New Mathematics and Natural Computation
Vol. 1, No. 2 (2005) 295–303

A HYBRID GENETIC ALGORITHM FOR THE MAXIMUM LIKELIHOOD ESTIMATION OF MODELS WITH MULTIPLE EQUILIBRIA: A FIRST REPORT

VICTOR AGUIRREGABIRIA
Department of Economics, Boston University, 270 Bay State Road
Boston, MA 02215, USA
vaguirre@bu.edu

PEDRO MIRA
CEMFI, Casado del Alisal, 5
Madrid, Madrid 28004, Spain
mira@cemfi.es

This paper presents a hybrid genetic algorithm to obtain maximum likelihood estimates of parameters in structural econometric models with multiple equilibria. The algorithm combines a pseudo maximum likelihood (PML) procedure with a genetic algorithm (GA). The GA searches globally over the large space of possible combinations of multiple equilibria in the data. The PML procedure avoids the computation of all the equilibria associated with every trial value of the structural parameters.

Keywords: Genetic algorithms; maximum likelihood estimation; multiple equilibria.

1. Introduction

This paper deals with the computation of maximum likelihood estimates of structural parameters in econometric models with multiple equilibria. Multiplicity of equilibria appears frequently in structural econometric models like empirical discrete games,[1] coordination games[2] or models with social interactions,[3] among others. Models with multiple equilibria do not provide unique predictions on the probability distribution of the endogenous variables conditional on the exogenous variables. This indeterminacy poses practical problems in maximum likelihood estimation. To obtain maximum likelihood estimates of the structural parameters one should maximize the likelihood function not only with respect to the structural parameters but also with respect to the multiple *equilibrium types* that can generate the observations in the data. There are two main reasons why optimization with respect to *equilibrium types* can be a very complicated task. First, computing all the equilibria associated with each trial value of the parameters can be computationally very demanding. And second, the number of possible combinations of equilibria in the data increases exponentially with the sample size and it becomes a huge number even for the simplest problems.

These problems have motivated the use of estimation methods which are computationally simpler than maximum likelihood. For instance, in the context of empirical discrete games, Tamer,[4] Pesendorfer and Smichdt-Dengler,[5] Aguirregabiria,[6] and Aguirregabiria and Mira[7] have proposed and applied pseudo maximum likelihood (PML) estimators. These PML methods proceed in two steps. The first step identifies nonparametrically the equilibrium (or equilibria) played in the data. The second step estimates the structural parameters of the model by maximizing a pseudo likelihood function based on best response functions evaluated at the equilibrium estimated in the first step. The main advantage of PML estimation is its computational simplicity. However, the method is not statistically efficient, and it can perform poorly in small samples because the nonparametric estimates in the first step can be very imprecise.

To deal with these limitations, Aguirregabiria and Mira[7] propose a recursive extension of the pseudo likelihood method, the so-called *nested pseudo likelihood* (NPL) algorithm. The NPL provides estimates which are statistically more efficient than the two-step PML, both asymptotically and in finite samples. Furthermore, the NPL algorithm can be used to search for the maximum likelihood estimator. In particular, the maximum likelihood estimator is a fixed point of the NPL procedure. However, the NPL algorithm can have other fixed points which are not the maximum likelihood estimator. Therefore, to compute the maximum likelihood estimator, we should search globally over the set of NPL fixed points.

In this paper, we propose a relatively simple but computationally effective algorithm to search for the maximum likelihood estimator. The algorithm combines the NPL procedure with a Genetic Algorithm (GA). The GA searches efficiently over the huge space of possible combinations of multiple equilibria in the data. The NPL procedure avoids the repeated computation of equilibria for each trial value of the structural parameters.

A Genetic Algorithm (GA) is a class of stochastic optimization method that exploits the concepts of natural selection and evolution to solve global optimization problems. GAs were first proposed by Holland.[8] Among their many applications, GA have been successfully used to search for global optima of discrete and step functions with very large search spaces. Mitchell[9] provides an excellent survey of the theory and applications of these algorithms. See also Mitchell, Holland and Forrest.[10] Although GAs have been extensively used in experimental and evolutionary economics, the application of GAs in econometrics has been more rare. Important exceptions are Dorsey and Mayer[11] and Beenstock and Szpirob.[12]

The rest of this paper is organized as follows. Section 2 presents the class of structural models that this paper is concerned with and provides a particular example of this class of models. Section 3 describes the estimation of structural parameters by maximum likelihood and by pseudo maximum likelihood. Section 4 describes our hybrid genetic algorithm. We summarize and conclude in Sec. 5.

2. Model

2.1. *General framework*

Let $y \in Y$ and $x \in X$ be two vectors of random variables where Y is a discrete and finite set. And let $p_0(y|x)$ be the probability distribution of y conditional to x in the population under study. Given that y is a discrete variable with a finite support, we can also represent the conditional probability function $p_0(\cdot|x)$ using the vector $P_0(x) \equiv \{p_0(y|x) : y \in Y\}$. A model can be described as a family of probability distribution functions $p(y|x, \theta)$, where $\theta \in \Theta$ is a vector of K parameters, and $\Theta \subset \mathbb{R}^K$ is a compact set. We can also represent a probability function $p(\cdot|x, \theta)$ using the vector $P(x, \theta) \equiv \{p(y|x, \theta) : y \in Y\}$.

An important feature of the class of models that we consider in this paper is that the model does not provide a closed form analytical expression for the probability functions $p(\cdot|x, \theta)$. Instead, these functions are only implicitly defined as fixed points of a mapping. More specifically, the vector $P(x, \theta)$ is a fixed point of a mapping $\Psi(x, \theta, P) = \{\psi(y|x, \theta, P) : y \in Y\}$, where $\psi(\cdot|x, \theta, P)$ is a probability function for y conditional on x. Therefore, $P(x, \theta) = \Psi(x, \theta, P(x, \theta))$ or, what is equivalent, for any $y \in Y$:

$$p(y|x, \theta) = \psi(y|x, \theta, P(x, \theta)). \tag{1}$$

The function ψ is twice continuously differentiable in P and θ, and for any (y, x, θ, P) the probability $\psi(y|x, \theta, P)$ is strictly greater than zero.

For some values of (x, θ) the mapping $\Psi(x, \theta, \cdot)$ can have more than one fixed point. That is, the model can have multiple equilibria for some values of the structural parameters. Let $\{P^\tau(x, \theta) : \tau = 1, 2, \ldots\}$ be the set of equilibria associated with (x, θ). The equilibria are indexed by the variable $\tau \in \{1, 2, \ldots\}$ that is called the *equilibrium type.*

2.2. *Example: Discrete game with incomplete information*

There are N players which are indexed by $i \in I = \{1, 2, \ldots, N\}$. Each player should choose an action from a set of choice alternatives $Y = \{0, 1, \ldots, J\}$. We represent the decision of player i by the variable $y_i \in Y$. The utility function of player i is:

$$U_i = u_i(y_i, y_{-i}, x) + \varepsilon_i(y_i) \tag{2}$$

where y_{-i} is the vector with the decisions of players other than i; x is a vector of players' exogenous characteristics which are common knowledge; and $\varepsilon_i \equiv (\varepsilon_i(0), \varepsilon_i(1), \ldots, \varepsilon_i(J))$ represents characteristics that are private information of player i.

Assumption 1. For any $i \in I$ the vector $\varepsilon_i \in R^{J+1}$ is: (1) independent of common knowledge variables x; and (2) independently distributed across players with distribution function $G_i(\cdot)$ that is absolutely continuous with respect to the Lebesgue measure.

Let $\sigma = \{\sigma_i(x, \varepsilon_i) : i \in I\}$ be a set of strategy functions where $\sigma_i : X \times R^{J+1} \rightarrow Y$. Associated with a set of strategy functions we can define a set of *choice probabilities* $P^\sigma(x) = \{P_i^\sigma(y_i|x) : (y_i, i) \in Y \times I\}$ such that:

$$P_i^\sigma(y_i|x) \equiv \int I\left\{\sigma_i(x, \varepsilon_i) = y_i\right\} dG_i(\varepsilon_i) \tag{3}$$

where $I\{\cdot\}$ is the indicator function. These probabilities represent the expected behavior of player i from the point of view of the other players (who do not know ε_i) when he follows his strategy in σ. Let $u_i^\sigma(y_i, x)$ be player i's expected utility if he chooses alternative y_i and the other players behave according to their respective strategies in σ. By the independence of private information in Assumption 1,

$$u_i^\sigma(y_i, x) = \sum_{y_{-i}} \left(\prod_{j \neq i} P_j^\sigma(y_j|x) \right) u(y_i, y_{-i}, x). \tag{4}$$

A *Bayesian Nash equilibrium* (BNE) is a vector of strategy functions σ^* such that for any player i and for any $(x, \varepsilon_i) \in X \times R^{J+1}$,

$$\sigma_i^*(x, \varepsilon_i) = \arg\max_{y_i \in Y} \left\{ u_i^{\sigma^*}(y_i, x) + \varepsilon_i(y_i) \right\}. \tag{5}$$

We can represent this BNE in probability space. Let σ^* be a set of BNE strategies, and let P^* be the choice probabilities associated with these strategies. By definition, $P_i^*(y_i|x) = \int I\{y_i = \sigma_i^*(x, \varepsilon_i)\}\, dG_i(\varepsilon_i)$. Solving the equilibrium condition (5) in this expression we get that for any $(y_i, i) \in Y \times I$:

$$P_i^*(y_i|x) = \int I\Big(y_i = \arg\max_{y \in Y} \left\{ u_i^*(y, x) + \varepsilon_i(y) \right\} \Big) dG_i(\varepsilon_i). \tag{6}$$

Notice that the function u_i^σ depends on players' strategies only through the choice probabilities P^σ associated with σ . To emphasize this point, we modify the notation and use the symbol u_i^P, instead of u_i^σ, to denote these expected utility functions. Therefore, the right-hand side in Eq. (6) is a function that we define as $\psi_i(y_i|x, P)$. We call the functions ψ_i *best response probability functions.* The vector of equilibrium probabilities $P^*(x) \equiv \{P_i^*(y_i|x) : (y_i, i) \in Y \times I\}$ is a fixed point of the best response mapping $\Psi(x, P) \equiv \{\psi_i(y_i|x, P) : (y_i, i) \in Y \times I\}$. Given Assumption 1, best response probability functions are continuous in the compact set of players' choice probabilities. By Brower's theorem, there exists at least one equilibrium. In general, the equilibrium is not unique.

The primitives of the model $\{u_i, G_i : i \in I\}$ can be described in terms of a vector of parameters $\theta \in \Theta \subseteq R^K$. Primitives are continuously differentiable in θ. We use $\Psi(x, \theta, P)$ to denote the equilibrium mapping associated with (x, θ). And $P(x, \theta) = \{P_i(y|x, \theta) : (y, i) \in Y \times I\}$ represents an equilibrium associated with (x, θ).

For the sake of exposition we will concentrate on this empirical discrete game for the rest of the paper. However, all our results can be applied to the general model in Sec. 2.1.

3. Estimation

3.1. *Data generating process*

Suppose that the game has been played at different moments in time or at different locations or markets. We have a random sample of T realizations of the game where we observe players' actions and common knowledge state variables $\{y_t, x_t : t = 1, 2, \ldots, T\}$ with $y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$. Let $\theta^0 \in \Theta$ be the true value of θ in the population under study. We are interested in the estimation of θ^0.

Let τ_t be the equilibrium type of observation t. These equilibrium types are unobservable to the researcher. And let $P_t^0(x_t)$ be the distribution of y_t conditional on x_t in the population that generates observation t. Since y_t comes from an equilibrium of the game, we have that $P_t^0(x_t) = P^{\tau_t}(x_t, \theta^0)$. The following assumption establishes some conditions on the data generating process that guarantee the identification of θ^0.

Assumption 2. (A) For every observation t the equilibrium type τ_t is determined by a function $\tau^0(\cdot)$ of the common knowledge state variables, i.e. $\tau_t = \tau^0(x_t)$. And (B) there is a unique pair (θ^0, τ^0) such that $P^0(x) = P^{\tau^0(x)}(x, \theta^0)$ for every $x \in X$.

Under Assumption 2(A), two sample points with the same values of the exogenous variables x should have the same equilibrium probabilities. The function τ^0 is called the *equilibrium selection mechanism*. Under this assumption, we can identify nonparametrically the equilibrium probabilities in the population under study. Assumption 2(B) establishes the joint identification of the structural parameters and the equilibrium selection mechanism.

3.2. *Maximum likelihood estimation*

The maximum likelihood estimator (MLE) of (θ^0, τ^0) maximizes the likelihood with respect to θ and with respect to the equilibrium types in the sample. Thus, the MLE is:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \left\{ \sup_{\tau_1, \ldots, \tau_T} \sum_{t=1}^{T} \sum_{i=1}^{N} \log P_i^{\tau_t}(y_{it} | x_t; \theta) \right\} \tag{7}$$

subject to the restriction that $\tau_t = \tau_{t'}$ when $x_t = x_{t'}$. This estimator is consistent and asymptotically efficient. However, its implementation can be computationally very costly. The problem is in the maximization with respect to the equilibrium types. First, we need to know all the equilibrium types that the model has for every trial value of θ and for every value x_t in the sample. This is impractical in many applications. And second, the number of possible values that the vector $(\tau_1, \ldots, \tau_T)$ can take is huge even for very simple models. For instance, if the number of equilibrium types is 3 and the number of observations is 200, we have $3^{200} \simeq 10^{95}$ possible values for the vector $(\tau_1, \ldots, \tau_T)$. It is clear that optimization with respect to the equilibrium types can be extremely costly. This problem has motivated the development of alternative methods, like pseudo maximum likelihood (PML) estimation, that avoid the search for the MLE of the equilibrium selection mechanism τ^0.

3.3. *Pseudo maximum likelihood estimation*

Define the *pseudo likelihood function*

$$Q\left(\theta, P_1, \ldots, P_T\right) \equiv \sum_{t=1}^{T} \sum_{i=1}^{N} \log \Psi_i(y_{it}|x_t, \theta, P_t). \tag{8}$$

This function is defined for arbitrary choice probabilities, not necessarily in equilibrium. The two-step PML estimator is the value of θ that maximizes the pseudo likelihood function $Q(\theta, \hat{P}_1, \ldots, \hat{P}_T)$, where $\{\hat{P}_1, \ldots, \hat{P}_T\}$ are nonparametric estimates of players' choice probabilities conditional on x. Notice that the nonparametric estimates of choice probabilities can be interpreted as estimates of the equilibrium selection mechanism τ^0. Therefore, this estimator avoids the search for the MLE of τ^0 by estimating it nonparametrically in a first step.

The nested pseudo likelihood (NPL) estimator is a recursive extension of the two-step PML. Given the initial nonparametric estimates of choice probabilities, the NPL generates a sequence of estimators $\{\hat{\theta}^K : K \geq 1\}$ where the *K-step* estimator is defined as:

$$\hat{\theta}^K = \arg\max_{\theta \in \Theta} Q\big(\theta, \hat{P}_1^K, \ldots, \hat{P}_T^K\big) \tag{9}$$

and the probabilities $\{\hat{P}_1^K, \ldots, \hat{P}_T^K : K \geq 2\}$ are obtained recursively as: $\hat{P}_t^{K+1} = \Psi(x_t, \hat{\theta}^K, \hat{P}_t^K)$. The algorithm iterates until convergence. For any given sample, Brower's fixed-point theorem guarantees the existence of at least one NPL fixed-point. Aguirregabiria and Mira[7] show that NPL is consistent and more efficient than the two-step PML. Monte Carlo experiments show that NPL has much better finite sample properties than the two-step estimator.

Furthermore, the NPL algorithm can be used to search for the MLE. The reason is that the MLE is a fixed point of the NPL algorithm (though not every fixed point of the NPL procedure is the MLE). To see this, notice that the MLE in (7) can be also described as:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \left\{ \begin{array}{c} \sup\limits_{P_1, \ldots, P_T} \; Q\left(\theta, P_1, \ldots, P_T\right) \\ \text{subject to: } P_t = \Psi(x_t, \theta, P_t) \text{ for every observation } t \\ P_t = P_{t'} \text{ for } x_t = x_{t'} \end{array} \right\}. \tag{10}$$

Therefore, if the NPL has a unique fixed point, this should be the MLE. Aguirregabiria and Mira[13] show that for a class of discrete choice dynamic programming models the NPL has a unique fixed point. Unfortunately, that is not the case in other models like empirical discrete games. For these models, the NPL can still be used to search for the MLE, but we need a method to search for the MLE within the set of NPL fixed points. It is in this context where the use of a Genetic Algorithm is particularly useful.

4. Hybrid Genetic Algorithm

Consider the following procedure to search for the MLE. We obtain M fixed points of the NPL procedure by applying this method to M different initial values of the choice probabilities. Given these M fixed-points of the NPL, the estimator of θ^0 will be the fixed point with the highest value of the pseudo likelihood $Q(\cdot)$. A limitation of this approach is that for some applications we will need a very large number of fixed points, M, to guarantee that this estimator is the MLE. To deal with this problem we combine the *NPL* procedure with a GA. At every step of the NPL we perform three operations on the M vectors of probabilities: crossover, mutation and selection. These operations, which are characteristic of GAs, permit a more global search over the space of $(\theta, P_1, \ldots, P_T)$. Here we describe this algorithm.

(0) *Initial population.* The initial "population" of probability vectors is $\Pi^1 = \{\hat{P}^1_{mt} : t = 1, \ldots, T; m = 1, \ldots, M\}$. This initial "population" may be arbitrarily chosen, or it may come from nonparametric estimates. For instance, the M probabilities can be obtained as M bootstrap nonparametric estimates of players' choice probabilities.

Our hybrid algorithm generates a sequence of probability vectors that we denote by $\{\Pi^K : K \geq 1\}$, where K is the index for the step or iteration of the algorithm. Associated with this sequence of probabilities the GA also generates a sequence of vectors of parameter estimates $\{\hat{\theta}^K_m : m = 1, \ldots, M;\ K \geq 1\}$. An iteration of the algorithm consists in the creation of a new generation with the offsprings of the existing generation. An iteration can be described in terms of four processes or steps that are followed sequentially: (1) mating or selection of parents; (2) crossover and mutation; (3) NPL iteration; (4) selection of offsprings.

(1) *Selection of parents.* We draw, with replacement, $O > M$ pairs of probability vectors from the population Π^K. The probability that a vector is chosen depends on its relative fitness. Fitness is a term from evolution theory. In our problem, the fitness of a probability vector is the Lagrangian function:

$$l(\hat{P}_1, \ldots, \hat{P}_T) = Q\big(\hat{\theta}, \hat{P}_1, \ldots, \hat{P}_T\big) - \lambda \sum_{t=1}^{T} \|\hat{P}_t - \Psi(x_t, \hat{\theta}, \hat{P}_t)\| \tag{11}$$

where $\hat{\theta}$ is the PML estimate associated with $(\hat{P}_1, \ldots, \hat{P}_T)$, and λ is a small and positive constant. Given the measures of fitness of the M elements of Π^K, the probability that the mth element is selected is:

$$S^K_m = \frac{\exp\big\{\sigma^{-1} l\big(\hat{P}^K_{m1}, \ldots, \hat{P}^K_{mT}\big)\big\}}{\sum_{j=1}^{M} \exp\big\{\sigma^{-1} l\big(\hat{P}^K_{j1}, \ldots, \hat{P}^K_{jT}\big)\big\}} \tag{12}$$

where $\sigma > 0$ is a parameter that measures the strength of the dependence of selection on fitness. If $\sigma = 0$, only the fittest individual is selected in the O "random" draws. If $\sigma = \infty$, every individual has the same probability of being selected.

(2) *Crossover and mutation.* Each couple generates one offspring. An offspring inherits "chromosomes" from its parents, but there maybe mutation as well. We represent this with two sets of binary variables: $\{z_t, d_t : t = 1, 2, \ldots, T\}$. z_t is the indicator of a mutation for chromosome t and it is i.i.d. over t with $\Pr(z_t = 1) = \gamma$, i.e. the mutation probability. d_t is the indicator for the identity of the parent who transmits the tth chromosome and it is i.i.d. over t with $\Pr(d_t = 1) = 1/2$. Let $\{\hat{P}_{m1}, \ldots, \hat{P}_{mT}\}$ and $\{\hat{P}_{m'1}, \ldots, \hat{P}_{m'T}\}$ be a couple. Then, the offspring from this couple is $\{\hat{P}'_1, \ldots, \hat{P}'_T\}$ where for any t:

$$\hat{P}'_t = d_t\big\{\hat{P}_{mt} + z_t\delta(\hat{P}_{mt} - U_t)\big\} + \big(1 - d_t\big)\big\{\hat{P}_{m't} + z_t\delta(\hat{P}_{m't} - U_t)\big\} \tag{13}$$

where U_t is a vector of N independent random draws from a $U(0, 1)$; and δ is a parameter that represents the magnitude of the mutation.

(3) *NPL iteration.* For each offspring we obtain its associated PML estimator of θ^0, i.e. the value of θ that maximizes the pseudo likelihood $Q\big(\theta, \hat{P}'_1, \ldots, \hat{P}'_T\big)$. Then, for each offspring $\{\hat{P}'_1, \ldots, \hat{P}'_T\}$ and its PML estimator $\hat{\theta}$, we obtain a new offspring $\{\hat{P}''_1, \ldots, \hat{P}''_T\}$ such that $\hat{P}''_t = \Psi(x_t, \hat{\theta}, \hat{P}'_t)$.

(4) *Selection of offsprings.* We calculate the fitness of each new O offsprings and select the M ones with highest fitness. This is the new population Π^{K+1}.

The algorithm iterates until convergence of the sequence of populations $\{\Pi^K\}$.

5. Conclusions

This paper proposes a new algorithm to obtain maximum likelihood estimates of structural parameters in econometric models with multiple equilibria. The algorithm combines the nested pseudo likelihood method proposed by Aguirregabiria and Mira[7,13] with a genetic algorithm. This method can be applied to the estimation of a broad class of econometric models with multiple equilibria. For instance, in the context of empirical Industrial Organization, the algorithm can be applied to obtain maximum likelihood estimates of parameters in empirical discrete games, e.g. models of market entry, spatial competition, adoption of new technologies, or investment in R&D, among many others.

Acknowledgments

This material is based upon the work supported by the National Science Foundation under Grant No. 0241943.

References

1. T. Bresnahan and P. Reiss, Empirical models of discrete games, *J. Econometrics* **48** (1991) 57–81.
2. A. Moro, The effect of statistical discrimination on black-white wage inequality: Estimating a model with multiple equilibria, *Int. Economic Review* **44** (2003) 457–500.

3. W. Brock and S. Durlauf, Discrete choice with social interactions, *Review of Economics Studies* **68** (2001) 235–260.
4. E. Tamer, Incomplete simultaneous discrete response model with multiple equilibria, *Review of Economic Studies* **70** (2003) 147–165.
5. M. Pesendorfer and P. Schmidt-Dengler, Identification and estimation of dynamic games, NBER Working Paper No. 9726 (2003).
6. V. Aguirregabiria, Pseudo maximum likelihood estimation of structural models involving fixed-point problems, *Economics Letters* **84**(3) (2004) 335–340.
7. V. Aguirregabiria and P. Mira, Sequential estimation of dynamic discrete games, CEMFI Working Paper No. 0413 (2004).
8. J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).
9. M. Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, 1996).
10. M. Mitchell, J. H. Holland and S. Forrest, When will a genetic algorithm outperform hill climbing?, in *Advances in Neural Information Processing Systems 6*, eds. J. D. Dowan, G. Tesauro and J. Alspector (Morgan Kauffman, 1994).
11. R. E. Dorsey and W. J. Mayer, Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features, *J. Business and Economic Statistics* **13** (1995) 53–66.
12. M. Beenstock and G. Szpirob, Specification search in nonlinear time-series models using the genetic algorithm, *J. Economic Dynamics and Control* **26** (2002) 811–835.
13. V. Aguirregabiria and P. Mira, Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models, *Econometrica* **70** (2002) 1519–1543.