Econometrics and Statistics 000 (2017) 1–19

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

A hyperplanes intersection simulated annealing algorithm for maximum score estimation☆

Kostas Florios

Department of Statistics, Athens University of Economics and Business, 76 Patission st., Athens GR-10434, Greece

ARTICLE INFO

Article history:
Received 20 October 2016
Revised 23 March 2017
Accepted 24 March 2017
Available online xxx

Keywords:
Maximum score estimator
Simulated annealing
Hyperplanes intersection
Local search
Comparative analysis
Metaheuristics

ABSTRACT

A new hyperplanes intersection simulated annealing (HISA) algorithm, based on a discrete representation of the search space as a combinatorial set of hyperplanes intersections, is developed for maximum score estimation of the binary choice model. As a prerequisite of the discrete space simulated annealing algorithm, also, a multi-start Hyperplanes Intersection Local Search algorithm (HILS) is devised. The implementation of the local search and simulated annealing algorithms searches the space of hyperplanes intersections combinations formulated by the regression's observations. A set of attributes that are equivalent to the hyperplanes whose intersections define potential maxima is selected as the solution representation. A swap move is introduced so that starting from an arbitrary set of attributes, nearby sets of attributes are generated and evaluated either using the steepest ascent or the Metropolis criterion. Applications include a work-trip mode choice application, for which the global optimum is known, and two labor force participation datasets with unknown global optima. Comparison is made to leading heuristic and metaheuristic approaches as well as to Mixed Integer Programming. Results show that multi-start HILS and especially HISA offer the best results for the two labor force participation datasets, and also discover the global optimum in the work-trip mode choice application.

1. Introduction

In this paper we are concerned with the estimation of the binary choice model (1)

$$y_i = 1\{X_i\beta' + \varepsilon_i \geq 0\}, \quad i = 1, 2, \dots, N, \tag{1}$$

where y_i is a binary dependent variable, X_i is a $1\times$ (K+1) vector of explanatory variables, β is a $1\times$ (K+1) vector of unknown parameters and ε_i is a random error term. The available sample of N observations $\{(y_i, X_i) : i = 1, 2, \dots, N\}$ is used to estimate β. It is assumed that $\beta = [\beta_0, \beta_1, \dots, \beta_K] \in \{-1, 1\} \times B$, where B is a compact subset of $\mathbb{R}^K$. Identification of the model also requires estimation of β vector is subject to the normalizing constraint $|\beta_0| = 1$.

Decisions that accept yes/no answers can be modeled with the binary choice model. For instance, participation or not in the labor force, commuting to work by car or not, willingness to pay a given fee in order to protect a forest etc.

Full distributional assumptions on the random error term permit estimation of β with the Maximum Likelihood approach. Frequently utilized parametric choices are the probit model when ε follows a normal distribution and the logit

☆ Fortran code to replicate the work-trip mode choice example and to use the methods for other models is available at http://www.runmycode.org/companion/view/2337.

E-mail addresses: cflorios@aueb.gr, kflorios@gmail.com

http://dx.doi.org/10.1016/j.ecosta.2017.03.005
2452-3062/

model when ε follows a logistic distribution. A third option is the complementary log-log model (Clog-log model), in which the probability of the binary event is equal to $F(z) = 1 - exp(-exp(z))$ with $z = X_i\beta'$. More recently, an increasing number of researchers have questioned the fully parametric specifications and have explored estimators of β based on less restrictive distributional assumptions.

Historically, the first semiparametric estimator proposed for the binary choice model is Manski's maximum score (MS) estimator (Manski, 1975; Manski and Thompson, 1986). Manski suggested finding the value of β that maximizes the number of correct predictions. This is attained by maximizing the so-called Score $S^{(1)}(\beta)$ objective function

$$\max_{\beta} S^{(1)}(\beta) = \frac{1}{N}\sum_{i=1}^{N}(-1+2y_i)\cdot \mathrm{sgn}(X_i\beta'). \tag{2}$$

It is assumed that $\beta = [\beta_0, \beta_1, \ldots, \beta_K] \in \{-1, 1\} \times B$, where B is a compact subset of $\mathbb{R}^K$. Identification of the model also requires maximization of the score objective function is subject to the normalizing constraint $|\beta_0| = 1$. The Score function measures the correct predictions according to the sign between $-1 + 2y$ and $X\beta'$. By the simple transformation $S^{(2)}(\beta) = (1 + S^{(1)}(\beta))/2$ the Score function can take values in [0,1] rather than in [-1,1] as above. In the experimental results we have chosen to work with the $S^{(2)}(\beta)$ form of the objective criterion.

The main distributional assumption behind the maximum score estimator is that median$(\varepsilon|X) = 0$. This is a very weak assumption compared to probit and logit. This robustness characteristic of the MS comes at some cost: The objective function $S^{(1)}(\beta)$ is (a) difficult to optimize numerically and (b) the rate of convergence to the true β is as low as $n^{1/3}$. These two reasons have kept the number of empirical applications of the maximum score estimator relatively low. For a list of empirical applications of the MS see Florios and Skouras (2008). In this paper we are concerned mainly with the computational aspects of the problem. It should be reminded that the consistency argument for the estimator depends on the global nature of the maximizer.

Florios and Skouras (2008) have proposed a technique for the exact computation of MS. This is based on the reformulation of the $S^{(1)}(\beta)$ function to a Mixed Integer Programming (MIP) model with disjunctive constraints (see Nemhauser and Wolsey (1999), p.12 for details), as follows:

$$\begin{aligned}
&\max_{z,\beta} \sum_{i=1}^{N} z_i \quad \text{st.}\\
&(1-2y_i)\cdot(X_i\beta') \le \sup_{\beta\in B}[|X_i\beta'|]\cdot(1-z_i), \quad i = 1,\ldots,N\\
&\text{and}\\
&z_i \in \{0,1\},
\end{aligned} \tag{3}$$

The above MIP problem can be solved for moderate sizes with available commercial and open-source solvers.

Model (3) is the first attempt to formalize MS computation as an optimization problem in the framework of mathematical programming and/or combinatorial optimization. While Model (3) converges in reasonable computation time ($\approx$ 5h) for a well known work-trip mode choice application (Horowitz, 1993; Florios and Skouras, 2008) to the exact mathematical optimum, in other situations (greater sample size, more estimated parameters, noisy data) the convergence of the MIP model (3) to the mathematical optimum is rather slow. Recently, there is also a growing interest in employing heuristics and/or metaheuristics in solving econometric problems, see (Winker and Gilli, 2004; Gilli and Winker, 2008). Along this trend, it is only natural to devise and apply optimization heuristics also for the maximum score estimation problem, since a Mixed Integer Programming formulation for MS has been already devised, and in real world problems its convergence is often slow. Especially interesting is the exact algorithm of Pinkse (1993) which locates the mathematical optimum for the maximum score estimator among the $C_p^N = \frac{N!}{p!(N-p)!}$ solutions of $p \times p$ linear systems which can be devised from a sample size of N observations for p estimated parameters. The main idea of the present paper is to devise first a local search (LS) and then a simulated annealing (SA) algorithm for maximum score estimator computation taking into advantage this $O(C_p^N) \approx O(N^p)$ complexity representation of the MS computation. So, instead of searching the whole of the search space involving $O(C_p^N) \approx O(N^p)$ trial points, starting from an arbitrary solution one can search only those which can be generated by the previous one through a series of transformations (or moves) that typically involve subtracting one attribute (or hyperplane) and adding a different attribute (or hyperplane). This transformation (or move) is called swap and defines a Neighborhood of size $O(p(N-p)) \approx O(N)$. Thus, searching for and selecting the best trial point in this neighborhood results in the Hyperplanes Intersection Local Search (HILS) algorithm which is a typical hill climber for the MS computation problem. On the other hand, if one samples stochastically the neighborhood and accepts new solutions via the Metropolis criterion (applying a slow cooling schedule), the obtained method results in hyperplanes intersection simulated annealing (HISA) algorithm.

The need to devise a practically feasible computation approach can be motivated by empirical applications in the issue of labor force participation. For instance Mroz (1987) and Gerfin (1996) study the determinants of married women's decision to participate or not in the labor market. Their regression models involve large sample sizes and a large number of explanatory variables, thereby rendering the estimation with MIP prohibitive.

The aim of this paper is to propose two new algorithms for the maximum score estimator computation: (a) Hyperplanes intersection simulated annealing (HISA) and (b) Hyperplanes Intersection Local Search (HILS). The effectiveness and effi-