Iran Agricultural Research (2015) **34(2)** 101-107

Shiraz University

A comparison of genetic algorithm and auto -regressive distributed lag model in determination of total factors productivity growth in the agricultural sector of iran

S. Negarchi[1*], M.R. Zare Mehrjerdi[1], H. Mehrabi Boshrabadi[1], H. Nezamabadi Pour[2]

[1]Department of Agricultural Economics, ShahidBahonar University of Kerman, Kerman, I. R. Iran.
[2]Department of Electrical Engineering, ShahidBahonar University of Kerman, Kerman, I. R. Iran.

* Corresponding Author:negarchi@gmail.com

ARTICLE INFO

Article history:
Received **10 November 2012**
Accepted **7 December 2015**
Available online **9 January 2016**

Keywords:
Genetic Algorithm
Auto-Regressive Distributed Lag
Total Factors Productivity
Agriculture
Iran

ABSTRACT- Due to the important role productivity plays in future decision making and programming, the productivity indexes should have accurate quantities. In this study, Auto-Regressive Distributed Lag (ARDL) and Genetic Algorithm (GA) methods are applied to time series of 1978-2008 to accurately measure total factor productivity (TFP) in the agricultural sector of Iran. The comparison of these two methods shows that GA method is more efficient than ARDL model. Also, the growth of TFP in the agricultural sector of Iran has had high fluctuations and annual average of productivity growth in this sector has been -0.16 during the period of the study. Therefore, it is necessary to emphasize the optimum use of available inputs, their appropriate combinations and increasing productivity in the agricultural sector of Iran.

INTRODUCTION

Improvingagriculturalproductivityhasavitalrole in the process of production and is crucial to economic development and can provide substantialguide for Iranian planners and policymakers. Total factor productivity (TFP) is an index that can be taken as a measure of an economy's long-run technological variation or dynamism. However, it cannot be measured directly. In most studies, production function technique is used to achieve the growth of TFP. The production function coefficients affect the accuracy of TFP growth. Therefore, production function should be estimated with a more precise method.

The aim of this study is to obtain the exact amount of TFP growth with the best and most accurate production function. For this purpose, econometric and heuristic algorithm methods are compared.

Over the last decades, engineers have tried to invent heuristic algorithms. These algorithms are good replacement tools to solve complex computational problems. Various heuristic approaches including genetic algorithm and tabu search have been adopted by researchers (Goldberg, 1989;Ceylan and Ozturk, 2004;HaldenbilenandCeylan, 2005;Ozturk et al., 2005). These methods are numerical methods that have been widely appliedin optimization problems during a short time. Genetic Algorithm (GA) is an optimization procedure based on the principles of natural selection (Goldberg, 1989;Holland,1992). Recently, GA is adopted by researchers around the world in estimation of economic function. Using genetic algorithm and Particle Swarm Optimization (PSO) algorithm, both linear and non-linear models were applied by Amjadi et al. (2010) to estimate electricity demand and provide ability of each algorithm for the matching and prediction of demand value with lowest error rate in the future. Ozturk et al. (2005) developed an electricity estimation model using the GA notion for industrial sector electricity consumptions based on the basic indicators of the gross national product, population, import and export figures.Ceylan and Ozturk (2004) developed the energy estimation using GA approach for energy consumption of the residential–commercial sectors and examined the effect of the design parameters on the energy consumption of the sectors. An agent-based model of multifunctional agricultural landscape was developed using GA for the Cache River basin in southern Illinois by Sethuram et al. (2008). Simmons and Cacho(1999) examined farm investment using a GA model in Australia.

MATERIALS AND METHODS

In this paper, we made use of the technique of production function to achieve the growth of total factors productivity. The production functions are a mathematical formalization of the relationship between the output and the inputs. The aggregate agricultural production function takes the Cobb-Douglas form, which is the most common specification used in studies. This function is easy to analyze, and appears to be a good approximation to productions (Romer, 2001). The data used in this study were country level agricultural output and inputs in estimation of the Cobb-Douglas production function of Iranian agriculture.

The dependent variable is a value added in agricultural sector. The main independent variables include three essential agricultural inputs; labor, capital and energy. In our analysis, the agricultural production function (Cobb-Douglas) is specified as follows:

$$VA = a.L^{\beta1}K^{\beta2}E^{\beta3} \quad (1)$$

In which VA is the value added and L is labor and K is capital stock and E is energy used in agricultural sector. The coefficients β_i (i=1,2,3) are the elastic ties of the respective variables with respect to agricultural production, with the assumption that β_i>0.

By taking log in both sides, Eq (1) would be as follow:

$$\ln VA = \ln a + \beta_1 \ln L + \beta_2 \ln K + \beta_3 \ln E \quad (2)$$

Total factor productivity (TFP) was measured using the production function approach. TFP growth is usually measured by the Solow (1956) residual model. This model is shown in Eq.3.

$$T\hat{F}P = \hat{v} - \alpha\hat{k} - \beta\hat{l} \quad (3)$$

In calculating TFP growth in Iran's agriculture sector,Eq(3) will convert to the following Eq.4:

$$T\hat{F}PA = \hat{q}_t - \alpha\hat{k} - \beta\hat{l} - \gamma\hat{e} \quad (4)$$

Where, $TF\hat{P}A$ is productivity growth rate and $\hat{e},\hat{l},\hat{k},\hat{q}_t$ are output growth rate (value added), capital, labor and energy in agriculture sector respectively. *α, β, γ*are input's coefficients that are same coefficient of production function.

In this paper, GA and ARDL are used to solve the problem and compare their ability to find the best solution. The 25-year available data from 1978 to 2003 are used to estimate weighting factors in both models. Then, the developed models are validated with the available actual data from 2004 to 2008. Required data are received from Iran's Annals, Central Bank of Islamic Republic of Iran and balance sheet of the department of energy (based on the price of year 1997).

Ardlmodel

In estimating a long-run and short-run economic function, there are various time series methods as cointegration, ECM and vector auto regression (VAR). In recent years, ARDL model has become increasingly popular in estimating economic function models due to the new approach developed by Pesaran and Shin (1999) and Pesaran and et al. (2001). An ARDL model is a general dynamic specification that uses the lags of the dependent variable and the lagged and contemporaneous values of the independent variables, through which the short-run effects and the long-run equilibrium relationship can be respectively estimated directly and indirectly.

The ARDL approach involves two steps for estimating the long-run relationship. The first step is to examine the existence of a long-run relationship among all variables in the equation under examination. Conditional upon cointegration is confirmed; in the second stage, the long-run and the short-run coefficients are estimated using the associated ARDL and ECMs (Atkins and Coe, 2002). To test cointegration in model (2) by the bounds test proposed by Pesaran et al.(2001), the following conditional ECM model, which is a variant of parameterization of the ARDL model, is constructed:\

$$\ln VA = \alpha_0 + \alpha_1 \ln VA_{t-1} + \alpha_2 \ln L_{t-1} + \alpha_3 \ln K_{t-1} + \alpha_4 \ln E_{t-1} + \beta_i \sum_{i=1}^{p} \Delta \ln VA_{t-i} + \gamma_i \sum_{i=0}^{p} \Delta \ln L_{t-i} + \delta_i \sum_{i=0}^{p} \Delta \ln K_{t-i} + \theta_i \sum_{i=0}^{p} \Delta \ln E_{t-i} + \varepsilon_t \quad (5)$$

For the bounds test, two separate statistics are employed to test the existence of a long-run relationship: an F-test for the joint significance of the coefficients of the lagged levels in Eq(3) , i.e., H_0 :$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$and a t -test for the null hypothesis of H_0 : $\alpha_1 = 0$.

After estimating the production function by using econometric method, production function will be estimated by applying genetic algorithm method, too.

Genetic algorithm

Genetic algorithm (GA) is based on the principles of Darwin's evolution theory. This algorithm is a tool for solving the optimization problems. Initial population is made of a set of chromosomes using a random process and each chromosome has two possible values, 0 and 1.The values of parameters are considered to be genes, which can be inherited by mating.

The fitness function is the index to evaluate the fitting ability of the chromosome. Selection of individuals is based on the fitness of the individuals with respect to an objective function. Individuals with high fitness values will have a higher chance of entering the mating population. Then, individuals with the best genes will have better chances of survival and mating.

Cross-over is done by mating and exchanging or recombining genes in the offspring. Next operation is called Mutation. This operation changes some randomly chosen genes. After cross-over and mutation, individuals will be selected for the next mating population. The process of selection, cross-over and mutation are repeated until an optimal solution is found. The structure of the algorithm in the optimization of the upper model is shown in fig. 1 (Amjadi et al., 2010).

The first step in the solution of optimum problems by GA is determination of initial parameters for this algorithm. The purpose of optimization is obtaining the weighting factors of the parameters defined in (2) with GA by optimizing the following objective or fitness function:

$$e = Y - F(x_i) \quad (6) \qquad \text{fitness} = \frac{1}{MSE} = \frac{1}{\left(\frac{\sum e^2}{n}\right)} \quad (7)$$

In the above function, $F(x_i)$, is considered production function and Y is value added to the production, MSE stands for Mean Square Error, and n is the number of observations. There are 4 weights in the considered model. Therefore, GA in logarithmic model has 4 genes. In this research, for the production of the new generation, crossover and mutation methods are used.

This process is repeated until the best answer is achieved.

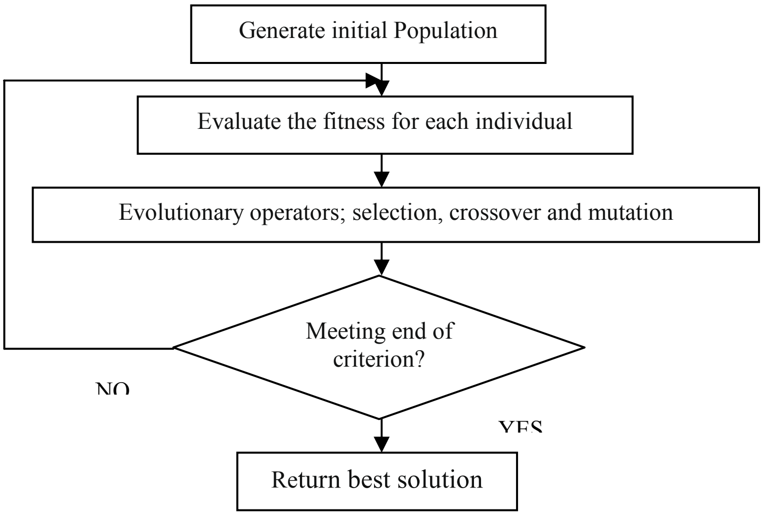

Fig. 1. General principle of GA.

Evaluation

In order to evaluate the performance of the two models, we can use related criteria. In this study, criteria of root error of mean square average (RMSE) and determination coefficient (R^2) are used. These statistics are calculated by Eqs 8 and 9:

$$RMSE = \sqrt{\sum_{k=1}^{N} (X_K - Y_K)^2 / N} \quad (8)$$

$$R^2 = \left[\frac{\sum_{K=1}^{N} (X_K - \overline{X})(Y_K - \overline{Y})}{\sum_{K=1}^{N} (X_K - \overline{X})^2 \sum_{K=1}^{N} (Y_K - \overline{Y})^2} \right]^2 \quad (9)$$

A model having the least value of RMSE is the best one. Also, a model that includes the most value for R^2 is better than other models.

RESULTS AND DISCUSSION

The application of the traditional methods in econometrics is based on the assumption of stationary of variables. Therefore, it is necessary to be sure of stationary of variables. For this purpose, Augmented Dickey – Fuller test is used and its results are shown in Table 1. The optimum order of ARDL, determined by the information criteria such as SBC, is found to be in order of (1, 0, 0, 0) according to Table 2.

Table 1. Unit root test results.

Variable	Stationary
Ln VA	I(1)
Ln L	I(1)
Ln K	I(1)
Ln E	I(0)

The Null hypothesis (the absence of a long- runs relation) studied by using of F-statistic. F-statistic is greater than the upper bound of the critical value obtained from Pesaran and et al. (2001); therefore, there would be a long-run relation between the variables of the model. Results from estimation of long-run relation are illustrated in Table 3.

Table 2. Estimation of the dynamic model, ARDL (1,0,0,0).

Variable	Coefficient	Std. error	t-Statistic
C	1.2875	1.9521	0.6595
Ln VA(-1)	0.5029	0.1292	***3.8916
Ln L	0.5580	0.4080	*1.3677
Ln K	0.1501	0.0990	*1.5151
Ln E	0.3780	0.1058	***3.5707
	R^2=0.98	F=494.045[0.00]	

*Significant at the 15 percent level.
***Significant at the 1 percent level

The long-run static solution of the estimated ARDL (1, 0, 0, 0) model is presented below:

$$Ln\ VA = 2.5904 + 1.1227\ Ln\ L + 0.3020 Ln\ K + 0.7606 Ln\ E \quad (10)$$

The short-run dynamics of the demand model, presented in Table 4, is estimated by the error correction representation of the ARDL (1, 0, 0,0) specification.

The estimated coefficient of the error correction term (-0.49) indicates fast speed of adjustment to equilibrium following a shock. In other words, nearly 49 percent of disequilibrium of the current year's shock converges back to the long-run equilibrium within the next two years.

Table 3. Long-run relation ARDL(1,0,0,0).

Variable	Coefficient	Std.error	t-Statistic
C	2.5904	4.0579	0.6383
Ln L	1.1227	0.6999	*1.6041
Ln K	0.3020	0.1949	*1.5495
Ln E	0.7606	0/0725	***10.4793

*Significant at the 15 percent level.
***Significant at the 1 percent level

Table 4. The ECM regression results

Variable	Coefficient	Std. error	t-Statistic
Dc	1.2875	1.9521	0.6595
dLn L	0.5580	0.4080	*1.3677
dLn K	0.1501	0.0990	*1.5153
dLn E	0.3780	0.1058	***3.5707

RESET test=0.041544[.838] Serial correlation (χ^2)=0.89021 [0.345]
Heteroscedasticity (χ^2)=.64961 [.420]

*Significant at the 15 percent level.
***Significant at the 1 percent leve

The ECM regression results in Table 4 seem to fit quite well and pass the diagnostic tests against serial correlation, heteroscedasticity, and functional form misspecification (RESET). The stability in the coefficients of the estimated model was also checked by using the cumulative sum (CUSUM) and cumulative sum of squares (CUSUMSQ) stability tests that utilize recursive residuals. The plots of CUSUM and CUSUMSQ statistics, presented in Fig. 2, are within the 95 percent critical bounds, representing that all coefficients in the estimated ECM model are stable over the sample period.

The GA model was run under the assumptions described above. Fig. 3 shows convergence trend to optimum answer well. Since they approach together in the last generation, the upper diagram shows the best answer up to that generation and the lower diagram shows average of the best answer up to that generation.

In genetic algorithm, by going from one generation to the next one, the objective function value is improved. According to the proposed model, the goalisto minimize the objective function. So, we expect that the chart has decreasing procedure after each generation,but the fitness function is the inverse of the objective function and the chart has increasing procedure. Thisshows that the algorithm operators work well.

Genetic algorithm was introduced by fitness function and we achieve to choose the best coefficient for production function by using considered parameters in Table 5.

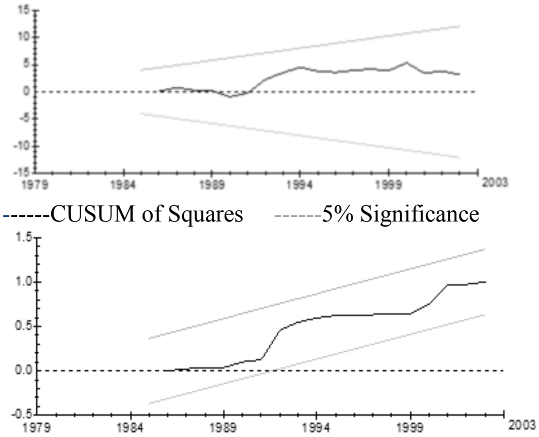

Fig. 2. CUSUM and CUSUMSQ plots for the estimated ECM model.

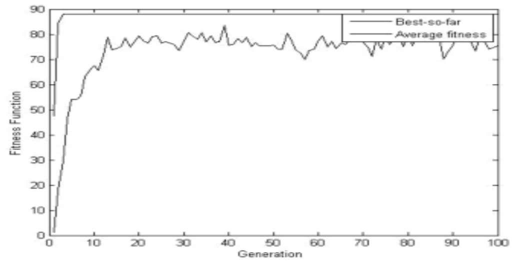

Fig. 3. The convergence trend of the algorithm in finding out the optimum answer.

Table 5. Parameters used in genetic algorithm.

Variable	Value
Population	100
crossover probability (p_c)	0.9
mutation probability (p_m)	0.005

The number of iterations is considered to be 100, which is the stopping criterion. The function obtained from genetic algorithm is as follow:

$$\text{Ln VA} = 2.5610 + 0.6471 \text{ Ln L} + 0.7702 \text{ Ln K} + 0.4369 \text{ Ln E} \quad (11)$$

In order to evaluate and choose the best estimated model, RMSE and R^2 criteria were examined to survey these methods. For this purpose, we made use of the data of the last 5 years. Table 6 shows the values of the two statistics for two methods of ARDL and described genetic algorithm.

Table 6. Comparison of ARDL method and genetic algorithm.

	ARDL	GA
RMSE	2.4890	0.1745
R^2	0.8215	0.9113

The criteria show that the genetic algorithm method has the least value for statistics RMSE after elimination. The value of R^2 in this method is also more than that of the other methods. Therefore, the best method for estimating the production function is the genetic algorithm method.

Now, by using production function estimated by the genetic algorithm method, we calculate the growth of TFP in the agricultural sector of Iran by using Eq(12) and the obtained result is presented in Table 7.

$$\widehat{TFP} = \hat{va}_t - 0/77028\hat{k} - 0/64711\hat{l} - 0/43695\hat{e} \quad (12)$$

Also, Fig. 4 shows the changing trend of TFP in the agricultural sector of Iran(The growth of TFP). On the base of these results, the growth of TFP in Iran's agriculture sector has had great fluctuations during the studied period and the productivity growth in this sector shows a negative rate of -0/16% annually for the period of 1978 to 2008. In Tahami Pour and Shahmoradi's (2008) study, the average annual TFP growth rate of agricultural sector of Iran in 1967-2003 was -3.08%, too. It should be noted that the productivity performance has deteriorated over time. Therefore, more attention should be paid to improving production productivity in the agricultural sector of Iran.

Table 7. Rate of TFP growth in agriculture sector of Iran.

Year	TFP growth	year	TFP growth	Year	TFP growth
1978-79	2.5737	1988-89	0.577401	1998-99	1.178829
1979-80	1.088022	1989-90	-2.60205	1999-00	-10.2141
1980-81	0.482979	1990-91	11.42556	2000-01	-3.17343
1981-82	-3.6539	1991-92	-5.49313	2001-02	-7.43593
1982-83	2.161844	1992-93	-2.21007	2002-03	9.171579
1983-84	-2.38983	1993-94	4.570034	2003-04	-4.88403
1984-85	4.275789	1994-95	-3.64933	2004-05	-9.18204
1985-86	3.540737	1995-96	1.347091	2005-06	-2.79972
1986-87	15.72112	1996-97	-2.3008	2006-07	1.373016
1987-88	-0.06699	1997-98	-4.21277	2007-08	-0.15294

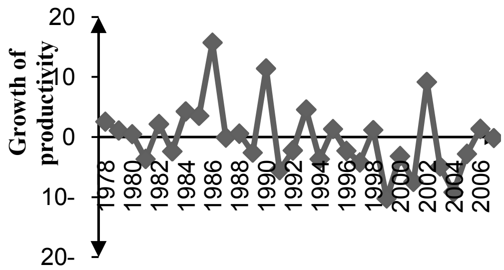

Fig. 4. The growth of TFP in agriculture sector of Iran.

CONCLUSIONS

This study was performed to comparethe performance of ARDL and genetic algorithm methods in order to measure the growth of TFP in the agricultural sector of Iran. Results imply that the genetic algorithm method has had high ability for estimating the production function. Therefore, the genetic algorithm method is used as accurate instrumentation for the optimization of the production function and obtaining more accurate values of productivity besides other methods. Due to the ability of the genetic algorithm method in optimizing economic functions, this method is recommended by agencies responsible for planning country's macro economy policy. Also, the growth of TFP in Iran's agriculture sector has had high fluctuation and annual average of productivity growth in this section was -0/16 during the studied period. The negative growth of productivity that is the same as more growth of inputs rather than product growth , implies that the problem of agriculture sector does not only result from the shortage of production inputs but also non-optimum use of inputs and their inappropriate combination are the basic problems of agriculture factor.

\

REFERENCES

Atkins, F.J., & Coe, P.J. (2002). An ARDL bounds test of the long-run Fisher effect in the United States and Canada, *Journal of Macroeconomics*, 24, 255–266.

Amjadi, M.H., Nezamabadi Pour, H., & Farsangi, M.M. (2010).Estimation of Electricity Demand of Iran Using Two Heuristic Algorithms.*Energy Conversion and Management*, 51, 493-497.

Central bank of Islamic Republic of Iran, Report and statistics; 2007 [in Pershian].

Ceylan, H., & Ozturk, H.K. (2004). Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. *Energy Converse Management*, 45, 2525–37.

Chen, P.C., YU, M.M., Chang, C.C., & HSU, S.H. (2008). Total factor productivity growth in China's agricultural sector, *China Economic Review*, 19, 580–593.

Energy balance-sheet: power ministry of Iran, Energy report and statistics; 2006 [in Persian].

Goldberg, D.E. (1989). Genetic Algorithm in Search, Optimization and Machine Learning.Addison-Wesley.

Haldenbilen, S., & Ceylan, H. (2005). Genetic algorithm approach to estimate transport energy demand in Turkey, *Fuel and Energy*, 46, 193-204.

Hamamoto, M. (2006).Environmental regulation and the productivity of Japanese manufacturing industries.*Journal of Resource and Energy Economics*, 604, 14-25.

Holland, J.H. (1992). Adaptation in natural and artificial systems. Cambridge, MA: MIT Press. (First edition, 1975, University of Michigan Press.)

Ozturk, H.K., Ceylan, H., Canyurt, O.E., & Hepbasli, A. (2005). Electricity estimation using genetic algorithm approach: a case study of Turkey. *Energy*, 30, 1003–12.

Pirasteh, H. (2003). The contribution of agriculture to economic and productivity growth of Iranian economy. Journal Iranian Economic Review, University of Tehran, *Faculty of Economic*, 8, 45-72.

Pesaran, M.H., Shin, Y., & Smith, R.J. (2001).Bounds testing approaches to the analysis of level relationships.*Journal Apply Econometr*, 16, 289–326.

Pesaran, H.M., & Shin, Y. (1999). Autoregressive distributed lag modeling approach to cointegration analysis. In: Storm S, editor. Econometrics and economic theory in the 20th century: the Ranger Frisch centennial symposium. Cambridge University Press; [chapter 1].

Romer, D. (2001).Advanced Macroeconomics. Shanghai University of Finance & Economics Press, 5–17.

Simmons, P., & Cacho, O. (1999). a genetic algorithm approach to farm investment.the*Australian Journal of Agricultural and Resource Economics*, 43,3,305-322.

Sethuram, S., Girmay, M., Steven, K., Jeffrey, B., & Lant Christopher.(2008). An agent based model of multifunctional agricultural landscape using genetic algorithms. The American agricultural economics association annual meeting, Orlando, fl, July 27-29.

Solow, R. (1956). A Contribution to the Theory of Economic Growth.*Quarterly Journal of Economics*, 70, 65-94.

تحقیقات کشاورزی ایران (۱۳۹۴) ۳۴(۲)۱۰۱-۱۰۷

مقایسه الگوریتم ژنتیک و روش خود توضیح با وقفه‌های گسترده در راستای تعیین رشد بهره‌وری کل عوامل تولید بخش کشاورزی ایران

سمانه نگارچی*[۱]، محمد رضا زارع مهرجردی[۱]، حسین مهرابی بشرآبادی[۱]، حسین نظام‌آبادی پور[۲]

[۱]بخش اقتصاد کشاورزی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان، کرمان، ج. ا. ایران
[۲]بخش برق، دانشکده فنی، دانشگاه شهید باهنر کرمان، کرمان، ج. ا. ایران

*نویسنده مسئول

اطلاعات مقاله

تاریخچه مقاله:
تاریخ دریافت: ۱۳۹۱/۸/۲۰
تاریخ پذیرش: ۱۳۹۴/۹/۱۶
تاریخ دسترسی: ۱۳۹۴/۱۰/۱۹

واژه های کلیدی:
الگوریتم ژنتیک
خود توضیح با وقفه‌های گسترده
بهره‌وری کل عوامل تولید
کشاورزی
ایران

چکیده– امروزه دستیابی به رشد اقتصادی از راه ارتقای بهره‌وری، از مهم‌ترین هدف‌های اقتصادی کشورها بشمار می‌آید با توجه به نقش مهم بهره‌وری در تصمیم‌گیری‌ها و برنامه‌ریزی‌های آینده، باید مقادیر دقیقی از این شاخص در اختیار داشت. در این مطالعه به منظور اندازه‌گیری دقیق بهره‌وری کل عوامل تولید در بخش کشاورزی، از دو روش ARDL و الگوریتم ژنتیک طی دوره زمانی ۸۶-۱۳۵۶ استفاده شده است. نتایج مقایسه‌ی این دو روش حاکیازآناستکه روش الگوریتم ژنتیک نسبت به روش ARDL از کارایی بسیار بالایی برخوردار است. همچنین نتایج حاصل از اندازه‌گیری بهره‌وری کل عوامل تولید ، نشان دهنده صعودی بودن روند آن تا سال۱۳۷۰ و از آن بعد روند نزولی ملایمی داشته است. همچنین رشد بهره‌وری کل عوامل تولید در بخش کشاورزی ایران طی دوره مورد بررسی نوسانات زیادی داشته است و میانگین سالانه رشد بهره‌وری در این بخش طی دوره مورد بررسی ۰/۱۶- می‌باشد. پس لازم است بر استفاده بهینه‌تر از نهاده‌های موجود و ترکیب مناسبتر آنها و افزایش بهره‌وری در بخش کشاورزی تأکید داشت.