Dumlupınar Üniversitesi Sosyal Bilimler Dergisi
Sayı 20 Nisan 2008

A Genetic Algorithm Approach To Parameter Estimation In Nonlinear Econometric Models

Harun ÖZTÜRKLER*
Şenol ALTAN**

Abstract: Genetic algorithm (GA) is a method based on the principle of evolution theory. It is widely used in stochastic optimization applications. In recent years, genetic algorithms have frequently been used in economics. The purpose of this study is to show that GA is not only used in solving optimization problems but also used as an alternative method in parameter estimation in solving nonlinear econometric models. By using a nonlinear trend model of the Turkish Statistical Institute's monthly CPI data for the period 1990.01 and 2000.10, we first estimated the parameter of an econometric model which is not linear in its parameters, and then we used GA method for parameter estimation of the same model. The results obtained from two methods are compared.

Keywords: Nonlinear econometric model, parameter estimation, genetic algorithm, consumer price index

Doğrusal Olmayan Ekonometrik Modellerde Genetik Algoritma Yaklaşımıyla Parametre Tahmini

Özet: Genetik algoritma (GA) evrim teorisi prensibi temelinde geliştirilmiş bir yöntemdir. Bu metot stokastik optimizasyon uygulamalarında yaygın biçimde kullanılmaktadır. Genetik algoritmalar son yıllarda ekonomide de yaygın biçimde kullanılmaya başlandı. Bu çalışmanın amacı GA'nın yalnızca optimizasyon problemlerinin çözümünde kullanılmadığını, ama doğrusal olmayan ekonometrik modellerin çözümünde parametre tahmininde kullanılan alternatif bir metot olduğunu da göstermektir. Biz önce 1990.01 ve 2000.10 dönemi için Türkiye İstatistik Kurumunun aylık TÜFE verilerinin doğrusal olmayan bir trend modelini kullanarak, parametrelinde doğrusal olmayan bir ekonometrik modelin parametrelerini tahmin ettik, daha sonra aynı modelin parametrelerini GA'yı kullanarak tahmin ettik. Sonuç olarak iki modelden elde edilen sonuçları karşılaştırdık.

Anahtar Kelimeler: Doğrusal olmayan ekonometrik model, parametre tahmini, genetik algoritma, tüketici fiyat endeksi

INTRODUCTION

Scholars have always been interested in solving real life problems and adapting these solutions to new problems. For this purpose, many deterministic and stochastic solution methods have been developed. However, since the deterministic solution tools necessitate certain

* Asst. Prof. Dr., Kocatepe University Department of Economics
** Asst. Prof. Dr., Gazi University Department of Econometrics

assumptions and restrictions, in some cases they are inadequate in solving real life problems. Therefore, alternative methods have been developed for the solution of complex processes. This study focuses on the usage of an alternative and widely used solution method in micro level, the genetic algorithm (GA) approach, in a macro framework.

GA is a method based on the principles of evolution theory and widely used in the solutions of stochastic optimization problems. Genetic algorithms (GAs), Evolutionary Programming (EP), and Evolution Strategies (ESs) are among the evolution algorithms used in the application stage of natural selection in stochastic optimization techniques. Today, GAs are the best known and the most used algorithms among evolutionary algorithms. Gen and Cheng (1996) declare that GAs generate better results than traditional optimization methods in solving real life problems.

NONLINEAR REGRESSION MODELS

The general form of nonlinear regression models can be written as:

$$Y = f(X,\theta) + \varepsilon \qquad (1)$$

where Y is the dependent variable, X is an (nx1) vector of independent variables, θ is a (kx1) (nonlinear) parameter vector, and ε is a random error.

One of the nonlinear regression models widely used in empirical studies is power regression model. A regression model of this type containing a single independent variable can be written as follows:

$$Y_i = \theta_0 \exp(\theta_1 * X_i) + \varepsilon_i \qquad (2)$$

where Y_i is the dependent variable, X_i is the independent variable, ε_i is a stochastic error term, and θ_0 and θ_1 are the parameters of the model.

According to Rawlings (1988), when compared with linear models, nonlinear models reflect real situations better, and it is possible to characterize a model's functional form with fever parameters. Therefore, nonlinear models are preferred to linear models. As in linear models, in estimating the parameters of a nonlinear model, least square error or maximum likelihood approach can be used. As Aksoy (1996) emphasizes, in the empirical work, error term should be a normally distributed independent stochastic variable with constant variance. The solution space for the nonlinear models is a curvature space for all possible values of parameter vector. On the other hand, curvature space of parameters in the linear models is approximated by a plane. Therefore, parameter estimation in nonlinear models is more challenging than it is in linear models. Since the deterministic methods used in linear models are not sufficient in the estimation of the parameters in nonlinear modes, iterative numerical methods must be used.

PARAMETER ESTIMATION IN NONLINEAR MODELS

In least square method, by minimizing S in equation (3) one can get the parameters of nonlinear regression model. In contrast with linear models, analytical solution methods are not sufficient in solving the parameters of nonlinear models, and therefore, we need to employ iterative numerical search methods.

$$S = \sum_{i=1}^{n} [Y_i - f(X_i, \theta)]^2 \tag{3}$$

In order to obtain the normal equations for the nonlinear regression model given by equation (1), we applied least square criteria. This was done by taking derivative of S in equation (3) with respect to θ.

GENETIC ALGORITHM

According to Goldberg (1989), GAs are based on the mechanism through which stochastic search technique is employed together with natural selection and genes. GA technique differs from traditional search techniques in that it starts with a randomly selected initial solution called population.

Operation of Genetic Algorithm Process

Yeniay (1999) establishes the following stages in solving problems by means of GA:

Stage 1: Before applying GA procedure, an appropriate set of codes compatible with
the nature of the problem is determined.

Stage 2: A randomly selected initial population is formed.

Stage 3: Combination value of each string in the initial population is calculated.

Stage 4: In order to change the population and create new generation, crossover and
mutation operators are used.

Stage 5: The new population is evaluated and genetic algorithm procedure is carried
out until the best solution value is reached.

Using Genetic Algorithms for Parameter Estimation

Chatterjee and Laudato (1997) mention the following problems for which GA is used as a tool for estimating parameters:

a) Deterministic problems,
b) Simple least square problems with single independent variable,
c) Nonlinear least square problems,
d) Linear regression problems in which parameter estimation is performed by minimizing absolute deviations of residuals, and
e) Multiple regression problems.

In addition to the problems mentioned above, GAs can be used in dimension reduction in linear regression and best subset selecting problems. The common feature of the results obtained from the GA applications is that they are similar to the results attained by using traditional techniques. These studies also show the power of GA, a heuristic approach, and the independence of GA code from the model chosen.

The Difference Between Genetic Algorithm and Direct Search Methods

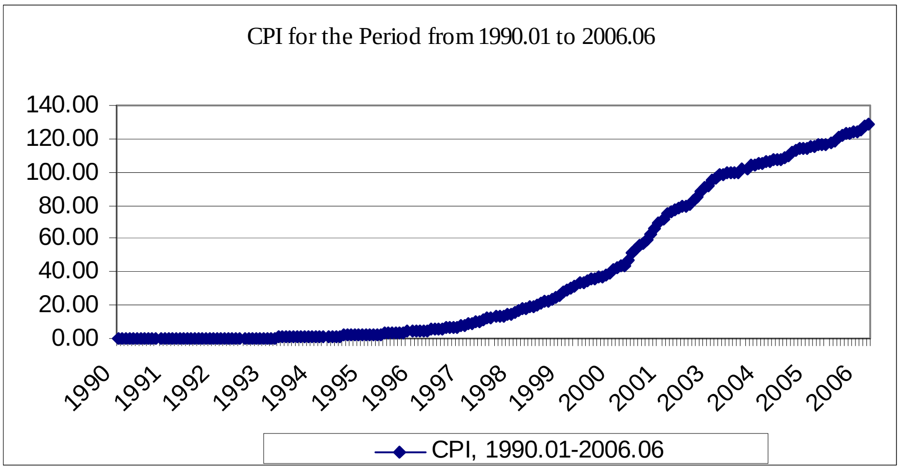
GA is also efficient in searching solutions to problems. GA can reach optimal solutions or solutions close to optimal solutions among the solution sets. Goldberg (1989) emphasizes the fact that by applying genetic operators step by step to produce new generations in an appropriate population, GAs can lead to the best solutions. Furthermore, without requiring assumptions, GA can attain proper solutions by scanning solution space from many different starting points.

EMPIRICAL IMPLEMENTATION

In this study, by using the Turkish Statistical Institute's monthly CPI (2003=100) data for the period 1990.01 and 2000.10, we estimate the parameter of an econometric model which is not linear in its parameters. We first take the logarithmic transformation of the model, and then we calculate the least square estimation of the parameters in Microfit. We then take these parameters as the initial solution values required for the estimation of nonlinear power model. Finally, the nonlinear model is solved by using GA and the results of the two solutions are compared.

The Structure of the Model Considered and the Data

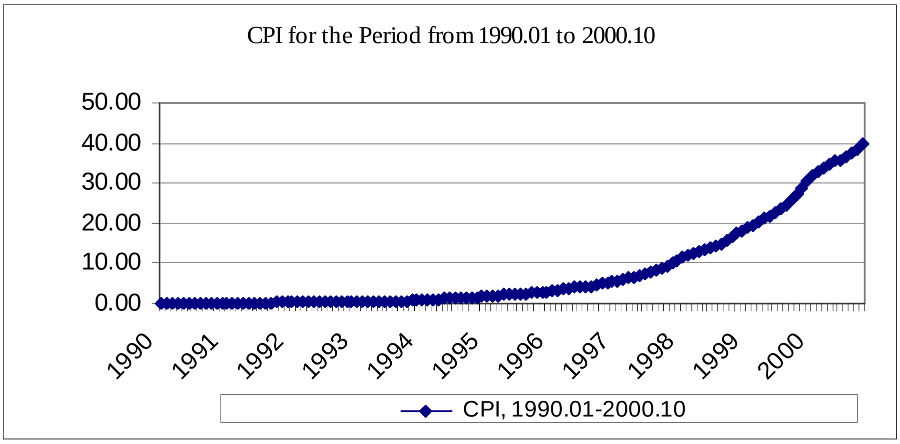
A visual plot of the data is usually the first step in the analysis of any time series. In order to detect an appropriate model we draw the graph of CPI given in Graph 1 below. An investigation of the CPI data for the period 1990.01-2006.06 reveals that a power growth curve can represent the data. However, a careful examination of Graph 1 also demonstrates that the radical structural policy change following the deep economic crises of 2000/2001 disturb the power growth shape of the data. On the other hand, Graph 1 shows that for the period 1990.01-2000.10 a power growth model fits the data.



Graph 1: CPI, 1990.01-2006.06

It is worth to remind that the purpose of this study is not to compare economic policies and their effects on consumer price index before and after the economic crises that occurred in November 2000-Februarry 2001. The plot of data studied is provided in Graph 2 below.

Graph 2 reveals clearly that a power growth curve can represent the CPI data for 1990.01-2000.10 period fairly well. Furthermore, the series has a deterministic structure with continuous increases. Therefore, we chose power growth curve as the model to represent the data.



Graph 2: CPI, 1990.01-2000.10

Then, we can write the model as;

$$y_t = \theta_0 e^{\theta_1 t} + \varepsilon_t \qquad (4)$$

where y_t is monthly CPI series, t is time (t = 1,2,…,130), ε_t is the random error term, and $\boldsymbol{\theta_0}$ ***and*** $\boldsymbol{\theta_1}$ are the parameters of the model.

Parameter Estimation of the Nonlinear Regression Model

The model given in equation (4) is nonlinear in its parameters. Because the error term is additive, the model cannot be transformed into a linear model. Therefore, in estimating parameters, we use nonlinear least square technique. As underlined by Zeltkevic (1998), we need starting values for being able to apply the proposed method. However, it is not easy to choose an initial value. Therefore, several trials need to be done. Table 1 provides some examples from such trials. By taking into account the fact that choosing initial values is not easy; in order to be able to find initial values required for the estimation of the parameters, we assume error term of power growth model as being multiplicative. Therefore, by the logarithmic transformation model can be turned into logarithmic form and least square method can be applied. We choose the resulting parameter estimations (θ_0 = 0.083014 and θ_1 = 0.049106) as the initial values. The nonlinear model can now be written as follows:

$$\underset{(0.01)}{y = 0.17979}\, \underset{(0.00047)}{e^{0.042198\, t}} \qquad (5)$$

R^2 = 0.99411, DW = 0.070758 Residual Sum of Squares (RSS) =90.6673

where the figures in parentheses are the estimated standard errors.

Table 1: A Sample of Results of Trials Run to Determine the Initial Values

Initial Parameter Values		Estimated Parameter Values			
θ_0	θ_1	θ_0	θ_1	RSS	Number of Iterations
0.17	0.020	0.0000000006134	0.224530	2.25E+7	3
	0.025	0.000002401	0.156910	1.04E+07	41
0.08	0.010	0.000002461	0.156710	1.04E+07	41
0.08	0.015	Initial value is not appropriate for solution			
0.08	0.020	Initial value is not appropriate for solution			
0.08	0.025	0	0.248940	2.75E+07	3
0.08	0.030	0.000002444	0.156760	1.04E+07	41
0.08	0.035	0.17979	0.042198	90.6673	10
0.08	0.040	0.17979	0.042198	90.6673	6
0.083014	0.049106	0.17979	0.042198	90.6673	6

We have determined that there is not a heteroscedasticity problem in the model, but error terms have autocorrelation. However, this study does not focus on the methods for solving autocorrelation problem because the advantage of using genetic algorithms is that it can produce the solution of the model without being constrained by model’s assumptions and theoretical restrictions. That is, existence of autocorrelation does not restrain the proceedings of the study with GA.

Estimation of the Parameters of the Power Growth Model Using Genetic Algorithm

We wrote a program in Matlab 7.2 and used Matlab 7.2’s GA tools menu in order to estimate the parameters using GA. The parameters in GA tools menu are chosen as follows:

Population Type: Double Vector, Population Size: 20, Creation Function: Uniform, Scaling Function: Rank, Selection Function: Stochastic Uniform, Crossover Fraction: 0.80, Mutation: Adaptive Feasible, Crossover Function: Heuristic, Ratio: 2, Migration Direction(Forward, Fraction:1.0, Interval: 20), Algorithm Settings(Initial Penalty: 10, Penalty Factor: 100), Hybrid Function: None, Stopping Criteria(Generations: 1000, Time Limit: lnf, Fitness Limit: -lnf, Stall Generations: 1000, Stall Time Limit: 20, Function Tolerance: 1e-006, Nonlinear Constraint Tolerance: 1e-006).

Population type and population size are determined as double vector and 20, respectively. On the other hand, initial population and starting scores are created randomly.

When we analyze the data we realize that increases in CPI are in decimal form while increases in time are one by one. Therefore, options seen as appropriate for GA must be used for both parameters. That is, the determined crossover strategy will work for both parameters with the same characteristics. However, the sensitivities of parameters over time are very different from each other. Therefore, it is crucially important to determine appropriate strategies without loosing the sensitiveness of the parameters. To be precise, the strategy must be chosen by considering choice, crossover, and mutation options together. In this study, rank choice is used.

In this study, in determining crossover and mutation operators, we used a strategy similar to choice strategy. Accordingly, mutation operator is chosen as adaptive feasible. When we examine the CPI data where large changes occur, we observe that GA finds the regions unchanging over generations and mutates them.

As mentioned above, GAs work independently of assumptions required by the model. Accordingly, we can write the model with the results obtained from GAs as follows:

$$y_t = 0.1798e^{0,0422t} \tag{6}$$

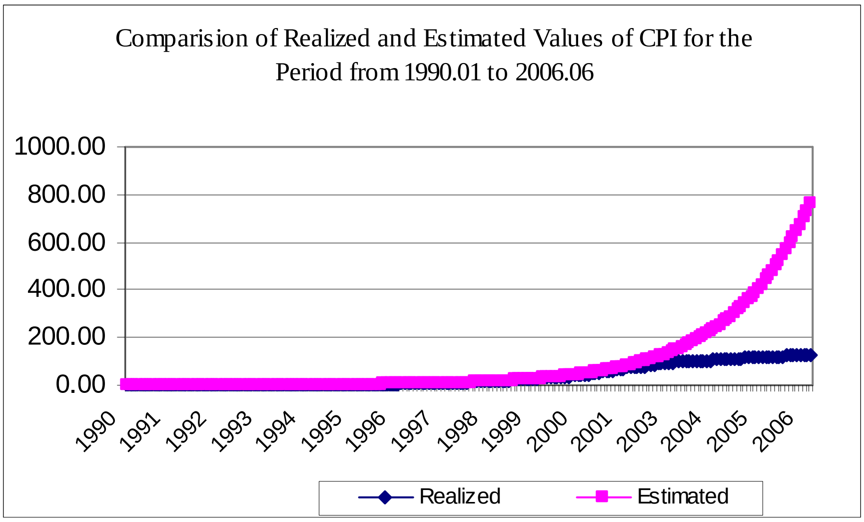
where Fitness Function is 90.66. Fitness function corresponds to residual sum of squares value in nonlinear model solution with Microfit program.

CONCLUSION

Genetic algorithms are used as an efficient solution tool in solving problems which are hard or impossible to be solved by deterministic solution methods. The purpose of this study is to show whether GA can be used as an alternative solution tool.

The study shows that for nonlinear least square solution, one needs initial values, which are hard to determine. However, GAs can produce the solution without requiring initial solutions by searching from many search points simultaneously. GAs, working independently of the assumptions of the models, can produce solutions close to solutions accepted as optimal solutions. This characteristic of GAs shows its difference with deterministic solution methods and leads to its usage increasingly in different fields. The difference between the solution of the nonlinear model parameters with Microfit and the model parameters obtained from GA solution is a result of cognitive characteristic of GA method. GA does not guarantee the optimum solution, but leads to solutions acceptably close to the optimal solution.

When the estimated and realized values of CPI are compared for the period studied, it is easily observed that these values are particularly close to each other for 1990.01-2000.10 period, while estimated values deviate from realized values upward following the 2000/2001 crises. Therefore, we can say that power growth model is not appropriate for CPI data after 2001.



Graph 3: Comparison of realized and estimated CPI values for the period from 1990 to 2006.

Finally, when nonlinear least square and genetic algorithm approaches are compared on the basis of power growth model of monthly CPI data for Turkey for the period from 1990 to 2006, we can suggest genetic algorithm approach as a tool for solving econometric problems involving complex nonlinear relations.

REFERENCES

Aksoy, S. (1996), Otokorelasyonlu Hata Terimli Doğrusal Olmayan Regresyon Modellerinde

Parametre Tahmini, (Yayınlanmamış Doktora Tezi), Gazi Üniversitesi, Ankara.

Chatterjee, S., and Laudato, M. (1997), "Genetic Algorithms in Statistics", Common Statist.

Simula., 26, 1617-1630.

Gen, M., and Cheng, R. (1996), Genetic Algorithms and Engineering Design, Wiley, New York.

Goldberg, D. E.(1989), Genetic Algorithms in Search, Optimization and Machine

Learning, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

Rawlings, J.O. (1988) Applied Regression Analysis, Wadsworth and Brooks / Cole

Advanced Books and Software, California.

Yeniay, .M.Ö.(1999), Taguchi Deney Tasarımı Problemlerine Genetik Algoritma Yaklaşımı,

Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Zeltkevic, M. (1998), "Nonlinear Models and Linear Regression"

http://web.mit.edu/on.001/Web/CourseNotes/StatisticsNotes/Correlation/node6.html