BORRADORES DE ECONOMÍA Y FINANZAS

Editor

Jhon James Mora

Jefe, Departamento de Economía

jjmora@icesi.edu.co

Gestión editorial

Departamento de Economía – Universidad Icesi

Contenido:

46 páginas

ISSN 1990-1568

Tercera Edición, Septiembre de 2010

Estimating dynamic models with aggregate shocks and an application to mortgage default in Colombia

Juan Esteban Carranza
Salvador Navarro*

Abstract

We estimate a dynamic model of mortgage default for a cohort of Colombian debtors between 1997 and 2004. We use the estimated model to study the effects on default of a class of policies that affected the evolution of mortgage balances in Colombia during the 1990's. We propose a framework for estimating dynamic behavioral models accounting for the presence of unobserved state variables that are correlated across individuals and across time periods. We extend the standard literature on the structural estimation of dynamic models by incorporating an unobserved common correlated shock that affects all individuals' static payoffs and the dynamic continuation payoffs associated with different decisions. Given a standard parametric specification the dynamic problem, we show that the aggregate shocks are identified from the variation in the observed aggregate behavior. The shocks and their transition are separately identified, provided there is enough cross-sectional variation of the observed states.

1 Introduction

In this paper we specify a dynamic model of mortgage default and estimate it using micro-level Colombian data spanning the years between 1998 and 2004. During this time, mortgage default rates in Colombia were unusually high due to an unprecedented economic downturn that was accompanied by a dramatic fall in home prices. The extent to which the fall in household incomes and the fall in home prices contributed separately to the unprecedented

*University of Wisconsin-Madison

rates of default is a relevant policy question that can be answered within the model we propose. In addition, we use the model to evaluate the impact of counterfactual policies which cannot be evaluated with a model that doesn't account for the dynamic concerns of debtors. We show that in the context of our data, the expectations of individuals regarding the evolution of relevant variables had a substantial impact on default behavior.

The estimation of discrete choice dynamic models is limited by the ability of standard microeconometric techniques to incorporate a rich pattern of unobserved heterogeneity affecting the choices of individuals. In the context of our data, accounting for common unobserved shocks is crucial for understanding the relationship between the observed states and the observed default behavior. The standard techniques for estimating such behavioral models are based on the assumption that all the unobserved heterogeneity is independent across individuals [1]. In this paper we develop a framework for estimating dynamic structural models under the presence of unobserved states that are both correlated across individuals *and* over time, due to the presence of unobserved common states.

The literature on the estimation of structural models that allow for correlated common shocks is scarce. For example, in the approach proposed by Altug and Miller (1998) the structure of the aggregate shocks is estimated separately and used as input into the dynamic model, which is then estimated using the technique developed by Hotz and Miller (1993). Such approach is only practical when the aggregate shocks can be estimated from a separate model (e.g. a macroeconomic model). A closer paper to ours is Lee and Wolpin (2006), in which the aggregate shocks are computed throughout the estimation algorithm using a general equilibrium model. In their case, the estimation is complicated by the need to solve the equilibrium throughout the estimation to obtain the aggregate shocks and their transition.

The methodological contribution of our paper is the incorporation, identification and estimation of the unobserved states that generate correlation both in the cross section and over time using a standard micro data set. In other words, we show that estimating a dynamic model with aggregate unobserved shocks doesn't require the solution of an aggregate model. In our model, in addition to a time invariant individual specific unoberved state, there is

[1] Early papers include Rust (1987), Wolpin (1984), Pakes (1986) and Hotz and Miller (1993); later papers as Keane and Wolpin (1994) incorporate unobserved states that vary systematically across individuals but stay constant over time. For a comprehensive review of the literature see, for example, Aguirregabiria and Mira (2002))

an unobserved correlated state that is common to all individuals and that is correlated over time. For simplicity, we refer to these unobserved correlated common states as aggregate shocks. Our specification of the dynamic model is based on a Markovian decision problem with finite horizon in which the payoffs depend on observed and unobserved state variables that vary systematically across individuals. As we show, in this particular formulation of the dynamic model, the micro data contains enough information to infer the aggregate shocks and their transition separately. The identification and estimation of these common correlated states exploits the variation in aggregate behavior, which is a piece of information that is not used directly by the existing literature. We show conditions under which these aggregate unobserved shocks and their transition probability are separately identified in a standard specification of a dynamic discrete choice model.

In the next section of the paper we describe our methodological framework. We formulate an optimal stopping problem with correlated unobserved heterogeneity, describe our estimation approach and discuss the identification of the different components of the model. In Section 3 of the paper we present the application of the model to the Colombian mortgage market. We describe the data, the estimation and the results. We perform counterfactual simulations to evaluate the impact of the policies adopted by the Central Bank and the Colombian government in the mid-1990s. The paper concludes with a discussion of the limitations of the proposed framework.

2 The framework

Consider the problem of a mortgage debtor who is deciding whether to default or continue making the mortgage payments on his home. This problem can be described as a discrete choice problem in which the choice of defaulting generates a payoff associated with the increased probability of foreclosure, a more restrictive access to the credit markets in the future, etc. Continuing making the mortgage payments generates a static payoff associated with the continued enjoyment of the home, plus the option of making the same decision the next period (i.e. the continuation value).

Formally, denote the flow utility that the individual i obtains from enjoying the home at time t as $\tilde{u}(\tilde{S}_{i,t})$ and the flow utility associated with the choice of default as $W(\tilde{S}_{i,t})$, where $\tilde{S}_{i,t}$ is the set of observed and unobserved (from the analyst's perspective) state variables that affect payoffs and that determine its expected evolution over time. For any t lower than the last period (T_i) of the mortgage, the problem of this individual can be described recursively as follows:

$$\tilde{V}(\tilde{S}_{i,t}) = \max_{\{\text{continue, default}\}} \left\{ \tilde{u}(S_{i,t}, \varepsilon_{i,t}) + E[\beta \tilde{V}(\tilde{S}_{i,t+1}) | \tilde{S}_{i,t}], W(\tilde{S}_{i,t}) \right\}, \tag{1}$$

such that at the terminal period the continuation payoff is a constant $\tilde{V}(\tilde{S}_{i,T_i}) = K_{i,t}$.

The specification of the optimal default problem in (1) highlights the importance of expectations in determining default decisions. The reason is that making mortgage payments is equivalent to purchasing an option to default in the future and the value of the option depends on the expected evolution of the relevant state variables. This is why debtors may choose not to default even if they have negative equity.

We are interested in inferring the relationship between the state variables $\tilde{S}_{i,t}$ and the observed behavior from individual-level data. The estimated model can then be used to simulate and evaluate counterfactual equilibria, exogenous policies and the impact of exogenous shocks. We are interested in a dynamic structural model like the one in equation (1) because it allows the evaluation of policies and shocks that cannot be evaluated with "reduced form" methods. In particular, we can evaluate policies that affect the expected evolution of the states but that do not affect the current values. For example, the introduction of adjustable rate mortgages introduced a dynamic feature into mortgage contracts that by definition cannot be accounted for by reduced form models, specially when these policies have not been observed in the past.

The specifics of the implementation of the model with real data are left for the application section below. For now, notice that the problem in (1) corresponds to an optimal stopping problem with an absorbing state. The main challenge associated with the identification and estimation of such models is accounting for a rich correlation over time and across individuals of the unobserved states contained in $\tilde{S}_{i,t}$. In the context of our mortgage default application it is important that we account for the potential correlation of the unobserved aggregate shocks that affect everyone's decisions because, as documented in earlier work by Carranza and Estrada (2007), most the variation of default over time in Colombia cannot be explained by micro-level factors.

Even if one accounts for the presence of aggregate shocks, for example by using time dummies, ignoring the potential serial correlation of the aggregate shocks might lead to estimation bias. For example, if individuals expect the unobserved benefits of defaulting to increase over time, they might choose to delay default even if current payoffs are negative.

A researcher that ignores such unobserved correlation would then overestimate the current payoffs.

In the next section, we discuss identification of structural dynamic models with serially correlated unobserved shocks and present a general method for their estimation. We show that the aggregate shocks are identified in micro-level data and can be estimated using a simple variation of the standard methods. We then estimate a dynamic model of optimal default with our Colombian data set using the arguments we present below.

2.1 A generic optimal stopping problem

Consider the standard optimal stopping problem of an individual i at time $t \leq T_i$, who has to choose action $j \in \{0, 1\}$ where $j = 0$ is an absorbing state over a finite horizon T_i which may be different across individuals. Each choice generates a static a payoff $\tilde{u}_{i,j,t} \equiv u(X_{i,j,t}) + \varepsilon_{i,j,t}$ with an observed component $u(X_{i,j,t})$ that depends on a vector of observable (to the econometrician) states $X_{i,j,t}$. It also depends on an additive unobserved state variable $\varepsilon_{i,j,t}$ that is correlated across individuals and time periods.

At time t, the problem of the individual is to maximize the flow of payoffs from $\tau = t, ..., T_i$:

$$max_{\{d_{i,t},..d_{i,T_i}\}} E_t \sum_{\tau=t}^{T_i} \beta^{\tau-t} \tilde{u}_{i,d_\tau,\tau}, \tag{2}$$

where $d_i = \{d_t, ...d_{T_i}\}$ is a sequence of feasible decisions such that once $d_{i,\tau} = 0$ is chosen, no other alternative can be chosen.

Normalize the payoff generated by the action $j = 0$ to zero and relabel $u_{i,1,t} \equiv u_{i,t}$. Let $\tilde{S}_{i,t} \equiv \{X_{i,t}, \varepsilon_{i,0,t}, \varepsilon_{i,1,t}\}$ be the set of relevant state variables for individual i at time t. The vector of observed states $X_{i,t}$ is assumed to follow a first order Markov process independently of the unobserved states and so it can be recovered directly from the data. The unobserved states $\{\varepsilon_{i,0,t}, \varepsilon_{i,1,t}\}$ are also assumed to be Markovian as described below.

We can use the Bellman representation to write recursively the problem for individual i who, as of time $t - 1 < T - 1$, has not chosen $j = 0$ as:

$$\tilde{V}_t(\tilde{S}_{i,t}) = max\{u(X_{i,t}) + \varepsilon_{i,1,t} - \varepsilon_{i,0,t} + \beta E_t \left[\tilde{V}_{t+1}(\tilde{S}_{i,j,t+1}) | \tilde{S}_{i,j,t}\right], 0\}, \tag{3}$$

where β is a known exogenous discount rate. At $t = T_i$ the continuation payoff of the problem is zero, so that:

$$E_{T_i}[\tilde{V}_{T_i+1}(\tilde{S}_{i,j,T_i+1}) | \tilde{S}_{i,j,T_i}] = 0. \tag{4}$$

It has been shown before that the model above is generically not identified non-parametrically [2]. Therefore, the mapping of the model above into data is based partly on parametric assumptions on the distribution of the unobserved states ε. In order to allow for a rich pattern of unobserved correlation, we decompose the unobserved states as follows:

$$\varepsilon_{i,1,t} - \varepsilon_{i,0,t} \equiv \xi_t + \mu_i + \epsilon_{i,t}, \tag{5}$$

where $\epsilon_{i,t}$ is an *iid* idiosyncratic disturbance, which we assume is distributed logit, a standard and convenient assumption. The term μ_i is an individual-specific unobservable state that stays constant over time and is distributed among the population of individuals according to a distribution $\Phi(\mu_i)$. The term ξ_t is a common aggregate unobserved shock that follows a first order Markov process. The individual heterogeneity distribution $\Phi()$ and the distribution (i.e. the transition of) ξ have to be estimated simultaneously with the whole model.

Notice that, under this specification, individual choices are correlated over time and across debtors even after conditioning on the observed states; in addition, this unobserved heterogeneity can be allowed to depend on $X_{i,t}$ which would be equivalent to a model with heterogenous coefficients. The model is similar to the standard dynamic discrete choice models except for the presence of the shock $\xi_t \neq 0$ which is allowed to be correlated over time. The importance of including this form of heterogeneity is that it permits individual choices to be correlated (in unobservable ways) in a given cross section (since all individuals face the same shock) and for this correlation to persist over time. We will refer to these shocks as aggregate shocks, but they more generally can be understood as the common component of the unobserved heterogeneity.

The model we specify nests the standard models in the literature. Specifically, if we set $\mu_i = \xi_t = 0$, all the unobserved heterogeneity in the model is *iid* and the model is similar to the models in Rust (1987), Wolpin (1987), Hotz and Miller (1993) and Pakes (1986). If we assume away the aggregate shocks so that $\xi_t = 0$, but account for a correlated individual shock $\mu_i \neq 0$ the model is similar to Keane and Wolpin (1994).

In contrast to the models by Altug and Miller (1998) and Lee and Wolpin (2006) we don't need to specify where the aggregate shocks stem from. In Section 2.3 we show that micro data alone is enough to identify the aggregate shocks and their transition separately. In a general equilibrium setup, the specification of a model for the determination of the aggregate shocks

[2] Rust (1994); see also Taber (2000) and Heckman and Navarro (2007) for conditions under which these models are semiparametrically identified

ξ and their transition would be necessary for the computation of counterfactual equilibria, but not for the estimation of the model.

Let $S_{i,t} \equiv \{X_{i,t}, \mu_i, \xi_t\}$ be the the set of state variables, excluding the idiosyncratic *iid* error. Define the expected value function as the expectation of the value function in (3) with respect to the idiosyncratic *iid* shock, conditional on the current states:

$$\begin{aligned} V_t(S_{i,t}) &= E_\epsilon \left(\tilde{V}_t(S_{i,t}, \epsilon_{i,t}) | S_{i,t} \right) \\ &= \ln \left(1 + e^{u(X_{i,t}) + \xi_t + \mu_i + \beta E_t[V_{1,t+1}(S_{i,t+1})|S_{i,t}]} \right), \end{aligned} \tag{6}$$

where the second equality is the standard "social surplus" equation which follows from the logit assumption.

For convenience, write the expectation of (6) as a function of the conditioning states as follows:

$$E_t[V(S_{i,t+1})|S_{i,t}] \equiv \Psi(S_{i,t}), \tag{7}$$

where the expectation is taken with respect to the dynamic states given their realization and their transition probabilities. For given state variables and transition probabilities, this value can be computed using standard numerical techniques starting at the terminal period.

Conditional on survival, the predicted probability that individual i chooses $j = 1$ at time t is given by:

$$\begin{aligned} Pr_{i,1,t} &= \Pr \left[u(X_{i,t}) + \mu_i + \xi_t + \epsilon_{i,t} + \beta E_t \left[V_{t+1}(S_{i,t+1}) | S_{i,t} \right] > 0 \right] \\ &= \frac{e^{u(X_{i,t}) + \xi_t + \mu_i + \beta \Psi(S_{i,t})}}{1 + e^{u(X_{i,t}) + \xi_t + \mu_i + \beta \Psi(S_{i,t})}}, \end{aligned} \tag{8}$$

where the continuation payoffs correspond to the expectation of (7). Notice that this probability depends on both the realization of the unobserved individual heterogeneity μ_i and the aggregate shock ξ_t.

Next, we define the probability of any given sequence of choices which we will use below. Given (8), let $\tilde{Pr}_i$ denote the probability of an individual history which can be computed as the product of probabilities over the given sequence of choices, conditional on the realization of the individual heterogeneity and the aggregate shocks:

$$\tilde{Pr}_i = \prod_{t=1}^{\tilde{T}_i} \int Pr_{i,1,t}^{d_{i,t}} (1 - Pr_{i,1,t})^{(1-d_{i,t})} d\Phi\left(\mu\right), \tag{9}$$

where $\bar{T}_i$ is the last time period at which the loan is observed to be outstanding either because it is defaulted on or because it reaches its maturity, i.e. either the time when individual i first chooses $j = 0$ or the final period T_i if it always chooses $j = 1$.

2.2 Estimation

Consider estimating the model above using a random sample of $i = 1, ..., N$ individuals who are observed solving the described optimal stopping problem during a sequence of $\bar{T} = max\{\bar{T}_1, ..., \bar{T}_N\}$ time periods. For simplicity we assume that all aspects of the model are parametric[3]. In Section 2.3 we discuss which aspects of the model can potentially be recovered nonparametrically.

Let γ denote the parameters of the utility function, ρ_ξ the parameters of the transition of the aggregate shocks and σ the parameters of the distribution of the individual unobserved heterogeneity (μ). For notational convenience, assume that all individuals start to solve the problem simultaneously but then have potentially different problem horizons T_i. For each individual, a matrix of potentially time-varying exogenous state variables $X_i = \{X^0_{i,1}, ..., X^0_{i,\bar{T}_i}\}$ is observed, as well as a sequence of decisions $d^0_i = \{d^0_{i,1}, ..., d_{i,\bar{T}_i}\}$.

Given the observed states, their transition probabilities can be estimated directly from the data before estimating the whole model if they are exogenous. The remaining parameters, the transition of the aggregate shocks and the shocks themselves $\xi = \{\xi_1, ..., \xi_{\bar{T}}\}$ have to be estimated jointly. The sample likelihood is given by: w

$$
\begin{aligned}
\ell(\theta, \xi) &= \prod_{i=1}^{N} \int \left[\prod_{\tau=1}^{\bar{T}_i} Pr_{i,1,\tau}^{d^0_{i,\tau}} (1 - Pr_{i,1,\tau})^{1-d^0_{i,\tau}} \right] d\Phi\left(\mu; \sigma\right) \\
&= \prod_{i=1}^{N} \int \tilde{Pr}_i(\gamma, \rho_\xi, \xi) d\Phi(\mu; \sigma), \qquad (10)
\end{aligned}
$$

where the choice probabilities are integrated with respect to the initial distribution of μ, Φ.

The model is estimated efficiently by maximizing the likelihood function over the parameter space. Notice that the estimation of the model we present is, in principle, identical to the estimation of standard dynamic models with unobserved heterogeneity. The key difference lies on the presence of the aggregate shocks ξ and their transition ρ_ξ. Depending on the

[3]In what follows, we emphasize the dependence of the probabilities on the parameters when helpful, but mostly we keep the dependence implicit for tractability of the notation.

case, maximizing (10) can be difficult, specially if the number of periods $\bar{T}$ is large, because each shock ξ_t has to be estimated for all t.

We show now how the estimation of these aggregate shocks can be concentrated out from the wider estimation algorithm by using aggregate information not commonly used in the estimation of dynamic discrete choice models. In other words, we show that the estimation of the model is identical to the estimation of a standard model with the addition of a restriction that arises from the likelihood itself that identifies the aggregate shocks. Specifically, take the derivative of (10) with respect to each ξ_t and set it equal to zero to obtain the following condition:

$$\frac{N_{d_{i,t}}}{N_t} \equiv \bar{s}_{1,t} = \left[\frac{1}{N_t}\sum_{i=1}^{N}\int Pr_{i,1,t}\frac{\tilde{Pr}_i}{\int \tilde{Pr}_i d\Phi(\mu)}d\Phi(\mu)\right] + \left[\frac{1}{N_t}\sum_{i=1}^{N}\int \beta\frac{\partial\Psi_{i,t}(S_{i,t})}{\partial\xi_t}(Pr_{i,t}(S_{i,t}) - d_{i,t})\frac{\tilde{Pr}_i}{\int \tilde{Pr}_i d\Phi(\mu)}d\Phi(\mu)\right], \tag{11}$$

where $N_{d_{i,t}}$ is the number of individuals in the sample who choose action $j = 1$ at time t.

The first term on the right hand side of (11) is the expected aggregate choice probability conditional on the observed individual histories. The second term is the sample covariance of the prediction error and the derivative of the expected continuation payoff with respect to the aggregate shock, again conditional on the observed histories. Notice that this condition *is not* the often used restriction matching the predicted and the observed aggregate choice probabilites *exactly*. This implies that an efficient estimation of the model *won't perfectly match* the predicted and the observed aggregate choice probabilities.

Equation (11) generates a set of $\bar{T}$ non-linear equations, which can be used to concentrate out the estimation of ξ from the problem of estimating θ. In other words, for any set of parameters θ_0, we can solve for the parameters ξ_0 that satisfy (11) as we look numerically for the estimator θ^* and its associated ξ^*.

Notice that, in general, (11) reduces to a set of intuitive average probabilities. Since the predicted choice probability and the expected continuation payoffs are conditioned on the same set $S_{i,t}$ of state variables, the covariance of the second term should converge to zero since this covariance is zero in the population. It follows then that, when N_t is large, the expression above can be approximated by the following expression:

$$\frac{N_{d_{i,t}=1}}{N_t} \equiv \bar{s}_{1,t} \approx \left[\frac{1}{N_t}\sum_{i=1}^{N}\int Pr_{i,1,t}\frac{\tilde{Pr}_i}{\int \tilde{Pr}_i d\Phi(\mu)}d\Phi(\mu)\right], \tag{12}$$

which might be an easier expression to use when concentrating out the estimation of ξ.

If we compute (11) in the population, we obtain a condition that we state as Lemma 1. This lemma can also be used to concentrate out the estimation of ξ when the population shares are observed and the second term on the RHS of equation (11) is zero. Denote the empirical distribution of the observed states as $F_t(x)$, which is (by assumption) independent of the distribution Φ of unobserved states. Let also $\mathrm{s}_{1,t}$ be the share of choice $j = 1$ at each time t among active agents.

Lemma 1 *Consider the estimation of the model described by the choice probabilities (8) and (9). At the true value of θ and ξ the following condition holds:*

$$\mathrm{s}_{1,t} = \int Pr_{i,1,t}(\theta,\xi)\frac{\tilde{P}r_i(\theta,\xi)}{\int \tilde{P}r_i(\theta,\xi)d\Phi(\mu)}d\Phi(\mu)dF_t(x) \equiv \tilde{s}_{1,t}(\theta,\xi). \tag{13}$$

This lemma states that, at the true value of the parameters, the observed aggregate choice probability has to be equal to a *weighted average* of the predicted choice probabilities. The weighting is equivalent to conditioning the predicted choice probabilities on the observed choice history of each individual up until the terminal period $\bar{T}_i$.

As a corollary of this lemma, we point out below that if there is no persistent unobserved heterogeneity the condition (13) reduces to a simple average. This condition is similar to the standard BLP-style market-level condition that is used to concentrate out the estimation of choice-specific shocks from the estimation of discrete demand systems, except that it only holds when the second term on the RHS of equation (11) is zero. The proof follows trivially from Lemma 1, by noting that when there is no persistent unobserved heterogeneity, the integrals in the expressions above vanish.

Corollary 1 *Consider the estimation of the model described by the choice probabilities (8) and (9). Let $\mu_i = \mu\ \forall i$ so that the distribution Φ is degenerate. At the true value of θ and ξ the following condition holds:*

$$\mathrm{s}_{1,t} = \int Pr_{i,1,t}(\theta,\xi)dF_t(x). \tag{14}$$

An interesting feature of (11) and (12) is that the average choice probabilities at any period t are not conditioned on the survival until $t-1$ but on the whole history until $\bar{T}_i$. This property is not a consequence of the dynamic structure of the problem, but of the presence of unobserved correlated shocks. In fact, this condition extends to static models (as in ?) in the sense that, whenever there are unobserved correlated shocks, efficient estimation with a finite sample would require that the observed aggregate choice behavior matches the predicted behavior, *conditional* on the observed choices. That is, when concentrating out the aggregate shocks under the presence of individual unobserved heterogenity, one should not exactly match the observed aggregate choice behavior to the simple predicted choice probability but rather to a weighted version of these probabilities.

When the population shares $\mathbf{s}_{1,t}$ are known exactly, so that the data set is a combination of micro-level and market-level information, Lemma 1 can be used to "concentrate out" the estimation of the aggregate shocks ξ from the estimation algorithm using the aggregate choice probabilities. Specifically, at each time t and for given parameters θ_0 and ξ_0, the model generates a vector of aggregate predicted choice probability $\tilde{s}_{1,t}(\theta_0, \xi_0)$. If the model is correctly specified and the sample is large (13) must hold:

$$\mathbf{s}_{1,t} = \tilde{s}_{1,t}(\theta, \xi) \; \forall t. \tag{15}$$

Given any value of θ^0, the expression in (15) generates a system of $\bar{T}$ non-linear equations, so that a unique value of $\xi(\theta^0)$ can be solved for directly. If the population shares $\mathbf{s}_{1,t}$ are not observed, but only the shares $\bar{s}_{1,t}$ in the sample, then (11) or (12) can be used instead.

The feasible computation of the model requires that for any set of feasible parameters θ_0, the vector ξ_0 that solves (11) be always defined. Moreover, the identification of the model will require that the vector ξ_0 be unique, at least around the true vector ξ^*. The following lemma establishes sufficient conditions under which the solution to (15) exists and is unique. The proof of this lemma, shown in the appendix, relies on the monotonicity of the average predicted default rates (13) on the aggregate shock.

Lemma 2 *Let $E_t[\xi_{t+1}|\xi_t] = h(\xi_t)$, such that $h(.)$ is strictly monotone and $-1 < h'(\xi_t) < 1$. Then, for the system of T equations implied by $\bar{s}_{1,t} = s_{1,t}(\theta^0, \xi)$ for $t = 1, ..., T$ has a unique solution $\xi(\theta^0)$, if the sample size N is large (so the second term on the RHS of equation* (11) *is zero).*

The sufficient conditions for the lemma to be true are very weak in the sense that they are far from necessary. Moreover, they imply restrictions that are usually natural in empirical environments. For example, if the aggregate shocks follow a linear autoregressive process, a sufficient condition for the lemma and the corollary to hold is that the process be stationary.

Lemmas 1 and 2 will be used to show our identification result below. For practical purposes, they imply that the model can be estimated using standard techniques. One can do estimation with the addition of (15) as a separate restriction, thereby reducing the computational dimension of the estimation algorithm if required. In other words, it is not strictly necessary to maximize the likelihood over all the parameters of the model, which is useful specially when the number of periods is large. Specifically, the model can be estimated maximizing the likelihood (10) over the parameters θ, solving numerically for ξ from (15) along the estimation algorithm:

$$max_{\theta}\ell(\theta,\xi(\theta)) \tag{16}$$

Before presenting an application of our methodology, in the following sections we discuss the identification of the components of the model and the applicability of the methodological framework to more general environments.

2.3 Identification of the model

We discuss now the identification of the model described above and show the conditions under which such identification is possible. The main problem lies in the separate identification of the aggregate shocks and their transition, which we show is possible only when micro level information is available. Importantly, the identification conditions that allow us to separate the transition from the value of the aggregate shocks are sufficient and necessary.

The choice probabilities in (8) are similar to the choice probabilities in standard empirical dynamic models with unobserved heterogeneity, except for the presence of the aggregate shocks ξ and their transition probabilities. Therefore, the identification of the utility function and the of the distribution of μ is based on similar arguments as in the standard literature. We provide a brief discussion of their identification and then discuss in detail the identification of the aggregate shocks ξ and their transition probabilities.

As pointed out by Taber (2000) and Heckman and Navarro (2007), the finite horizon of the problem facilitates the nonparametric identification of the dynamic discrete choice models. We briefly describe how their argument works. Notice that since at T_i the continuation payoffs of the problem are zero, the probability that individual i chooses $j = 1$, obtained from (8), doesn't contain a continuation value and therefore does not include the transition of the aggregate shocks:

$$Pr_{i,1,T_i} = \Pr\left(u(X_{i,T_i}) + \xi_{T_i} + (\mu_i + \epsilon_{i,T_i})\right). \tag{17}$$

Notice that in this terminal period ξ_{T_i} is simply the constant in the model. In limit sets where one can control for the dynamic selection (survival up to T_i) one can use standard arguments, i.e., Matzkin (1992), to identify nonparametrically the utility function $u\left(\right)$, the constant ξ_{T_i} and the nonparametric distribution of $(\mu_i + \epsilon_{i,T_i})$. Once this distribution has been identified at different periods (since T_i represents different periods for different individuals) one can use deconvolution arguments (Kotlarski (1967)) to recover the distribution of μ_i from the repeated observations of the marginal distribtuion of $(\mu_i + \epsilon_{i,T_i})$ over time.

The novel part of this paper is the separate identification of the aggregate shocks and their transition. Intuitively, the identification of the aggregate shocks comes from the variation in the data on the aggregate behavior, a feature which is not fully exploited in the standard literature. Notice that, in practice, our estimation approach is equivalent to a standard estimation of a Markovian decision model, with the "addition" of the "aggregate" restriction (15), which directly identifies the aggregate shocks.

The separate identification of the levels ξ of the aggregate shocks and their transition probabilities has to be explained in detail. >From inspecting (8) it can be seen that both the aggregate shocks ξ and their transition probabilities enter the continuation payoffs. Moreover, ξ enters additively the instant payoffs, so that it can potentially happen that changes in ξ that are offset by changes in their expected serial correlation generate identical predictions, so that they would not be separately identified.

We have two sources for the separate identification of the two set of unobservables. On one hand, notice from (17) that as we go over groups if individuals with different terminal periods $\{T_1, ..., T_N\}$ the transition probabilities for the aggregate shocks don't enter the choice probabilities and therefore the aggregate shocks are identified up to the constant of the utility function. Therefore, if we observe individuals who face their terminal period at each time period of our sample, ξ will be identified. Since we can identify ξ for different periods we can, in principle, recover their transition probabilities, $f\left(\xi_t|\xi_{t-1}\right)$ nonparameterically in the domain of the recovered ξ.

The second, and more general source of identification, comes from of the choice probabilities themselves. ξ and ρ_ξ (the parameters of the transition probabilities) will be separately identified even in a sample of individuals who all face the same terminal period. To see this, notice that at the true value ρ_ξ^* of the transition parameters, our estimation algorithm looks for the unique vector ξ^* that satisfies (15) which we can rewrite as follows:

$$\mathbf{s}_{1,t} = \int Pr_{i,1,t}(.;\xi^*,\rho_\xi^*)\frac{\tilde{P}r_i(.;\xi^*,\rho_\xi^*)}{\int \tilde{P}r_i(.;\xi^*,\rho_\xi^*)d\Phi(\mu)}d\Phi(\mu)\,dF_t(x)$$

$$= \int \tilde{P}r_{i,t}(.;\xi^*,\rho_\xi^*)dF_t(x) \tag{18}$$

where $\mathbf{s}_{1,t}$ is the observed proportion of individuals who choose $j=1$ at time t and where $\tilde{P}r_{i,t}$ is the choice probability integrated over the distribution of individual heterogeneity, conditional on each choice history:

$$\tilde{P}r_{i,t}(.;\xi^*,\rho_\xi^*) = \int Pr_{i,1,t}(.;\xi^*,\rho_\xi^*)\frac{\tilde{P}r_i(.;\xi^*,\rho_\xi^*)}{\int \tilde{P}r_i(.;\xi^*,\rho_\xi^*)d\Phi(\mu)}d\Phi(\mu)$$

The key thing to notice is that, as we change ρ_ξ, the algorithm will find new vectors of ξ consistent with (18). The implicit function theorem implies that the variation of ξ as ρ_ξ changes is given by:

$$\frac{\partial\xi_t}{\partial\rho_\xi} = -\frac{\int(\partial\tilde{P}r_{i,1,t}/\partial\rho_\xi)dF_t(x)}{\int(\partial\tilde{P}r_{i,1,t}/\partial\xi)dF_t(x)} \tag{19}$$

If such variation in ξ leads to the same choice probabilities as in (19), then the two sets of parameters are not separately identified. Notice, though, that at any given ρ_ξ and for every agent i, the implicit variation of ξ as ρ_ξ changes such that $\tilde{P}r_{i,1,t}$ is constant is given by:

$$\frac{\partial\xi_t}{\partial\rho_\xi} = -\frac{(\partial\tilde{P}r_{i,1,t}/\partial\rho_\xi)}{(\partial\tilde{P}r_{i,1,t}/\partial\xi)} \tag{20}$$

which is in general different than (19), as long as the predicted choice probabilities vary across individuals. Consequently if this is the case, the predicted choice probabilities will change as the transition parameters change.

In other words, if there is variation in the observed states across individuals, the derivative of the individual choice probabilities with respect to the ρ_ξ is different from zero. Therefore, the sample likelihood will necessarily fall around the estimated parameter ρ_ξ^* so that ξ and ρ_ξ are separately identified as formally established in the following proposition, which we prove in the appendix. Put it differently, if there is no individual variation on the predicted choice probabilities then equations (19) and (20) will be the same.

Proposition 1 *Consider the model with sample likelihood $\ell(\gamma^0,\sigma^0,\rho_\xi)$ given by (10) with known parameters γ^0 and σ^0. Assume that the conditions in Lemmas 1 and 2 hold. The parameter vector ρ_ξ is identified if and only if the states $X_{i,t}$ vary across individuals for at least one individual i for all t.*

The proposition establishes the identification of ρ_ξ, conditional on the utility function and the distribution of individual unobserved heterogenity, whose identification was explained before. Moreover, the identification of ρ_ξ is formally independent from the identification of ξ. That is, if we were to estimate ρ_ξ using the estimated ξ (for example, by taking the estimated ξ and running a regression of ξ_t against ξ_{t-1}) we might find substantial discrepancies with the estimated ρ_ξ obtained from the estimation above, specially in short samples.

This implies that additional restrictions can be added to (16) to guarantee the consistency of both (the transition implied by the estimated ξ and the one estiamted from the choice probabilities), which might be desirable in long panels. More importantly, however, it also implies that the choice probabilities contain enough information to distinguish the individuals perceptions about how the aggregate shocks transitions from the actual transition implied by the realized ξ.

The identification of the parametric model is not surprising. The more important result is the *nonidentification* of the model when no micro level data (i.e., with no individual variation in the choice probabilities) is available. There is a growing literature on the estimation of structural dynamic models of demand using market-level data (e.g. Carranza (2007) and Gowrisankaran and Rysman (2006)). Our result highlights the limits of the identification for this general class of models.

2.4 Further remarks on the methodology

For illustrative purposes, we have described our methodological framework using a simple binomial optimal stopping problem. The general approach extends naturally to more general dynamic Markov decision problems with multiple repeated choices.

For example, if instead of an absorbing state, we let individuals choose $j = 0$ repeatedly, the only difference is that a continuation payoff has to be computed for both $j = 0$ and $j = 1$. This adds to the computational burden of the algorithm, but the fact that we would observe the same individuals making the same choices repeatedly over time would also strengthen the identification of the individual-level unobserved heterogeneity.

In addition, we can allow for multiple choices each with its associated continuation payoff. The computation of multiple continuation payoffs along the estimation algorithm is feasible but computationally costly. In addition, the data requirements are stronger, as the identification of the aggregate shocks relies on the computation of choice-specific aggregate