Artificial Intelligence Analysis of the Impact of Tax Policies on the Underground Economy

Mohamad B. Shaaf *
Professor of Economics
University of Central Oklahoma, U.S.A.

Abstract

This study used artificial intelligence method to investigate the effect of all tax policies on the underground economy. The results confirm earlier findings, with the order of impact of tax policies on the underground economy as follows: corporate tax rate, individual and corporate audit rate, individual penalty rate, and individual marginal tax rate. The model explained over 87 percent of the changes of the relative size of the underground economy, somewhat better than those of other studies using regression.

Key Words:

Neural Networks, Tax Evasion, Underground Economy

JEL Classifications: C45; H26

---

* Correspondence to: Department of Economics, University of Central Oklahoma, Edmond, OK 73034 U.S.A. E-mail: mshaaf@ucok.edu. Telephone: (405) 974-5326. Fax: (405) 974-1111

Artificial Intelligence
Analysis of the Impact of Tax
Policies on the Underground Economy

1. Introduction

Underground economy (UE), also called shadow economy, parallel, or informal economy includes income unreported, and thus not accounted for and not added to the measurement of GDP. It consists of illegal transactions, such as prostitution, pornography, illegal gambling, and production and distribution of illegal substances. It also consists of legal activities unreported for the purpose of tax avoidance. One of the indicators of the robust development of the UE is the fact that per capita holding of currency rose from $160 in 1960 to $1,450 in 1994 (Feige, 1994)

Some studies indicate that the size of the informal economy is rising around the world (Lippert and Walker, 1997; Schneider and Enste, 2002). A growing UE reduces the government tax base, reduces tax revenue and/or compliance. It also biases the official statistics of GDP, income, consumption, labor force, employment, and unemployment. Studies of a sample of 84 countries concluded that countries with relatively low tax rates, fewer laws and regulations, and a well-established rule of law tend to have a smaller UE (Schneider and Enste, 2002).

Researchers used direct and indirect approaches to measure the size of the UE. The direct method uses data from sample surveys and questionnaires and tax audits of undeclared taxable income (Fridland, 1982; Spicer and Thomas, 1982; Benjamini and Maital, 1985; Aam, Jackson, and McGee, 1992). The indirect approach uses official data such as national income accounting, labor force statistics, and other statistics (Slemrod, 1985; Pommerehne and Weck-Hannemann, 1989; Erard and Feinstein, 1994).

A large volume of theoretical models have focused on the determinants of the tax evasion portion of the unreported economy (Falkinger, 1988; Pyle, 1989; Klepper, Nagin, and Spurr, 1991; Das-Gupla, 1994; Pestieau, Possen, and slutsky, 1994). Extensive empirical research shows that the size of the UE may be affected by governmental tax policies. For example, Clotfelter (1983), Slelmrod (1985), and Cegula (1997) found that the higher the individual and corporate tax rate, the greater the benefit from and the incentive for not reporting taxable income, thus resulting an increase in the size of the UE.

Empirical findings also suggest that as the risk involved in participation in the UE increases, the likelihood that economic agents would choose not to report or to underreport taxable income decreases (De juan, 1989; Thurman, 1991; Alm, Jackson, and Mckee, 1992; and Cegula, 1997). This risk is reflected in the likelihood of the individual tax audit rate (number of audits per number of returns) in the future and/or imprisonment, corporate audit rate (number of audits per number of returns), and tax penalties such as fines and interest on unpaid past tax liabilities per adjusted gross income.

Investigators have used parametric regression to establish empirical relationships between governmental tax policies and the size of the UE. However, each of those studies investigated only a subset of tax policies due to the complexity of the interactions, rather than an inclusive evaluation of all tax policies on the UE. The aim of this study is to use a new methodology, called Artificial Neural Network (ANN), to predict the combined effect of various tax measures on the UE. In addition, the aim is to estimate the relative importance of those tax measures contributing to the size of the UE. The ANN is a branch of the artificial intelligence model that recently has gained in popularity for two reasons. First, the predictive power of ANN is as good as or better than those of the regression technique. Second, the methodology of the ANN is nonparametric, which does not need the assumptions required by regression method.

The computation of the ANN starts with loading a set of values of independent variables into the input layer of the network for each observation. Then, each hidden node calculates the weighted summation of the inputs and incorporates the threshold, or bias, weight. Next, each hidden node computes the sigmoid function of the summation, which represses the value of the summation to a range between one and zero. After that, each hidden node sends the outcome to the output nodes. In the output layer, computations similar to those in the hidden layer take place, and finally, in that layer, the value of the dependent variable (s) is determined. The networks then compute the error as the difference between the estimated and the actual dependent variable (s). The output node propagates the amount of error on each observation back through the links to the hidden nodes. This process continues and, based on this error minimization, the weights are determined.

The findings of ANN confirm those obtained earlier by other methods, and suggest that corporate tax rate and audit rate, individual tax audit rate and tax penalty rate, and individual marginal income tax rate influence the size of the tax evasion, and thus UE.[1] The following section explains the model of the artificial intelligence. The nature of data, their form in the model, the composition of inputs and output, and the estimates of the weights of the model are described in the empirical section of the paper, followed by summary and conclusion of the study.

2. Artificial Intelligence Model

One of the branches of the study of artificial intelligence is the Artificial Neural Network (ANN). ANN is a multiple-layer configuration similar to the structure of the brain, consisting of simple processing elements or nodes that interact with each other through weighted connections in the system. This study utilizes the most widely used ANN model termed back propagation, which is a non-parametric algorithm for adjusting connection weights in a multiple-layer network. Back propagation is a learning design by which the multi-layer network is set for pattern recognition utilizing actual cross section or time series data as the external teacher. A typical ANN consists of:

1) one input (independent-variable) layer.

2) one or more intermediate or hidden layers.

3) one output (dependent-variable) layer.

Similar to the simultaneous equations in econometric modeling, ANN architecture can have several independent variables and several dependent variables. Each connection from the input layer to the hidden layer or from the hidden layer to the output layer has a weight. These weights represent the coefficients or the parameters of the model. The size of each weight represents the relative strength of the connection. Each node computes a weighted sum of the incoming values and passes this sum through a nonlinear or linear function as output. The output of one layer serves as the input to the next.

The hidden layer(s) is a processor of information and a bridge between the independent variable(s) and dependent variable(s). That layer(s) computes the information and does most of the work in the network. Since this layer is connected to independent variables (inputs) on one hand and dependent variables (output) on the other, and is hidden from the outside world, it is called the hidden layer. Besides its link to the hidden layer, the input layer can be directly connected to the output layer.

The network is fully connected. That is, there are links among all the nodes in adjacent layers. These links are the weights that can be strong or weak depending on their size. The weights are adjusted for the minimization of the mean square error as the objective function. Similar to the econometric method, the back propagation design is an optimization technique for finding optimum values of the weights as parameters. Furthermore, there is a bias-processing element that is similar to the intercept notion in equations of econometric models.

The three main phases in the operation of a network are learning, recall, and testing. In the learning phase, the ANN recognizes a pattern between independent variable(s) and dependent variable(s) and estimates the final value at the output layer. Subsequently, this estimated value is compared with the actual output, and errors are calculated. These errors become the factor used to adjust the weights and, subsequently, to reduce the errors and readjust the weights of the connections.

This process of weight adjustment continues further until the error declines to an acceptable level, when possible. Through this process, the ANN learns the rules and adapting patterns for processing the information. The resulting pattern used for testing the accuracy of the networks. In addition, the pattern can be used for prediction of the dependent variable(s), given the independent variable(s).

To explain the mathematical equations of the neural network, assume there are I number of independent variables, J hidden nodes in the hidden layer, and k output nodes in the output layer. Accordingly, back propagation computes the summation of multiplication of independent variables, X_I, by their corresponding weights (from the input layer to the hidden layer), and adds bias weights (intercept) to it as follows:

$$u_J = w_{0j} + \sum_{i}^{I} w_{ij} * X_i \qquad (1)$$

Where w_{0j} is the bias weight j, and w_{ij} are the weights from the input i to hidden node j.

The weighted summations, u_j, in equation (1) are then transformed by a transfer function in the hidden layer. It is called the transfer function because it acts to transfer the internally generated sum to a potential output value. Three common nonlinear transfer functions used in ANN are the sigmoid, hyperbolic tangent, and sine functions. The sigmoid transfer function, considered appropriate and used in this study, is a continuous monotonic mapping of the input into a limited-range value between zero and one. The sigmoid function of *u* variable is a smooth version of {0,1} step-shape function, and is defined as:

$$y_j = f(u) = (1 + e^{u})^{1} \qquad (2)$$

Here *e* is the base of natural logarithms. This output, y_j, from the hidden layer enters the output node as input, and subsequently computes the output of the network with *k* output nodes, z_k, as:

$$z_k = g(v_k), \qquad k = 1,2,......K \qquad (3)$$

$g(v_k)$ is the sigmoid function of v_k as:

$$v_k = w_{0k} + \sum_{j=1}^{J} w_{jk}\, y_j \,, \qquad (4)$$

Where w_{jk} are weights on the links from hidden node *j* to the output *k*, and w_{ok} is the bias weight of output node *k*. Notice that at the output node the process works the same way as in the hidden nodes, and the output *v* in equation (4) enters the sigmoid function and transforms the information in the form of:

$$z = f(v) = (1 + e^{v})^{1} \qquad (5)$$

If the ANN is directly connected from the input nodes to the output nodes, the v_k function in equation (4) is modified as follows:

$$v_k = w_{0k} + \sum_{j=1}^{J} w_{jk}\ y_j + \sum_{j=J+1}^{J+I} w_{jk}\ X_{j-J} \qquad (6)$$

The third term in equation (6) represents the weighted summation of input nodes by the weights directly connecting them to the output nodes.

The next step in the ANN is the calculation of the error. Suppose the estimated final dependent variable at the output layer which is produced by the network with N observation is z_{nk}, and the actual dependent variable is A_{nk} , then the mean square error is computed as:

$$E = \sum_{n=1}^{N} \sum_{k=1}^{K} (Z_{kn} - A_{kn})^2 / 2NK \qquad (7)$$

Here, E is the global error function which is differentiable of all the connection weights in the network.

The subsequent stage in the ANN is to adjust the weights (parameters) of the model. The strength of the connections (weights) changes continuously for the minimization of the error as the objective function. There are two major categories of weights in the ANN. The first category connects the hidden nodes to the output nodes. The second connects the input nodes to the hidden nodes. The equations of these two types of weight adjustment are explained next.

2.1 Hidden-to-Output-Node Weights

When the error E is determined, the network back propagates it for the increase or decrease of the weights. First, it selects some arbitrary point in weight space and computes the slope of the error surface at that point. Then, it computes the change or derivative of the error, *E*, with respect to the weight from the hidden node *J* to the output k, $\delta E/\delta w_{Jk}$, through a chain-rule multiplication of three terms as follows:

$$\delta E/\delta w_{jk} = \sum_{k=1}^{K} \sum_{j=1}^{J} (\delta E/\delta z_k)(\delta z_k/\delta v_J)(\delta v_J/\delta w_{jk}) \qquad (8)$$

The first term on the right side of the equation, $\delta E/\delta z_k$, is the derivative of the error with respect to the dependent variable *k*. The second term in the chain is the derivative of that nodes output with respect to its weighted sum of inputs from the hidden layer to the output layer. The last term, $\delta v_J/\delta w_{jk}$, is the derivative of that sum with respect to the weight connected to the output node *k*. According to equation (4), the last term in equation (8) depends on the output in the hidden node, y_j, as:

$$\delta v_J/\delta w_{jk} = y_j. \qquad (9)$$

2.2 Input-to-Hidden-Node Weights

Similarly, the ANN computes the relative change of the error, *E*, with respect to a change in the weights (from input nodes to the hidden nodes), $\delta E/\delta w_{jk}$, as a chain-rule multiplication of five terms as follows:

$$\delta E/\delta w_{ij} = \sum_{i=1}^{I}\sum_{j=1}^{J} \sum_{k=1}^{K} (\delta E/\delta z_k)(\delta z_k/\delta v_k)(\delta v_k/\delta y_j)] (\delta y_j/\delta u_j)(\delta u_j/\delta w_{ij}) \qquad (10)$$

The first-two terms on the right-side of equation (10) are similar to those in equation (8). The third, $\delta v_k/\delta y_j$, term is the derivative of the input from the hidden layer to the output layer with respect to the output *j* to the hidden node. The fourth term, $\delta y_j/\delta u_j$, is the derivative of the output in the hidden node with respect to the weighted sum of inputs in that node. The fifth term is the derivative of that sum with respect to the weight connecting inputs to the hidden node *j*.

The next task of the ANN is to find the direction of the weight adjustment in which the error surface declines the most. Given a point in the weight space, including those connecting the input nodes to the hidden nodes and those connecting the hidden nodes to the output nodes, the network produces a certain mean square error. The direction and magnitude of the weight change is computed according to the following equations:

$$w_p = w_{p-1} - lc \sum_{n=1}^{N} (\delta E / \delta w_p) \qquad (11)$$

or

$$\Delta w_p = - lc \sum_{n=1}^{N} (\delta E / \delta w_p). \qquad (12)$$

Where w_p, is the new weight at process number p, Δw is the weight change and lc is the learning coefficient as a parameter which can be set by the user. It actualizes the proportion and the speed of the weight change. Its best value is determined by trial and error. This approach is called *gradient descent rule*.

The last term on the right side of equations (11) and (12) is the total derivative of error with respect to each weight in the weight surface. The negative sign in those equations changes the direction in which the error declines rather than increases. With the new weight values, the process continues until the magnitude of the error approaches the minimum point in the error surface.[2] Figure1 shows the ANN used in this study, having six independent variables (inputs), thirteen hidden layers, one dependent variable (output), and a threshold or bias.

3. Empirical Results

Earlier studies used the size of the UE relative to the Adjusted Gross Income (AGI) reported to the Internal Revenue Service, as the dependent variable. Therefore, in this ANN model (UE/AGI) from the study by Edgar Feige (1994) was used as the output. This study focuses on the legal part of the underground economy and assumes that the illegal part is given. Accordingly, the inputs (variables) in the model only influence the legal part of the underground economy. In the previous theoretical and empirical research, economists used individual marginal income tax rate (MITR), average corporate income tax rate (ACTR), individual audit rate (IAR), corporate tax audit rate (CAR), and individual tax penalty per adjusted gross income (IP) to estimate the determinants of tax evasion. Because the (UE/AGI) ratio has a downward trend between 1973 and 1985, some researchers added a time trend (YR) to the independent variables. Accordingly, in this study the time trend (YR) was included as an additional input (independent variable) in the model.

The Augmented Dickey-Fuller (ADF) test suggested that MITR and UE/AGI were not stationary, but were stationary in first differences. Therefore, the first-difference of these two data was used. The ADF test results further revealed that the remaining variables, ACTR, CAR, IAR, and IP, were stationary. Thus, in this research the first difference of UE/AGI, d (UE/AGI), as output (dependent variable), and dMITR, ACTR, CAR, IAR, YR, and IP as inputs (independent variables) were used as follows:

$$d\,(UE/AGI) = f\,(ACTR, CAR, IAR, dMITR, YR, IP).$$

Figure1 provides a diagrammatic representation of the ANN used in this study.

Using annual data from 1973 to 1994 the results of the ANN analysis of the impact of tax policies on the tax evasion and UE are presented in Table 1. The model has six inputs including the time-trend, and employs 13 hidden layers. The resultant R^2 of this input mix was 0.877, the average error of the estimate of d (UE/AGI) was 0.496, the correlation coefficient of the actual and estimated output was 0.937, the mean square error was 0.507, and the root mean square error was 0.712.

The weight of each input variable captured in the learning process characterizes the relative importance of each input. These weights were estimated and are shown in Table 1 and Figure 2. According to the results, average corporate tax rate, ACTR, with the weight of 31.1% was the most important contributor to the changes in the relative size of the UE. The order of the five remaining variables was as follows: individual audit rate, IAR, 24.5%, time trend, YR, 18.4%, corporate audit rate, CAR, 8.1%, and changes in marginal individual tax rate, dMITR, 1.8%.

To test the accuracy and the validity of the ANN prediction, the results of the estimated (learned) output, d (UE/AGI), are compared with the actual, and are shown in Figure 3. The estimated output appeared to be very close to the actual as shown in that figure, and confirmed by a high R^2 of 0.877. The figure shows that the actual and estimated output had downward trends between 1973 and 1985, and stationary between 1985 and 1994. This suggests that although the size of the UE is increasing, as shown in Figure 4, its relative size has been declining from 1973 to 1985, and stationary from1985 to 1994. It should be noted that the ANN method clearly estimates these two trends.

Furthermore, Figure 3 shows that both the actual and estimated value of d (UE/AGI) has cyclical variations. These variations are mainly due to the variations of the AGI rather than that of UE. As shown in Figure 4, the size of the UE has an upward trend between 1973 and 1994. Between 1973 and 1981 this measure grows steadily. From 1981 to 1983 and 1984 to 1987 it was flat, and from 1991to 1993 it was declining. This suggests that the cyclical variation of d

(UE/AGI) is mainly related to the variation of AGI rather than UE during that period. Again, the ANN method clearly predicts and follows these cyclical variations.

4. Summary and Conclusions

Economists have utilized parametric regression methods to study the effect of governmental tax policies on the underground economy (UE). Those studies could not include all tax-policy variables due to the constraints inherent in the model. Using the artificial neural networks, a branch of the artificial intelligence method, this study reinvestigated the effect of these policies and their relative importance to the size of the UE. The outcome of the neural network analysis found that tax policies of the government have had an impact on the size of the UE. The importance of the influence of each tax policy, ranked from the greatest effect to the least as follows: average corporate tax rate, individual audit rate, corporate audit rate, individual penalty rate, and changes in marginal individual tax rate. The model explained over 87 percent of the variation in the changes of the relative size of the UE, and predicted the trends and the cyclical variations somewhat more accurately than studies using regression methodology.

References

Alm, J., Jackson, B. and McKee, M., (1992), "Institutional Uncertainty and Taxpayer Compliance," *American Economic Review*, 82:4, 1018-26.

Benjamini, Y. and Maita, S, (1985), "Optimal Tax Evasion and Optimal Tax Evasion Policy: Behavioral Aspect," in W. Gaertner and A. Wenig (Eds.), *The Economics of the Shadow Economy*, Berlin: Springer-Verlag.

Cegula, R. J. (1997), "An Empirical Analysis of the Impact of Government Tax and Auditing Policies on the Size of the Underground Economy: The Case of the United States, 1973-94," *American Journal of Economics & Sociology*, 56:2 April, pp, 173-187.

Clotfelter, C. T., (1983), "Tax Evasion and Tax Rates: An Analysis of Individual Returns," *Review of Economics and Statistics*, 65:2, 363-73.

De Juan, A, Lasheras, M., and Mayo, R., (1994), "Voluntary Tax Compliant Behavior of Spanish Tax Payers," *Public Finance/Finances Publiques*, 49:4, 90-105.

Erard, B., and Feistein, J., (1994), "The Role of Moral Sentiments and Audit Perceptions in Tax Compliance," *Public Finance/Finances Publiques*, 49:4, 70-89.

Falkinger, J., (1988), "Tax Evasion and Equity: A Theoretical Analysis," *Public Finance/Finances Publiques*, 43:3, 388-95.

Feige, E., (1994), "The Underground Economy and the Currency Enigma," *Public Finance/Finances Publiques*, 49:4, 119-136.

Internal Revenue Service, *(1972-2000), Annual Report of the Commissioner and Chief Counsel of the Internal Revenue Service*, Washington: GPO.

Klepper, S, Nagin, D., and Spurr, S., (1991), "Tax Rates, Tax Compliance, and the Reporting of

Long-Term Capital Gains," *Public Finance/Finances Publiques,* 46:2, 236-51.

Lippert, M., and Walker, I., (1997), *The Underground Economy: Global Evidence of its Size and Impact,* Vancouver, BC., The Frazer Institute.

Pestieau, P., Possen, U., and Steve Slutsky, (1994), "Optimal Differential Taxes and Penalties," *Public Finance/Finances Publiques,* 49:4, 15-27.

Pommerehne, W., and Weck-Hannemann, H., (1989), "Tax Rates, Tax Administration and Income Tax Evasion in Switzerland," paper presented at the Conference of the Applied Econometrics Association, Rome, Italy.

Pyle, D., (1989), *Tax Evasion and the Black Economy*, New York: St. Martins Press.

Schneider, F., and Enste, D., (2002), "Hiding in the Shadows: The Growth of the Underground Economy," *IMF Staff Papers*, 49:30, April, 1-12.

Shaaf, M., (2000), "An Artificial Intelligence Approach to the Predicting Recession Using the Yield Spread," *Eastern Economic Journal*, 26:2, Spring, 171-190.

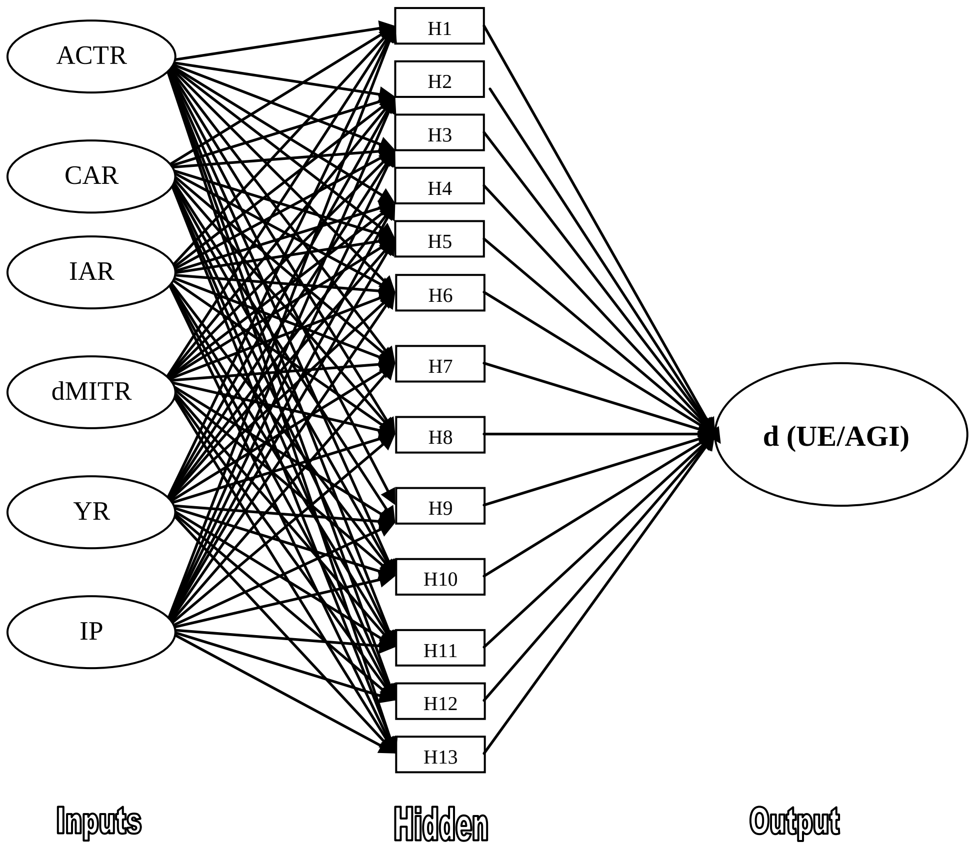
Slelmrod, J, (1985), "An Empirical Test for Tax Evasion," *Review of Economics and Statistics,* 67:2, 232-67.

Spicer, M., and Lundstedt, S., (1976), "Understanding Tax Evasion," *Public Finance/Finances Publiques,* 31:2, 295-305.

Thurman, Q., (1991), "Taxpayer Noncompliance and General Prevention: An Expansion of the Deterrence Model," *Public Finance/Finances Publiques,* 46:2, 289-98.

Figure 1

Artificial Neural
Network of Underground Economy



----

* YR = Year representing the trend, ACTR = Average Corporate Tax Rate, IAR = Individual Audit Rate, CAR = Corporate Audit Rate, IP = Individual Penalty per return, and dMITR = the first difference of Marginal Individual Tax Rate.

Figure 2

Relative Importance of Vaviables
Influencing Underground Economy, d (UE/GDP)

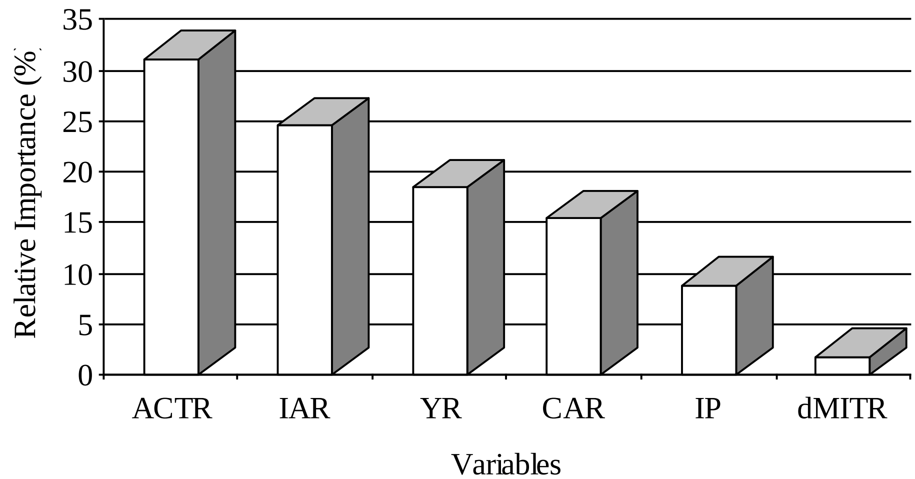


Figure 3

Actual vs. Predicted Underground Economy,
d(UE/GDP), vaviables: ACTR, IAR, YR, CAR, IP, AND dMITR

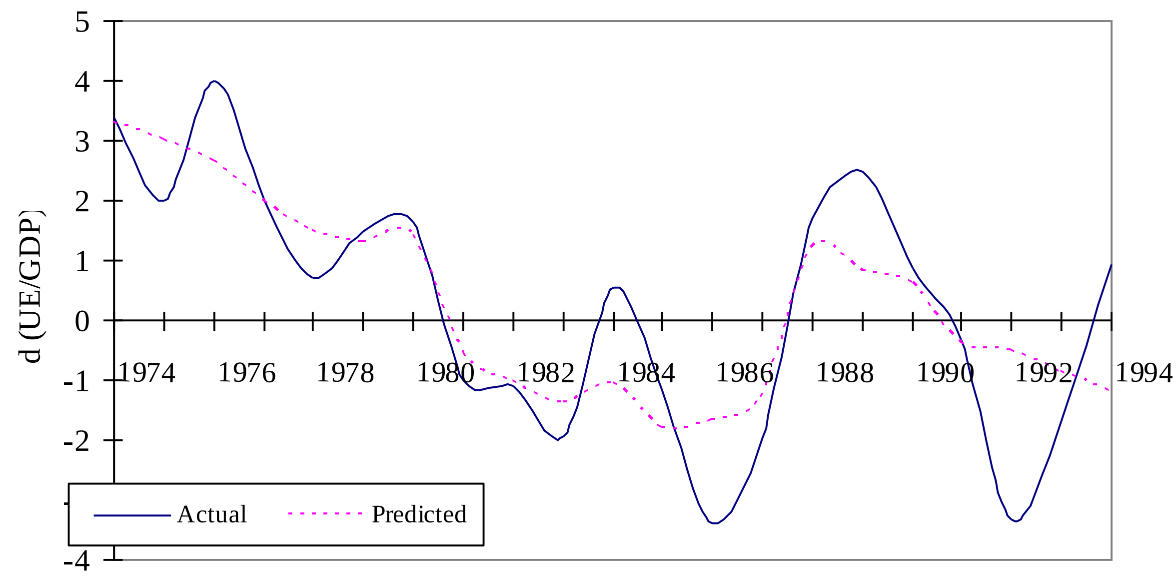


Figure 4

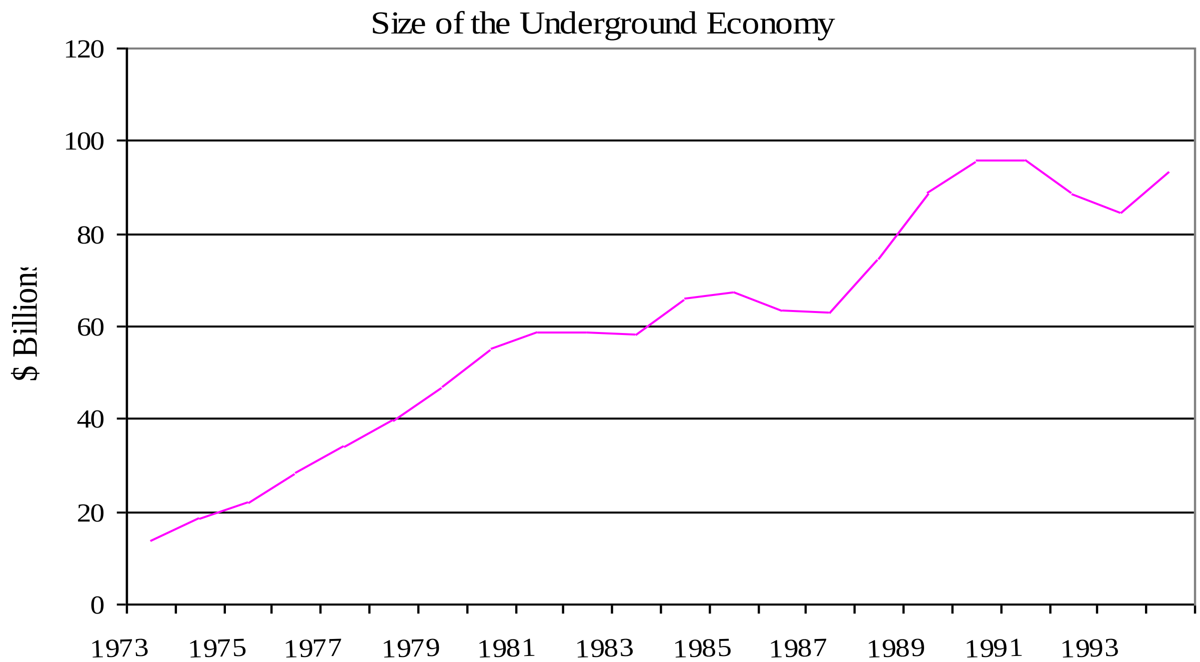

Size of the Underground Economy
$ Billions
0
20
40
60
80
100
120
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993

TABLE 1

Results of the Artificial Neural
Network: the Impact of Average Corporate
Tax Rate, Tax Audit Rate, Tax Penalty, and Marginal
Individual Tax Rate on the Relative Size of the Underground Economy

Number of Inputs	6
Number of Hidden Neurons Trained	13
Optimal Number of Hidden Neurons	13
R^2	0.877
Average Error	0.496
Correlation	0.937
Mean Square Error	0.507
Root Mean Square Error	0.712
Variables	Relative Importance (%)
ACTR	31.1%
IAR	24.5
YR	18.4
CAR	15.4
IP	8.8
dMITR	1.8
Total	100

* YR = Year representing the trend, ACTR = Average Corporate Tax Rate, IAR = Individual Audit Rate, CAR = Corporate Audit Rate, IP = Individual Penalty per return, and dMITR = the first difference of Marginal Individual Tax Rate.