Structural models involving highly dimensional fixed point problems: An asymptotically efficient two-stage estimator

Victor Aguirregabiria
University of Chicago

Pedro Mira
CEMFI, Madrid

January 2000
PRELIMINARY AND INCOMPLETE VERSION

Abstract

In this paper we consider a class of structural econometric models in which the distribution of the endogenous variables is implicitly defined as the solution of a fixed-point problem. We propose a simple two-stage estimator which does not require the econometrician to solve the fixed point problem, and provide sufficient conditions for its consistency and asymptotic efficiency. We show that these sufficient conditions hold if a Newton operator is used to solve the fixed point problem. Finally, we study the finite sample performance of this estimator in the context of several structural models: a dynamic discrete choice model, a static game with incomplete information, a model of irreversible investment, and a model of oligopolistic competition in a differentiated product market.

1 Introduction

The concept of equilibrium in economics is intimately related to the mathematical concept of fixed point. Many interesting economic models involve fixed point problems in highly dimensional spaces. Bellman equations characterizing the value function of a dynamic decision model, or best response functions defining the equilibrium of a game are two examples. During the last two decades techniques for the estimation of structural econometric versions of these models have been developed.[1]

[1]Recent literature on this topic is very extensive. In the context of dynamic discrete choice structural models see the seminal papers by Wolpin (1984), Miller (1984), Pakes (1986), Rust (1987), Hotz and Miller (1993), Keane and Wolpin (1996), and the survey by Rust (1994). For static game theoretic models see Bresnahan and Reiss (1991), Berry (1994), Berry, Levinshon and Pakes (1996), and the survey by Pakes (1994). For the estimation of general equilibrium models using micro data see Heckman, Lochner and Taber (1999) and the recent survey by Browning, Hansen and Heckman

Although methods which make full use of the restrictions embodied in the theoretical model, such as full maximum likelihood estimation, are desirable for efficiency and, sometimes, identification reasons, these methods require the use of computationally burdensome nested solution-estimation algorithms which are often not feasible in all but the simplest models.

In this paper we consider a class of structural econometric models in which the distribution of the endogenous variables can be defined as the solution of a fixed-point problem. We propose a simple two-stage estimator which avoids repeated solution of the fixed point problem. The first stage involves non-parametric estimation of the distribution of endogenous variables. In the second stage a pseudo-likelihood function is maximized. Given parameter values, the pseudo-likelihood function is computed by a single iteration on the fixed point operator, using the first stage non-parametric estimates as starting values. Our two-stage estimator is the value of the structural parameters that maximizes this pseudo-likelihood function. We state regularity conditions for consistency of this estimator and provide additional sufficient conditions under which the two-stage estimator can be used instead of a maximum likelihood estimator with no loss of asymptotic efficiency. We also show that these sufficient conditions will be satisfied if a Newton operator is used to solve the fixed point problem. Finally, we study the performance of this estimator in finite samples for several specific examples: a dynamic programming discrete choice model, a static game with incomplete information, a model of irreversible investment, and a model of oligopolistic competition in a differentiated product market.

The idea behind our estimator builds on previous work by Hotz and Miller (1993), Manski (1993) and Aguirregabiria and Mira (1999) in the context of dynamic discrete choice models. This paper extends this previous literature in several directions. First, we consider a general class of structural models where the distribution of the endogenous variables is implicitly defined as a fixed point. Second, we show that the two-stage estimator is asymptotically efficient under conditions that seem quite general.

The rest of the paper is organized as follows. Section 2 presents the general econometric model and illustrate it with four examples. In section 3 we define the two-stage estimator and present our main results of consistency and asymptotic efficiency. In Section 4 [NOT INCLUDED IN THIS VERSION] we present our Monte Carlo experiments. The method used to obtain the second stage estimator can be extended to define *K-stage estimators* (i.e., $K > 2$) which, although asymptotically equivalent, may have better finite sample properties. Our experiments will compare the finite sample performance performance of ML, 2 and K-stage estimators in the different example models.

(1999). Finally, Ericson and Pakes (1995), Pakes and McGuire (1995), and Pakes, Gowrisankaran and McGuire (1995) present a very interesting and promising framework for empirical analysis of Markov-Perfect equilibria in dynamic games.

2 Econometric model

Let $y \in Y \subseteq N^L$ and $x \in X \subseteq N^M$ be two vectors of discrete random variables, where N is the space of the natural numbers. Let $p^*(y_0, x_0)$ be the *true probability* (i.e., probability in the population) of $y = y_0$ conditional on $x = x_0$, and define $\tilde{p}^*$ as the vector with all the true probabilities, i.e., $\tilde{p}^* \equiv \{p^*(y_0, x_0) : y_0 \in Y; x_0 \in X\}$. A parametric model for $\tilde{p}^*$ is a family of probability distributions

$$\{\tilde{p}(\theta) : \theta \in \Theta\}, \tag{1}$$

where $\tilde{p}(\theta) = \{p(y_0, x_0; \theta) : y_0 \in Y; x_0 \in X\}$ and θ is a finite vector of parameters. We are interested in models with the following properties.

Property 1: For any $\theta \in \Theta$, $\tilde{p}(\theta)$ is the unique fixed point in π of the mapping:

$$\pi = \tilde{\Psi}(\theta; \pi) \tag{2}$$

where $\tilde{\Psi}(\theta; \pi) \equiv \{\Psi(y_0, x_0; \theta; \pi) : y_0 \in Y; x_0 \in X\}$, and $\Psi(.)$ is twice continuously differentiable in (θ, π).

Property 2: For any pair (θ, π), $\tilde{\Psi}(\theta; \pi)$ is a probability distribution for y conditional on x.

Property 3: There is a unique vector in Θ, say θ^*, that maximizes in θ the function

$$Q^*(\theta) \equiv \sum_{x_0 \in X} p^*(x_0) \sum_{y_0 \in Y} p^*(y_0, x_0) \;\ln p(y_0, x_0; \theta)$$

where $\{p^*(x_0)\}$ is the true marginal distribution of x.

Given a random sample of y and x, $\{x_i, y_i : i = 1, 2, ..., n\}$, we are interested in the estimation of θ^*. The conditional log-likelihood function is:

$$l(\theta) \;=\; \sum_{i=1}^{n} \ln p(y_i, x_i; \theta) \tag{3}$$

where the probabilities $p(y_i, x_i; \theta)$ satisfy $\tilde{p}(\theta) = \tilde{\Psi}(\theta; \tilde{p}[\theta])$.

2.1 Example 1: Discrete choice dynamic programming model

Consider a dynamic programming model where $y \in Y = \{1, 2, ..., J\}$ is the discrete decision variable, and $s \in S$ is the vector of state variables (see Rust 1994a, 1994b). Time is discrete and indexed by t. Utility is time separable, the discount factor is β, and $U(y_t, s_t)$ represents the one-period utility function. The time horizon of the decision problem is infinite. An agent's beliefs about uncertain future states can be represented by a Markov transition probability $f_s(ds_{t+1}|s_t, y_t)$. According to these assumptions, the value function and the optimal decision rule are time invariant (Blackwell's theorem). Let $V(y, s)$ be the value function conditional on the (hypothetical) choice of alternative y. These conditional choice value functions are implicitly defined by the Bellman equations

$$V(y,s) = U(y,s) + \beta \int \max_{j \in Y} \{V(j,s')\} \; f_s(ds'|s,y) \qquad for \; y = 1,2,...,J \tag{4}$$

We partition the state vector as follows: $s = (x, \varepsilon)$, where x includes the state variables which the econometrician observes and ε those that are unobservable. We assume *Conditional Independence;* that is, the transition probability of the state variables factors as:

$$f_s(x_{t+1}, \varepsilon_{t+1}| \; x_t, \varepsilon_t, y_t) = f_\varepsilon(\varepsilon_{t+1}| \; x_{t+1}) \; f_x(x_{t+1}| \; x_t, y_t) \tag{5}$$

where $f_\varepsilon(.)$ has finite first moments and is continuous and twice differentiable in ε. Let $EU(y, x)$ be the expected (one-period) utility conditional on x and on the hypothetical choice of alternative y, and define $\mu(y, x, \varepsilon) \equiv U(y, s) - EU(y, x)$. Similarly, let $EV(y, x)$ be the expected value function conditional on x. Using these definitions and the conditional independence assumption, it is simple to verify that:

$$V(y,s) = EV(y,x) + \mu(y,x,\varepsilon)$$

Notice that, by construction, $\tilde{\mu}(x, \varepsilon) \equiv (\mu[1, x, \varepsilon], \mu[2, x, \varepsilon], ..., \mu[J, x, \varepsilon])'$ is mean independent of x.

Suppose the primitives of the model $\{U, f_x, f_\varepsilon, \beta\}$ are known up to a finite vector of parameters θ. Let $p(y_0, x_0; \theta)$ be the probability that alternative y_0 is the optimal choice given $x = x_0$ and given the vector of structural parameters θ. According to this model:

$$p(y_0, x_0; \theta) = \int I\left(y_0 = \arg\max_{j \in Y} \{EV(j,x;\theta) + \mu(j,x,\varepsilon)\}\right) f_\varepsilon(d\varepsilon|x;\theta) \tag{6}$$

where $I(.)$ is the indicator function. Given the conditional independence assumption, and the mean independence of $\tilde{\mu}(x, \varepsilon)$ and x, it is possible to obtain an expression for $EV(y, x; \theta)$ in terms of the primitives U f_x, f_ε and the set of choice probabilities $\tilde{p}(\theta)$ (see Hotz and Miller, 1993, and Aguirregabiria and Mira, 1999). We denote this function by $\sigma(y, x; \theta; \tilde{p}[\theta])$. This function $\sigma(.)$ can be evaluated at any arbitrary vector of choice probabilities π, "optimal" or not. In general, $\sigma(y, x; \theta; \pi)$ represents the expected value of choosing alternative y today if future decisions are based on the choice probabilities in π. If x is discrete, i.e., $x \in \{x^1, x^2, ..., x^M\}$, it is possible to show that $\sigma(j, x^m; \theta; \pi)$ is the $m - th$ component of the vector:

$$E\tilde{U}_x(j) + \beta F_x^j (I_M - \beta F_x^U[\pi])^{-1} \left\{ \sum_{k=1}^{J} \pi^k * \left[\tilde{U}(k) + e(k,\pi)\right] \right\} \tag{7}$$

where $E\tilde{U}_x(j)$ is the vector of one period utilities conditional on $y = j$ and all states; F_x^j is the matrix of transition probabilities of x conditional on $y = j$; π^j is the subvector of π associated with alternative j; $F_x^U(\pi)$ is the matrix of unconditional transition probabilities induced by π, i.e., $F_x^U(\pi) = \sum_{j=1}^{J} \pi^j * F_x^j$; and $e(j,\pi)$ is a vector of functions of π: the form of these functions depends on the distribution of ε (e.g., if $\tilde{\mu}(x,\varepsilon)$ has an extreme value distribution $e(j,\pi) = -\ln \pi^j$).[2]

Therefore, we can write the set of conditional choice probabilities $\tilde{p}(\theta)$ as the unique fixed point of the mapping $\pi = \tilde{\Psi}(\theta;\pi)$, where,[3]

$$\Psi(y,x;\theta;\pi) = \int I\left(y = \arg\max_{j\in Y}\{\sigma(j,x;\theta;\pi) + \mu(j,x,\varepsilon)\}\right) f_\varepsilon(d\varepsilon|x;\theta) \tag{8}$$

For instance, if $\tilde{\mu}(x,\varepsilon)$ has an extreme value distribution,

$$\Psi(y,x;\theta;\pi) = \frac{\exp\{\sigma(y,x;\theta;\pi)\}}{\sum_{j=1}^{J}\exp\{\sigma(j,x;\theta;\pi)\}.} \tag{9}$$

2.2 Example 2: Static game with incomplete information[4]

Consider an N-player game, and let $i \in \{1,2,...,N\}$ be the index that denotes a player. The payoff function for player i is $U_i(y_i, y_{-i}, x, \varepsilon_i)$, where y_i represents his own action and y_{-i} is the vector of actions of the other players. The set of choice alternatives is discrete, i.e., $y_i \in Y = \{1,2,...,J\}$. x is a vector of exogenous characteristics of all players and/or the environment which is known by all players, whereas ε_i represents characteristics of player i which are private information. Each player knows the other players' payoff functions up to the private information components; furthermore, every player has a subjective probability distribution over the $\varepsilon' s$ of the other players, $G_i(\varepsilon_{-i}; x, \varepsilon_i)$, where $\varepsilon_{-i} = \{\varepsilon_j : j \neq i\}$.

Notice that, without further assumptions, players are unable to calculate equilibria on their own because each player i is ignorant of the subjective beliefs of the others. Here we assume that it is known to all players that conditional on x the private information components $\varepsilon_i' s$ are independently and identically distributed over individuals with cumulative distribution function $F(\varepsilon; x)$; therefore, $G_i(\varepsilon_{-i}; x, \varepsilon_i) = \prod_{j\neq i} F(\varepsilon_j; x)$, which is the same for all players.

Let $\alpha_i(x,\varepsilon_i)$ be a strategy function for player i. Define $V_i(y_i, x, \varepsilon_i; \alpha_{-i},)$ as player i's expected payoff from choosing action y_i, conditional on x and ε_i and given that the other players have strategy functions $\alpha_{-i} = \{\alpha_j : j \neq i\}$.

$$V_i(y_i, x, \varepsilon_i; \alpha_{-i}) = \int U(y_i, \alpha_{-i}[x, \varepsilon_{-i}], x, \varepsilon_i)\ G(d\varepsilon_{-i}; x) \tag{10}$$

[2] More specifically, $e(j,\pi) = \{e(j, x^m, \pi) : m = 1,2,...,M\}$, where $e(j,x,\pi)$ is the expected value of $\mu(j,x,\varepsilon)$ conditional on x and on the choice of alternative j, if choices are made according to the choice probabilities in π.

[3] See Lemma 1 in Aguirregabiria and Mira (1999) for a proof of the existence and uniqueness of this fixed point.

[4] This example is based on Rust (1994b).

Player i chooses y_i in order to maximize his expected payoff . A *Bayesian Nash Equilibrium* in this game is a collection of strategies $\{y_i^*(x,\varepsilon_i) : i = 1,2,...,N\}$ such that for all i:

$$y_i^*(x,\varepsilon_i) = \arg\max_{j\in Y} \left\{V_i(j,x,\varepsilon_i;y_{-i}^*)\right\} \tag{11}$$

where $y_{-i}^* = \{y_j^*(x,\varepsilon_j) : j \neq i\}$.

Now suppose the econometrician observes x and actions $\{y_i\}$. Players are assumed to choose equilibrium actions and to have rational expectations; i.e., $\prod_j F(\varepsilon_j; x)$ is the true distribution of ε given x. The payoff functions and the probability distribution of ε are known up to a finite vector of parameters θ. Define the equilibrium probabilities $p(y,x;\theta)$ as:

$$p_i(y,x;\theta) = \Pr(y_i^*[x,\varepsilon_i] = y \mid x;\theta) = \int I\{y = \arg\max_{j\in Y}\ V_i(j,x,\varepsilon_i;y_{-i}^*)\}\ F(d\varepsilon_i;x) \tag{12}$$

Let $\tilde{p}(\theta)$ be the vector $\{p_i(y,x;\theta) : y \in Y; x \in X; i = 1,2,...,N\}$. Notice that these probabilities are the ones that a player uses to predict the behavior of the other players. Therefore, we can use these probabilities to obtain an alternative expression for the expected payoff functions $V_i(.)$ evaluated at the equilibrium strategies of the other players, y_{-i}^*. We denote these functions by $\delta_i(y_i,x,\varepsilon_i;\theta;\tilde{p}[\theta])$.

$$\delta_i(y_i,x,\varepsilon_i;\theta;\tilde{p}[\theta]) = \sum_{y_{-i}} \left(\prod_{m\neq i} p_m(y_m[y_{-i}],x;\theta)\right) U(y_i,y_{-i},x,\varepsilon_i) \tag{13}$$

where $y_m[y_{-i}]$ represents the choice of individual m in the vector y_{-i}. The functions $\delta_i(.)$ can be evaluated at any vector of choice probabilities, equilibrium or not. For an arbitrary vector of choice probabilities, $\delta_i(.)$ represents the expected payoff function under the hypothesis that the decision of the other players will be based on these probabilities. We can combine equations (12) and (13) to write the set of equilibrium choice probabilities $\tilde{p}(\theta)$ as a solution of the fixed point equation $\pi = \tilde{\Psi}(\theta;\pi)$, where $\tilde{\Psi}(\theta;\pi) \equiv \{\Psi_i(y,x,\theta;\pi) : y \in Y; x \in X; i = 1,2,...,N\}$, and

$$\Psi_i(y,x,\theta;\pi) = \int I\left(y = \arg\max_{j\in Y} \{\delta_i(j,x,\varepsilon_i;\theta;\pi)\}\right) F(d\varepsilon_i|x;\theta) \tag{14}$$

As an example of this rather general setup, consider the market of a differentiated product as in Anderson, de Palma and Thisse (1991) or Berry, Levinshon and Pakes (1995). There are N firms in this market, and each of these firms sells one variety of the product. The demand is the result of a logit model, i.e., the demand for the $i-th$ variety is:

$$D(d_i,d_{-i},x_i,x_{-i}) = M\frac{\exp\{x_i'\beta - \alpha d_i\}}{\sum_{j=0}^{N}\exp\{x_j'\beta - \alpha d_j\}} \tag{15}$$

where M is the market size; (α, β) are parameters; variety 0 is the outside alternative; x_i represents a vector of exogenous characteristics of variety i that affect consumers' utility; d_i is the price of variety i, and d_{-i} is the vector of prices for the other varieties. We assume that for the outside alternative $x_0'\beta - \alpha d_0 = 0$.

The unit cost of producing variety i is c_i, that depends on characteristics x_i but also on a firm specific effciency component ε_i.

$$c_i = \exp\{x_i'\gamma + \varepsilon_i\} \tag{16}$$

Therefore, the profit function is:

$$U(d_i, d_{-i}, x_i, x_{-i}, \varepsilon_i) = [d_i - \exp(x_i'\gamma + \varepsilon_i)]\ D(d_i, d_{-i}, x_i, x_{-i}) \tag{17}$$

The characteristics $\{x_i\}$ are known by all the firms in the market, but the relative efficiency of a firm $\{\varepsilon_i\}$ is private information of that firm. We assume that it is known to all players that conditional on x the private information components $\varepsilon_i's$ are independently and identically distributed over individuals with zero mean, constant variance μ^2, and normal distribution; therefore, $G_i(\varepsilon_{-i}; x, \varepsilon_i) = \prod_{k \neq i} \Phi(\varepsilon_k/\mu)$, which is the same for all players. Finally, the set of possible prices is discrete, i.e., $d_i \in D = \{d[1], d[2], ..., d[J]\}$.

Let θ be the vector of model parameters, $\theta = (\alpha, \beta', \gamma', \mu)'$. Let $p_i(d, x; \theta)$ be the probability that firm i chooses price d in the equilibrium of a market with characteristics x and parameters θ. Define $\tilde{p}(\theta)$ as the vector of probabilities $\{p(d, x_i, x_{-i}; \theta) : d \in D; x_i \in X, x_{-i} \in X^{N-1}\}$. It is possible to show that the expression for the mapping $\widetilde{\Psi}(.)$ in this model is:

$$\Psi(d, x_i, x_{-i}, \theta; \pi) = \int I\left(d = \arg\max_{j \in D} \{\delta(j, x_i, x_{-i}, \varepsilon_i; \theta; \pi)\}\right) \Phi(d\varepsilon_i/\mu) \tag{18}$$

where:

$$\delta(d, x_i, x_{-i}, \varepsilon_i; \theta; \pi) = \sum_{d_{-i}} \left(\prod_{m \neq i} \pi_m(d_m[d_{-i}], x_m, x_{-m})\right) [d - \exp(x_i'\gamma + \varepsilon_i)]\ D(d, d_{-i}, x_i, x_{-i}) \tag{19}$$

2.3 Example 3: Model of irreversible investment[5]

Consider a firm that produces a good using labor and physical capital as inputs. The firm purchases the capital that it employs and it operates in a competitive market both for the output and for the inputs. There are no adjustment costs associated with labor. Let $F(k_t, \eta_t)$ be the production function net of labor costs (once the

[5]See Abel and Eberly (1996) and Abel et al (1996) for a version of this model in continuous time. Bentolila and Bertola (1990) present a similar model in the context of labor demand with linear hiring and firing costs.

optimal amount of labor has been solved), where k_t is the stock of physical capital at the beginning of period t, and η_t is a productivity shock. $F(k, \eta)$ is continuous, differentiable, strictly concave in k, and $\lim_{k \to 0} F(k, \eta) = \infty$. The firm's current profit function is:

$$U(k_t, \eta_t, c_t, i_t) = F(k_t, \eta_t) - c_t \; I(i_t > 0) i_t - c_t^u \; I(i_t < 0) i_t, \tag{20}$$

where i_t is investment in physical capital; c_t is the price that the firm should pay when buying new equipment; and c_t^u is the price that the firm receives when it sells its used capital in the second hand market. Due to asymmetric information and/or firm-specific equipment used by the firm, the selling price of capital is lower than its purchasing price. In particular, we assume that $c_t^u = \lambda c_t$, where $\lambda < 1$. Installed capital depreciates geometrically, $k_{t+1} = \delta k_t + i_t$, where $\delta \in (0, 1)$. The productivity shock follows a first order Markov processes with transitional density $\phi_\eta(\eta_{t+1}; \eta_t)$. The purchasing price of capital has two components: an aggregate component, $\tilde{c}_t$, common to all firms in this market; and an idiosyncratic component, ε_t, such that $c_t = \tilde{c}_t \varepsilon_t$. For the sake of simplicity we assume here that the aggregate price is constant, but it is simple to obtain similar results when $\tilde{c}_t$ follows a first order Markov process. The idiosyncratic component is *iid* over time and firms.

The Bellman equation of this problem is:

$$V(k_t, c_t, \eta_t) = \max_{\{i_t\}} \; U(k_t, \eta_t, c_t, i_t) + \beta \; EV(\delta k_t + i_t, \eta_t) \tag{21}$$

where $\beta \in (0, 1)$ is the discount factor, $V(.)$ is the value function and

$$EV(\delta k_t + i_t, \eta_t) = \int \int V(\delta k_t + i_t, c_{t+1}, \eta_{t+1}) \; \phi_\eta(d\eta_{t+1}; \eta_t) \; \phi_c(dc_{t+1}) \tag{22}$$

It is possible to show that this problem has a unique solution with the following optimal decision function:

$$i^*(k_t, c_t, \eta_t) = \begin{cases} k^P(c_t, \eta_t) - \delta k_t & if \quad \ln c_t < \ln[\beta \; EV_k(\delta k_t, \eta_t)] \\ 0 & if \quad \ln[\beta \; EV_k(\delta k_t, \eta_t)] \leq \ln c_t \leq \ln[\beta \; EV_k(\delta k_t, \eta_t)] - \ln \lambda \\ k^S(c_t, \eta_t) - \delta k_t & if \quad \ln c_t > \ln[\beta \; EV_k(\delta k_t, \eta_t)] - \ln \lambda \end{cases} \tag{23}$$

where $k^P(.)$ and $k^S(.)$ are the optimal capital stocks when the firm decides to purchase new capital and when it sells used capital, respectively; and $EV_k \equiv \partial EV / \partial k$. $k^P(.)$ and $k^S(.)$ are implicitly defined by the equations:

$$\begin{aligned} \beta \; EV_k(k_t^P, \eta_t) &= c_t \\ \beta \; EV_k(k_t^S, \eta_t) &= \lambda c_t \end{aligned} \tag{24}$$

Notice that if we knew the function $EV_k(.)$ we could obtain a closed form expression for the optimal decision rule. However, $EV_k(.)$ has to be obtained recursively. In particular, it is possible to show that this function is the unique fixed point of the following functional equation:

$$
\begin{aligned}
EV_k(k_{t+1}, \eta_t) &= E\left(F_k[k_{t+1}, \eta_{t+1}] \mid k_{t+1}, \eta_t\right) \\
&+ \delta \Pr\left(i^*_{t+1} > 0 \mid k_{t+1}, \eta_t\right) \; E\left(c_{t+1} \mid i^*_{t+1} > 0, k_{t+1}, \eta_t\right) \\
&+ \delta\lambda \Pr\left(i^*_{t+1} < 0 \mid k_{t+1}, \eta_t\right) \; E\left(c_{t+1} \mid i^*_{t+1} < 0, k_{t+1}, \eta_t\right) \\
&+ \beta\delta \Pr\left(i^*_{t+1} = 0 \mid k_{t+1}, \eta_t\right) \; E\left(EV_k(\delta k_{t+1}, \eta_{t+1}) \mid i^*_{t+1} = 0, k_{t+1}, \eta_t)\right)
\end{aligned} \tag{25}
$$

Now, let y_t be the sign of the investment at period t, i.e., $y_t \equiv sign(i_t) \in \{-1, 0, +1\}$, and define x_t as the vector (k_t, η_t). We define $p(y, x; \theta)$ as the probability, conditional on x, that the optimal sign of investment is y, where θ is the vector of structural parameters. Taking into account the optimal decision rule in (18), it is clear that:

$$
\begin{aligned}
p(1, x; \theta) &= \Phi_c(\ln\{\beta \; EV_k[\delta k, \eta]\}) \\
p(-1, x; \theta) &= 1 - \Phi_c(\ln\{\beta \; EV_k[\delta k, \eta]\} - \ln \lambda)
\end{aligned} \tag{26}
$$

Furthermore, the conditional expectations of c_{t+1} on the right hand side of equation (20) can be written as a function of the probabilities $p(\theta)$; if we do this and solve for EV_k we obtain an alternative expression for EV_k in terms of the probabilities $p(\theta)$ and the primitives ϕ_η, ϕ_c and F_k. We denote this function by $\sigma(x; \theta, p[\theta])$. As in previous examples, the function $\sigma(.)$ can be evaluated at any vector of probabilities π, not just at the probabilities associated with optimal behavior. Therefore, substituting the function $\sigma(x; \theta, p[\theta])$ for EV_k in (21) we can write $p(\theta)$ as the solution of the fixed point equation $\pi = \tilde{\Psi}(\theta; \pi)$, where

$$
\begin{aligned}
\Psi(1, x, \theta; \pi) &= \Phi(\ln\{\beta \; \sigma[\delta k, \eta; \theta, \pi]\}) \\
\Psi(-1, x, \theta; \pi) &= 1 - \Phi(\ln\{\beta \; \sigma[\delta k, \eta; \theta, \pi]\} - \ln \lambda)
\end{aligned}
$$

3 Two–stage estimator

Let $\{y_i, x_i : i = 1, 2, ..., n\}$ be a random sample of y and x. The *Maximum Likelihood Estimator (MLE)* of θ^* is:

$$
\hat{\theta}_n^{MLE} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \ln p(y_i, x_i; \theta) \tag{27}
$$

where the probabilities are known to satisfy $\tilde{p}(\theta) = \tilde{\Psi}(\theta; \tilde{p}(\theta))$. The MLE can be obtained using a nested fixed point solution -estimation. Consider instead the following *Two-stage Estimator* of θ^*. Let $\hat{p}_n$ be a nonparametric estimator of $\tilde{p}^*$. Then,

$$
\hat{\theta}_n^{2S} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \Psi(y_i, x_i; \theta, \hat{p}_n) \tag{28}
$$

For given θ the pseudo-likelihood in (23) can be computed with a single evaluation of the operator $\widetilde{\Psi}$, whereas computing the likelihood in (22) requires iterating in $\widetilde{\Psi}$ until convergence. It is clear that using the two-stage estimator may result in large savings in the computational cost of estimation, extending the class of estimable models. In Proposition 1 below we give sufficient conditions for consistency of the two-stage estimator, and in Proposition 2 we provide additional conditions under which the two-stage estimator can be used instead of the Maximum Likelihood estimator at no cost in terms of asymptotic efficiency.

PROPOSITION 1 (Consistency):
Let $\Gamma \equiv \Theta \times [0,1]^{(J-1)M}$. Consider the following regularity conditions.
(i) Θ is a compact sets.
(ii) $\Psi(y,x;\theta,\pi)$ is continuous and twice continuously differentiable in (θ,π).
(iii) $\Psi(y,x;\theta,\pi) > 0$ for any $(y,x) \in Y\text{x}X$ and for any $(\theta,\pi) \in \Gamma$.
(iv) $\{y_i, x_i\}$ for $i = 1,2,...,n$ are independently and identically distributed.
(v) There is a $\theta^* \in \Theta$ such that, for any $(y,x) \in YxX$, $p(y,x;\theta^*) = p^*(y,x)$, and for any $\theta \neq \theta^*$ the set $\{(y,x) : \Psi(y,x;\theta,f^*_{y|x}) \neq p^*(y,x)\}$ has positive probability.
(vi) $(\theta^*, \tilde{p}^*) \in int(\Gamma)$.
(vii) $\hat{p}_n$ is a consistent estimator $\tilde{p}^*$.
Under these conditions the two-stage estimator $\hat{\theta}_n^{2S}$ converges a.s. to θ^*.

Proof: See Appendix.

PROPOSITION 2 (Asymptotic Efficiency):
Under regularity conditions (i)-(vii), in Proposition 1, and
(viii) $\hat{p}_n$ is such that

$$\left[\frac{1}{\sqrt{n}}\sum_i \frac{\partial \ln \Psi(y_i,x_i;\theta^*,\tilde{p}^*)}{\partial \theta'}, \sqrt{n}(\hat{p}_n - \tilde{p}^*)'\right]' \rightarrow_d N(0,\Omega)$$

(ix) For any $\theta \in \Theta$, $\partial\tilde{\Psi}(\theta;\tilde{p}[\theta])/\partial\pi' = 0$,
the two-stage estimator $\hat{\theta}_n^{2S}$ is asymptotically normal and asymptotically equivalent to the conditional maximum likelihood estimator.

Proof: See Appendix 1.

Notice that whereas consistency of the first-stage non-parametric estimators (together with regularity conditions) is enough to obtain consistency in the second stage, sufficient conditions for asymptotic efficiency include an additional restriction on the model. This is condition (ix) which states that the Jacobian matrix of the fixed point operator should be zero at the fixed point. If this is the case, the proof of Proposition 2 shows that the asymptotic variance matrix of the first stage estimators does not affect the asymptotic variance of the second stage estimators.

Aguirregabiria and Mira (1999) have proved that this condition is satisfied for the class of problems and for the fixed point operator considered in Example 1. However, in general working with condition (ix) will require a deep understanding of the model at hand. Even when .we are able to write the distribution of endogenous variables as a solution of a fixed point problem, verifying condition (ix) may be non-trivial, or we may find that it is not satisfied. We now show that, provided a fixed point operator has been found, Newton's algorithm can be used to define another valid fixed point operator which will satisfy condition (ix).

Consider a model with fixed point mapping $\tilde{\Psi}(.,.)$. Suppose that for given θ we solve the fixed-point problem $\tilde{p}(\theta) = \tilde{\Psi}(\theta; \tilde{p}(\theta))$ using Newton's method to find a zero of $\pi - \tilde{\Psi}(\theta; \pi)$. Newton iterations have the following form:

$$\pi_{k+1} = \Gamma(\theta; \pi_k) \equiv \pi_k - \left(I - \frac{\partial \tilde{\Psi}(\theta; \pi_k)}{\partial \pi'} \right)^{-1} \left(\pi_k - \tilde{\Psi}(\theta; \pi_k) \right)$$

The following Lemma establishes two relevant porperties of the Newton fixed point operator $\Gamma()$

LEMMA 1:

(1) A fixed point $\pi^*$of $\tilde{\Psi}(\theta; .)$ is a fixed point of $\Gamma(\theta; .)$ if the inverse of $I - \dfrac{\partial \tilde{\Psi}(\theta; \pi^*)}{\partial \pi'}$ exists, and a fixed point of $\Gamma(\theta; .)$ is a fixed point of $\tilde{\Psi}(\theta; .)$.

(2) Let π^* be a fixed point of $\Gamma(\theta, \cdot)$; then, $\dfrac{\partial \Gamma(\theta; \pi^*)}{\partial \pi} = 0$

Proof: See Appendix 1.

This suggests that, if a fixed operator $\tilde{\Psi}()$ has been found which does not satisfy condition (ix), an asymptotically efficient estimator based on the associated Newton operator may be available.

APPENDIX 1. Proofs

Proof of Proposition 1 (Consistency):

Let n denote sample size. Consider the sample functions:

$$\tilde{Q}_n(\theta,\pi) = \frac{1}{n}\sum_{i=1}^{n} \ln \Psi(y_i, x_i; \theta, \pi) \qquad ; \qquad Q_n(\theta) = \tilde{Q}_n(\theta, \hat{p}_n)$$

And define:

$$\tilde{Q}_\infty(\theta,\pi) \equiv E(\ln \Psi[d_i, x_i; \theta, \pi])$$

By Property 24.2 in Gourieroux and Monfort (vol. II, page 387), if: (I) $Q_n(\theta)$ converges a.s. and uniformly in θ to $Q_\infty(\theta)$; and (II) $Q_\infty(\theta)$ has a unique maximum in Θ at θ^*; then $\hat{\theta}_n^{2S} \equiv \arg\max_{\alpha\in\Theta} Q_n(\theta)$ converges a.s. to θ^*. Regularity condition (v) implies (II) by the information inequality. We have to prove (I).

By Lemma 24.1 in Gourieroux and Monfort (vol. II, page 392), we have that if: (a) $\tilde{Q}_n(\theta,\pi)$ converges a.s. and uniformly in (θ,π) to $\tilde{Q}_\infty(\theta,\pi)$; (b) $\tilde{Q}_\infty(\theta,\pi)$ is uniformly continuous in (θ,π); and (c) $\hat{p}_n$ converge a.s. to $\tilde{p}^*$; then $\tilde{Q}_n(\theta,\hat{p}_n)$ converges a.s. and uniformly in θ to $\tilde{Q}_\infty(\theta,\tilde{p}^*) \equiv Q_\infty(\theta)$. Condition (c) holds directly from (vii). By regularity conditions (i) and (ii), $\tilde{Q}_\infty(\theta,\pi)$ is continuous on a compact set, so it is uniformly continuous, i.e., (b) holds. Now, we prove that condition (a) holds.

Let $H_*(.,.)$ be the true probability distribution of (y,x), and let $H_n(.,.)$ be the empirical distribution of (y,x) in a sample with size n. By definition, for any (θ,π):

$$\begin{aligned} |\tilde{Q}_n(\theta,\pi) - \tilde{Q}_\infty(\theta,\pi)| &= |\sum_{x\in X}\sum_{j=1}^{J} \ln \Psi(j,x;\theta,\pi)\ [H_n(j,x) - H_*(j,x)]| \\ &\leq \sum_{x\in X}\sum_{j=1}^{J} |\ln \Psi(j,x;\theta,\pi)|\ |H_n(j,x) - H_*(j,x)| \end{aligned}$$

By condition (iv) $H_n(j,x) \rightarrow_{a.s.} H_*(j,x)$. Furthermore, by conditions (i)-(iii) $|\ln \Psi(j,x;\gamma)|$ is bounded. Therefore,

$$\begin{aligned} & \Pr\left(\lim_{n\to\infty} \sup_{(\theta,\pi)\in\Gamma} |\tilde{Q}_n(\theta,\pi) - \tilde{Q}_\infty(\theta,\pi)| = 0\right) \\ \geq\ & \Pr\left(\lim_{n\to\infty} \sup_{\theta,\pi)\in\Gamma} \sum_{x\in X}\sum_{j=1}^{J} |\ln \Psi(j,x;\theta,\pi)|\ |H_n(j,x) - H_*(j,x)| = 0\right) \\ \geq\ & \Pr\left(\forall(j,x) : \lim_{n\to\infty} |H_n(j,x) - H_*(j,x)| = 0\right) = 1 \end{aligned}$$

i.e., $\tilde{Q}_n(\theta,\pi)$ converges a.s. and uniformly in (θ,π) to $\tilde{Q}_\infty(\theta,\pi)$.

Proof of Proposition 2 (Asymptotic efficiency):

Given conditions (ii) and (vii) and the definition of $\hat{\theta}_n^{2S}$, the first order conditions of optimality imply that with probability approaching one $\partial \tilde{Q}_n(\hat{\gamma}_n)/\partial \theta = 0$ where $\hat{\gamma}_n \equiv (\hat{\theta}_n^{2S\prime}, \hat{p}_n')'$. By condition (ii), $\tilde{Q}_n(.)$ is twice continuously differentiable and we can apply the stochastic mean value theorem to $\partial \tilde{Q}_n(.)/\partial \theta$ between $\hat{\gamma}_n$ and $\gamma^* \equiv (\theta^{*\prime}, \tilde{p}^*)'$. There are K vectors $\{\bar{\gamma}_n^1, \bar{\gamma}_n^2, ..., \bar{\gamma}_n^K\}$ which are convex combinations of $\hat{\gamma}_n$ and γ^* such that:

$$\frac{\partial \tilde{Q}_n(\hat{\gamma}_n)}{\partial \theta} - \frac{\partial \tilde{Q}_n(\gamma^*)}{\partial \theta} = \begin{pmatrix} \partial^2 \tilde{Q}_n(\bar{\gamma}_n^1)/\partial \theta_1 \partial \gamma' \\ ... \\ \partial^2 \tilde{Q}_n(\bar{\gamma}_n^K)/\partial \theta_K \partial \gamma' \end{pmatrix} (\hat{\gamma}_n - \gamma^*)$$

Given that any $\bar{\gamma}_n^j$ is a convex linear combination of $\hat{\gamma}_n$ and γ^*, and given that $\hat{\gamma}_n \rightarrow_{a.s.} \gamma^*$, then $\bar{\gamma}_n^j \rightarrow_{a.s.} \gamma^*$. Furthermore, $\partial^2 \tilde{Q}_n(\gamma)/\partial \theta \partial \gamma'$ converges in probability and uniformly in γ to $\partial^2 \tilde{Q}_\infty(\gamma)/\partial \alpha \partial \gamma'$ and therefore $\partial^2 \tilde{Q}_n(\bar{\gamma}_n^j)/\partial \theta_j \partial \gamma' \rightarrow_p \partial^2 \tilde{Q}_\infty(\gamma^*)/\partial \theta_j \partial \gamma'$ (see Amemiya, Thm 4.2.2 and Thm.4.1.5). We can now rewrite the previous mean value theorem as follows:

$$-\frac{\partial \tilde{Q}_n(\gamma^*)}{\partial \theta} = \left(\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \theta'} + o_p(1) \right) (\hat{\theta}_n^{2S} - \theta^*) + \left(\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \pi'} + o_p(1) \right) (\hat{p}_n - \tilde{p}^*)$$

Or:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{2S} - \theta^*) &= -\left(\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \theta'} + o_p(1) \right)^{-1} \\ & \qquad \left(I; \frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \pi'} + o_p(1) \right) \begin{pmatrix} \sqrt{n} \partial \tilde{Q}_n(\gamma^*)/\partial \theta \\ \sqrt{n}(\hat{p}_n - \tilde{p}^*) \end{pmatrix} \end{aligned}$$

Notice that condition (vi) implies that $\partial^2 \tilde{Q}_\infty(\gamma^*)/\partial \theta \partial \theta'$ is a non singular (negative definite) matrix. By condition (viii) and the Mann-Wald Theorem, it is straightforward that $\sqrt{n}(\hat{\theta}_n^{2S} - \theta^*) \rightarrow_d N(0, V^*)$, where:

$$V^* = \left(\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \theta'} \right)^{-1} \left(I; \frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \pi'} \right) \Omega \left(I; \frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \pi'} \right)' \left(\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \theta'} \right)^{-1}$$

Now, using condition (ix) it is possible to simplify this expression. First, notice that,

$$\frac{\partial^2 \tilde{Q}_\infty(\gamma^*)}{\partial \theta \partial \pi'} = E\left(\frac{\partial^2 \ln \Psi(y_i, x_i; \gamma^*)}{\partial \theta \partial \pi'} \right) = E\left(\frac{\partial \ln \Psi(y_i, x_i; \gamma^*)}{\partial \theta} \frac{\partial \ln \Psi(y_i, x_i; \gamma^*)}{\partial \pi'} \right)$$

By condition (ix), for any pair (y, x): (a) $\Psi(y, x; \gamma^*) = p(y, x; \theta^*)$; (b) $\partial \Psi(y, x; \gamma^*)/\partial \pi = 0$; and (c) $\partial \Psi(y, x; \gamma^*)/\partial \theta = \partial p(y, x; \theta^*)/\partial \theta$. Therefore, $\partial^2 \tilde{Q}_\infty(\gamma^*)/\partial \theta \partial \pi' = 0$ by equivalence of the information matrix and we get:

$$V^* = \Omega_{00}^{-1}$$

where Ω_{00} is Fisher's information matrix,

$$\Omega_{00} = E\left(\frac{\partial \ln p(y_i, x_i; \theta^*)}{\partial \theta} \frac{\partial \ln p(y_i, x_i; \theta^*)}{\partial \theta'}\right)$$

Proof of Lemma 1:

The fixed point problem is $\pi = \widetilde{\Psi}(\theta, \pi)$. The associated Newton operator was defined as

$$\Gamma(\theta; \pi) \equiv \pi - \left(I - \frac{\partial \tilde{\Psi}(\theta; \pi)}{\partial \pi'}\right)^{-1} \left(\pi - \tilde{\Psi}(\theta; \pi)\right)$$

where π, Ψ and Γ are vectors of dimension M. The proof of (1) trivial given the definition of Γ. As for (2), let π^* be a fixed point of Γ; taking derivatives we get

$$\begin{aligned} D\Gamma(\pi^*) &= I - \sum_i^M [\pi_i^* - \Psi_i(\pi^*)] \frac{\partial \{[I - D\Psi(\pi^*)]^{-1}\}_{\bullet i}}{\partial x'} - [I - D\Psi(\pi^*)]^{-1}[I - D\Psi(\pi^*)] \\ &= I - 0 - I = 0 \end{aligned}$$

since π^* is also a fixed point of Ψ.

REFERENCES:

Abel, A., and J. Eberly (1996): "Optimal investment with costly reversibility", *Review of Economic Studies*, 63, 581-594.

Abel, A., A. Dixit, J. Eberly, and R. Pindyck (1996): "Options, the value of capital, and investment", *Quarterly Journal of Economics*, 111,

Aguirregabiria and Mira (1999): "Swapping the nested fixed point algorithm: A class of estimators for Markov decision models". CEMFI Working Paper 9904.

Anderson, S., A. de Palma, and J.-F. Thisse (1991): "Discrete choice theory of product differentiation", Cambridge, MA. MIT Press.

Bentolila, S., and G. Bertola (1990): "Firing costs and labour demand: How bad is Eurosclerosis?", *Review of Economic Studies*, 57, 381-402.

Berry, S., J. Levinshon and A. Pakes (1996): "Automobile rpices in market equilibrium", *Econometrica*, 64,

Bresnahan, T., and P. Reiss (1991): "Empirical models of discrete games", *Journal of Econometrics*, 48, 57-81.

Browning, M., L. Hansen, and J. Heckman (1999): "Micro data and general equilibrium models", in *Handbook of Macroeconomics*, J. Taylor and M. Woodford (eds.) North Holland.

Ericson, R., and A. Pakes (1995): "Markov-Perfect industry dynamics: A framework for empirical work", *Review of Economic Studies,* 62, 53-82.

Gourieroux, C., and A. Monfort (1995): "Statistics and Econometric Models: Volume I", Cambridge University Press.

Heckman, J., L. Lochner, and C. Taber (1998): "Explaining rising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents", *Review of Economic Dynamics*, 1, 1-58.

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

Hotz, J., R.A. Miller, S. Sanders, and J. Smith (1994): "A simulation estimator for dynamic models of discrete choice", *Review of Economic Studies* 61, 265-89.

Keane, M. P., and K. I. Wolpin (1994): "The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence", *The Review of Economics and Statistics* 76, pp. 648-672.

Manski, C. (1991): "Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty", in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, W. Barnett, J. Powell, and G. Tauchen (eds.). Cambridge University Press.

Manski, C. (1993): "Dynamic choice in social settings: Learning from the experience of others", *Journal of Econometrics* 58, 121-36.

McFadden, D. (1973): "Conditional logit analysis of qualitative choice behavior", in P. Zarembka (ed.) *Frontiers of Econometrics*, New York, Academic Press, 105-42.

Miller, R. (1984): "Job matching and occupational choice", Journal of Political Economy, 92, 1086-1120.

Newey, W. K. (1984): "A Method of Moments Interpretation of Sequential Estimators", *Economic Letters*, 15, 201-206.

Pakes, A. (1986): "Patents as options: Some estimates of the value of holding European patent stocks", *Econometrica*, 54, 755-784.

Pakes, A. (1994): "Dynamic structural models, problems and prospects", in C. Sims (ed.) *Advances in Econometrics. Sixth World Congress*, Cambridge University Press.

Pakes, A. and P. McGuire (1994): "Computing Markov-Perfect Nash equilibria: Numerical implications of a dynamic differentiated product model, *Rand Journal of Economics*, 555-589.

Pakes, A, and P. McGuire (1996): "Stochastic Algorithms for Dynamic Models: Markov Perfect Equilibrium ad the 'Curse of Dimensionality"', manuscript, Department of Economics, Yale University.

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

Rust, J. (1988): "Maximum Likelihood Estimation of Discrete Control Processes", *SIAM Journal of Control and Optimization*, 1006-1024

Rust, J. (1994): "Structural estimation of Markov decision processes", in R. E. Engle and McFadden (eds.) *Handbook of Econometrics Volume 4*, North-Holland.

Rust, J. (1994b): "Estimation of dynamic structural models, problems and prospects: discrete decision processes", in C. Sims (ed.) *Advances in Econometrics. Sixth World Congress*, Cambridge University Press.

Wolpin, K. (1984): "An estimable dysnamic stochastic model of fertility and child mortality", *Journal of Political Economy*, 92, 852-874.