\Pi_i(k, s) = V_i(0, s) + \log [\sigma_i(k|s) / \sigma_i(0|s)] - \beta E [V_i(0, s') - \log \sigma_i(a_i = 0|s') | s, a_i = k],$$

which directly establishes the identification of  $\Pi_i(k, s)$  from the knowledge of  $V_i(0, s)$ .

The rest of identification proof can then follow exactly as in equations (6)-(8), as once  $\Pi_i(k, s)$  is known, the argument with exclusion restrictions of section 2.1 applies to identify  $\Pi_i(k, a_{-i}, s)$ . We simply need to construct the  $\Pi_i(a_i, a_{-i}, s_i)$  from the static choice specific value functions  $\Pi_i(a_i, s)$  by imposing exclusion restrictions.

**Theorem 2** *Suppose that Assumptions 1-2 hold. Also suppose that for each  $s_i$ , there exist  $(K+1)^{n-1}$  points in the support of the conditional distribution of  $s_{-i}$  given  $s_i$ . Assume that the  $(K+1)^{n-1}$  equations defined by (8) are linearly independent. Then the latent utilities  $\Pi_i(a_i, a_{-i}, s_i)$  are identified.*

### 3.6 Nonparametric Shock Distribution

Our results has thus far focused on nonparametric identification and semiparametric estimation of the payoff function taking the parametric known distribution of the shocks as given. As is well known in the discrete choice model literature, without imposing strong identification at infinity assumptions, it is clearly not possible to identify both the mean utility functions and the shock distributions entirely nonparametrically. In Nekipelov, Bajari, and Hong (2010), we show that if we are willing to impose a parametric structure on the mean utility functions, then it is possible to identify and estimate a nonparametric specification of the shock distributions, which is assumed to be fully independent of the state variables. These results apply to both the static game and the dynamic discrete game.



### 3.7 Identification-based estimation procedures

Under the assumption that one has access to a data set from a collection of independent markets  $m = 1, \dots, M$  with at least two periods of observations each, a nonparametric multi-step estimator can be constructed by using the empirical analogue of our identification strategy. The translation from identification arguments to the nonparametric estimator essentially only requires replacing the appropriate conditional expectations with analog sample projections. There are many possible local and global nonparametric smoothing techniques to estimate conditional expectations. For example, series expansions have been a popular choice as most of the recent literature (e.g. Newey (1994) and Chen, Linton, and Van Keilegom (2003)).

In step 1,  $\hat{V}_i(k, s) - \hat{V}_i(0, s)$  can be estimated using (16) given a flexible estimator  $\hat{\sigma}_i(a_i|s)$  of the equilibrium choice probabilities  $\sigma_i(a_i|s)$ , in

$$\hat{V}_i(k, s) - \hat{V}_i(0, s) = \log(\hat{\sigma}_i(k|s)) - \log(\hat{\sigma}_i(0|s)).$$

We also need to construct an estimate of  $g(s'|s, a_i, a_{-i})$ . The details of estimating  $g(s'|s, a_i, a_{-i})$  will vary with the application. In many problems, the law of motion for the state variable is deterministic and therefore does not need to be directly estimated. Another common case is when  $g(s'|s, a_i, a_{-i})$  is defined by a density, which can be modeled by a flexible parametric density or estimated nonparametrically.

In step 2, an estimate of  $\hat{V}_i(0, s)$ , the baseline choice specific value function for  $k = 0$ , can be constructed by iterating on the empirical analogue of equation (19), using a numerical quadrature method possibly in combination with discretization of the state spaces. See e.g. section 5.2. If the numerical quadrature used to implement (19) is integrating against an estimated conditional density function, then Blackwell's sufficient condition for contraction mapping will be automatically satisfied. If the conditional expectation in (19) is approximated by least square projections then iterating on the least square projection might not ensure the contraction mapping property. However, the semiparametric conditional moment estimator that we propose below will still be consistent and does not require iterating on the contraction mapping relation.

In step 3, we evaluate the empirical analogue of (20) to estimate the static choice specific payoff function which we denote as  $\hat{\Pi}_i(k, s)$ . From the previous step, we have constructed

an estimate of  $\widehat{V}_i(0, s)$  and from step 1 we have constructed an estimate of  $\widehat{V}_i(k, s) - \widehat{V}_i(0, s)$ . Putting these two steps together implies that we have an estimate of  $\widehat{V}_i(k, s)$  for all  $i, k, s$ .

The empirical analogue of equations (17) and (20) is then

$$\widehat{\Pi}_i(a_i = k, s) = \widehat{V}_i(a_i = k, s) - \beta \int \left( \log \sum_{k=0}^K \exp(\widehat{V}_i(k, s)) \right) \widehat{g}(s' | s, a_i = k) ds'. \quad (21)$$

Standard methods from numerical integration can be used to compute the integrals.

The final step is to perform the empirical analogue of inverting the linear system (8) in order to estimate the nonparametric mean utilities  $\widehat{\Pi}_i(a_i, a_{-i}, s_i)$ . Recall that the state has to be partitioned as  $s = (s_i, s_{-i})$  and the variables  $s_{-i}$  are assumed not to enter into  $i$ 's mean utilities. This allows us to write  $i$ 's utility as  $\Pi_i(a_i, a_{-i}, s_i)$ . One approach to inverting this system will be to run a local linear regression (see Fan and Gijbels (1992)). Local linear regression is essentially a weighted least squares regressions where the weights are defined using a kernel distance between the observations. The exclusion restrictions guarantee that the standard rank condition from the theory of regression is satisfied.

Without a sufficiently large sample, nonparametric estimators suffer from a curse of dimensionality and may be poorly estimated. Therefore, it might be desirable to have a semiparametric approach to the problem where the transition density is specified nonparametrically while the utility functions  $\Pi_i(a, s_i, \theta)$  are specified to depend on a finite number of parameters. Frequently, applied researchers will assume that utility is linear in the structural parameters:

$$\Pi_i(a, s_i) = \Phi_i(a, s_i)' \theta_{i,a}. \quad (22)$$

Here,  $\Phi_i(a, s_i)$  is a known vector valued function and  $\theta$  is used to weight the elements of the basis function.

In the semiparametric model, steps 1-3 of the nonparametric section are left unchanged. It is only necessary to modify step 4 to include the parametric restrictions in (22).

An advantage of the semiparametric estimator is that it can be shown that  $\widehat{\theta}$  converges to the true parameter value at a rate proportional to the square root of the sample size and has a normal asymptotic distribution. This is a common result in semiparametric estimation. Even though the nonparametric part of our model,  $\widehat{\sigma}_{-i}(a_{-i} | s_{m,t})$  and  $\widehat{\Pi}_i(a_i, s_{m,t})$  can only be estimated at nonparametric slower rates, the payoff parameters  $\theta$  converge at the faster

parametric rates. Derivation of the limit distribution of the multi-step semiparametric estimator can be found in a previous online working paper version of this article.

### 3.8 Unobserved heterogeneity

Unobserved heterogeneity is an important concern for dynamic discrete choice models. A recent insight from this literature (Hu and Shum (2008), Kasahara and Shimotsu (2008)) is that it is sufficient to estimate a reduced form model of conditional choice probabilities and transition probabilities that account for the presence of the unobserved heterogeneity. A variety of such methods are available in the recent literature, some allowing for a fixed number of support points in the distribution of unobserved state variables while others allowing for continuous unobserved state variables. For each of the discrete and continuous support cases of the unobserved state variables, some methods are limited to only non time varying unobserved state variables while other methods might allow for serially correlated unobserved state variables.

In the following, we will take as given the ability to estimate a first stage model of conditional choice probabilities and conditional transition probabilities that incorporate the presence of general (discrete and continuous, time invariant and serially correlated) state variables. Therefore, we will assume that it is possible to use one of the methods available in the existing literature to estimate a reduced form model of  $\hat{\sigma}_i(k|s), \forall i, k$  and  $\hat{g}(s'|s, a)$ , where now  $s'$  and  $s$  include both observed and unobserved state variables that can be either discrete or continuous, either time-invariant or serially correlated.

We now note that the entire nonparametric identification process in section 3.4 and the entire estimation procedure, both nonparametric and semiparametric, described in section 3.7, depend only on the first stage  $\hat{\sigma}_i(k|s), \forall i, k$  and  $\hat{g}(s'|s, a_i)$ . Therefore, as long as the state transition process is assumed to be common across individuals, we can follow exactly the same procedures outlined in sections 3.4 and 3.7 to estimate the primitive mean utility functions  $\Pi_i(a, s_i)$  and  $\Phi_i(a, s_i)' \theta$ . Perhaps the best way to understand this argument is through simulations. Given knowledge of  $\hat{\sigma}_i(k|s), \forall i, k$  and  $\hat{g}(s'|s, a_i)$ , a researcher can generate a data set with as many markets and as many time periods as desired, and apply the estimation procedures described in the previous subsections of section 3.7 to the simulated

data set.

## 4 Efficient Semiparametric and Nonparametric Estimation

In the previous section we described a multi-stage procedure which allows us to estimate both a finite-dimensional and an infinite dimensional specification of the profit function. This procedure is very intuitive because it follows directly from the identification argument. The asymptotic distribution of this estimator also has an explicit analytic structure. However, this approach inherits the disadvantages of many multi-stage estimation techniques. First of all, the standard errors are hard to compute because of propagation of errors from the previous steps of the procedure which will depend on the degree of smoothness of the unknown functions of the model. Second, this multistage estimation procedure is not efficient. It is well known that it is difficult to design multistage estimation procedures that can achieve efficiency bounds, because each subsequent step has to compensate the estimation errors that will arise from previous estimation errors.

In this section we will propose an efficient one step estimation procedure using the framework of conditional moment models. It has the advantage that given the choice of instrument functions and the weighting matrix, practical inference can be performed using standard parametric methods as if a finite dimensional linear parametric model of  $V_i(k, s)$  and  $\Pi(a, s_i)' \theta$  is estimated, as long as the estimation noise in the estimation of  $\sigma_i(k|s)$  is appropriately accounted for.

By formulating the model in a conditional moment framework and making use of the stationary controlled Markov process structure, we can avoid direct estimation of the transition density of the state variable. This simplifies the derivation of the efficiency bound of the model and the statement of the regularity conditions for the efficient estimator. The efficient estimation procedure is applicable to both semiparametric and nonparametric models.

## 4.1 Semiparametric efficient estimation

The conditional moment formulation is derived from the Bellman equations for individual players. Recall the Bellman equations of interest:

$$V_i(k, s) = \Pi_i(k, s; \gamma) + \beta \int \sum_{a_{-i} \in A_{-i}} \sigma_{-i}(a_{-i}|s) \log \left[ \sum_{l=0}^K \exp(V_i(l, s')) \right] g(s'|s, a_i = k, a_{-i}) ds',$$

where

$$\sigma_i(a_i = k|s) = \frac{\exp(V_i(k, s))}{\sum_{l=0}^K \exp(V_i(l, s))},$$

for  $i = 1, \dots, n$  and  $k = 0, \dots, K$ . Denote  $d^{i,l}$  the dummy for choice  $l$  by player  $i$ . We can use the second equation to substitute it into the first one, which leads to  $n \times K$  conditional moment equations for each  $(T - 1) \times M$  observations:

$$E \left[ d_{m,t}^{i,k} (V_i(0, s_{m,t}) - \beta V_i(0, s_{m,t+1}) + \beta \log \sigma_i(0|s_{m,t+1})) - d_{m,t}^{i,k} (1 - d_{m,t}^{i,0}) [\Pi_i(a_i, a_{-i}, s_{m,t}; \gamma) + \log \sigma_i(0|s_{m,t}) - \log \sigma_i(a_i|s_{m,t})] \middle| s_{m,t} \right] = 0. \quad (23)$$

Together with the following  $n \times (K + 1)$  moment conditions for each  $T \times M$  observations,

$$E \left( d_{m,t}^{i,k} | s_{m,t} \right) = \sigma_i(k|s_{m,t}), \quad (24)$$

(23) and (24) form a system of conditional moment restrictions that fully characterize the implications from the structural dynamic discrete choice model. This system of conditional moment restrictions can be used to obtain asymptotically normal semiparametric estimators that can achieve the semiparametric efficiency bound by adapting the recipe prescribed in Ai and Chen (2003). In their notation of  $E[\rho(w_{m,t}, \gamma, V(\cdot), \sigma(\cdot)) | s_{m,t}] = 0$ , where  $w_{m,t}$  are all the random variables in the model,  $\gamma$  are the finite dimensional parameters,  $V(\cdot)$  and  $\sigma(\cdot)$  are the infinite dimensional unknown parameters, we can write, for  $h(\cdot) = (V(\cdot), \sigma(\cdot))$ :

$$\rho(w_{m,t}, \gamma, h(\cdot)) = (\rho_1(w_{m,t}, \gamma, V(\cdot), \sigma(\cdot)))', \rho_2(w_{m,t}, \gamma, V(\cdot), \sigma(\cdot))',$$

where  $\rho_1$  is the  $T \times m \times n \times K$  dimensional collection of  $d_{m,t}^{a_i, k} - \sigma_i(k|s_{m,t})$ , and

$$\rho_2(w_{m,t}, \gamma, V(\cdot), \sigma(\cdot)) = d_{m,t}^{i,k} (V_i(0, s_{m,t}) - \beta V_i(0, s_{m,t+1}) + \beta \log \sigma_i(0|s_{m,t+1})) - d_{m,t}^{i,k} (1 - d_{m,t}^{a_i, 0}) [\Pi_i(a_i, a_{-i}, s_{m,t}; \gamma) + \log \sigma_i(0|s_{m,t}) - \log \sigma_i(a_i|s_{m,t})].$$

Strictly speaking, the original efficiency bound in Ai and Chen (2003) requires the same conditioning variables in the moment restrictions. However, for each  $m = 1, \dots, M$ , the collection of moment conditions in (23) and (24) involve different conditioning variables  $s_t, t = 1, \dots, T - 1$ . In addition, the moment condition (24) can also be used at time  $t = T$ . Fortunately, the Markov structure of the model implies that conditioning on  $s_t$  is equivalent to conditioning on  $s_1, \dots, s_t$ . Therefore the conditioning sets form an increasing sequence of sigma-algebra, and the sequential conditional moment model of Ai and Chen (2009) applies to substantiate the semiparametric efficiency bound.

The conditional moment restrictions in (23) and (24) can be transformed into unconditional moments by forming an instrument matrix  $z_{m,t}$  using the state variables  $s_{m,t}$ , its lags  $s_{m,t-\tau}$  and polynomial powers  $s_{m,t}$  and its lags, such that the number of instruments in  $z_{m,t}$  increases at appropriate rates as the sample size increases to infinity. Equations (23) and (24) implies the following moment vectors with elements

$$E \left[ d_{m,t}^{i,p} z_{m,t} \left( V_i(0, s_{m,t}) - \beta V(0, s_{m,t+1}) + \log \frac{\sigma_i(a_i | s_{m,t})}{\sigma_i(0 | s_{m,t})} - \beta \log \frac{\sigma_i(a_i | s_{m,t+1})}{\sigma_i(0 | s_{m,t+1})} \right. \right. \\ \left. \left. - \left( 1 - d_{m,t}^{i,0} \right) [\Pi_i(a_i, a_{-i}, s_{m,t}; \gamma) - \beta \log \sigma_i(a_i | s_{m,t+1})] \right. \right. \\ \left. \left. + d_t^{i,0} \beta \log \sigma_i(0 | s_{m,t+1}) \right) \right] = 0,$$

and  $E z_{m,t} \left( d_{m,t}^{i,p} - \sigma_i(p | s_{m,t}) \right) = 0$ . To estimate  $\gamma$  we can follow two steps.

**Step 1**

We approximate the conditional choice probabilities using orthogonal series:

$$\sigma_i(a_i = p | s) = q^{k_1(MT)'}(s) b_{i,p}^1 + \Delta_{k_1(MT)},$$

and approximate the value function similarly

$$V_i(0, s) = q^{k_2(MT)'}(s) b_i^2 + \Delta_{k_2(MT)},$$

where  $\Delta_{k_1(M)}$  and  $\Delta_{k_2(MT)}$  are numerical approximation errors that decrease to zero as  $k_1(MT)$  and  $k_2(MT)$  increase to infinity with  $MT$  at appropriate rates.

**Step 2**

Next we form an instrument  $z_{m,t}$  by stacking an orthogonal series of functions of the state

variables  $s_{m,t}$ ,  $(q_0(s_{m,t}), \dots, q_{k_3(MT)}(s_{m,t}))$ . This produces an over-identified empirical moment vector with the elements, for  $b = (b_1^{i,p}, b_2^i, \forall i, p)$ ,

$$\widehat{\varphi}(\gamma, b) = \sum_{m,t} \varphi_{m,t}(\gamma, b) \quad \text{where} \quad \varphi_{m,t}(\gamma, b) = \rho(w_{m,t}, \gamma, b) \otimes z_{m,t}.$$

Then we introduce a weighting matrix  $\mathcal{W}$  with both row and column dimensions  $\dim(z_{m,t}) \times \dim(\rho)$ . In the simplest case we can use the identity matrix in lieu of  $\mathcal{W}$ . Using a given weighting matrix we form a GMM objective and minimize it with respect to parameters of interest  $\gamma$  as well as the parameters of the expansion of the value function  $\min_{\gamma, b} \widehat{\varphi}(\gamma, b)' \mathcal{W} \widehat{\varphi}(\gamma, b)$ . In particular, as shown in Akerberg, Chen, and Hahn (2011), if we let  $Z \equiv (z_{m,t}, \forall m, t)'$  denote the data matrix for the instruments, the following choice of the weighting matrix

$$\mathcal{W} = I \otimes (Z'Z)^{-1} \left( \sum_{m,t} \Omega_{m,t}^{-1} \otimes z_{m,t} z_{m,t}' \right) I \otimes (Z'Z)^{-1}.$$

yields the semiparametric minimum distance estimators of Ai and Chen (2003). In the above  $\Omega_{m,t}$  is a candidate estimate of the conditional variance covariance matrix of  $\rho(w_{m,t}, \gamma, h(\cdot))$  given  $s_{m,t}$ . When  $\Omega_{m,t} \equiv I$  an identity matrix, the estimator becomes a nonlinear two stage least square estimator. When  $\Omega_{m,t} = \Omega$  is homoscedastic across observations, this becomes a nonlinear three stage least square estimators. Semiparametric efficiency bound is achieved when  $\Omega_{m,t}$  is a consistent estimate of  $Var(\rho(w_{m,t}, \gamma, h(\cdot)) | s_{m,t})$ , in which case it becomes a heteroscedasticity weighted nonlinear three stage least square estimator. When  $\rho(\cdot)$  is a scalar, the semiparametric efficient minimum distance estimator is a weighted nonlinear two stage least square estimator.

**Remark 1:**

By appropriate choices of the instrument functions and the weighting matrix, the conditional moment framework also incorporates the multistage procedure in the previous section as special cases. If the same orthogonal series is used in approximating  $V_i(0, s_{m,t})$ ,  $\sigma_i(p|s_{m,t})$  and in obtaining the instruments, and if  $k_1(MT) = k_2(MT) = k_3(MT)$ , the instrumented moment conditions (24) are exactly identifying, and  $\sigma_i(p|s)$  are computed from least square

regressions:

$$\widehat{\sigma}_i(a_i = p | s) = q^{k(MT)'}(s) \left( \sum_{m,t} q^{k(M)}(s_{m,t}) q^{k(MT)'}(s_{m,t}) \right)^{-1} \sum_{m,t} q^{k(MT)}(s_{m,t}) d_t^{a_i, p}.$$

Given the estimate of  $\widehat{\sigma}_i(a_i = p | s)$ , the component of the instrumented moment condition  $\rho_2(w_{m,t}, \cdot)$  that corresponds to  $k = 0$  is also exactly identifying and depends only on  $V_i(0, s)$ . Hence  $V_i(0, s)$  can be estimated by a single equation two stage least square regression with dependent variables  $\beta \log \widehat{\sigma}_i(0 | s_{m,t+1})$ , independent variables  $q^{k(MT)}(s_{m,t}) - \beta q^{k(MT)}(s_{m,t+1})$  and instrument matrix  $q^{k(MT)}(s_{m,t})$ . Subsequently, given estimates of  $\widehat{V}_i(0, s)$  and  $\widehat{\sigma}_i(a_i = p | s)$ , the parameters

$$\gamma_i^p = \left( \gamma_{i, a_{-i}}^p, \forall a_{-i} \right),$$

for  $i = 1, \dots, n$ ,  $p = 1, \dots, K$  in a linear profit function specification  $\Pi_i(p, a_{-i}; \gamma) = \Phi_i(p, a_{-i})' \gamma_i^p$ , can be estimated by single equation linear two stage least square regression methods when  $\Omega(x_i) \equiv I$ , with dependent variables

$$Y_{i,p,m,t} = \widehat{V}_i(0, s_{m,t}) - \beta \widehat{V}_i(0, s_{m,t+1}) + \log \frac{\widehat{\sigma}_i(a_i | s_{m,t})}{\widehat{\sigma}_i(0 | s_{m,t})} - \beta \log \frac{1}{\widehat{\sigma}_i(0 | s_{m,t+1})}$$

and the vector of independent variables  $\mathbf{X}_t$  with elements

$$X_{i,p,m,t} = - \left( 1 - d_{m,t}^{i,0} \right) \Phi_i(p, a_{-i}),$$

and instrument matrix  $Z = (z_{m,t}, \forall m, t)'$ . Efficiency can be improved by weighted 2SLS or weighted 3SLS by choosing  $\widehat{\Omega}(x_i)$  appropriately.

**Remark 2:**

The semiparametric efficient minimum distance estimator of Ai and Chen (2003) can be interpreted both in light of weighted nonlinear three stage least square estimator and the efficient instrument method of Newey (1990) for finite dimensional parameters. The semiparametric minimum distance objective function can be equivalently rewritten as

$$\sum_{m,t} \widehat{\rho}(s_{m,t}, b, \gamma)' \widehat{\Omega}_{m,t}^{-1} \widehat{\rho}(s_{m,t}, b, \gamma),$$



where  $\hat{\rho}(s_{m,t}, b, \gamma)$  is an estimate of  $E(\rho(w_{m,t}, b, \gamma) | s_{m,t})$ ,

$$\hat{\rho}(s, b, \gamma) = z (Z'Z)^{-1} \sum_{m,t} z_{m,t} \rho(w_{m,t}, b, \gamma)' .$$

Its first order condition resembles the efficient instrument estimator of Newey (1990):

$$\sum_{m,t} \frac{\partial}{\partial (b, \gamma)} \hat{\rho}(s_{m,t}, b, \gamma)' \hat{\Omega}_{m,t}^{-1} \hat{\rho}(s_{m,t}, b, \gamma) .$$

The efficient instrument estimator of Newey (1990) only differs in using  $\rho(s_{m,t}, b, \gamma)$  in place of the second  $\hat{\rho}(s_{m,t}, b, \gamma)$  in light of the law of iterated expectation, and instead uses the first order condition of

$$\sum_{m,t} \frac{\partial}{\partial (b, \gamma)} \hat{\rho}(s_{m,t}, b, \gamma)' \hat{\Omega}_{m,t}^{-1} \rho(s_{m,t}, b, \gamma) .$$

The following theorem adapts the semiparametric efficiency bound in Ai and Chen (2003) to our model. In our model the unknown function  $V(\cdot)$  enters non-linearly as a function of the state variable in the period  $t$  and in the period  $t + 1$ .

Using the results from Ai and Chen (2003) we can provide the semiparametric efficiency bound for estimating the parameter  $\gamma$  of the payoff function. Denote

$$\Sigma_0(s_{m,t}) = \text{Var}(\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}) .$$

The semiparametric efficiency bound expressed in theorem 3 will depend on the functional derivatives of the moment conditions  $\rho_1$  in (23) and  $\rho_2$  in (24) on the unknown functions  $h_1^i(\cdot) = V_i(0, \cdot)$  and  $h_2^{i,k}(\cdot) = \sigma_i(k|\cdot)$ . The functional derivative of the conditional moment functions with respect to these unknown functions can be expressed using the linear expectation operator

$$\mathcal{P}_i^k \circ f = E[f(s_{m,t+1}) | s_{m,t} = s, a_{m,t}^i = k] ,$$

where expectation is defined for the conditional density

$$\sum_{a_{-i}} g(s_{m,t+1} | s_{m,t} = s, a_i = k, a_{-i}) \sigma_{-i}(a_{-i} | s_{m,t} = s) .$$

The operator  $\mathcal{P}_i^k \circ f$  is assumed to have a discrete spectrum with eigenfunctions  $\left\{ \Theta_j^{i,k}(s) \right\}_{j=0}^{\infty}$  and eigenvalues  $\left\{ \lambda_j^{i,k} \right\}_{j=0}^{\infty}$  different from zero. Then we can find that

$$\frac{dE \left[ \rho_1^{i,k}(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t} \right]}{dh_1^i} [\psi] = \sigma_i(k|s) \sum_{j=0}^{\infty} \psi_j \left( 1 - \beta \lambda_j^{i,k} \right) \Theta_j^{i,k}(s),$$

for all sequences of real numbers  $\psi$  which belong to  $\mathcal{H} = \left\{ \psi \mid \sum_{j=0}^{\infty} |\psi_j| \left\| \Theta_j^{i,k}(s) \right\| < \infty \right\}$ ,

Furthermore, we also calculate that

$$\frac{dE \left[ \rho_1^{i,k}(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t} \right]}{dh_2^{i,0}} [\psi] = \beta E \left[ d_{m,t}^{i,k} \frac{1}{\sigma_i(0|s_{m,t+1})} h_2^{i,0}(s_{m,t+1}) | s_{m,t} \right].$$

and for  $k \neq 0$  the linear derivative of,

$$\frac{dE \left[ \rho_1^{i,k}(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t} \right]}{dh_2^{i,k}} [\psi] = h_2^{i,k}(s_{m,t}).$$

Finally, for all  $k$ ,

$$\frac{dE \left[ \rho_2^{i,k}(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t} \right]}{dh_2^{i,k}} [\psi] = -h_2^{i,k}(s_{m,t}).$$

The functional derivatives in the direction of the unknown functions  $\frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{dh} [\psi]$  are formed by stacking the above individual components together.

Then for each component of  $\gamma$  solve the minimization problem

$$\min_{\psi^{(j,0)} \in \mathcal{H}} E \left\{ \left( \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{d\gamma_j} - \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{dh} [\psi^{(j,0)}] \right) \Sigma_0(s_{m,t})^{-1} \right. \\ \left. \times \left( \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{d\gamma_j} - \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{dh} [\psi^{(j,0)}] \right) \right\}.$$

Form the vector

$$D_{\psi^{(0)}}(s_{m,t}) = \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{d\gamma'} - \frac{dE[\rho(w_{m,t}, \gamma_0, h_0(\cdot)) | s_{m,t}]}{dh} [\psi^{(0)}].$$

The following theorem follow directly from the result provided in Ai and Chen (2003):

**Theorem 3** *The semiparametric efficiency bound for estimation of  $\gamma$  in equation (23) can be found as*

$$V(\gamma) = E \left[ D_{\psi(0)}(s_{m,t})' \Sigma_0(s_{m,t})^{-1} D_{\psi(0)}(s_{m,t}) \right]^{-1}.$$

## 4.2 Asymptotic distribution for semiparametric estimator

We impose the following regularity assumptions on the functions in the model to assure that the two-stage conditional moment-based estimation method delivers consistent estimates for the Euclidean parameter in the per period payoff function as well as the non-parametric estimate of the continuation value of players.

### Assumption 3

1. *Parameter space  $\Gamma$  is a convex compact set. Profit function  $\Pi_i(a_i, a_{-i}, s; \gamma)$  is continuous in  $\gamma$  for each  $(a_i, a_{-i}) \in \mathcal{A}$ . Moreover, for each  $\gamma \in \Gamma$  profit function is bounded:*

$$\sup_{a \in \mathcal{A}, s \in \mathcal{S}} |\Pi_i(a_i, a_{-i}, s; \gamma)| < \infty.$$

2. *The data  $\left\{ \{a_{1t}, \dots, a_{nt}, s_t, s_{t+1}\}_{t=1}^{T-1} \right\}_{m=1}^M$  are i.i.d. generated by the stationary distribution determined by Markov transition kernel for the state variable.*
3. *The approximating series expansion  $\{q^{k(MT)}\}$  forms a basis in  $\mathcal{C}^{k(MT)}(\mathcal{S})$ , such that the eigenvalues of  $E[q^{k(MT)}(s_{t+1}) q^{k(MT)'}(s_{t+1}) | s_t = s]$  are bounded away from zero for all  $s \in \mathcal{S}$ . The operator*

$$\mathcal{P}_i \circ f = E[f(s_{m,t+1}) | s_{m,t} = s, a_i],$$

where expectation is defined for the conditional density

$$\sum_{a_{-i}} g(s_{m,t+1} | s_{m,t} = s, a_i = k, a_{-i}) \sigma_{-i}(a_{-i} | s_{m,t} = s),$$

which has a discrete spectrum with eigenfunctions  $\left\{ \Theta_j^{i,k}(s) \right\}_{j=0}^{\infty}$  such that for each  $j$  we can find  $j' \leq j$  for which  $\langle q_j^{k(MT)}, \Theta_j^{i,k} \rangle \neq 0$ . In addition,

$$\limsup_{m \rightarrow \infty} E \left[ (TM)^{-1/2} \left( 1 + \beta \Lambda_j^{i,k} \right) q_j^{k(MT)} \right] < \infty.$$

4. The value function  $V_i(s_t)$  is piece-wise continuous on  $\mathcal{S}$  and bounded. Moreover, for each  $V_i(\cdot) \in \mathcal{V}$  there exists a vector  $\mu \in \mathbb{R}^{k(n)}$  such that  $E \left[ (V(s_t) - \mu' q^k)^2 \right] = o(1)$ .
5. For a given  $V(\cdot)$  and transition density, there exists a unique solution  $\gamma \in \Gamma$  to the system of equations

$$E[\varphi_i(s_t, s_{t+1}, a; V_i, \gamma) \mid s_t] = 0,$$

for  $i = 1, \dots, n$ .

These assumptions allow us to apply the results from Newey and Powell (2003) for each order of approximation  $k(MT)$ . By the appropriate choice of basis we can guarantee that the approximation error is negligible as compared to the estimation error. The estimation problem is linear in parameters: expansion coefficients for  $V(\cdot)$  and the Euclidean parameter  $\gamma$ . For each finite approximation order  $k(MT)$  we can assure that the estimated parameters are consistent estimates for the functions given the order of approximation. When  $MT \rightarrow \infty$  approximation error approaches zero and the estimated coefficients will be consistent for the true coefficients. Given that by assumption the value function admits consecutive approximations in the basis  $\{q^K(s)\}$  for each  $K \in \mathbb{N}$ , the fitted values  $\widehat{b}^{K'} q^K(s)$  will be consistent for the true value function in the limit.

We can provide a similar set of assumptions that will assure the asymptotic normality of the estimates.

**Assumption 4** 1. There exists a metric  $\|\cdot\|_s$  such that the product space  $\mathcal{V} \times \Gamma$  is compact. Moreover, the space  $\{q^\infty(s)\} \times \Gamma$  is dense in  $\mathcal{V} \times \Gamma$  for the chosen metric.

2. For the covering number in the family of the moment functions defined by consecutive series approximations

$$\log N \left( \varepsilon, \{q^{k(MT)}(s)\} \times \Gamma, \|\cdot\|_s \right) \leq Ck(MT) \log \left( \frac{k(MT)}{\varepsilon} \right).$$

3. The weighting matrix can be estimated consistently such that

$$\left\| \widehat{\mathcal{A}}(s, d^a) - \mathcal{A}(s, d^a) \right\| = o_p \left( (MT)^{-1/4} \right).$$

Moreover for each  $\|\mu - \mu_0^{k(MT)}\| < C(MT)^{-1/4}$  and each  $\|\gamma - \gamma_0\| < C(MT)^{-1/4}$

$$\left\| \left( \widehat{\mathcal{A}}(s, d^a) - \mathcal{A}(s, d^a) \right) \varphi \left( s', s, a; \mu' q^{k(m)}(s), \gamma \right) \right\| = o_p \left( (MT)^{-1/4} \right)$$

4. The variance of the moment function  $\text{Var} \left( \varphi \left( s', s, a; V_0, \gamma_0 \right) \mid s \right)$  is positive definite for all  $s \in \mathcal{S}$ .

5. For each direction  $h \in \mathcal{C}^{k(m)}(\mathcal{S})$  we define the directional derivative of the moment function as a vector  $\partial_h \varphi = \left( \frac{\partial \varphi}{\partial \gamma}, \left( \frac{\partial \varphi}{\partial V} \right)_h \right)$ , where

$$\frac{\partial \varphi_i}{\partial \gamma} = \frac{\partial \Pi_i(a_i, a_{-i}, s; \gamma)}{\partial \gamma}, \quad \text{and} \quad \left( \frac{\partial \varphi_i}{\partial V_i} \right)_h = \sum_{j=0}^{\infty} h_j \left( 1 - \beta \lambda_j^{i,k} \right) \Theta_j^{i,k}(s).$$

We assume that in the ball of radius  $C(MT)^{-1/4}$  around the true value  $(V_0, \gamma_0)$  in  $\mathcal{V} \times \Gamma$  the directional derivative  $\partial_h \varphi$  is Hölder-continuous with respect to norm  $\|\cdot\|_s$  and bounded above by a linear functional of  $h$ ,  $F[h]$  such that  $E[F[h]] < \infty$ . Choose  $h^*$  such that

$$E \left\{ \partial_h \varphi(V_0, \gamma_0)' E \left[ \mathcal{A}(s_{t+1}, d^a) \mathcal{A}(s_{t+1}, d^a)' \mid s_{t=s} \right] \partial_h \varphi(V_0, \gamma_0) \right\}$$

is minimized with respect to  $h$ . Then uniformly in the chosen ball

$$E \left( \|\partial_{h^*} \varphi(V_0, \gamma_0) - \partial_{h^*} \varphi(V, \gamma)\|^2 \right) = o \left( (MT)^{-1/4} \right),$$

where we use a standard Euclidean norm.

The following theorem is an immediate consequence of Ai and Chen (2003), which we state without proof.

**Theorem 4** Under assumptions 3 and 4, for  $\hat{\gamma}$  defined in steps 1 and 2 of the previous section,  $\hat{\gamma} \xrightarrow{p} \gamma_0$ , and for  $V(\gamma)$  given in theorem 3

$$\sqrt{MT} (\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, V(\gamma)).$$

### 4.3 Nonparametric estimation

The moment equation (23) in general does not depend on the dimensionality of the payoff parameter  $\gamma$ . While making  $\gamma$  infinite-dimensional will cost the loss of the parametric convergence rate, a fully nonparametric estimation procedure of the per-period payoff function is feasible because of the identification results, and is implementationally essentially equivalent to the efficient estimation procedure in the semiparametric case.

**Step 1** Estimate conditional choice probabilities non-parametrically using the orthogonal series representation:

$$\widehat{\sigma}_i(a_i = p | s) = q^{k(MT)'}(s) \left( \sum_{m,t} q^{k(MT)}(s_{m,t}) q^{k(MT)'}(s_{m,t}) \right)^{-1} \sum_{m,t} q^{k(MT)}(s_{m,t}) d_t^{i,p}.$$

#### Step 2

Consider a series approximation for the value function

$$V_i(a_i = p, s) = q^{k(MT)'}(s) b^{i,p} + \Delta_{k(MT)},$$

where  $\Delta_{k(MT)}$  is a numerical approximation error, and consider a similar expansion for the payoff function

$$\Pi_i(a_i = p, a_{-i}, s) = q^{k(MT)'}(s) \gamma^{i,p,a_{-i}} + \Delta'_{k(MT)}.$$

For implementability of the procedure at this step we need the payoff function to be continuous (or, at least, to have a finite set of points of first-order discontinuity). Next we form an instrument  $z_{m,t}$  by stacking the state variables  $s_{m,t}$  across the markets to form vectors  $s_t$ , and then choosing the linearly independent subset of vectors from the collection

$$(q_0(s_{m,t-\tau}), \dots, q_k(M)(s_{m,t-\tau})),$$

for all  $0 \leq \tau \leq t-1$ . Additional instruments come from other functions of  $a_{m,t}$  and the estimated choice probabilities  $\widehat{\sigma}_i(j|s_{m,t+1})$ . This produces an empirical moment vector with

$2k(M)$  unknown expansion coefficients with the elements

$$\begin{aligned} \widehat{\varphi}_{i,p}(\gamma, b) = & \frac{1}{T} \sum_{t=1}^{T-1} d_t^{i,p} z_t \left( b^{i,p'} (q^{k(M)}(s_{m,t}) - \beta q^{k(M)}(s_{m,t+1})) \right. \\ & - \left( 1 - d_t^{i,0} \right) [q^{k(M)'}(s) \gamma^{i,p,a-i} - \beta \log \widehat{\sigma}_i(a_i | s_{m,t+1})] \\ & \left. + d_t^{i,0} \beta \log \left( 1 - \sum_{j=1}^K \widehat{\sigma}_i(j | s_{m,t+1}) \right) \right). \end{aligned}$$

Then we introduce a weighting matrix  $\mathcal{W}$  with dimensions  $n \times n$ . In the simplest case we can use the identity matrix in lieu of  $\mathcal{W}$ . For this weighting matrix we form a GMM objective and minimize it with respect to parameters of interest  $\gamma$  as well as the parameters of the expansion of the value function

$$\min_{\gamma, b} \widehat{\varphi}(\gamma, b)' \mathcal{W} \widehat{\varphi}(\gamma, b).$$

In this estimation procedure the object of interest is the entire surface of the profit function, which can be computed as

$$\widehat{\Pi}_i(a_i = p, a_{-i} s) = q^{k(M)'}(s) \widehat{\gamma}^{i,p,a-i}.$$

We need to determine the conditions that assure consistency and non-degeneracy of the asymptotic distribution of the pointwise estimate of the payoff function as well as find the rate of convergence of the estimator. Previous we imposed conditions that assure convergence of the semiparametric estimator. We can supplement them with additional assumptions which will provide consistency and asymptotic normality in the non-parametric case.

- Assumption 5**
1. *The payoff function  $\Pi_i(a_i, a_{-i}, \cdot)$  belongs to the functional space  $\mathcal{C}^p(S)$  for  $p > 1$ . Moreover, the orthocomplement of projecting the payoff function onto some Hilbert space  $\mathcal{H}$ , defined by the set of basis functions  $\{q_t(\cdot)\}_{t=0}^p$  with the scalar product  $\langle \cdot, \cdot \rangle$  has a norm in  $\mathcal{C}^\infty(S)$  decreasing in  $p$ . Moreover its projection on the first  $p$  basis vectors converge absolutely, uniformly in the argument as  $p \rightarrow \infty$ .*
  2. *For a truncation sequence  $k(m) < m^{2r}$  the error of approximation of  $\Pi_i(\cdot)$  and  $V(\cdot)$  by the basis function  $\{q_t(\cdot)\}_{t=0}^{k(m)}$  is  $o(m^{-2r})$  with respect to the norm implied by the scalar product in  $\mathcal{H}$ .*

3.  $\widehat{\sigma}_i(a_i|a_i, \cdot)$  is asymptotically normal pointwise in  $\Omega$  and converges at rate  $q$ . The truncation sequence  $k(m)'$  giving the convergence rate  $m^q$  is  $o(k(m)')$ . The approximation error of  $h_i(\cdot)$  with respect to the norm in  $\mathcal{H}$  is of order smaller than  $m^q$ .

This set of assumptions allows us to formulate the following theorem, which is proven in the appendix. The asymptotic variances  $\omega_v^2$  and  $\omega_\pi^2$  are also defined in the appendix.

**Theorem 5** *Given assumption 3, 4 and 5,*

$$m^{\min\{q,r\}} \left( \widehat{V}_i^{k(m)}(s) - V_i(s) \right) \xrightarrow{d} N(0, \omega_v^2),$$

and

$$m^{\min\{q,r\}} \left( \widehat{\Pi}_i^{k(m)}(a_i, a_{-i}, s) - \Pi_i(a_i, a_{-i}, s) \right) \xrightarrow{d} N(0, \omega_\pi^2).$$

## 5 Simulations

To demonstrate the performance of proposed estimators in finite samples, we conduct two sets of numerical simulations. The first set is a simple two by two entry game with discrete state variables and the second set is a single agent dynamic discrete choice model with continuous state variables.

### 5.1 Simulation setup

In the first set of numerical simulations, each of the two players has one state variable that takes two possible values. Each player simultaneously decides whether to enter a market. The payoff to not entering into the market is normalized to zero regardless of the action of the competing player. We construct the payoff function as  $\Pi_i(a_i, a_{-i}) = U_1 \mathbf{1}(a_i = 1) + U_2 \mathbf{1}(a_{-i} = 1)$  where  $U_1$  and  $U_2$  for each Monte Carlo sample are taken as independent draws from the uniform distribution between -2 and 2. The payoffs of players are independent across both the combination of the states and across the actions of the competing players. Therefore, we do not impose restrictions on how the action of the competing player affects the payoff to entering into the market, and allow the actions of both players to be either substitutes or complements. The transition probability matrices



for a new state condition on the previous state and the actions of both players are also randomly generated from uniform distributions between 0 and 1. They are normalized so that the transition probability matrix is a proper stochastic matrix. The discount rate is set to 0.9.

Once generated, the payoff matrix and the transition probability matrices are held constant across the simulation runs. Following the recipe described in the estimation section, we first estimate the entry probabilities from independently generated data on the entry indicators, and then invert out the choice specific continuation value function and the choice specific static expected utility function. Finally, the primitive payoffs are recovered from the choice specific static expected utility functions.

Tables 1 to 4 report the results across 1000 simulation runs. The number of markets (nmarket), reported in the following tables refer to the number of observations (markets) generated for each combination of the state variables. The columns labelled “1st quartile”, “median”, “mean”, “3rd quartile” and “std” refer to the deviation of the estimates from the true parameters.

These tables show that the estimator performs well in finite sample, and that the amount of estimation error decreases monotonically as the sample size increases.

In the second set of numerical simulations for a single agent dynamic discrete choice model. We construct this exercise in a “reverse” way. We generate the data from a reduced form system of choice probabilities. Given that we know the functional form of the reduced-form probabilities, we can also use the methods of numerical integration to recover the per-period payoff function that corresponds to the specific choice probabilities and a specific state transition process. The goal of this empirical exercise is to compare the payoff function estimated from the sample, generated by the state variable and the policy function using our one-stage estimation method and the utility function that we can recover by numerically solving the first-order condition for the player. We keep indexing variables by  $i$  to maintain the coherence with our theoretical analysis. The state variable follows a continuous distribution and evolves continuously according to a normal AR(1) process:

$$s_t = \varphi(a_i) s_{t-1} + \sigma \varepsilon_t,$$

where  $\varepsilon_t$  is a standard normal random variable and  $\varphi(a_i) = 0.81(a_i = 0) + 0.31(a_i = 1)$ .

The probability of choosing action 1 is assumed to take the following flexible functional form:

$$\sigma_i(a_i = k | s_t) = \alpha_{0ik} + \alpha_{1ik}s_t + \alpha_{2ik}s_t^2 + \sum_{j=1}^J [\beta_{0j} + \beta_{1j} \cos(p_j s_t) - \beta_{2j} \sin(p_j s_t)],$$

where parameters  $\alpha$  are fixed. Now for known state transition and choice probabilities we can recover the corresponding per period payoff function using a high-order finite-point approximation formulas for integrals. We normalize the payoff  $\Pi_i(\cdot, a_i = 0) = 0$  and aim at recovering the function  $\Pi_i(\cdot, a_i = 1)$ . We begin with describing the numerical computation algorithm.

## 5.2 Determining the base value function

For each player  $i$  the value function associated with choice 0 can be expressed as

$$V_i(0, s) = \beta \int_{-\infty}^{+\infty} \log \left[ \sum_{r=0}^K \exp(V_{i,r}(s')) \right] \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s' - \varphi(0)s)^2}{2\sigma^2}} ds'.$$

Using the relation  $\sigma_i(k | s) = \frac{\exp(V_{i,k}(s))}{\sum_{r=0}^K \exp(V_{i,r}(s))}$ , this expression can be written as a functional relation to solve for the continuation value function:

$$V_i(0, s) = \beta \int_{-\infty}^{+\infty} [V_0(0, s') - \log \sigma_i(0 | s')] \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s' - \varphi(0)s)^2}{2\sigma^2}} ds'.$$

The value function will be approximated on a discrete uniform grid using linear extrapolation and the integral will be approximated by a Gauss-Hermite Gaussian quadrature method. The value function for the grid points will be solved from a system of linear equations. In particular, by a change of variables

$$\begin{aligned} V_i(0, s) &= \frac{\beta}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \left[ V_i\left(0, \sqrt{2}\sigma x + \varphi(0)s\right) - \log \sigma_i\left(0 \mid \sqrt{2}\sigma x + \varphi(0)s\right) \right] e^{-x^2} dx \\ &\approx \frac{\beta}{\sqrt{\pi}} \sum_{n=1}^N \omega_n \left[ V_i\left(0, \pm\sqrt{2}\sigma x_n + \varphi(0)s\right) - \log \sigma_i\left(0 \mid \pm\sqrt{2}\sigma x_n + \varphi(0)s\right) \right], \end{aligned}$$

where  $\omega_n$  are the weights and  $x_n$  are the points of  $2N$ -point Gauss-Hermite quadrature approximation for the integral of interest. We aim to solve for the value function at a uniform grid  $\mathbf{S}_G = \{s_1, s_2 \dots, s_G\}$  for the state variable:  $V_i(0, s_g) = V_{i,0,g}$ . For numerical computations we will use linear interpolation. The intermediate values of the value function will be approximated by linear interpolation: for instance, if  $s \in [s_g, s_{g+1}]$  then  $V_i(0, s) \approx V_{i,0,g} + \frac{V_{i,0,g+1} - V_{i,0,g}}{s_{g+1} - s_g} (s - s_g)$ . Let  $\xi_{g,n,p}$  correspond to the index of the grid point that is not further from the point  $(-1)^p \sqrt{2} \sigma x_n + \varphi(0) s_g$  than the cell length and has the smallest absolute value. Then the discretized Bellman equation can be written as  $G$  linear equations for the grid function:

$$\begin{aligned} V_{i,0,g} &- \frac{\beta}{\sqrt{\pi}} \sum_{n=1}^N \sum_{p=0}^1 [a_{g,n,p} V_{i,0,\xi_{g,n,p}} + b_{g,n,p} V_{i,0,\xi_{g,n,p+1}}] \\ &= -\frac{\beta}{\sqrt{\pi}} \sum_{n=1}^N \sum_{p=0}^1 \omega_n \log \sigma_i(0 \mid (-1)^p \sqrt{2} \sigma x_n + \varphi(0) s_g) \end{aligned}$$

Denote  $\Delta$  the step of the grid. Then we can express the above coefficients as

$$\begin{aligned} a_{g,n,p} &= \frac{\beta}{\Delta \sqrt{\pi}} \omega_n [s_{\xi_{g,n,p+1}} - (-1)^p \sqrt{2} \sigma x_n - \varphi(0) s_g], \\ b_{g,n,p} &= \frac{\beta}{\Delta \sqrt{\pi}} \omega_n [(-1)^p \sqrt{2} \sigma x_n + \varphi(0) s_g - s_{\xi_{g,n,p}}]. \end{aligned}$$

### 5.3 Simulation Results

We compare the utility function that we obtain from a numerical solution of the Bellman equation with the estimated payoff that we obtain using our method. The following table tabulates the integrated difference between the utility function that is numerically computed and the utility function that is estimated from a randomly generated sample. We use the stationary density of the state variable for the comparison. Specifically, if  $\widehat{\Pi}_{i,T}(\cdot, a_i = 1)$  is the estimated utility from sample of size  $T$  and  $\Pi_i(\cdot, a_i = 1)$  is the numerical solution, the reported criterion is

$$Q_T = \sqrt{T} \int_{\mathcal{S}} \left( \widehat{\Pi}_{i,T}(s, a_i = 1) - \Pi_i(s, a_i = 1) \right) \pi(s) ds,$$

where  $\pi(\cdot)$  is the stationary density of the state variable. We obtain this integral using the Monte-Carlo integration technique. To do so we make the joint draws from the state

variable transition and the decision rules using a preliminary draw of the state variable. We generate the state variable as well as the policy rule as a Markov chain until it reaches the stationary distribution (we determine that by the behavior of the distribution mean across the blocks of consecutive draws). Then if  $N_s$  is the number of draws from the stationary distribution, we compute the approximate criterion

$$Q_T^{N_s} = \frac{\sqrt{T}}{N_s} \sum_{t=1}^{N_s} \left( \hat{\Pi}_{i,T}(s_t, a_i = 1) - \Pi_i(s_t, a_i = 1) \right).$$

This object converges to the integral of interest as the number of draws increases. For our purposes we use 2.5 million draws.

As table 5 shows, the nonparametric procedure for recovering the primitive utilities works well in finite samples. In particular, just to give the reader a visual sense of the shape of the value function that is being recovered in the simulation, the following figure illustrates the median of numerically recovered utility with top and bottom 10% quantiles for 600 Monte-Carlo draws. The horizontal axis represents the value of the state variable, while the vertical axis represents the value of the recovered utilities.

## 6 Empirical Application

### 6.1 Data

We apply our identification and estimation results to analyze an empirical model of consumer choices using an IRA scanner data set of supermarket purchases of potato chips. The scanner data include multiple supermarket locations in two separate geographic areas. It contains over 900 different product types, some of which are not available in all areas. The panel data structure cover 312 weeks of purchases at 54 supermarkets in Pittsfield, MA and Eau Claire, WI. The products are categorized into 20 major brands that account for over 97% of the total market. The remaining brands are combined into a single brand category “other brands”. The market is dominated by “Lays” (the label of Pepsico inc.), which has a 45% share, and the second largest brand is “Pringles” (the label of Procter & Gamble). The empirical question we investigate is the demand elasticity with respect to price discounts. Since potato chips have long expiration periods and can be purchased

outside vendors outside of the supermarkets, we expect that demand reacts elastically to price discounts due to substitution and stockpiling effects.

The purchase prices in this data set vary substantially from 5 cents to \$6.99, and the average time between purchases varies from 1 to 285 weeks for returning consumers (Table 8). On average, the market share of a brand increases by 20% when it is offered at a discounted price. For most brands, market shares are highly correlated with the shares of inventory on sale. Volume discounts are prevalent. The price per ounce for the largest packages is almost half the price per ounce for the smallest packages. These reduced-form evidence indicates a high correlation between price discounts and purchase behaviors.

The market shares of various Potato chip products are persistent over time. In Figure 2, the cross product variation at a given time substantially exceeds the variation of product share over time for the same product. In the data price promotions occur regularly. Some brands having price discounts as frequently as 50 % of the time. Consumers appear to be responding to the posted price drops. Table 6 displays summary sale statistics and frequency of promotions across products. The reported statistics correspond to the parameters of distribution of sales aggregated by time and markets and sales aggregated by time and brand, to the consumers surveyed in the panel. The sales averaged over time and market range from 11 cents to \$1795. The sales averaged over brand and time range from 25 cents to \$1849 per week. The sale distribution has a long tail with a visible concentration at the bottom. This is due to small sales in some markets and small sales of certain brands.

Table 7 compares the market shares of brands following a week with price promotion and the market shares of brands during the same number of weeks when no price promotion has previously occurred. The mean as well as the 25, 50 and 75% quantiles of the market share distribution are all significantly higher for the promoted products. Figure 3 also shows that the histogram of log-market shares has visibly lower mode and mean for periods following a week without price promotion. In contrast, log-market shares concentrate at higher levels following a period of price promotions. The effect of promotion effect can also be visualized by the time path of log-market share and promotion timing for particular products. Figure 4 shows the market share of Classic Lay's chips over time with a spline-smoothed graph. The market share of Classic Lay tends to increase following the price promotion, and it

decreases when the price promotion is absent.

Additional evidence on the effect of price promotion on the transition of the market shares over time is reported in Figure 5, which summarizes the results of a nonparametric regression of the log-market share on the lagged log-market shares, separately for the cases where there was a price promotion in the previous period and where there were no promotions. It is clear from figure 5 that log-market share tends to be higher in the periods following the price promotion. Even the periods without the price promotion, we observe a strong positive relationship between the past and the present log-market shares, suggesting persistent non-price product-specific fixed effects. While a similar pattern is also observed in the period following a price promotion, the effect of price promotion is very large and outweighs the effect of the product characteristics.

Table 12 reports the results of a set of models in which we regress log market shares on price, a dummy variable for promotion and various product characteristics. Product characteristics include package size, fat content, sodium content, cooking method, brand name and shape. The three columns in table 12 correspond to OLS, a linear IV regression using the sums of product characteristics of the other products and the demographic information of consumers as instrument, and nonlinear BLP IV demand function with the same instruments. The largest value for the price coefficient was obtained in the BLP model with fixed coefficients. To estimate the BLP we used the same instruments as in the IV setup. The price promotion dummy has a similar range for all of the estimated models.

Brand level demand estimates accounting for time lags from the previous discount promotion are reported in table 13. The demand elasticity is estimated at 1.6 for the entire sample which varies from 1.7 for estimates in the sale periods only to 1.5 in the periods not on sale. In addition, the overall demand is lower for the products that are on sale less frequently, although the effect of sale frequency on demand is diminishing. The estimates of the price elasticity are compatible with those obtained using the BLP approach on the product level. As expected, the fraction of purchases of potato chips on sale is higher for households with lower income and higher for larger households. The share of products purchased on sale is higher for unemployed individuals as well as for students and retirees.

## 6.2 The Empirical Model

To account for forward-looking consumers and their stockpiling behavior, we need to estimate a dynamic choice model to distinguish between short term demands and long term demands. Applying the nonparametric and semiparametric estimators discussed earlier in the paper, however, is complicated by the presence of unobserved heterogeneity in consumer demand at the individual level. The presence of unobserved heterogeneity is reflected in the higher price elasticity in the household level demand than in the aggregate brand level demand. Unobserved consumer characteristics that affects their tendency to stockpile are likely to be persistent but can also vary over time.

Methods to account for the presence of unobserved heterogeneity are developed by Aguirregabiria and Mira (2007), Arcidiacono and Miller (2006), Kasahara and Shimotsu (2008) among others. In particular, Hu and Shum (2008) allows for nonparametric identification of continuously distributed and serially correlated unobserved heterogeneity. These methods are not only more advantageous in nonparametric identification but also more computationally feasible. It is prohibitive to compute the consumer decision rules in the nested fixed point maximum likelihood method due to the rich state space in most consumer choice models. In our data of price promotion with 21 brands over 1000 households, storing 1000 value functions defined over 20 continuous variables of market shares on a discretized grid of 100 points in single precision will require a total memory of exceeding  $10^{200000}$  Terabytes.

We augment an infinite horizon heterogeneous consumer model using the nonparametric identification and estimation method cited above. Each period  $t = 1, \dots, \infty$  is associated with the visit of a particular consumer to the store. Each product  $j = 1, \dots, J$  is characterized by a vector of observable characteristics  $x_{jt}$ , price  $p_{jt}$ , and scalar product-consumer specific unobservable characteristic  $\nu_{ijt}$ . The utility for consumer  $i$  from purchasing product  $j$  in period  $t$  is given by

$$u_{ijt} = x'_{jt}\beta - \alpha p_{jt} + \nu_{ijt} + \varepsilon_{ijt} = v_j(p_{jt}, x_{jt}) + \nu_{ijt} + \varepsilon_{ijt},$$

where  $p_{jt}$  is the price of product  $j$  and  $\varepsilon_{ijt}$  is an i.i.d. idiosyncratic preference shock according to the extreme value distribution. We assume that  $p_{jt}$  follows a first-order Markov process the stationary distribution being continuous on compact support. Consumers are

forward looking and maximize the expected lifetime discounted utility. We also assume that conditional on  $\xi_t, \delta_t, d_t$ ,  $(\xi_{t+1}, \delta_{t+1})$  are independent of  $(p_{t+1}, x_{t+1})$ .

Denote  $p_t = (p_{1t}, \dots, p_{Jt})$  the vector of current period prices and  $\nu_{it} = (\nu_{i1t}, \dots, \nu_{iJt})$  the vector of product-consumer-specific characteristics. The ex ante continuation value of consumers, denoted  $V_i(p_t, \nu_{it})$ , only depends on the current state variables because of the Markov transition assumption. The choice-specific value function of choosing product  $j$  is

$$V_{ij}(p_t, \nu_{it}) = v_j(p_{jt}, x_{jt}) + \nu_{ijt} + \beta E [V(p_{t+1}, \nu_{i,t+1}) | p_t, \nu_{it}, d_{it} = j].$$

The static utility of the outside option of no purchase, indexed by  $k = 0$ , is normalized to zero. The corresponding choice specific value function of outside purchase is

$$V_{i0}(p_t, \nu_{it}) = \beta E [V(p_{t+1}, \nu_{i,t+1}) | p_t, \nu_{it}, d_{it} = 0].$$

We consider a model in which  $\nu_{ijt} = \xi_{jt} + \delta_{it}$ , where the unobserved product characteristic has both a product-specific component  $\xi_{jt}$  that is common across the consumers at the same instance of time, and a consumer-specific component  $\delta_{it}$  that is common across different brands for the same consumer. We assume that  $\xi_{jt}$  and  $\delta_{it}$  both have finite support with  $K$  points, denoted by  $\{z_k\}_{k=1}^K$ , and that they follow first-order Markov processes according to the choice  $d$ .

In the presence of unobserved product characteristics both the value functions and the choice probabilities depend on the unobserved components. For example,

$$\begin{aligned} V_{ij}(p, x, \xi, \delta) &= v_j(p_j, x_j) + \xi_j + \delta - \beta E [\log \sigma_{ij}(p_{t+1}, x_{t+1}, \xi_{t+1}, \delta_{i,t+1}) | p, x, \xi, \delta] \\ &\quad + \beta E [V_{ij}(p_{t+1}, x_{t+1}, \xi_{t+1}, \delta_{i,t+1}) | p, x, \xi, \delta], \end{aligned}$$

for choices  $j = 1, \dots, J$ , and

$$\sigma_{ij}(p, x, \xi, \delta) = \frac{\exp(V_{ij}(p, x, \xi, \delta))}{\sum_{k=0}^J \exp(V_{ik}(p, x, \xi, \delta))}, \quad j = 0, \dots, J.$$

While  $\sigma_{ij}(p, x, \xi, \delta)$  are not observed, they are related to the observed choice probabilities through the relation

$$\sigma_{ij}(p, x) = \sum_{k=1}^K \sum_{p=1}^K \sigma_{ij}(p, x, \xi = z_k, \delta = z_p) \pi_{i,\delta}(z_p) \pi_{j,\xi}(z_k).$$



where  $\pi_{i,\delta}(z_p)$  and  $\pi_{j,\xi}(z_k)$  correspond to the stationary distribution of  $\delta_i$  and  $\xi_j$ . Also define matrixes  $\Pi^1$  with elements  $\Pi_{lmnp}^1 = \Pr(\delta_{i,t} = z_l, \xi_{i,t} = z_m, \delta_{i,t+1} = z_n, \xi_{i,t+1} = z_p)$ , and  $\Pi^2 = \Pi^1 \Pi^1$ . In the above,  $\pi_{i,\delta}(z_p)$  and  $\pi_{j,\xi}(z_k)$  are the corresponding marginal distribution of  $\delta_{i,t}, \xi_{i,t}$  implied by  $\Pi^1$ .

As shown in Hu and Shum (2008), in addition to the conditional choice probabilities  $\sigma_{ij}(p, x)$ , the conditional covariance between choices over different periods also contain useful information to identify the unknown parameters in  $\sigma_{ij}(p, x, \xi, \delta)$ . In particular, consider the choice correlations between one and two periods apart:

$$\begin{aligned} \gamma_{ijh}(p', x', p, x) &= E [d_{ij,t+1} d_{iht} \mid p_{t+1} = p', x_{t+1} = x', p_t = p, x_t = x] \\ &= \sum_{p,k,l,m=1}^K \sigma_{ij}(p', x', z_p, z_k) \sigma_{ih}(p, x, z_p, z_k) \Pi_{pklm}^1, \end{aligned}$$

and

$$\begin{aligned} \kappa_{ijh}(p', x', p, x) &= E [d_{ij,t+2} d_{iht} \mid p_{t+2} = p', x_{t+2} = x', p_t = p, x_t = x] \\ &= \sum_{p,k,l,m=1}^K \sigma_{ij}(p', x', z_p, z_k) \sigma_{ih}(p, x, z_p, z_k) \Pi_{pklm}^2. \end{aligned}$$

These provide additional moment conditions that we will use to estimate the distribution of the unobserved heterogeneity components.

We use  $\zeta_k, k = 1, \dots, K^2$  to denote all the support points of  $\delta, \xi$ . To implement the model we represent the choice probabilities given both the observed and unobserved components using a polynomial:

$$\hat{\sigma}_{ij}(p, x, \zeta_k) = a_{ij}^{N,k} H^N(p, x),$$

where  $H^N(p, x)$  are basis functions and  $a_{ij}^{N,k}$  are unknown coefficients to be estimated. Similarly define the observable conditional value function

$$\hat{V}_{ij}(p, x) = \sum_{k=1}^K \sum_{p=1}^K V_{ij}(p, x, z_k, z_p) \pi_{i,\delta}(z_p) \pi_{j,\xi}(z_k),$$

and represent it with a polynomial expansion  $\hat{V}_{i0}(p, x) = v_i^N H^N(p, x)$ . Using instruments  $Z_t = H^N(p_t, x_t)$ , we form two sets of moment conditions to estimate the model. The first

set of moments relate to the conditional correlation of the choices over time:

$$\rho^{1,jh}(A, v(\cdot)) = E \left[ Z_t \left( d_{ij,t+1} d_{iht} - \sum_{k,p=1}^{K^2} \Pi_{kh} a_{ij}^{N,k} H^N(p_{t+1}, x_{t+1}) a_{ij}^{N,p} H^N(p_t, x_t) \right) \right]$$

$$\rho^{2,jh}(A, v(\cdot)) = E \left[ Z_t \left( d_{ij,t+2} d_{iht} - \sum_{k,p,r=1}^{K^2} \Pi_{kr} \Pi_{rp} a_{ij}^{N,k} H^N(p_{t+2}, x_{t+2}) a_{ij}^{N,p} H^N(p_t, x_t) \right) \right].$$

The second set of moment conditions are the conditional choice probabilities in relation to the Bellman equations:

$$\rho^3(A, v(\cdot)) = E \left[ Z_t \left( v_i^N H^N(p_t, x_t) + \beta \sum_{k,p=1}^{K^2} \Pi_{kp} \log \left( a_{i0}^{N,k} H^N(p_{t+1}, x_{t+1}) \right) - \beta v_i^N H^N(p_{t+1}, x_{t+1}) \right) \right],$$

and

$$\rho^{4,j}(A, v(\cdot)) = E \left[ Z_t \left( v_i^N H^N(p_t, x_t) + \beta \sum_{k,p=1}^{K^2} \Pi_{kp} \log \left( a_{i0}^{N,k} H^N(p_{t+1}, x_{t+1}) \right) - \beta v_i^N H^N(p_{t+1}, x_{t+1}) - v_j(x_{jt}, p_{jt}) + \sum_{k,p=1}^{K^2} \Pi_{kp} \log \left( \frac{a_{ij}^{N,k} H^N(p_t, x_t)}{a_{i0}^{N,k} H^N(p_t, x_t)} \right) \right) \right].$$

We estimate the parameters in this system using a conventional GMM method.

### 6.3 Empirical Results

The additive individual-specific unobserved heterogeneity component with serial correlation over time can represent the reactions of consumers purchasing potato chips to purchasing potato chips can react to price variations over time by either stocking up potato chips, or using the alternative retail locations such as gas stations. While consumer individual unobserved heterogeneity can be interpreted either as their stocking behavior or as their tendency

to switch to alternative outlets, and may have other interpretations, the data pattern of the clear responses of consumer purchases to the timing of large price discounts is consistent with the stock piling interpretation of unobserved individual components. We implement a model specification where the unobserved heterogeneity component has a support of 20 points, and we use second-order approximations to the choice probabilities. As a result, we estimate the period utility as a function of demographic and brand characteristics for each of the 20 brands (excluding the “combined” brand #21). The obtained structural estimates are given in Tables 17- 20. The structural elasticity estimates reported in these tables are much higher than those in the static BLP model. This suggests that the presence of consumer-level persistent unobserved heterogeneity has a substantial impact on the coefficient estimates. The estimates that we obtain in the structural model exceed the estimates reported in the static BLP model by a factor of two. We also find that an increase in the package size has a positive impact on consumer utility. Consumer demographic information also plays important roles in the utility function. Higher income individuals tend to extract smaller utility from purchasing potato chips uniformly over brands. Larger family size makes consumer more prone to purchasing potato chips. Furthermore, education tends to decrease the utility from purchasing potato chips. In addition, the families where the oldest male works tend to value potato chips less than those where the oldest male stays at home. Finally, we find that Hispanic households tend to value potato chips more than other households.

Our findings show that the elasticity of demand in the dynamic model is very different from that obtained in a static BLP style demand model. These findings reinforce the insights in Hendel and Nevo (2006) that short term and long term demand elasticities can be substantially different because of unobserved stockpiling behaviors by consumers. Qualifying this difference between short term and long term demand is difficult, however, using conventional demand estimation techniques in the BLP style. A semiparametric dynamic discrete choice model that is computationally flexible and attractive, provides a powerful tool to distinguish long term behavior from short term demands, and can be useful to obtain meaningful demand elasticity estimates for consumers.

By comparison, we also estimate the structural discrete choice without the serially cor-

related unobserved heterogeneity component. The brand-by-brand estimates are reported in tables 21-24. One can immediately see that the price coefficients are lower in the model with serially correlated unobserved component, which provides evidence of consumption smoothing behavior by consumers who purchase larger quantities of products on sale.

We also present a direct comparison between the results from the two models with and without unobserved heterogeneities in Table 25. To construct this table we consider how the price elasticity of demand evolve over time in response to an unanticipated price decline for Lay's potato chips by 10% in Week 1 (10% price discounts are typical in our data). Only the model with unobserved heterogeneity allows for time-varying elasticity. A direct comparison between the OLS and IV results shows that price endogeneity indeed creates a serious bias in the point estimates. Then when we compare the fixed and random coefficient BLP model, we notice that the random coefficient BLP produces a higher price elasticity. Furthermore, when we compare elasticity between the structural models with and without serially correlated unobserved heterogeneity we can see that, the model without serially correlated unobserved heterogeneity overstates both the short term and the long-term price elasticity. The price elasticity tends to decline over time following the weeks after the one-time price drop for Lay's.

## 7 Conclusion

We study nonparametric identification of a dynamic discrete game model of incomplete information, and develop nonparametric and semiparametric estimators that have flexible computational properties and desirable statistical properties. Our identification analysis provides a unified framework for both discrete and continuous state variables, and suggests a natural implementation of a nonparametric estimator. In addition, we derive the semiparametric efficiency bound and propose a one-step semiparametric efficient estimator under the assumptions that the transition process is nonparametrically specified while the static payoff functions are parametric. The properties of the model are illustrated in a set of numerical simulations and by an empirical application. Similarly, the

The identification and estimation framework in this paper is not without limitations. Assumption 1 requires the independence and identical extreme value distribution of the error

terms. Additionally, these errors enter additively into the agent's static payoffs. While this is a common assumption in the literature, it is also very strong. It is possible to relax some of these restrictions if we are willing to compromise the other components of the model. For example, if a linear index functional form is imposed on the static payoffs, the error term distribution can be identified nonparametrically and does not need to be assumed entirely known. This is well known in static linear index discrete choice models, but requires more complex deconvolution arguments in dynamic models through nonlinear functional relations. Similarly, the normalization assumption is a necessary identification condition and is innocuous in a static model. But it might not be innocuous in a dynamic model. Adding a function of the state variables to the payoffs of all actions will not change the static choice probabilities, but might change the dynamic choice probabilities, especially if one of the choices tends to shift the distribution of the future state variables towards a higher utility area. In these cases assumption 2 is not necessarily more appealing than making a parametric static utility assumption. These can be interesting future research directions.

## References

- ACKERBERG, D., X. CHEN, AND J. HAHN (2011): "A practical asymptotic variance estimator for two-step semiparametric estimators," *Cowles Foundation Discussion Paper No. 1803*.
- AGUIRREGABIRIA, V., AND P. MIRA (2002): "Swapping the nested fixed point algorithm: a class of estimators for discrete Markov decision models," *Econometrica*, 70(4), 1519–1543.
- (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1.
- AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.
- (2009): *Semiparametric efficiency bound for models of sequential moment restric-*

- tions containing unknown functions.* Yale University, Cowles Foundation for Research in Economics, forthcoming, *Journal of Econometrics*.
- ANDERSON, S., A. DEPALMA, AND J. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press.
- ARCIDIACONO, P., AND R. MILLER (2006): “CCP estimation of dynamic discrete choice models with unobserved heterogeneity,” *Manuscript, Duke University, forthcoming Econometrica*.
- BAJARI, P., C. BENKARD, AND J. LEVIN (2007): “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75(5), 1331–1370.
- BERRY, S., A. PAKES, AND M. OSTROVSKY (2003): “Simple estimators for the parameters of dynamic games (with entry/exit examples),” Technical Report, Harvard University.
- BRESNAHAN, T., AND P. REISS (1991): “Empirical Models of Discrete Games,” *Journal of Econometrics*, 48, 57–81.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of Semiparametric Models when the Criterion Function Is Not Smooth,” *Econometrica*, 71(5), 1591–1608.
- DUNFORD, N., AND J. SCHWARTZ (1958): *Linear operators. Part 1: General theory*. pp285–305, Wiley, New York.
- FAN, J., AND I. GIJBELS (1992): “Variable bandwidth and local linear regression smoothers,” *The Annals of Statistics*, pp. 2008–2036.
- FERSHTMAN, C., AND A. PAKES (2009): “Finite state dynamic games with asymmetric information: A framework for applied work,” SSRN working paper.
- HENDEL, I., AND A. NEVO (2006): “Measuring the implications of sales and consumer inventory behavior,” *Econometrica*, 74(6), 1637–1673.
- HOLMES, T. J. (2011): “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1), 253–302.

- HOTZ, J., AND R. MILLER (1993): “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60, 497–529.
- HU, Y., AND M. SHUM (2008): “Nonparametric identification of dynamic models with unobserved state variables,” *Johns Hopkins University, Dept. of Economics working paper*, 543, 2009.
- JENKINS, M., P. LIU, D. MCFADDEN, AND R. MATZKIN (2004): “The Browser War: Econometric Analysis of Markov Perfect Equilibrium in Markets with Network Effects,” UC Berkeley, working paper.
- JIA, P. (2008): “What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry,” *Econometrica*, 76(6), 1263–1316.
- JOFRE-BONET, M., AND M. PESENDORFER (2003): “Estimation of a dynamic auction game,” *Econometrica*, 71(5), 1443–1489.
- KASAHARA, H., AND K. SHIMOTSU (2008): “Nonparametric identification of finite mixture models of dynamic discrete choices,” *Econometrica*, 77(1), 135–176.
- MAGNAC, T., AND D. THESMAR (2002): “Identifying dynamic discrete decision processes,” *Econometrica*, 70(2), 801–816.
- NEKIPELOV, D., P. BAJARI, AND H. HONG (2010): “Identification and Inference in Discrete Games with Unknown Distributions of Private Shocks,” manuscript, UC Berkeley, Minnesota and Stanford University.
- NEWKEY, W. (1990): “Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics*, 5(2), 99–135.
- (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349–82.
- NEWKEY, W., AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle, and D. McFadden, pp. 2113–2241. North Holland.

- NEWKEY, W., AND J. POWELL (2003): “Instrumental variable estimation of nonparametric models,” *Econometrica*, pp. 1565–1578.
- PESENDORFER, M., AND P. SCHMIDT-DENGLER (2010): “Sequential estimation of dynamic discrete games: A comment,” *Econometrica*, 78(2), 833–842.
- PESENDORFER, M., P. SCHMIDT-DENGLER, AND H. STREET (2008): “Asymptotic least squares estimators for dynamic games,” *Review of Economic Studies*, 75(3), 901–928.
- RUST, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999–1033.
- SOMAINI, P. (2011): “Competition and Interdependent Costs in Highway Procurement,” .
- STOKEY, N., R. LUCAS, AND E. PRESCOTT (1989): *Recursive methods in economic dynamics*. Harvard Univ Pr.

## A Proof of theorem 3

First we need to characterize the tangent set of the model. The likelihood of the model will be determined by the choice probabilities and the transition density for the state variable. Given that choices of players are observed by the econometrician, the log-likelihood of the model can be written as

$$\mathcal{L}(s, s', d) = \sum_{i=1}^n \sum_{k=0}^K d^{i,k} \log \sigma_i(a_i = k | s) + \sum_{a \in \mathcal{A}} d^a \log g(s | s', a) + \log p(s'),$$

where  $g(\cdot | s', a)$  is the transition density of the state variable,  $d^a$  is the indicator of the action profile  $a$ , and  $p(\cdot)$  is the stationary density of the state variable. We choose a particular parameterization path  $\theta$  for the model and compute the score by differentiating the model along the path:

$$S_\theta(s, s', d) = \sum_{a \in \mathcal{A}} d^a s_{1\theta}(s | s', a) + s_{2\theta}(s') + \sum_{i=1}^n \sum_{k=0}^{K-1} \left( \frac{d^{i,k}}{\sigma_i(k|s)} - \frac{d^{i,K}}{\sigma_i(K|s)} \right) \dot{\sigma}_i(k|s),$$

where  $E[s_{1\theta}(s | s', a) | s', a] = 0$ ,  $E[s_{2\theta}(s')] = 0$ ,  $E[|s_{1\theta}(s | s', a)|^2 | s', a] < \infty$ ,  $E|s_{2\theta}(s')|^2 < \infty$ , and  $E|\sigma_i(k|s)|^2 < \infty$ . Then we characterize the tangent set as

$$\mathcal{T} = \left\{ \sum_{a \in \mathcal{A}} d^a \eta_1(s | s', a) + \eta_2(s') + \sum_{i=1}^n \sum_{k=0}^{K-1} \eta_3(s) \left( \frac{d^{i,k}}{\sigma_i(k|s)} - \frac{d^{i,K}}{\sigma_i(K|s)} \right) \right\},$$



with  $E[\eta_1(s|s', a)|s', a] = 0$ ,  $E[\eta_2(s')] = 0$ ,  $E\left[|\eta_1(s|s', a)|^2|s', a\right] < \infty$ ,  $E|\eta_2(s')|^2 < \infty$ , and  $E|\eta_3(s)|^2 < \infty$ . We will derive the semiparametric efficiency bound for this model under the absence of parametric restrictions on the state transition density. To derive the bound we find the parametric and the non-parametric parts of the score of the model using a particular parametrization path for the non-parametric component. For the chosen parametric path  $\theta$  we denote

$$\frac{\partial V_i(k, s)}{\partial \theta} = \zeta_i(k, s) \quad \text{and} \quad \frac{\partial V_i(k, s)}{\partial \gamma'} = \tilde{\zeta}_i(k, s).$$

Also denote  $\pi_i(k, s) = \frac{\partial \Pi(k, s; \beta)}{\partial \gamma'}$ . We form vectors  $V^i = (V_i(1, s), \dots, V_i(K, s))'$ ,  $V = (V^1, \dots, V^n)'$ ,  $\sigma^i = (\sigma_i(1|s), \dots, \sigma_i(K|s))'$  and  $\zeta = (\zeta_1(1, s), \dots, \zeta_1(K, s), \dots, \zeta_n(K, s))'$ . First of all, we note that we can transform the original moment equation. Consider the operator

$$\mathcal{P}_i \circ f = E[f(s') | s, a_i],$$

where expectation is defined for the conditional density  $\sum_{a_{-i}} g(s'|s, a_i = k, a_{-i}) \sigma_{-i}(a_{-i}|s)$ . This operator has a discrete spectrum with eigenfunctions  $\{\Theta_j^{i,k}(s)\}_{j=0}^{\infty}$  and eigenvalues  $\{\lambda_j^{i,k}\}_{j=0}^{\infty}$  different from zero. This follows directly from the properties of the Hilbert-Schmidt operators which can be found in Dunford and Schwartz (1958). Then we can represent the value function as

$$V_i(k, s) = \sum_{j=0}^{\infty} a_j^{i,k} \Theta_j^{i,k}(s).$$

Then we can transform the moment equation to

$$\begin{aligned} \tilde{\varphi}(s, s', a; \gamma, V_i, \sigma_i) &= \sum_{j=0}^{\infty} a_j^{i,k} \left(1 - \beta \lambda_j^{i,k}\right) \Theta_j^{i,k}(s) \\ &+ (1 - d^{i,0}) [-\Pi_i(a_i, a_{-i}, s; \gamma) + \beta \log \sigma_i(a_i|s')] + d^{i,0} \beta \log \left(1 - \sum_{j=1}^K \sigma_i(j|s')\right). \end{aligned}$$

Then we can define a directional derivative of the moment function with respect to  $V_i$  in the direction  $h$  as

$$\left(\frac{\partial \varphi_i}{\partial V_i}\right)_h = \sum_{j=0}^{\infty} h_j \left(1 - \beta \lambda_j^{i,k}\right) \Theta_j^{i,k}(s),$$

for all  $h$  with  $\sum_{j=0}^{\infty} |h_j| \|\Theta_j^{i,k}(s)\| < \infty$ . Differentiating the unconditional moment equation with respect to the parametrization path we obtain

$$\begin{aligned} E[\mathcal{A}(s, d^a) \pi(s) (1 - d^{a,0})] \dot{\gamma} - E\left[\mathcal{A}(s, d^a) \left(\frac{\partial \varphi}{\partial V}\right)_h\right] \dot{h} \\ = \beta E\left[\mathcal{A}(s, d^a) \left(\frac{d^{a \neq 0}}{\sigma(a|s')} - \frac{d^{a,0}}{\sigma(0|s')}\right)\right] + E[\mathcal{A}(s', d^a) \varphi_{s1\theta}]. \end{aligned}$$

We consider the right-hand side and try to find a function  $\tilde{\Psi}$  such that the expression on the right-hand side can be represented as  $\langle \Psi, S_\theta \rangle$ . This function can be obtained as

$$\tilde{\Psi} = \mathcal{A}(s, d^a) \left\{ (\varphi - E[\varphi | s, a]) + \frac{d^{a \neq 0} - \sigma(a|s)}{\sigma(a|s')} - \frac{d^{a,0} - \sigma(0|s)}{\sigma(0|s')} \right\}.$$

We note that conditional moment equation (23) holds and we can differentiate it with respect to the parameterization path. Then we can substitute the expression for the derivative into the expression for the unconditional moment. This allows us to express the directional derivative of  $\gamma$  and, consequently, the efficient influence function for a fixed instrument matrix:

$$\Psi = E \left[ \mathcal{A}(s, d^a) \left( \pi(s) (1 - d^{a,0}) - \left( \frac{\partial \varphi}{\partial V} \right)_h \right) \right]^{-1} \tilde{\Psi}.$$

The semiparametric efficiency bound as a minimum variance of the influence function. Denoting

$$\Omega(s, a) = \text{Var} \left( \varphi + \frac{d^{a \neq 0}}{\sigma(a|s')} - \frac{d^{a,0}}{\sigma(0|s')} \mid s, a \right).$$

Using standard GMM arguments, we can express the efficiency bound for fixed instrument as

$$V_h(\beta) = \left( \left( \pi(s) (1 - d^{a,0}) - \left( \frac{\partial \varphi}{\partial V} \right)_h \right) \zeta(d^a, s)' \Omega(s, a)^{-1} \zeta(d^a, s) \left( \pi(s) (1 - d^{a,0}) - \left( \frac{\partial \varphi}{\partial V} \right)_h \right) \right)^{-1}.$$

The efficiency bound overall can be found as  $V_{h^*}(\beta)$  for  $h^*$  solving

$$\inf_h \left( \pi(s) (1 - d^{a,0}) - \left( \frac{\partial \varphi}{\partial V} \right)_h \right) \zeta(d^a, s)' \Omega(s, a)^{-1} \zeta(d^a, s) \left( \pi(s) (1 - d^{a,0}) - \left( \frac{\partial \varphi}{\partial V} \right)_h \right).$$

The optimal instrument matrix can be explicitly written as

$$\mathcal{M}(s) = E \left[ \left( \pi(s) (1 - d^{a,0}) - \sum_{j=0}^{\infty} h_j^* (1 - \beta \lambda_j) \Theta_j(s) \right) \zeta(d^a, s)' \Omega(s, a)^{-1} \mid s \right].$$

*Q.E.D.*

## Proof of theorem 5

We can use the Bellman equation to express the estimate of the payoff function in terms of the estimate of the value function. We use a series projection estimator to estimate  $V_i(k, s) - V_i(0, s)$ . To evaluate the elements of the Bellman equation for player  $i$  we need to analyze the right hand side function

$$h_i(s) = E \left\{ \log \sum_{k=0}^K (V_i(k, s') - V_i(0, s')) \mid s \right\}.$$

Function  $h_i(s)$  admits a series representation  $h_i(s) = \sum_{j=1}^{k(m)} q_j(s) \lambda_{i,j}^{k(m)} + o(\|q_{k(m)}(s)\|)$ , where we use the standard Sobolev norm. The coefficients for this representation can be obtained from the coefficients for  $V_i(k, s) - V_i(0, s)$ . This result can be used to find a series representation for  $V_i(0, s)$  which needs to be estimated. To do that we proceed by analyzing the nonparametric conditional expectation estimation component of step two, which takes the form of

$$V_i(s, 0) = \beta \int V_i(s', 0) g_i(s'|s, 0) ds' + h_i(s) = (\mathcal{K}_i \circ V_i)(s, 0) + h_i(s), \quad (25)$$

where  $g_i(s'|s, 0) = \sum_{a_{-i} \in \mathcal{A}_{-i}} g(s'|s, 0, a_{-i}) \sigma(a_{-i}|s)$ .

This is an integral equation for  $V_i(\cdot, 0)$ . We assume that the integral operator  $\mathcal{K}_i$  and the term  $h_i(\cdot)$  satisfy the standard assumptions assuring the existence of a smooth solution of this equation. In particular  $s \in S$ ,  $V_i : S \mapsto \mathbb{R}_+$ , both the kernel function  $g_i(\cdot)$  and the function  $h_i(\cdot)$  have derivatives up to order  $p \geq k(m)$ , which assures a high degree of smoothness of the value function. Thus  $V_i \in C^p(S)$ , and  $\mathcal{K}_i : C^p(S) \mapsto C^p(S)$ . A standard method for solving this equation is to represent solution by a series expansion over a particular basis in  $C^p(S)$ . We will use the basis  $q^{k(m)}(s) = (q_1(s), \dots, q_{k(m)}(s))'$  for these purposes. Then the approximation for the value function can be written as:

$$V_i(\cdot, 0) = q^{k(m)}(s)' \theta_i^{k(m)}.$$

We endow the space  $C^p(S)$  with an inner product  $\langle \cdot, \cdot \rangle$  and introduce matrices

$$\Gamma = (\langle q_t(s), q_j(s) \rangle)_{t,j=1}^{k(m)} \quad \text{and} \quad G_i = (\langle \mathcal{K} q_t(s), q_j(s) \rangle)_{t,j=1}^{k(m)}.$$

We define the inner product for two functions  $f, g \in C^p(S)$  as:

$$\langle f, g \rangle = \int_S f(s)g(s)\pi(ds) = E[f(s)g(s)],$$

where  $\pi(\cdot)$  is a stationary distribution measure for the state space  $S$ . In general, this measure is not available. For this reason, we substitute it with the empirical measure  $\pi^m(\cdot)$ , which we require to be weakly converging to  $\pi(\cdot)$ . We call the space associated with the inner product generated by  $\pi^m(\cdot)$  by  $C^{pm}(S)$ . This space is only a semi-Hilbert space as the inner product in it might have a non-empty kernel (and, thus, the associated norm is only a seminorm). We will use the same basis in  $C^{pm}(S)$  as before.

We can use the expansion for  $h_i(\cdot)$  to derive the series approximation for the value function  $V_i(\cdot, 0)$ . In this case the vector of coefficients in the series representation of the value function can

be found as:

$$\theta_i^{k(m)} = (\Gamma - \beta G_i)^{-1} \Gamma \lambda_i^{k(m)}.$$

This result is obtained from substituting series expansions for  $h_i(\cdot)$  and  $V_i(\cdot, 0)$  into equation (25) and projecting both sides of this equation on the basis vectors  $q^{k(m)}(\cdot)$ .

These coefficients allow us to obtain an approximation for the value of the function  $V_i(\cdot, 0)$  which can be expressed as:

$$V_i^{k(m)}(s, 0) = q^{k(m)}(s)' (\Gamma - \beta G_i)^{-1} \Gamma \lambda_i^{k(m)}$$

For sufficiently smooth coefficients of the original integral equation, this expression will provide an approximation of order  $k(m)$  such that the norm of the deviation of the approximation from the true solution will be bounded from above by  $\frac{L}{k(m)!} \sup_{s, s' \in \Omega} \|s - s'\|^{k(m)}$ , where  $\Omega \subset \mathcal{S}$  is a subset of the state space where the value function is approximated by the series expansion. Note that all components of this formula are exactly known, although the matrices are specific to a particular basis.<sup>3</sup>

We estimate coefficients in the series representation of the value function from the data. To do so, first, we estimate the state transition probability. We assume that an estimator with the rate  $r \in (0, 1/2]$  is available which produces the estimate that is point-wise asymptotically normal at  $s'$  uniformly over  $s$  in  $\Omega$ :

$$n^r (\hat{g}_i(s' | s, 0) - g_i(s' | s, 0)) \xrightarrow{d} N(0, \sigma_g^2(s', s)).$$

We assume for convenience that this estimate is obtained using an estimation procedure which can be approximated by a series expansion with the order of precision at least  $o_p(n^{-r})$ . To estimate the vector of coefficients  $\lambda^{k(m)}$  we use the data from the observed states and values of  $h_i(\cdot)$  to estimate it. Note that the values of  $h_i(\cdot)$  are obtained from the Hotz-Miller-type inversion and thus contain noise. By the nature of this inversion we can in principle evaluate  $\hat{h}_i(\cdot)$  at any point of  $\Omega$ . Although the probabilities of actions are estimated non-parametrically, by Delta-method we can assure that for some  $q \in (0, 1/2]$  we obtain a point-wise asymptotically normal estimator of  $h_i(\cdot)$  in  $\Omega$ . In particular we use a spectral representation of  $h_i(\cdot)$  to estimate it non-parametrically and obtain the coefficients  $\lambda^p$ . Thus

$$n^q (\hat{h}_i(s) - h_i(s)) \xrightarrow{d} N(0, \sigma_h^2(s)).$$

---

<sup>3</sup>For instance, if  $q^{k(m)}(\cdot)$  is a system of Legendre polynomials then  $\Gamma = \text{diag} \left\{ \sqrt{\frac{2}{2k(m)+1}} \right\}$ .

We consider the properties of the pointwise approximation error for the value function:

$$\begin{aligned} \hat{V}_i^{k(m)}(s, 0) - V_i(s, 0) &= q^{k(m)}(s)' (\Gamma - \beta G_i)^{-1} \langle \hat{h}_i(s) - h_i(s), q^{k(m)}(s) \rangle \\ &\quad + \beta q^{k(m)}(s)' (\Gamma - \beta G_i)^{-1} \langle (\hat{\mathcal{K}}_i - \mathcal{K}_i) q^{k(m)}(s), q^{k(m)}(s)' (\Gamma - \beta G_i)^{-1'} \Gamma \lambda^{k(m)} + \Delta^{k(m)} \rangle. \end{aligned}$$

In this expression  $\Delta^{k(m)}$  is a residual function. In the expression for the error in the estimate of the value function the matrices only play the role of normalization while the asymptotic behavior of the error is governed by the integrated error in the estimated components of the Bellman equation. This normalization does not change the rate of convergence of the estimators, and the order of polynomial expansion is determined only by the degree of smoothness of the function approximation. Assumption 5 restricts the operator  $\mathcal{K}$  to be bounded. Consider the transformation  $\lambda \mapsto \Gamma^{1/2} \lambda$  and  $q^{k(m)}(\cdot) \mapsto \Gamma^{-1/2} q^{k(m)}(\cdot)$ . This a rotation of the basis which does not change the asymptotic properties. In fact, indicating the rotated variables by tildes we get:

$$m^q \left( \tilde{q}^{k(m)}(s)' \tilde{\lambda}_i^{k(m)} - \tilde{q}^{k(m)}(s) \tilde{\lambda}_i^{k(m)} \right) \xrightarrow{d} N(0, \sigma_\psi^2).$$

Specifically,  $\sigma_V^2 = \lim_{m \rightarrow \infty} \text{trace} \left\{ m^{2q} \tilde{\Omega}_\lambda \tilde{q}^{k(m)}(s) \tilde{q}^{k(m)}(s)' \right\} = \lim_{m \rightarrow \infty} \text{trace} \left\{ m^{2q} \Omega_\lambda q^{k(m)}(s) q^{k(m)}(s)' \right\} = \sigma_\psi^2(s)$ .

Next, note that  $I_{k(m)} \leq (I_{k(m)} - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2})^{-1} \leq (1 - \beta)^{-1} I_{k(m)}$ , where inequality should be treated as the difference between the two matrices is a positive semi-definite matrix. We can show that the last inequality is valid in two steps. First, the matrix  $\Gamma - G_i$  is positive semi-definite because the operator  $\mathcal{K}$  is defined by a density function. Second, the matrix  $(1 - \beta)^{-1} (I_{k(m)} - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2}) - I_{k(m)}$  is positive semi-definite. To see that, consider decomposition

$$I_{k(m)} - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2} = (1 - \beta) I_{k(m)} + \beta \Gamma^{-1/2} (G_i - \Gamma) \Gamma^{-1/2} \geq (1 - \beta) I_{k(m)}.$$

As a result:

$$\begin{aligned} &\text{trace} \left\{ m^{2q} (\Gamma - \beta G_i)^{-1} \Gamma \Omega_\lambda \Gamma (\Gamma - \beta G_i)^{-1} q^{k(m)}(s) q^{k(m)}(s)' \right\} \\ &= \text{trace} \left\{ m^{2q} (I_k - \beta \Gamma^{-1/2} G \Gamma^{-1/2})^{-1} \tilde{\Omega}_\lambda (I_{k(m)} - \beta \Gamma^{-1/2} G \Gamma^{-1/2})^{-1} \tilde{q}^{k(m)}(s) \tilde{q}^{k(m)}(s)' \right\}. \end{aligned}$$

This means that

$$\omega_1^2 = \lim_{m \rightarrow \infty} \text{trace} \left\{ m^{2q} (\Gamma - \beta G_i)^{-1} \Gamma \Omega_\lambda \Gamma (\Gamma - \beta G_i)^{-1} q^{k(m)}(s) q^{k(m)}(s)' \right\} < \frac{\sigma_\psi^2}{(1 - \beta)^2},$$

and it does not vanish. This proves that the rate of convergence of the non-parametric estimate for  $V_i(\cdot, 0)$  is the same as the rate for  $h_i(\cdot)$ .

The approximation for the value function can be expressed in terms of subsequent projections. From the Bellman's equation it follows that

$$\widehat{V}_i^{k(m)}(s, 0) - V_i(s, 0) - \beta E \left[ \widehat{V}_i(s', 0) - V_i(s', 0) | s \right] = \beta \left( \widehat{E} [V_i(s', 0) | s] - E [V_i(s', 0) | s] \right) + \Delta, \quad (26)$$

with the residual  $\Delta$ . Using the spectral representation for the expectation in the basis  $q^{k(m)}(\cdot)$  (where the coefficients of  $V_i(\cdot, 0)$  in this basis are denoted  $\theta^{k(m)}$ ) we obtain that up to the error of order smaller than  $\Delta$ :

$$\widehat{E} [V(s', 0) | s] - E [V(s', 0) | s] = q^{k(m)}(s)' \Gamma^{-1} \left( \widehat{G}_i - G_i \right) \theta^{k(m)},$$

$$E \left[ \widehat{V}(s') - V(s') | s \right] = q^{k(m)}(s)' \Gamma^{-1} G_i \left( \widehat{\theta}^{k(m)} - \theta^{k(m)} \right).$$

From spectral representation of the Bellman's equation it follows that (up to the series approximation error):

$$\theta^{k(m)} = (\Gamma - \beta G_i)^{-1} \Gamma \lambda^{k(m)}.$$

Substitution of these expressions into (26) gives:

$$\begin{aligned} \widehat{V}_i^{k(m)}(s, 0) - V_i^{k(m)}(s, 0) &= \beta q^{k(m)}(s)' \Gamma^{-1/2} \left( I - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2} \right)^{-1} \Gamma^{-1/2} \\ &\times \left( \widehat{G}_i - G_i \right) \Gamma^{-1/2} \left( I - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2} \right)^{-1} \Gamma^{1/2} \lambda_i^{k(m)}. \end{aligned}$$

This suggests that the method of approximating value function by consecutive conditional expectations (26) is equivalent to the spectral approach up to approximation error.

Now we will discuss the case where we substitute the stationary measure  $G_i(\cdot)$  by its empirical analog. In this case for the sample  $\{s_l\}_{l=1}^m$  the inner product for  $f, g \in C^p(S)$  can be defined as:

$$\langle f, g \rangle^m = \sum_{l=1}^m f(s_l) g(s_l).$$

We can describe the quality of approximation only outside the kernel of the seminorm in  $C^{pm}(S)$ . In that part of the subspace the norm of the elements of the basis is well-defined. For this reason, we can write the same expressions for the coefficients for expansion of the value function in the basis  $q^{k(m)}(\cdot)$  but in terms of matrices  $\Gamma^m$  and  $G^m$  defined by the inner product in  $C^{pm}(S)$ . In this case, the problem of evaluation of the difference between the estimate of the value function obtained from  $\Gamma^m$  and  $G^m$  and the true value reduces to two separate problems. The first one is evaluation of the error due to series approximation, which was considered above. The second one is evaluation of the quality of approximation when using empirical measure instead of the true stationary measure. The

general results regarding these properties are given, for instance, in (Billingsley, 1968). Here we will consider a special case when the stationary and empirical measures have densities. We can evaluate the quality of approximation of the value function as:

$$V_i^{m,k(m)}(s) - V_i^{k(m)}(s) = \beta q^p(s) (\Gamma - \beta G_i)^{-1} (G_i^m - G_i) (\Gamma - \beta G_i)^{-1} \Gamma \lambda^p \\ + \beta q^{k(m)}(s) (\Gamma - \beta G_i)^{-1} (\Gamma^m - \Gamma) \left[ I - (\Gamma - \beta G_i)^{-1} \Gamma \right] \lambda_i^{k(m)} + o(\|\Gamma^m - \Gamma\|, \|G_i^m - G_i\|),$$

where the norm in the residual term is a standard matrix norm. This expression has similar structure as the expression for the errors due to estimation of  $h_i(\cdot)$ . From Assumption 5 it follows that traces of matrices  $\Gamma^m - \Gamma$  and  $G_i^m - G_i$  approach to zero faster than  $m^{\max\{q,r\}}$ . This means that in the asymptotic expansion the corresponding term vanishes as well.

This result proves that we can, in general, substitute the matrices  $G_i$  and  $\Gamma$  by their sample versions without affecting the asymptotic variance. The estimate of the value function will take the form:

$$\hat{V}_i^{k(m)}(s, 0) = q^{k(m)}(s)' \left( \hat{\Gamma} - \beta \hat{G}_i \right)^{-1} \hat{\Gamma}' \hat{\lambda}_i^{k(m)},$$

where  $\hat{\Gamma}$  and  $\hat{G}_i$  are sample averages for estimating  $\Gamma$  and  $G$ . For example:

$$\hat{G}_i = \frac{1}{m} \sum_{j=1}^m \frac{1}{T} \sum_{t=1}^{T-1} q^{k(m)}(s_{j,t+1}) q^{k(m)}(s_{j,t})'.$$

In the previous step we have estimated  $V_i(s, l) - V_i(s, 0)$  non-parametrically as  $q^{k(m)'} \gamma_{i,l}^{k(m)}$ . This means that the non-parametric estimate for the choice-specific value function is a combination of the obtained estimate for  $V_i(s, 0)$  and this difference and:

$$\hat{V}_i^{k(m)}(s, l) = q^{k(m)'} \left( \hat{\theta}_i^{k(m)} + \hat{\gamma}_{i,l}^{k(m)} \right).$$

This variable will be normal as it is non-degenerate and computed as a sum of two asymptotically normal estimates. This fact becomes straightforward if we explicitly express coefficients  $\theta_i^{k(m)}$  in terms of  $\gamma_{i,l}^{k(m)}$ . Let  $\gamma_i^{k(m)} = \left( 0, \gamma_{i,1}^{k(m)}, \dots, \gamma_{i,K}^{k(m)} \right)$  be the stacked matrix of coefficients in the expansions for  $V_i(s, l) - V_i(s, 0)$ . We introduce the following vector of logit probabilities:

$$\Lambda = \left( \frac{\exp(V_i(s, l) - V_i(s, 0))}{\sum_{j=0}^K \exp(V_i(s, j) - V_i(s, 0))} \right)_{l=1, \dots, K}$$

Then we can express  $\lambda_i^{k(m)}$  (up to the error of approximation) as:

$$\lambda_i^{k(m)} = \Gamma^{-1} G_i \gamma_i^{k(m)} \Lambda.$$

Therefore, the corresponding coefficients for the value function can be expressed as:

$$\theta_i^{k(m)} = (\Gamma - \beta G_i)^{-1} G_i \gamma_i^{k(m)} \Lambda.$$

Value function can be explicitly estimated from coefficients  $\hat{\gamma}_i^{k(m)}$  and matrices  $G_i$  and  $\Gamma$  as:

$$\hat{V}_i^{k(m)}(s, l) = q^{k(m)'} \left( \hat{\gamma}_{i,l}^{k(m)} + \left( \hat{\Gamma} - \beta \hat{G}_i \right)^{-1} \hat{G}_i \hat{\gamma}_i^{k(m)} \hat{\Lambda} \right).$$

From this estimate one can see that the estimate for the value function is obtained from the estimates for the choice-specific probabilities by permuting them by bounded linear transformations (as  $\sum_t \Lambda_t = 1$  and  $\Lambda_t > 0$ , while the operator represented by the matrix  $I - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2}$  is bounded as shown above). This motivates asymptotic normality with non-degenerate distribution for their estimates. Estimated profit will be, again, a non-degenerate linear combination of the estimates for the choice-specific probabilities, and pointwise normality of the estimate with the rate of convergence, corresponding to the minimum of the convergence rate for the choice specific probability or transition density.

To formalize this recall that we can compute the profit function from the value function by the formula:

$$\Pi_i(s, l) = V_i(s, l) - \beta E [V_i(s') | s, a_i = l].$$

Let  $G_i^{(l)}$  be the matrix corresponding to the state transition density  $g_i(s' | s, l)$  such that  $G_{i,lr}^{(l)} = \int \int g_i(s' | s, l) q_t^{k(m)}(s') q_r^{k(m)}(s) \pi(ds) ds'$ . We can then express the spectral representation for the profit as:

$$\Pi_i^{k(m)}(s, l) = q^{k(m)'}(s) \left( \gamma_{i,l}^{k(m)} + \left\{ [I_{k(m)} - \Gamma^{-1} G_i^{(l)}] [(\Gamma - \beta G_i)^{-1} G_i + I_{k(m)}] - I_{k(m)} \right\} \gamma_i^{k(m)} \Lambda \right).$$

Then we can transform the expression for the profit function as:

$$\begin{aligned} \Pi_i^{k(m)}(s, l) &= q^{k(m)'}(s) \gamma_{i,l}^{k(m)} + \tilde{q}^{k(m)'}(s) \left\{ [I_{k(m)} - \Gamma^{-1/2} G_i^{(l)} \Gamma^{-1/2}] \right. \\ &\quad \left. \times \left[ (I_{k(m)} - \beta \Gamma^{-1/2} G_i \Gamma^{-1/2})^{-1} + \Gamma^{1/2} G_i^{-1} \Gamma^{1/2} \right] \Gamma^{-1/2} G_i \Gamma^{-1/2} - \Gamma^{-1/2} \right\} \tilde{\gamma}_i^{k(m)} \Lambda. \end{aligned}$$

In this expression tildes denote the rotation of the basis considered before. The matrix in the second expression represents a bounded linear transformation due to assumption 3. Therefore the estimate for the profit function is a bounded transformation of the estimate of the choice probabilities. Then we can evaluate the variance-covariance matrix as

$$\begin{aligned} \omega_\pi^2 &= \lim_{m \rightarrow \infty} \text{trace} \left\{ q^{k(m)'}(s) \left( \gamma_{i,l}^{k(m)} + \left\{ [I_{k(m)} - \Gamma^{-1} G_i^{(l)}] [(\Gamma - \beta G_i)^{-1} G_i + I_{k(m)}] - I_{k(m)} \right\} \gamma_i^{k(m)} \Lambda \right) \right. \\ &\quad \left. \times \left( \gamma_{i,l}^{k(m)} + \left\{ [I_{k(m)} - \Gamma^{-1} G_i^{(l)}] [(\Gamma - \beta G_i)^{-1} G_i + I_{k(m)}] - I_{k(m)} \right\} \gamma_i^{k(m)} \Lambda \right)' q^{k(m)'}(s) \right\} \end{aligned}$$



Table 1: Simulation summary for entry utilities, nmarket=100

i	$a_{-i}$	state	1st quartile	median	mean	3rd quartile	std	true Pi
1	1	1	-0.238	-0.006885	-0.0028	0.225	0.36	-0.52
1	1	2	-0.22	0.00004	0.0071	0.2304	0.343	0.749
1	2	1	-0.33	-0.011	-0.0349	0.28	0.48	-1.023
1	2	2	-0.21	0.0038	0.017	0.24	0.34	0.81
2	1	1	-0.25	0.027	0.023	0.31	0.44	0.53
2	1	2	-0.36	-0.021	-0.005	0.35	0.57	-1.005
2	2	1	-0.31	-0.022	0.0032	0.31	0.482	1.15
2	2	2	-0.38	0.013	-0.021	0.36	0.619	-1.600

Table 2: Simulation summary for entry utilities, nmarket=500

i	$a_{-i}$	state	1st quartile	median	mean	3rd quartile	std	true Pi
1	1	1	-0.109	-0.006	-0.002	0.102	0.161	-0.52
1	1	2	-0.095	0.0017	0.0011	0.097	0.146	0.749
1	2	1	-0.15	-0.0044	-0.0068	0.13	0.211	-1.02
1	2	2	-0.092	-0.0059	0.00004	0.1	0.146	0.812
2	2	1	-0.109	0.013	0.009	0.12	0.18	0.53
2	2	2	-0.164	0.0085	-0.0013	0.15	0.23	-1.005
2	2	1	-0.13	0.0017	0.0043	0.14	0.203	1.15
2	2	2	-0.15	0.0001	0.002	0.16	0.24	-1.60

Table 3: Simulation summary for entry utilities, nmarket=1000

i	$a_{-i}$	state	1st quartile	median	mean	3rd quartile	std	true Pi
1	1	1	-0.073	-0.00002	0.001	0.077	0.109	-0.52
1	1	2	-0.075	-0.0049	-0.0032	0.072	0.106	0.749
1	2	1	-0.107	-0.003	-0.005	0.092	0.14	-1.023
1	2	2	-0.066	0.0053	0.004	0.075	0.108	0.812
2	2	1	-0.078	0.0023	0.0046	0.086	0.127	0.537
2	2	2	-0.11	0.0024	-0.0033	0.107	0.166	-1.005
2	2	1	-0.098	-0.0021	-0.001	0.091	0.14	-1.60

Table 4: Simulation summary for entry utilities, nmarket=2000

i	$a_{-i}$	state	1st quartile	median	mean	3rd quartile	std	true Pi
1	1	1	-0.05	0.005	0.0038	0.059	0.0772	-0.52
1	1	2	-0.053	-0.00017	-0.00037	0.051	0.074	0.749
1	2	1	-0.078	-0.0066	-0.0062	0.0603	0.1007	-1.023
1	2	2	-0.045	0.0018	0.0017	0.05	0.075	0.812
2	2	1	-0.055	0.0017	0.0039	0.064	0.089	0.537
2	2	2	-0.088	-0.005	-0.0049	0.07	0.119	-1.005
2	2	1	-0.066	-0.004	-0.0021	0.059	0.097	1.150
2	2	2	-0.079	0.011	0.0051	0.086	0.124	-1.600

Table 5: Simulation summary for entry utilities

sample size	mean	variance	median	90% quantile/10% quantile ratio
50	-0.2075	1.0898	-0.2371	0.0007
100	-0.2064	1.1111	-0.1776	0.0012
150	-0.2075	0.9341	-0.1935	0.0011
200	-0.2056	1.0461	-0.1934	0.0010
250	-0.2047	1.0346	-0.1936	0.0009
300	-0.2041	0.9111	-0.1851	0.0010

Table 6: Summary statistics for sales and promotions

Variable	Obs	Mean	Std. Dev.	Min	Max	25%	50%	75%
Total sales across brands								
in dollars	68102	15.997	51.492	0.11	1795.4	2.39	5.38	12.88
in units	68103	9.025	24.566	1	721	1	3	8
Total sales across markets								
in dollars	7499	145.282	216.266	0.25	1848.5	6.56	33.8	196.1
in units	7499	81.958	120.995	0.25	156	5	23	107
Frequency of promotions	5166	0.283	0.322	0	1	0	0.16	0.5

Table 7: Product market shares following the weeks with a price promotion and without a promotion

Variable	Obs	Mean	Std.Dv.	Min	Max	25%	50%	75%
<i>All sample</i>								
log-market share	149460	-4.134	1.178	-7.355	0	-4.997	-4.241	-3.423
<i>Promotion</i>								
log-market share	22496	-3.916	1.144	-7.355	0	-4.751	-3.985	-3.135
<i>No promotion</i>								
log-market share	126964	-4.173	1.179	-7.355	0	-5.037	-4.290	-3.481

Table 8: Summary of purchases

	Obs	Mean	Std. Dev.	Min	Max
Price per purchase	335230	1.875	0.820	0.059	6.99
# of items purchased	345952	1.278	0.737	0.25	70
Average size of item	345952	0.591	0.251	0.031	3
Time between purchases, weeks	335608	5.588	10.552	1	285
Average number of brands purchased	345952	1.114	0.348	1	6

Table 9: Summary of price promotions

Variable	Obs	Mean	Std.Dv.	Min	Max	25%	50%	75%
<i>All sample</i>								
log-market share	149460	-4.134	1.178	-7.355	0	-4.997	-4.241	-3.423
<i>Promotion</i>								
log-market share	22496	-3.916	1.144	-7.355	0	-4.751	-3.985	-3.135
<i>No promotion</i>								
log-market share	126964	-4.173	1.179	-7.355	0	-5.037	-4.290	-3.481

Table 10: Summary of package sizes

volume, rounded	# of observations	quantity on discount, %	% time on sale	Average discount
0	563	69%	13%	41%
0.1	22253	13%	8%	13%
0.2	3787	8%	11%	12%
0.3	88309	34%	23%	35%
0.4	78056	30%	23%	28%
0.5	15491	32%	18%	16%
0.6	27189	19%	13%	22%
0.7	88105	47%	41%	33%
0.8	116989	51%	33%	34%
0.9	20376	39%	38%	25%
1	17960	12%	14%	19%
1.1	319	18%	11%	27%
1.2	238	30%	32%	16%
1.3	10040	19%	17%	21%
1.4	63	56%	42%	6%
$\geq 1.5$	413	6%	13%	-1%

Table 11: Characteristics of brands in the sample

Rank	Brand	Company	Share (vol.)	% on sale	Share (rev.)
1	LAYS	PEPSICO INC	0.4561	0.4325	0.5018
2	PRINGLES	PROCTER & GAMBLE	0.1304	0.2964	0.0939
3	PRIVATE LABEL	PRIVATE LABEL	0.1191	0.2180	0.0964
4	RUFFLES	PEPSICO INC	0.0630	0.2306	0.0917
5	WISE	PALLADIUM EQUITY PARTNERS	0.0536	0.4183	0.0456
6	OLD DUTCH	OLD DUTCH FOODS INC	0.0398	0.4739	0.0481
7	UTZ	UTZ QUALITY FOODS	0.0311	0.5347	0.0247
8	CAPE COD	CAPE COD POTATO CHIP	0.0276	0.2984	0.0350
9	JAYS	UBIQUITY BRANDS	0.0197	0.5002	0.0224
10	BARREL O FUN	KLN ENTERPRISE	0.0193	0.4562	0.0173
11	POORE BROTHERS	THE INVENTURE GROUP INC	0.0094	0.5740	0.0070
12	TERRA	THE HAIN CELESTIAL GROUP INC	0.0029	0.1980	0.0045
13	STATE LINE	STATE LINE SNACKS CORP	0.0025	0.0755	0.0026
14	KETTLE CHIPS	KETTLE FOODS INC	0.0024	0.1995	0.0031
15	HERRS	HERR FOODS INC	0.0013	0.0109	0.0010
16	GIBBLES	MARTIN SNACKS	0.0012	0.1403	0.0015
17	COTTAGE FRIES	PALLADIUM EQUITY PARTNERS	0.0008	0.1023	0.0007
18	BACHMAN GOLDEN RIDGES	BACHMAN CO	0.0004	0.1734	0.0004
19	GRANDMA UTZ	UTZ QUALITY FOODS	0.0003	0.1203	0.0003
20	UTZ DELITES	UTZ QUALITY FOODS	0.0002	0.4479	0.0003

Table 12: Estimated demand for products in potato chips category

	OLS	IV	BLP
log-price	-0.29 [74.69]***	-1.2 [56.14]***	-1.979 [44.59]***
promotion	0.153 [19.67]***	0.137 [13.99]***	0.144 [7.32]***
Observations	149460	125401	125401

t-statistics are in the braces

Table 13: Aggregate demand estimates

	All sample		On sale		Not on sale	
	log(quantity)	log(quantity)	log(quantity)	log(quantity)	log(quantity)	log(quantity)
log(price)	-1.594 [0.01]***	-1.293 [0.01]***	-1.741 [0.03]***	-1.627 [0.02]***	-1.539 [0.01]***	-1.176 [0.01]***
time	-0.006 [0.0004]***	-0.006 [0.0004]***			-0.008 [0.0004]***	-0.005 [0.0004]***
time <sup>2</sup>	0.00005 [0.000005]***	0.00006 [0.000005]***			0.00006 [0.000005]***	0.00005 [0.000005]***
Constant	2.126 [0.01]***	1.736 [0.01]***	2.215 [0.03]***	2.098 [0.02]***	2.09 [0.02]***	1.557 [0.01]***
brand FE	Yes	No	Yes	No	Yes	No
Observations	139058	139058	37709	37709	101033	101033
R-squared	0.16	0.2	0.11	0.16	0.11	0.16

Table 14: Fraction of products purchased on sale as function of household characteristics

	Fraction of purchase made on sale			
Pre-tax income	-0.001 [0.000]*	-0.001 [0.000]**	-0.001 [0.000]	-0.001 [0.000]**
Family size	0.014 [0.001]***	0.012 [0.001]***	0.014 [0.001]***	0.012 [0.001]***
Education male	0.001 [0.001]	0.001 [0.001]	0 [0.001]	0.001 [0.001]*
Age male	-0.002 [0.001]**	0.001 [0.001]	-0.002 [0.001]*	0.001 [0.001]
Education female	-0.001 [0.001]**	-0.002 [0.001]***	-0.002 [0.001]***	-0.002 [0.001]***
Age female	0.008 [0.001]***	0.007 [0.001]***	0.007 [0.001]***	0.008 [0.001]***
I(not employed, male)	0.095 [0.010]***	0.006 [0.010]	0.023 [0.010]**	0.094 [0.010]***
I(Part time, male)	0.022 [0.024]	-0.018 [0.024]	0.002 [0.024]	0.006 [0.025]
I(Full time, male)	0.018 [0.024]	-0.017 [0.024]	-0.003 [0.024]	0.009 [0.025]
I(Retired, male)	0.118 [0.028]***	0.089 [0.028]***	0.094 [0.028]***	0.118 [0.029]***
I(Retired, male)	0.061 [0.003]***	-0.005 [0.004]	0.002 [0.004]	0.067 [0.003]***
I(Student, male)	0.028 [0.019]	-0.036 [0.019]*	-0.022 [0.018]	0.027 [0.019]
I(not employed, female)	0.067 [0.009]***	-0.023 [0.010]**	-0.016 [0.010]*	0.079 [0.010]***
I(Part time, female)	-0.019 [0.021]	-0.024 [0.021]	-0.045 [0.021]**	0.009 [0.021]
I(Full time, female)	-0.021 [0.021]	-0.026 [0.021]	-0.048 [0.021]**	0.008 [0.021]
I(Retired, female)	-0.049 [0.025]**	-0.048 [0.025]*	-0.08 [0.024]***	-0.007 [0.025]
I(Retired, female)	0.029 [0.008]***	-0.027 [0.008]***	-0.023 [0.008]***	0.037 [0.009]***
I(Student, female)	0.053 [0.015]***	0.021 [0.015]	0.002 [0.014]	0.086 [0.015]***
Market FE	Yes	No	Yes	No
Time FE	No	Yes	Yes	No
Marital status FE	Yes	Yes	Yes	Yes
Occupation, male FE	Yes	Yes	Yes	Yes
Occupation, Female FE	Yes	Yes	Yes	Yes
HH race FE	Yes	Yes	Yes	Yes
Demographics variables: I(N/A)	Yes	Yes	Yes	Yes
Constant	0.211 [0.036]**	0.216 [0.036]***	-0.128 [0.039]***	0.195 [0.037]***
Observations	354380	354380	354380	354380
R-squared	0	0	0.04	0



Table 15: Demand as a function of sale parameters

	log(quantity)				
log(price)	-0.05 [0.003]***	-0.038 [0.003]***	-0.022 [0.003]***	-0.046 [0.003]***	-0.023 [0.003]***
I(sale)	0.13 [0.004]***	0.154 [0.004]***	0.197 [0.004]***	0.149 [0.005]***	0.194 [0.004]***
I(sale) X log(price)	-0.086 [0.004]***	-0.096 [0.004]***	-0.136 [0.004]***	-0.09 [0.004]***	-0.137 [0.004]***
Size	2.018 [0.003]***	2.118 [0.003]***	2.128 [0.003]***	2.115 [0.003]***	2.128 [0.003]***
Household FE	Yes	No	No	No	No
Time FE	No	No	Yes	Yes	No
Market FE	No	Yes	No	Yes	No
Flavour FE	Yes	Yes	Yes	Yes	Yes
Fat content FE	Yes	Yes	Yes	Yes	Yes
Produser FE	Yes	Yes	Yes	Yes	Yes
Cooking stile FE	Yes	Yes	Yes	Yes	Yes
Salt/sodium content	Yes	Yes	Yes	Yes	Yes
Cut type FE	Yes	Yes	Yes	Yes	Yes
Packegae type FE	Yes	Yes	Yes	Yes	Yes
Constant	-0.044 [0.327]	-0.083 [0.349]	-0.16 [0.351]	0.643 [0.372]*	-0.171 [0.352]
Observations	471953	471953	471953	471953	471953
R-squared	0.62	0.67	0.67	0.68	0.68

Table 16: Brands of potato chips in the considered markets

Brand rank	Brand Name	Brand owner
1	LAYS	PEPSICO INC
2	PRINGLES	PROCTER & GAMBLE
3	PRIVATE	PRIVATE LABEL
4	RUFFLES	PEPSICO INC
5	WISE	PALLADIUM EQUITY PARTNERS
6	OLD DUTCH	OLD DUTCH FOODS INC
7	UTZ	UTZ QUALITY FOODS
8	CAPE CODE	CAPE COD POTATO CHIP
9	JAYS	UBIQUITY BRANDS
10	BARREL	KLN ENTERPRISE
11	POORE	THE INVENTURE GROUP INC
12	TERRA	THE HAIN CELESTIAL GROUP INC
13	STATE	STATE LINE SNACKS CORP
14	KETTLE	KETTLE FOODS INC
15	HERRS	HERR FOODS INC
16	GIBBLES	MARTIN SNACKS
17	COTTAGE	PALLADIUM EQUITY PARTNERS
18	BACHMAN	BACHMAN CO
19	GRANDMA	UTZ QUALITY FOODS
20	UTZ	UTZ QUALITY FOODS
21	OTHER	OTHER

Table 17: Coefficient estimates for consumer choice model with serially correlated consumer and brand-specific unobserved heterogeneity: part 1

Variables	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
log(Price)	-3.093084	-2.839459	-2.731196	-3.163044	-2.692065
Pkg. quantity (oz)	0.023648	0.027227	0.113665	0.052785	0.056427
Packaging: bag	1.083336	2.257118	2.076922	0.160061	1.120758
Box	0.167489	0.710799	0.989843	0.443545	0.412596
Canister	2.126507	1.08645	0.918042	2.380979	0.004586
other	2.130894	0.857171	3.322598	0.313321	1.393228
Flavor: Barbeque	2.516947	1.121407	0.927539	2.578106	0.830753
Cheddar	0.002148	2.123562	2.887561	0.81756	1.116835
Classic	0.081069	3.319681	0.056023	1.339931	0.726215
Dill	0.09442	1.97255	0.983903	2.120143	0.336933
Other	0.894347	1.443588	3.086589	3.301995	0.839248
Missing	0.401883	1.362309	0.945448	3.126884	0.871189
Cut: Flat	17.504062	11.922788	16.878971	5.546945	18.11449
Rippled	0.563384	6.881071	2.475236	4.990139	4.607974
Wavy	2.054767	1.529476	3.308461	2.575336	0.23931
Other	1.893328	2.60554	3.370363	2.50533	2.519205
Missing	0.767867	1.453247	3.402031	3.001584	2.016039
Fat indication: Reduced	0.412106	0.660561	1.31856	1.450738	2.816783
Regular	3.331358	1.393665	1.095798	1.610804	2.088087
Missing	2.954277	3.163212	0.911828	2.750444	3.085395
Cooking: Crispy	3.358041	1.506079	3.125307	1.668809	0.502346
Kettle cooked	0.87759	2.51677	3.264619	3.215015	2.202708
Other	0.588135	0.294594	2.573136	1.256161	1.134427
Missing	2.564087	2.952459	1.128398	0.632139	0.373934
Sodium indication: Reduced	1.055897	3.10707	2.399652	3.10766	3.029386
Regular	1.993188	3.427846	1.26847	0.947693	2.653605
Missing	1.902018	2.761925	1.884527	2.683362	3.241812
Pre-tax income	-2.771089	-1.819382	-2.430073	-2.297347	-0.248018
Family size	5.349905	4.103607	5.144125	5.241447	0.725615
Age male	-0.422976	-0.751685	-0.173533	-0.101178	-0.522588
Education male	-12.245675	-4.573244	-2.860647	-7.009571	-30.733799
Age Female	-5.205421	-2.352651	-5.136977	-2.208746	-2.728967
Education Female	-8.045068	-1.045667	-15.690651	-8.1239	-2.483252
Work male	-2.347099	-0.738439	-0.881929	-2.716539	-2.984811
Work female	0.077111	0.070068	0.06911	0.06611	0.003634
Hispanic	0.97801	0.246425	2.268229	2.124053	1.485854

Table 18: Coefficient estimates for consumer choice model with serially correlated consumer and brand-specific unobserved heterogeneity: part 2

Variables	Brand 6	Brand 7	Brand 8	Brand 9	Brand 10
log(Price)	-2.911434	-2.998795	-2.950773	-2.672683	-2.473143
Pkg. quantity (oz)	0.077634	0.156119	0.090197	0.175843	0.192358
Packaging: bag	1.381033	1.113762	0.708142	0.975481	0.81713
Box	0.177706	0.102814	0.658632	0.757247	0.039964
Canister	1.570065	2.636845	0.598895	3.094523	0.452339
other	0.431822	0.309739	2.399878	2.287465	0.130452
Flavor: Barbeque	0.62645	1.018675	0.734392	0.183648	1.669077
Cheddar	2.235641	2.613372	3.331874	2.859479	1.202645
Classic	2.038993	2.156949	1.807655	1.38874	0.774996
Dill	0.487649	0.054597	2.637733	2.900615	2.676727
Other	2.965642	0.104623	0.129827	1.592186	0.989739
Missing	1.134811	0.427577	2.267329	0.306437	0.172047
Cut: Flat	13.151739	2.881501	1.738296	18.662324	22.868188
Rippled	5.639751	4.017684	6.745104	3.429067	5.441019
Wavy	2.818587	1.890329	3.394326	0.61501	3.239276
Other	0.912543	3.296327	2.683909	2.52803	0.541938
Missing	1.983328	2.267515	1.093936	1.991945	1.574072
Fat indication: Reduced	1.241271	2.117671	0.777772	0.544591	2.787941
Regular	2.018017	1.582632	0.866518	1.31806	0.939611
Missing	1.531191	0.998216	2.905236	0.728097	1.663594
Cooking: Crispy	3.004646	0.930495	1.357602	3.364759	2.223727
Kettle cooked	2.932436	0.864896	0.136	1.778929	0.213116
Other	1.932747	1.154702	0.164381	2.008287	2.578632
Missing	2.235834	2.947569	2.862703	2.566581	2.177511
Sodium indication: Reduced	3.339568	0.92091	0.371203	1.002449	0.12617
Regular	0.857643	0.135918	1.918062	3.303149	1.49087
Missing	1.587419	2.271327	2.661004	1.24907	2.247563
Pre-tax income	-2.265123	-2.31245	-3.10701	-1.603886	-3.144913
Family size	2.361773	2.349088	2.489815	2.285943	1.438528
Age male	-0.253109	-0.62194	-0.731417	-1.028493	-1.187941
Education male	-31.260136	-24.93648	-29.108916	-8.994038	-31.168224
Age Female	-3.391082	-5.489202	-1.12166	-4.483172	-5.019412
Education Female	-5.893146	-0.327465	-23.465545	-33.513486	-33.660915
Work male	-3.223599	-1.585419	-2.77831	-3.06848	-2.867074
Work female	0.060679	0.05034	0.011888	0.066718	0.033457
Hispanic	2.097809	2.375445	2.443606	1.414577	1.602941

Table 19: Coefficient estimates for consumer choice model with serially correlated consumer and brand-specific unobserved heterogeneity: part 3

Variables	Brand 11	Brand 12	Brand 13	Brand 14	Brand 15
log(Price)	-2.832483	-3.377132	-2.771604	-3.497284	-2.529089
Pkg. quantity (oz)	0.049913	0.230882	0.053747	0.058217	0.041049
Packaging: bag	1.600841	0.118706	0.617699	2.056933	0.869879
Box	0.66783	0.353583	0.037915	0.630748	0.320821
Canister	0.139967	1.525342	1.816851	2.932369	2.517147
other	1.064437	1.152173	0.740411	2.18755	1.210693
Flavor: Barbeque	2.461822	3.416663	2.033453	0.193975	3.045213
Cheddar	4.808022	2.221313	3.301636	0.329845	3.980536
Classic	1.07844	1.917593	1.685969	0.338605	0.856008
Dill	2.687815	2.200718	1.338211	2.977077	1.352452
Other	3.090167	1.616635	0.548024	1.587683	0.538777
Missing	1.208416	1.091454	1.163744	1.782808	2.062348
Cut: Flat	0.569647	13.65379	4.337904	0.914032	9.027001
Rippled	9.962334	7.662772	6.120877	7.448297	7.093168
Wavy	1.224002	1.545203	2.729351	0.558106	0.695127
Other	3.228049	2.306742	0.349161	2.128965	1.702242
Missing	1.025563	1.182387	1.9929	1.314187	0.275021
Fat indication: Reduced	0.494604	0.618838	2.900136	0.315187	2.544821
Regular	0.977317	1.664291	3.103324	1.754553	1.279596
Missing	0.453452	1.897675	2.550097	0.366025	2.723793
Cooking: Crispy	1.282562	1.199485	0.76986	0.922236	2.201934
Kettle cooked	0.846958	2.254751	0.684318	1.126301	2.962456
Other	0.490823	1.409014	0.613642	0.085044	2.266251
Missing	3.066277	2.290534	1.426655	2.805567	0.900059
Sodium indication: Reduced	0.220772	2.03641	2.992886	0.761481	2.594839
Regular	3.278297	0.896235	0.785438	0.877279	2.438604
Missing	1.769905	0.108077	0.465667	1.745102	3.205335
Pre-tax income	-3.107872	-2.114851	-0.482997	-0.726054	-1.734766
Family size	5.085033	4.436701	5.358412	4.696524	0.313327
Age male	-0.308294	-0.933368	-0.518613	-0.942868	-1.336357
Education male	-11.639638	-10.194708	-20.207561	-14.761104	-21.938081
Age Female	-3.958591	-2.246844	-0.810577	-5.343278	-0.145376
Education Female	-22.979569	-15.827254	-11.254127	-39.739514	-12.535587
Work male	-2.465584	-2.559206	-0.994668	-1.205902	-1.376537
Work female	0.073555	0.005894	0.063267	0.048208	0.080662
Hispanic	1.636374	1.335034	0.645398	1.331692	2.231147

Table 20: Coefficient estimates for consumer choice model with serially correlated consumer and brand-specific unobserved heterogeneity: part 4

Variables	Brand 16	Brand 17	Brand 18	Brand 19	Brand 20
log(Price)	-2.65456	-2.671703	-3.402789	-2.876268	-2.610249
Pkg. quantity (oz)	0.214722	0.069933	0.023916	0.117554	0.080076
Packaging: bag	0.329038	2.010122	1.392058	0.508031	0.265844
Box	0.456768	0.808673	0.767574	0.344338	0.916292
Canister	2.430506	1.36155	0.801319	1.837389	1.128458
other	3.304824	2.336067	1.889204	1.383907	3.423793
Flavor: Barbeque	2.318563	2.898738	0.661945	3.182186	2.134064
Cheddar	1.248369	3.641905	3.100562	4.430765	2.862633
Classic	0.740903	3.341088	2.813865	1.194648	2.614328
Dill	2.418467	0.927911	1.732779	3.373826	3.333253
Other	2.925802	1.586381	2.942008	1.538393	3.146978
Missing	1.738911	1.001665	2.457159	1.28895	2.800001
Cut: Flat	5.157646	8.127552	9.455298	1.931442	8.526636
Rippled	8.278694	4.803991	6.586358	7.938665	1.443765
Wavy	1.759582	1.992636	3.14824	1.00224	3.244111
Other	0.073455	0.367029	1.069979	3.27423	0.015846
Missing	1.698027	0.919257	1.581047	2.866553	0.445962
Fat indication: Reduced	3.12346	0.740612	1.941488	1.099642	2.53422
Regular	1.667674	2.186504	0.131235	0.039267	2.979704
Missing	0.862517	2.932301	1.131967	2.435796	1.551027
Cooking: Crispy	2.948857	2.600496	1.024544	2.865601	2.281548
Kettle cooked	0.591628	0.549288	2.21809	1.650761	2.90468
Other	3.078843	0.158544	1.708176	2.033724	1.019161
Missing	2.520156	0.240018	2.243487	2.816472	2.67039
Sodium indication: Reduced	0.188314	3.335747	1.835417	0.158818	1.007104
Regular	1.760145	3.061989	2.362086	1.47624	3.291899
Missing	1.216696	1.309844	1.665115	1.864185	0.148474
Pre-tax income	-3.150911	-0.298756	-1.584276	-0.723553	-1.788233
Family size	4.497552	0.147249	1.090846	4.11252	3.563815
Age male	-0.243806	-0.640892	-0.126559	-0.497055	-1.641209
Education male	-7.518827	-20.366765	-32.768399	-22.227705	-23.662763
Age Female	-4.674357	-3.098578	-0.587867	-3.512758	-3.791617
Education Female	-38.776416	-18.424189	-30.783284	-0.613146	-39.625269
Work male	-3.174855	-3.153605	-0.373659	-0.98848	-0.507682
Work female	0.076362	0.019372	0.015398	0.052889	0.049528
Hispanic	1.311957	1.313039	0.871052	1.5276	1.721466

Table 21: Coefficient estimates for consumer choice model without unobserved heterogeneity: part 1

Variables	Brand 1	Brand 2	Brand 3	Brand 4	Brand 5
log(Price)	-4.165822	-5.386564	-5.365204	-4.193513	-4.300462
Pkg. quantity (oz)	0.110578	0.158264	0.068733	0.010223	0.011204
Packaging: bag	0.326362	0.341004	0.864795	0.2872	0.292761
Box	0.021068	0.650912	0.374105	0.37723	0.080592
Canister	0.13326	0.476179	0.473303	1.558845	1.417243
other	2.212735	0.813927	2.457802	0.536493	1.168099
Flavor: Barbeque	0.309188	0.715059	1.573114	0.651692	0.565978
Cheddar	1.926521	0.746351	2.479831	0.163897	1.361593
Classic	0.178167	1.317739	2.055443	1.609769	2.047781
Dill	2.423557	0.80767	1.831116	1.547685	0.127566
Other	0.783219	0.675258	1.011162	0.922373	0.884892
Missing	0.547649	0.359212	0.936272	1.351815	1.139046
Cut: Flat	17.424891	2.951845	8.680732	14.586069	8.236349
Rippled	2.459214	8.556563	4.837721	6.878325	4.889042
Wavy	1.788228	1.713201	0.039203	0.199366	1.706493
Other	0.562744	0.725479	1.555318	2.634525	2.449236
Missing	1.102998	0.022778	0.68523	0.808989	0.654261
Fat indication: Reduced	0.588252	2.558548	1.009842	1.806997	0.704029
Regular	0.279152	1.561708	0.329483	0.764724	1.387526
Missing	1.378011	2.576356	1.370222	0.298344	0.339283
Cooking: Crispy	0.790333	1.125594	1.361933	1.990712	0.328233
Kettle cooked	1.506478	1.301525	1.977683	2.078281	0.896909
Other	2.493808	0.426577	0.866939	1.715311	1.046856
Missing	0.516302	0.22664	2.414835	1.236038	1.472778
Sodium indication: Reduced	1.525073	1.469333	1.995248	1.673107	2.143054
Regular	1.019638	1.781308	1.90269	0.744342	2.298365
Missing	0.161072	1.434877	1.625611	0.744006	0.474184
Pre-tax income	-0.855846	-4.250458	-4.463546	-1.127139	-5.5819
Family size	2.691766	2.314595	4.08801	7.095724	2.540808
Age male	-1.041996	-2.443184	-1.534026	-1.110189	-0.93047
Education male	-6.242541	-40.394791	-19.719089	-16.976177	-29.850739
Age Female	-6.627981	-0.727685	-6.96913	-8.556765	-7.504168
Education Female	-40.128482	-26.834798	-61.137031	-1.390044	-53.885656
Work male	-4.449358	-1.87242	-3.109071	-1.798329	-3.875555
Work female	0.052662	0.03255	0.108231	0.023519	0.094144
Hispanic	0.880216	0.282617	0.579385	3.555073	0.895571

Table 22: Coefficient estimates for consumer choice model without unobserved heterogeneity: part 2

Variables	Brand 6	Brand 7	Brand 8	Brand 9	Brand 10
log(Price)	-5.942396	-6.631553	-5.33314	-5.463779	-4.585297
Pkg. quantity (oz)	0.100403	0.015798	0.046498	0.062104	0.015571
Packaging: bag	1.528673	1.262471	0.397429	0.052525	0.42474
Box	0.262017	0.248657	0.11899	0.140372	0.084643
Canister	2.208332	0.935361	1.631955	0.604184	1.933457
other	1.198789	0.458909	0.613683	1.63443	1.100732
Flavor: Barbeque	1.640518	0.963101	1.434067	0.596472	0.835032
Cheddar	2.138886	1.101339	0.476418	0.622793	1.272544
Classic	1.720011	0.016727	0.56878	1.235842	0.899407
Dill	0.19335	1.558747	0.680782	0.835989	2.193451
Other	0.198886	1.760965	2.352432	1.586522	1.508106
Missing	1.228729	1.48904	1.226602	0.543439	1.694876
Cut: Flat	10.755164	12.723075	11.819334	9.378905	1.896329
Rippled	5.560848	3.575512	1.021045	4.216397	1.677433
Wavy	1.122226	0.519879	1.279616	1.113999	1.249871
Other	0.96761	0.947765	0.868797	0.934202	1.361492
Missing	0.090624	1.813063	0.23638	2.639069	0.572035
Fat indication: Reduced	1.008904	0.377457	1.356393	1.069813	0.394479
Regular	1.053043	0.139044	1.877635	1.168561	0.61018
Missing	0.270797	1.558468	0.532695	1.197255	1.424094
Cooking: Crispy	0.408829	1.690843	1.941648	0.603736	1.931615
Kettle cooked	1.650217	1.177204	0.948795	2.162084	0.329338
Other	0.284677	1.255097	0.860667	0.244123	2.039686
Missing	0.992637	0.273154	0.341603	2.016055	1.901628
Sodium indication: Reduced	0.727254	0.928646	0.406623	1.989706	0.056746
Regular	1.524695	1.783301	1.887261	1.243915	1.597559
Missing	1.388155	0.185051	1.613794	1.575446	2.259512
Pre-tax income	-4.487869	-5.50511	-0.798736	-1.827621	-3.104033
Family size	6.77549	9.872027	6.009347	5.543223	8.759527
Age male	-1.537945	-1.975047	-1.061209	-2.063721	-1.270058
Education male	-45.872031	-30.453402	-42.889598	-25.072204	-13.63651
Age Female	-6.204733	-4.661933	-5.370452	-1.054304	-6.491553
Education Female	-15.772796	-63.238631	-40.137163	-63.13209	-55.651209
Work male	-0.062776	-3.253028	-0.825571	-3.887503	-2.180464
Work female	0.113847	0.110697	0.045031	0.048243	0.051391
Hispanic	2.772715	2.507646	1.522741	0.235443	2.912699



Table 23: Coefficient estimates for consumer choice model without unobserved heterogeneity: part 3

Variables	Brand 11	Brand 12	Brand 13	Brand 14	Brand 15
log(Price)	-5.701579	-6.153897	-5.569636	-4.179997	-6.387933
Pkg. quantity (oz)	0.052347	0.044234	0.065901	0.123	0.004293
Packaging: bag	0.599797	1.20975	0.585687	0.570026	0.71747
Box	0.539036	0.117165	0.400723	0.48628	0.452487
Canister	1.328678	2.239394	0.685635	1.858454	0.569346
other	1.003705	0.999802	1.294024	2.080618	0.485416
Flavor: Barbeque	0.707111	0.099615	0.710088	1.994047	1.773578
Cheddar	0.781812	1.745907	2.550642	2.369324	3.359982
Classic	0.036601	1.054822	1.081916	0.555373	2.8331
Dill	0.140996	1.009944	1.145985	1.509845	1.1288
Other	0.094312	1.269004	0.766857	0.341079	0.458858
Missing	1.173771	0.206495	1.116473	1.512532	1.621733
Cut: Flat	5.477097	4.056816	4.275265	6.66029	3.848013
Rippled	3.444869	3.747512	3.94062	2.655449	6.212754
Wavy	1.92152	0.341824	2.036878	1.138402	0.472894
Other	1.011258	1.269357	1.062805	1.50636	1.926623
Missing	0.825069	0.065132	0.967002	0.363923	1.275612
Fat indication: Reduced	1.483487	1.72401	0.301453	0.96603	1.035122
Regular	0.226998	0.878857	0.60496	2.126524	1.648472
Missing	0.005091	1.563122	0.359839	0.829744	1.657967
Cooking: Crispy	0.336888	0.320954	0.735975	1.530435	1.16523
Kettle cooked	2.398844	0.459408	0.514702	2.343202	2.022477
Other	2.569468	2.021763	0.302024	1.608572	1.834448
Missing	0.028339	0.764783	0.058601	0.243557	1.648182
Sodium indication: Reduced	0.511914	1.184127	1.130265	0.835812	2.112263
Regular	1.41527	0.619621	0.74064	1.438851	0.617883
Missing	0.237487	2.2337	2.702513	1.726688	1.460125
Pre-tax income	-0.471282	-2.858001	-0.793913	-1.134544	-3.011794
Family size	2.840776	2.882248	5.98188	5.830896	4.483393
Age male	-2.444957	-1.277594	-2.089676	-2.057892	-0.212153
Education male	-30.61355	-2.350366	-6.891827	-14.514786	-11.433809
Age Female	-2.084504	-6.290976	-1.038558	-0.632368	-8.778083
Education Female	-0.098408	-44.789743	-41.493357	-33.137318	-60.581205
Work male	-3.686434	-4.003346	-4.008495	-5.977308	-4.636364
Work female	0.017135	0.008615	0.049143	0.060251	0.148519
Hispanic	2.782203	3.744826	3.835797	2.177668	0.887173

Table 24: Coefficient estimates for consumer choice model without unobserved heterogeneity: part 4

Variables	Brand 16	Brand 17	Brand 18	Brand 19	Brand 20
log(Price)	-3.657458	-5.087297	-5.721577	-5.394077	-5.820561
Pkg. quantity (oz)	0.072216	0.096999	0.005196	0.12646	0.101312
Packaging: bag	0.323703	1.503906	0.791686	0.708138	0.55918
Box	0.130689	0.643636	0.62666	0.627598	0.002174
Canister	2.34006	0.310726	1.182337	1.183339	1.397798
other	0.651848	0.510269	0.123232	0.399809	0.872859
Flavor: Barbeque	1.537395	1.627388	0.584187	0.763447	2.562344
Cheddar	2.241224	1.9829	1.292326	3.110279	1.123984
Classic	1.650107	0.098588	0.210854	0.777324	2.025672
Dill	0.04128	2.20342	0.390171	1.445992	0.368976
Other	1.270501	1.698583	0.205198	2.352648	2.173152
Missing	0.977024	2.53051	0.448205	0.086543	0.735897
Cut: Flat	15.983062	17.587212	9.795252	15.078067	11.286216
Rippled	7.974686	2.273874	0.534869	0.222243	3.023763
Wavy	0.918974	0.797805	0.756303	2.344447	0.462558
Other	0.036814	1.513505	0.217719	1.01393	2.10527
Missing	0.634043	1.121361	1.650162	2.136423	1.058599
Fat indication: Reduced	0.550543	0.402566	2.045991	1.852582	0.301347
Regular	0.306531	0.495048	0.741667	1.402807	1.923906
Missing	2.011747	1.592642	0.263512	0.151334	1.203223
Cooking: Crispy	1.040639	0.526655	0.264302	1.35717	1.548327
Kettle cooked	0.542647	0.891644	0.51518	2.021249	1.438555
Other	0.867668	1.614956	1.297056	0.286831	1.274577
Missing	0.579038	0.933107	1.665178	1.621085	0.555817
Sodium indication: Reduced	1.744876	1.80043	1.284818	0.372325	0.590017
Regular	0.189598	0.499006	1.413911	0.750868	1.271185
Missing	0.103778	0.658885	0.133986	1.628518	0.977413
Pre-tax income	-5.897316	-3.518929	-1.44832	-3.32489	-3.572895
Family size	3.071121	6.79469	0.691731	6.581674	1.513708
Age male	-1.217371	-0.225644	-2.478639	-0.362583	-0.681375
Education male	-26.642239	-22.698188	-42.233385	-29.881693	-15.795049
Age Female	-7.719932	-0.72764	-0.979365	-4.417174	-7.825331
Education Female	-30.140097	-23.681847	-43.966854	-46.336628	-31.340084
Work male	-5.456659	-2.276498	-2.56086	-2.113586	-1.737959
Work female	0.055232	0.14906	0.014044	0.060431	0.013366
Hispanic	0.428278	4.600718	1.283886	0.585398	3.790626

Table 25: Impact of an unexpected 10% price drop for “Lay’s potato chips” on the price elasticity

Week	Demand model				
	OLS	IV	BLP (random coefficients)	BLP (no random coefficients)	Dynamic decision model (random coefficients + serial correlation)
1	-0.29134	-1.21362	-2.92362	-3.69832	-2.85327
2	-0.29134	-1.21362	-2.92362	-3.69832	-2.83287
3	-0.29134	-1.21362	-2.92362	-3.69832	-2.82933
4	-0.29134	-1.21362	-2.92362	-3.69832	-2.63647
5	-0.29134	-1.21362	-2.92362	-3.69832	-2.51385
6	-0.29134	-1.21362	-2.92362	-3.69832	-2.49583
7	-0.29134	-1.21362	-2.92362	-3.69832	-2.47564
8	-0.29134	-1.21362	-2.92362	-3.69832	-2.46541
9	-0.29134	-1.21362	-2.92362	-3.69832	-2.45695
10	-0.29134	-1.21362	-2.92362	-3.69832	-2.44564

Figure 1: Median and Percentiles of numerically recovered utility

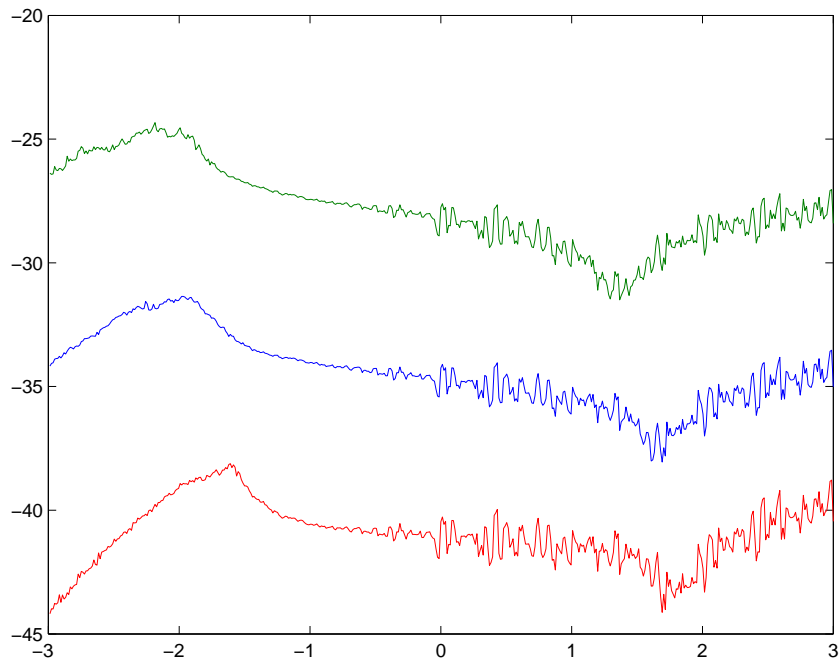


Figure 2: Histogram of variances of log-market shares over products and over markets

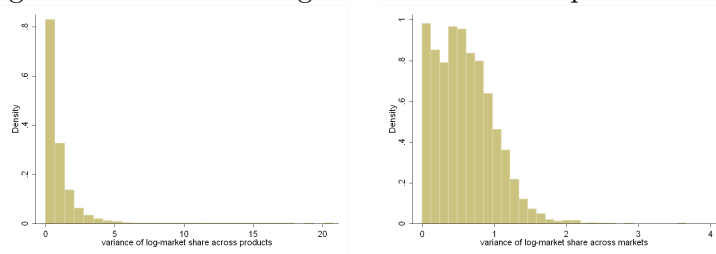


Figure 3: Histogram of log-market shares following the weeks with and without price promotions

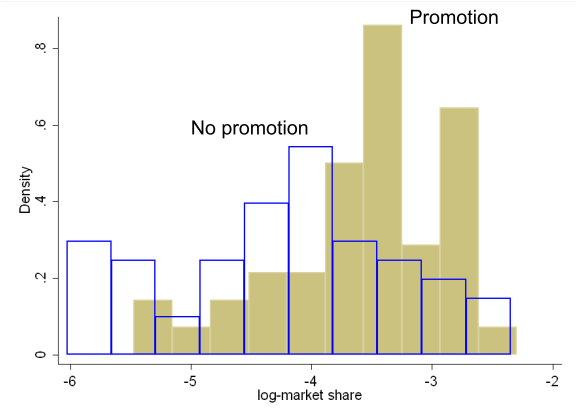


Figure 4: Histogram of log-market shares following the weeks with and without price promotions

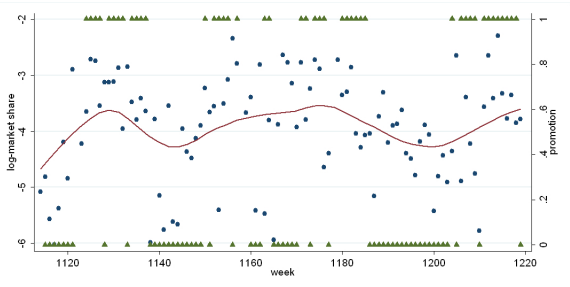


Figure 5: Dependence of the current log-market share from lagged log-market share the weeks with and without price promotions

