XVIII

Volume XVIII, Number 2, November 2015

# Journal of Applied Economics

**Mostafa Tahmasebi**
**Michel Rocca**

A fuzzy model to estimate the size of the underground economy applying structural equation modeling

UCEMA

Edited by the Universidad del CEMA
Print ISSN 1514-0326
Online ISSN 1667-6726

*Journal of Applied Economics. Vol XVIII, No. 2 (November 2015), 347-368*

# A FUZZY MODEL TO ESTIMATE THE SIZE OF THE UNDERGROUND ECONOMY APPLYING STRUCTURAL EQUATION MODELING

**Mostafa Tahmasebi*** **and Michel Rocca**
*CREG, Université de Grenoble*

Submitted December 2013; accepted September 2014

The underground economy is an ambiguous concept: the literature presents a wide variety of definitions about it; the activities it encompasses are mobile and dynamic; and its structure has displayed several variations as time goes by. The present work aims to estimate a fuzzy number (a possible interval) for the size of the underground economy by applying structural equation modelling with fuzzy data. The proposed fuzzy model applied here involves two main steps, changing the structural equation model to a reduced form, then making a non-linear model from reduced-form equations applying fuzzy linear regression concepts and solutions. Finally, the time series of the underground economy are obtained using the GAMS mathematical optimization software and compared with the findings of two MIMIC models and a microeconomic method.

*JEL classification codes*: O17, H26, C22, C51, E26
*Key words*: underground economy, structural equation modelling, fuzzy linear regression, estimation methods.

## I. Introduction

Despite the government regulations in developing and developed countries, there are large numbers of market transactions (mostly in cash) in various activities and services which are not reported and declared to the government and make a phenomenon known as the underground economy. Most governments restrict

* Mostafa Tahmasebi (corresponding author): CREG, Université de Grenoble (UPMF), BP 47 - 38040 Grenoble cedex 9, Grenoble, France; tel.:+33 (0) 75 14 72 509; fax: +33 (0)4 76 82 56 76; e-mail: Mostafa.Tahmasebi@etu.upmf-grenoble.fr. Michel Rocca: e-mail: Michel.Rocca@upmf-grenoble.fr.

these activities by prosecuting, punishing and educating culprits, but the changes in the form of such activities have been considerable over the course of time (Jie et al. 2011).

Frey and Schneider (2000) discuss three major concerns about the existence and development of underground activities: reliance on official statistics may be misleading; underground activities avoid taxation, which reduces tax revenue; the existence of underground activities is an indicator of the unhealthy relations between citizens and government.

Various methods have been proposed and applied to measure the size of the underground economy. Schneider (2005) classifies these methods in three categories: direct methods, indirect methods and modelling methods. Direct methods or micro-methods build up estimates of the size of the underground economy based on the results of surveys, questionnaires, interviews and tax auditions of firms and/or households. These methods only yield estimates for a specific point in time; they do not provide a time series. They are correct at a specific point in time and can be regarded as low-end estimations. Difficulty in selecting appropriate samples, the possibility of selection bias and measurement errors regarding the surveys are among the defects of direct methods (Albu 1995; Elgin and Oztunali 2012).

Indirect or macroeconomic methods focus on the discrepancy between the official and actual criteria, such as differences between national income and expenditure, transactions and national income, and electricity consumption and GDP. As an example, if total labour force participation is assumed to be constant, a decrease in the official rate of participation can indicate an increase in underground activities. However, these methods are generally criticized for employing various simplifying and limiting assumptions, such as focusing on a specific aspect or indicator and neglecting the other causes which may also have a predominant effect on the underground economy (Albu 1995; Elgin and Oztunali 2012).

The most popular method used to estimate the underground economy is Multiple Indicators Multiple Causes (MIMIC). Similar to the other two methods, this approach is based on the use of simplified econometric conditions, making it vulnerable to statistical errors. Another deficiency of this approach is that it does not rely on any micro-foundations (Breusch 2005). On the whole, the evidence presented has failed to achieve a consensus on a reliable method for measuring the size of the underground economy without strong assumptions.

On the other hand, it can be claimed that many dynamic systems become naturally fuzzy due to uncertain initial conditions and parameters. Fuzzy logic can be applied in the case of an insufficient number of observations or samples; an unclear relationship between response and explanatory variables; difficulties verifying distribution assumptions, the ambiguity of events or the degree to which they occur; and inaccuracy and distortion due to linearity assumptions. The criteria cited here correspond very closely to the initial conditions of the underground economy, offering further encouragement for using fuzzy logic to model the underground economy. Until now, only one fuzzy model, composed of three steps — fuzzification, fuzzy inference and defuzzification — has been applied to estimate the underground economy in different countries (Draeseke and Giles 2002; Ene 2010; Yu et al. 2006). The criticisms of this model are: a weak method (mean and variance of the values) used in the fuzzification step, some shortcomings in defining the fuzzy inference table based on the Lindström (1997) rules, and determining the wrong value for the underground economy in the defuzzification step (unlike the fuzzy concept, a range of 0 to 100 percent of GDP for the size of the underground economy is adopted for all countries).

Unlike other works which provide an underground economy time series, this work aims to estimate an interval (more exactly, a fuzzy number consisting of a mode and the spreads) for the size of the underground economy. A method is proposed to solve the structural equation model with fuzzy raw data which improves the results compared to structural equation modelling without fuzzy data. The goal of this paper is consequently threefold: (i) to propose a model to measure the size of the underground economy based on structural equation modelling with fuzzy data; (ii) to propose a solution for the mentioned model applying fuzzy regression concepts; and, (iii) to obtain the fuzzy time series of the underground economy from the results.

The following section describes the underground economy concept and some theoretical considerations such as causes and indicators. Section III illustrates the fitting methods of fuzzy linear regression. Section IV presents structural equation modelling with fuzzy data, the resultant non-linear model and fuzzy time series of the underground economy. The results are compared with the findings of two MIMIC models (Dell'Anno et al. 2007; Schneider and Buehn 2012) and of a model proposed by Elgin and Oztunali (2012). Finally, conclusions are provided.

## II. Some theoretical considerations about the underground economy

### A. Definition of the underground economy

Various terms are used in the literature to describe underground economy activities, such as non-observed, irregular, unofficial, shadow, black, grey, hidden and unobserved economy (Gylys 2005), but there is no consensus about the definition in the literature (Öğünç and Yilmaz 2000). The confusion regarding terminology and concept in this ambiguous socio-economic category could entail huge cognitive losses, due to the complications in naming a concept with different content and comprehension (Gylys 2005).

The most frequently referenced definitions relate the underground economy to officially measured national income generated by non-reporting productive or value-added activities which should have been included in gross national product (GNP) (Jie et al. 2011). Smith (1997), in his definition, refers to market-based production of goods and services, whether legal or illegal, which escapes being detected in the official estimates of gross domestic product (GDP). Draeseke and Giles (2002) state that the underground economy involves legal or illegal transactions, activities and services which are not measured as they are not reported in order to evade tax liabilities. Ihrig and Moe (2004) mention a similar specification of the underground economy, escaping from reporting and recording, and define the underground economy as a sector which produces legal goods but does not comply with government regulations. On the whole, these definitions consider three main specifications for the underground economy: the underground economy consists of transactions, activities and services; these activities are either legal or illegal; and finally, these activities are not reported or recorded in spite of government regulations due to unconsciousness, evasion or avoidance of government regulations, or the shortcomings of government agents.

Gerxhani (2004) underlines that a universal definition for the underground economy is not necessary. She states that recognizing and classifying the activities is a reliable method for providing a schema of the underground economy. As the nature of activities is a determinant parameter for categorizing them, Table 1, which is commonly used in the literature, may complement the definitions above, presenting the various categories of underground activities (Jie et al. 2011).

**Table 1. Taxonomy of types of underground activities (Schneider and Enste 2000)**

| | Monetary transactions | | Non-monetary transactions | |
|---|---|---|---|---|
| Illegal activities | Trade with stolen goods; drug dealing and manufacturing; prostitution; gambling; smuggling; fraud; etc. Counterfeiting, computer system hacking; trading stolen information; identify theft; spamming. | | Barter of drugs, stolen goods, smuggling etc. Produce or growing drugs for own use. Theft for own use. | |
| Legal activities | Tax evasion | Tax avoidance | Tax evasion | Tax avoidance |
| | Unreported income from self-employment; Wages, salaries and assets from unreported work related to legal services and goods | Employee discounts, fringe benefits | Barter of legal services and goods | All do-it-yourself work and neighbour help |

## B. Causes and indicators of the underground economy

### Causes

Various factors, referred to as causes, affect growth of the underground economy. The causes most frequently applied and cited include the burden of direct and indirect taxation and the social security system, government regulation, forced reduction in weekly working hours, the unemployment rate, early retirement, prohibitions and corruption (Schneider and Enste 2000; Tanzi 2009; Jie et al. 2011).

Direct and indirect taxation are the causes most often cited, introduced as the burden of direct and indirect taxation, the effective direct and indirect tax rate and the ratio of total tax revenue to GDP. These taxes can affect labour leisure choices and labour supply in the underground economy which is known as an untaxed market. Different amounts of total revenue and labour cost in the official economy and in the underground economy give more incentives for participating in underground activities. Since the mentioned differences depend broadly on the social security system and the overall tax burden, they play a key role in the existence and rise of underground activities.

The countries with more general regulations in the official economy usually confront a bigger size of the underground economy (Schneider 2009). On the other

hand, the stricter the regulations against underground activities, the less individuals and firms turn to the underground economy. However, when the governments cannot guarantee absolute fair regulations for business communities, people seek other careers which may eventually involve them in the underground economy (Risteski 2009).

The government regulations considered in this study include the regulations and required licenses concerning the labour force – for example, supportive rules on hiring and firing, labour restrictions on foreigners, minimum wages and ages, safety regulations about health, fire, environmental hazards and so on which increase the labour and overhead costs in some official activities (Schneider et al. 2010) — and the trade barriers restricting the trade and commerce based on their concepts. These types of regulations increase the motivation of people to participate in the underground economy. In brief, more government regulations without any effective supervision on how well they are applied is the main reason for firms and individuals to participate in underground activities, which results in a higher share of the underground economy in total GDP.

Two other factors are also broadly discussed in the literature: forced reduction in official working hours and the unemployment rate. The reduction in working hours in the official economy was proposed by governments (e.g., France) and/or labour unions (e.g., Germany) in order to reduce unemployment. The main idea behind this policy is that there is a limited quantity of work, but it overlooks the fact that a forced reduction of working hours, contrary to employee preferences, increases the potential hours they can work in the underground economy (Schneider and Enste, 2000). In most OECD countries, rising unemployment is mostly due to the economic crisis in 2008 and high labour costs in hiring and firing the labour force. Giles and Tedds (2002) explain that two opposite forces determine the relationship between the unemployment rate and the underground economy, which means that this relationship is ambiguous. Also, Tanzi (1999) admits that "...for OECD countries there seems to be a broad relation between the panel data of the size of the underground economy and the official unemployment rates". Other factors such as early retirement and part-time work also provide opportunities for the labour force to work in untaxed unregulated activities.

From another point of view, the rate of self-employment as a percentage of the labour force is considered a determinant of the underground economy. Bordignon and Zanardi (1997) believe that the development of small firms, and the large proportion of professionals and self-employed with respect to the total

workforce, has a predominant effect on the Italian production system. Also, they explain that "…a large proportion of professionals and self-employeds implies greater possibilities for transferring expenses from consumption to production (to be deducted from taxes), simplified accounting and easier path collusion with customers".

Prohibitions can be treated as a form of regulation, so, they may be put in the same category as regulations. Prohibitions are related to the destination countries. Some activities are prohibited in some countries whereas they may be allowed in others. Germany, for example, allows prostitution. The Netherlands allows the sale and consumption of some drugs which are illegal elsewhere. Such prohibition can play a major role in pushing the labour force to participate in underground activities. Also, some activities are banned by the regulations in force, but being highly profitable, are much sought after. Those who engaged in these activities are consequently involved in illegal or even criminal activities such as production and distribution of illegal drugs, illegal gambling, lending money at usury rates, production and sale of weapons, prohibited biological substances, etc.

The proposed model in this paper is composed of six causes (see online Appendix B): the tax on income and profits as the share of direct taxation in GDP; the tax on goods and services as the share of indirect taxation in GDP; the ratio of long term unemployment to labour force; a proxy of government regulation which is defined as the ratio of government consumption and expenditure to GDP (Yu et al. 2006); the unemployment rate and the ratio of self-employment to labour force.

## Indicators

The dependent variables which indicate the situation of the latent variable — here, the underground economy — are called indicators. Schneider et al. (2010) suggest a classification for indicators and divide them into three groups: indicators which indicate the state of the official economy, monetary indicators and labour market indicators. There is no doubt that the underground economy has an effect on the state of the official economy. To take into the account these effects, real GDP per capita and its growth are considered as the indicators which reflect the state of the official economy. People trade underground goods and services in cash to prevent transactions being traced, while the effects of cash transactions are reflected in overuse of cash or currency. The currency ratio and $M_0$ over $M_1$ are used as the

monetary indictors to consider these effects of the underground economy. The labour force participation rate and growth rate of the total labour force are also regarded as the labour force indicators most frequently used in research.

In most analysis using structural equation modelling, real GDP per capita serves as an indicator with its coefficient set to unity. GDP indicates the amount of recorded gross production in a country. Thus, a sudden reduction in GDP can indicate that part of productions has not been recorded, for many reasons such as manufacturers moving to the underground sector or not declaring the precise amount of production in order to avoid taxation.

Dell'Anno and co-authors used real GDP (as the variable of scale), participation ratio of labour force and currency in circulation outside of banks as the indicators to estimate the underground economy (Dell'Anno et al. 2007; Dell'Anno and Schneider 2009; Dell'Anno 2007b). In another study, the real GDP index, Population Activity Rate and $M_1$ share in $M_2$ were chosen as the indicators for estimating the underground economy of Romania (Ene and ğefğnescu 2011). Buehn and Schneider (2008) selected the GDP volume index and monetary aggregate as indicators for estimating the unofficial economy in France.

Gerxhani (2004) explains that there are great differences between the significance of underground activities in developed and developing countries. In developing countries these activities are identified with low income, lacking any capacity for accumulation, whereas in developed countries the possibilities for accumulation and incomes are often comparable to formal activities. She also explains that the reasons for participating in the underground economy differ between these countries; survival demands and savings are the main reasons for participating in the underground economy in developing and developed countries, respectively. The amount of household saving can consequently also be considered as an indicator of participation in the underground economy of developed countries.

The following indicators are selected as best representing the size of the underground economy and applied in the proposed model: real GDP per capita, which is ratio of the real GDP to population in working age (15-64 years old); participation ratio of the Labour force which is calculated by dividing the total labour force by population in working age and; the household net saving rate. It must be mentioned that the causes and indicators are taken from the OECD statistical database (OECD 2014).

## III. Fuzzy linear regression

Fuzzy set theory, proposed in 1965 by Lotfi Zadeh (Arabacioglu 2010), is a form of many-valued logic which deals with reasoning which is approximate rather than fixed and exact. In contrast with traditional logic theory in which the elements of a crisp set definitely belong to the set, in a fuzzy set, elements of the set have a degree of membership. This membership value can range from 0 (not an element of the set) to 1 (a full member of the set). Fuzzy set theory is also extended in other mathematical methods such as the fuzzy regression model. The first studies of the possible use of fuzzy theory in regression analysis were done by Tanaka et al. (1982), who proposed the fuzzy linear regression model. The basic idea in fuzzy linear regression was to minimize a model's fuzziness by minimizing the total spread of fuzzy coefficients, while including all the given data (Shapiro 2005; Tanaka et al. 1982). This suggestion was the start of the application of fuzzy linear regression in various topics and domains. Diamond (1988) subsequently completed Tanaka's approach, which will be completely discussed next.

The degree of fitting and vagueness of the model were the two factors which Tanaka et al. (1982) considered to estimate the coefficients in different equations during the first studies applying fuzzy theory to regression analysis. Diamond (1988) completed Tanaka's approach by proposing measurement of the best fit based on the residuals under fuzzy consideration. He proposed the fuzzy least-squares approach which is a fuzzy extension of the ordinary least-squares approach based on a new defined distance on the space of fuzzy numbers. On the whole, the fuzzy regression methods can be roughly classified into two main categories: the Tanaka possibilistic regression approach, also called the possibilistic model, and Diamond's fuzzy least-squares approach (Shapiro 2005).

The first approach, the possibilistic model, aims to minimize the fuzziness of the model by reducing the total spreads of fuzzy coefficients by setting the data points of each sample within a specified feasible data interval (see online Appendix A). The objective of the second approach, the least-squares model, is to minimize the distance between the output of the model and the observed output, based on the modes and spreads of the model's output (Shapiro 2005; Arabpour and Tata 2008). Here, the possibilistic model is applied as it is easy to use, does not contain complicated equations, coding of this approach is simple, and the resulting linear/non-linear model can be solved using several softwares. The possibilistic fuzzy

regression approach focuses on minimizing the spread of the output (dependent variables) by considering adequate containment of data, which implies that:

$$\min[S_0 + \sum_{j=1}^{k} s_j \mid x_{ij} \mid], \quad s_j \geq 0 \ . \tag{1}$$

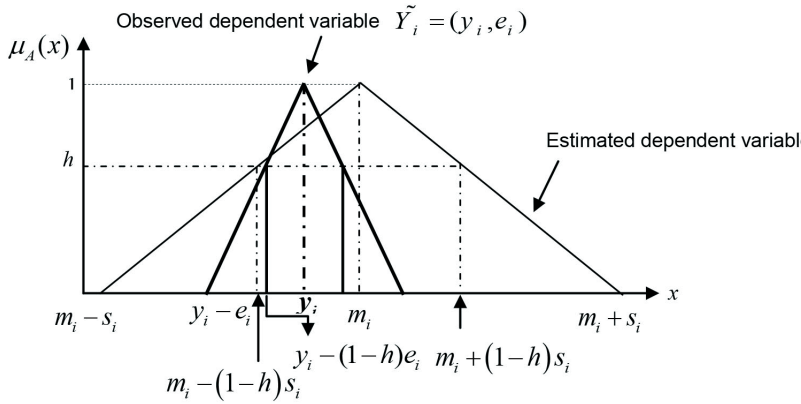
Now, putting the observed fuzzy output together with the fuzzy output of the model (Figure 1), it can be seen how the estimated fuzzy output may be fitted to the observed fuzzy data. The key point of this approach is that the observed fuzzy data, adjusted for an *h*-certain factor (see online Appendix A), is contained within the estimated fuzzy outputs, adjusted for that *h*-certain factor, which means that:

$$m_i + (1-h)s_i > y_i + (1-h)e_i \quad and \quad m_i - (1-h)s_i < y_i - (1-h)e_i \ , \quad \text{or} \tag{2}$$

$$a_0 + \sum_{j=1}^{k} a_j x_{ij} + (1-h)[S_0 + \sum_{j=1}^{k} S_j \mid x_{ij} \mid] > y_i + (1-h)e_i \ , \tag{3}$$

Chang and Ayyub (2001) showed that by increasing the *h*-certain factor, the confidence interval expands and it is consequently more probable that out-of-sample values fall within the model. This is comparable to increasing the confidence in statistical regression by extending the confidence interval.

**Figure 1. Fitting the observed output to the estimated output**

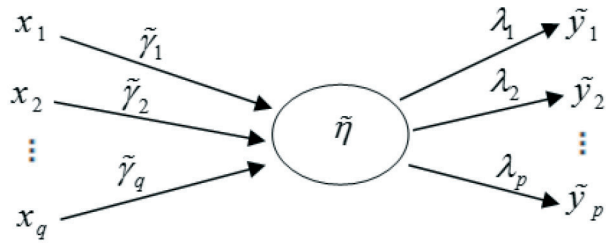


## IV. Structural equation model with fuzzy data

### A. Proposed model

The model presented here is based on structural equation modelling (Giles and Tedds, 2002) which contains a latent variable (the underground economy) and two main sets of variables, the causal variables which affect the latent variable and the indicator variables which represent the latent variable variations as depicted in Figure 2.

Figure 2. Definition of different elements of fuzzy model and their relations




Also, the model comprises two sets of equations, one between the causal variables and the latent variable and the other between the latent variable and the indicator variables:

$$\tilde{y}_t = \lambda \tilde{\eta}_t \ , \tag{4}$$

$$\tilde{\eta}_t = \tilde{\gamma}' x_t \ , \tag{5}$$

where $\tilde{y}\ (p \times 1)$ is the vector of indicator variables with a fuzzy value, $x\ (q \times 1)$ is the vector of causal variables or independent variables, which affect the model without being affected by it, and $\tilde{\eta}$ (scalar) is the latent variable. The latent variable is necessarily a fuzzy variable, so, if we assume that $x$ is crisp, the coefficients of the second set of equations (Equation 5) must be fuzzy ($\tilde{\gamma}(q \times 1)$). Depending on the fuzzy nature of the latent variable ($\tilde{\eta}$), $\tilde{y}$ must be fuzzy. Consequently, the coefficients of the first set (Equation 4) of equations $(\lambda(p \times 1))$ can be either fuzzy or crisp. In this study we assume they are crisp. The vector $\tilde{\gamma}'(1 \times q)$ is the transpose of the vector $\tilde{\gamma}$.

The model comprising equations (4)–(5) cannot specify the scale of all of the parameters, so a normalization condition is required. To do so, Giles and Tedds (2002) propose the convention of setting the first element of $\lambda$ to unit, as $\lambda_1 = 1$. Equation (6) is written as the reduced form of equations (4)–(5):

$$\tilde{y}_t = \tilde{\Pi} x_t \ , \quad \tilde{\Pi} = \lambda \tilde{\gamma}' . \tag{6}$$

The uncertainty in the fuzzy regression model becomes fuzziness, not randomness. So random errors do not appear in fuzzy regression and the uncertainty of the model is expressed as the spreads or fuzziness of variables (Kazemi et al. 2011).

For more clarification, the case of three indicators ($p$=3) and six causes ($q$=6) will be discussed. The coefficient $\lambda$ is a vector (3x1) with the first element set to one ($\lambda_1 = 1$) based on Giles and Tedds' convention. The restricted reduced form equations of the model can be written as equation (7), or equations (8)–(9) in full:

$$\tilde{y}_{jt} = \tilde{\Pi}_j x_t \ , \quad j = 1,2,3 \qquad \lambda = \begin{bmatrix} 1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}, \tilde{\Pi} = \lambda \times \tilde{\gamma}' = \begin{bmatrix} \tilde{\gamma}' \\ \lambda_2 \tilde{\gamma}' \\ \lambda_3 \tilde{\gamma}' \end{bmatrix}, \tag{7}$$

$$\tilde{y}_{1t} = \tilde{\gamma}' x_t \ , \tag{8}$$

$$\tilde{y}_{it} = \lambda_i \tilde{\gamma}' x_t \quad i = 2,3 . \tag{9}$$

The possibilistic regression model which minimizes the spread of the dependent variable will now be used to fit the data in the equations of the fuzzy model. Based on the possibilistic regression model, each equation has an objective function and two constraints. Here, there are three objective functions and consequently six constraints for the restricted, reduced-form equations. To obtain the objective function of the final non-linear model, the weighted sum method for Multi-Objective optimization (Marler and Arora, 2004) will be used where $w_i$ is the weight of the i$^{th}$ objective function ($F_i(x)$). There is no priority among these objective functions, thus, the equal weights for objective functions are considered:

$$=\sum_{i=1}^{k} w_i \times F_i(x) \xrightarrow{W_i = m(\text{constant})} \sum_{i=1}^{k} m \times F_i(x) = m \times \sum_{i=1}^{k} F_i(x) \ . \tag{10}$$

On the other side, based on optimization rules from Operation Research (OR), a constant coefficient in the objective function of a model has no effect on the result and can be ignored. So, the final objective function can be written as the summation of the objective functions (Equation 11). The constraints of the final model are the collection of conditions which were written for each equation.

$$=\sum_{i=1}^{k} F_i(x) \ . \tag{11}$$

Finally, GAMS software, a high-level modelling system for mathematical programming and optimization, will be used to optimize the resulting model.

## B. GAMS model and results

The possibilistic regression model is used in this study to obtain the coefficients of the fuzzy model. This approach focuses on minimizing the spread of the dependent variable by considering adequate containment of data. The objective function of the model which is the sum of the objective functions (objective functions of equations 8-9) is written as follows:

$$\min\{\sum_{\mathrm{j}=1}^{3} \lambda_j * [\sum_{t} \sum_{i=1}^{6} s_i \ |x_{it}|]\}, \quad s_i \geq 0 \ . \tag{12}$$

There are three equations in the restricted reduced form and each of them has two constraints which can be written as follows:

$$\lambda_j * \sum_{i=1}^{6} a_i \, x_{it} + (1-h) * \{\lambda_j * \sum_{i=1}^{6} s_i \, x_{it} \} > y_{jt} \qquad j = 1, 2, 3 \ , \tag{13}$$

$$\lambda_j * \sum_{i=1}^{6} a_i \, x_{it} - (1-h) * \{\lambda_j * \sum_{i=1}^{6} s_i \, x_{it} \} < y_{jt} \qquad j = 1, 2, 3 \ . \tag{14}$$

In the resulting model, 162 main mathematical expressions must be considered when optimizing the objective function. Some additional mathematical expressions are also added as constraints of the model; for example, the expressions which are imposed on the model due to the positive and negative variables in the objective function.

Using GAMS, the coefficients of equation (4)–(5) are calculated: $\lambda_{(3x1)}$ and $\tilde{\gamma}_{(6x1)}$ $(or\ \tilde{\gamma}'_{(6x1)})$. It is noteworthy that each element of $\tilde{\gamma}_{(6x1)}$ is a fuzzy number written as $\tilde{a}_i = (a_i, s_i) = (a_i - s_i, a_i, a_i + s_i)$, (see online Appendix A). The time series of the underground economy based on various *h*-certain factors can be written as equation (5) which is the relation between the causal variables $(x_{(6\times1)})$ and the latent variable $(\tilde{\eta})$.

Equation (5), which is a fuzzy number of underground economy $(\tilde{\eta}_t)$ in each year (*t*), can be written regarding the definitions presented:

$$\tilde{\eta}_t = \tilde{\gamma}' x_t = \tilde{a}_1 x_{1t} + \ldots + \tilde{a}_6 x_{6t} = \sum_{i=1}^{6} \tilde{a}_i x_{it} = \sum_{i=1}^{6} (a_i, s_i) x_{it} = \sum_{i=1}^{6} (a_i - s_i, a_i, a_i + s_i) x_{it}$$

$$= (\sum_{i=1}^{6} (a_i - s_i) x_{it},\ \sum_{i=1}^{6} a_i x_{it},\ \sum_{i=1}^{6} (a_i + s_i) x_{it}), \quad t = 1, \ldots, T_f. \tag{15}$$

Based on Figure 1 and equation (15), the central value and the left and right bounds of the underground economy in year (t) are:

*Central value in year t which makes central time series:* $\sum_{i=1}^{6} a_i x_{it}, \ t = 1, \ldots, T_f$, (16)

*Left values in year t which make lower bound time series:*

$$\lambda_j * \sum_{i=1}^{6} a_i\, x_{it} - (1-h) * \{\lambda_j * \sum_{i=1}^{6} s_i\, x_{it}\} < y_{jt} \qquad j = 1, 2, 3. \tag{17}$$

*Right values in year t which make upper bound time series:*

$$\sum_{i=1}^{6} (a_i + s_i) x_{it}, \ t = 1, \ldots, T_f. \tag{18}$$

The time series obtained by varying *h*-certain factor are discussed in the following subsection.

### C. Results and discussion

There are different criteria for evaluating fuzzy regression models. In this study we have used the Index of Confidence (Wang and Tsaur, 2000) to evaluate different results based on the *h*-factor:

$$IC = 1 - \frac{SSE}{SST}, \quad SSE = 2\sum_{t=1}^{T_f}(Y_{mim,t} - \tilde{Y}^{a}_{Fmim,t})^2, \tag{19}$$

where $Y_{mim,t}$ is the underground economy time series obtained from the MIMIC model (Dell'Anno et al. 2007; Schneider and Buehn 2012) and $(\tilde{Y}^{L}_{Fmim,t}, \tilde{Y}^{a}_{Fmim,t}, \tilde{Y}^{R}_{Fmim,t})$ are the fuzzy numbers of the underground economy time series obtained from the fuzzy model.

The average absolute error percentage *(AAEP)* is also applied to select the most suitable *h*-certain factor yielding the best estimate. The *AAEP* is calculated using equation (20) (Kazemi et al. 2011):

$$AAEP = \frac{1}{n}\sum_{t=1}^{T_f}\left|\frac{\tilde{Y}^{a}_{Fmim,t} - Y_{mim,t}}{Y_{mim,t}}\right|, \tag{20}$$

where $\tilde{Y}^{a}_{Fmim,t}$ is the central estimated value of the underground economy time series gained from the fuzzy model and $Y_{mim,t}$ is the underground economy time series obtained from MIMIC.

There is a trade-off between the minimum *IC* and the minimum *AAEP* and the best time series will be selected on the basis on these two items (Table 2). Regarding *AAEP*, in cases 2 and 5 the average absolute error percentage is minimum and between these two cases, case 2, with the lowest index of confidence, is more favourable. So, the *h*-factor must be set to 0.4 for the data set of France.

**Table 2. Evaluation of model based on different *h*-certain values**

| *Case* | *h*-certain | *SSE* | *SST* | *IC* | *AAEP* |
|---|---|---|---|---|---|
| 1 | 0.2 | 0.166155 | 0.241169 | 0.311043 | 0.28854 |
| 2 | 0.4 | 0.174953 | 0.404872 | 0.567882 | 0.278663 |
| 3 | 0.5 | 0.173016 | 0.434624 | 0.601919 | 0.285188 |
| 4 | 0.6 | 0.165269 | 0.404741 | 0.591666 | 0.282714 |
| 5 | 0.8 | 0.14623 | 0.436384 | 0.664906 | 0.274314 |

The final time series for France's underground economy, resulting from the model for a selected *h*-certain (*h=0.4*), is scaled in Figure 3. Other time series estimated by MIMIC models (Schneider and Buehn 2012; Dell'Anno et al. 2007) and a time series estimated by Elgin and Oztunali (2012) are shown to compare the trend of the four results. For France, the trend and turning points of Dell'Anno results and fuzzy curves are nearly similar between 1980 and 1990. Also, a similar trend is observed between Schneider-Buehn's result and the fuzzy result from 2001 to 2010 with an increasing trend for the subsequent years. Elgin-Oztunali's result does not show any significant variation in 1985-99, whereas it reveals a generally downward trend for the following years. Also, the magnitude of France's underground economy generally increases after 2004 based on Schneider-Buehn's and the fuzzy estimates (unlike the Elgin-Oztunali results), which is meaningful and reasonable. It means that without changing the policies and regulations, the working situation in the underground economy will be favourable and it may have undesirable effects on the official economy.

Schneider and Frey (2000) indicated that the expected interval for the size of France's underground economy was between 13% and 16% of GDP. So, the results of the fuzzy model in the central graph (the more likely value of the underground economy or the mode of the fuzzy number of the underground economy for each year) are within this limit, as are Dell'Anno's (1990-99), Schneider-Buehn's and Elgin-Oztunali's time series. On the whole, the central time series of the fuzzy method coincides with Dell'Anno's between 1990 and 1998 and Schneider-Buehn's estimates from 1999 to 2010. However, it should be borne in mind that the true time series of the underground economy is unknown; there is no way of saying which of these methods is more accurate (Schneider and Buehn 2012; Dell'Anno et al. 2007).

Figure 3 also shows the time series of the underground economy for other countries. For Canada and Italy all three models estimate a downward trend in the magnitude of the underground economy, with average sizes of respectively 16.36% and 28.34% of official GDP in the fuzzy model, 16.82% and 28.81% in the Elgin-Oztunali model, and 15.61% and 26.93% in the Schneider-Buehn estimate. For the United States, the trend for the fuzzy results and those of Schneider-Buehn are nearly similar, with an increase in the magnitude of the underground economy after 2007, whereas the Elgin-Oztunali model shows a downward trend.

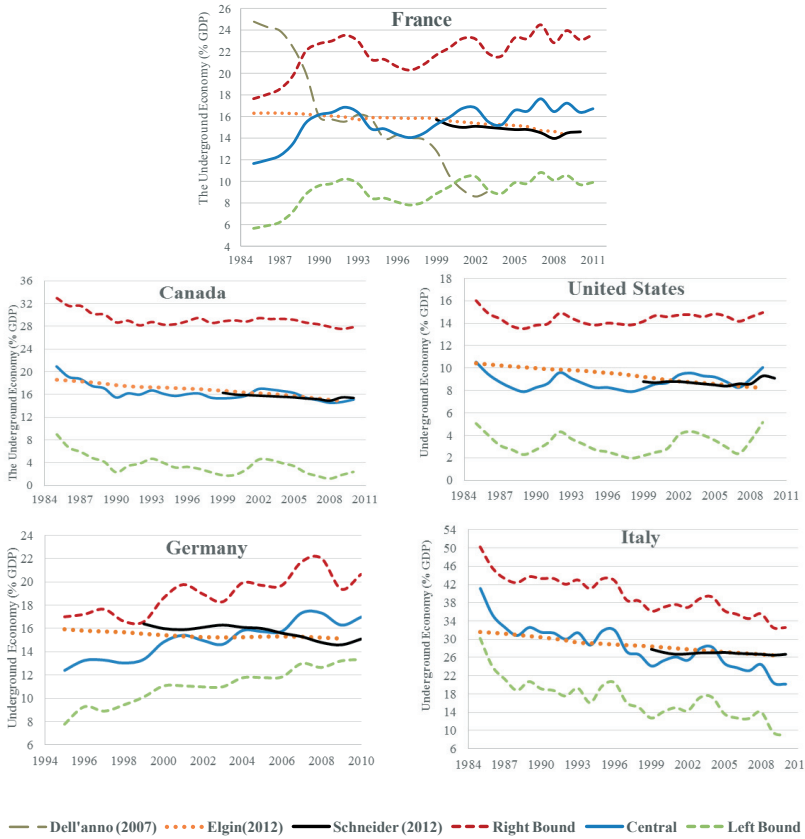
For Germany, the Elgin-Oztunali model shows a nearly constant trend unlike Schneider-Buehn's and the fuzzy results which find increasing trends after 2007. Also, the MIMIC estimate of Buehn and Goethel (2010) for Germany is very similar in trend and size to the result of the fuzzy method unlike the MIMIC estimate of Schneider and Buehn (2012). It can be seen that in most countries with GDP growth in 2006-7 (the United States, Canada and France), the size of the underground economy increased in 2009 as a result of the world financial and economic crisis.

Although it is not possible to determine the true series of values for the size of the underground economy or to determine which model is more accurate, comparing the fuzzy model with MIMIC, we can say that both methods are reliable when the raw data are adequate. But the MIMIC method is very sensitive to the size of the raw data and does not give reliable results with an inadequate number of observations (small data set). Various factors – data set too small, difficulty verifying the normal distribution of the term error, unclear relationships between variables, ambiguity associated with the event, inappropriate linearity assumption – pose problems for the statistical regression underpinning the MIMIC method. In contrast, the proposed fuzzy model is not based on strong assumptions and can be applied despite difficulties verifying distribution assumptions, unlike the MIMIC method which makes strong assumptions for the distribution of term errors, etc. Nor is the proposed fuzzy model sensitive to data specification, dimension, etc. (see Breusch 2005 for more on the weakness of MIMIC in this respect).

Comparing the fuzzy model and the Elgin-Oztunali model, it is noteworthy that despite some similarities in these models, they are completely different. The fuzzy model focuses on the macroeconomic aspects of the underground economy whereas the Elgin-Oztunali model focuses on the micro-economic aspects (behaviour of actors of the underground economy). It seems that modeling actors' behaviour in economic events is not a reliable method if the same model is applied

to all countries. For example in France welfare services, and regulatory sanctions and deterrents, are the main factors which prevent people from working in the underground economy. But the Elgin-Oztunali model makes no allowance for these factors. Only the tax rate is applied as a regulation. Also, it seems that it is not a good idea to use the same model ( or similar constants in the model) for two different countries.

**Figure 3. The fuzzy model, Elgin-Oztunali and MIMIC time series of the underground economy**



## V. Conclusion

The underground economy is unobservable; the causes of the underground economy are often vague and there is very limited knowledge about the nature of the relationships between these causes. Among the methods proposed to model the underground economy, fuzzy modelling allows for rapid modelling, even with imprecise and incomplete data, as it is not necessary to know all the details about the relations before starting. It also allows us to model non-linear functions of arbitrary complexity and achieve simplicity and flexibility using a fuzzy modelling strategy.

We modelled the underground economy using structural equation modelling with the fuzzy raw data. We then converted the structural equation model to reduced-form equations and used an optimization method to obtain two sets of coefficients (equations 4-5) in the structural equation model. Finally, the time series of the underground economy was scaled using the equations of the second set (equation 5). Briefly, what makes this proposed method different from previous fuzzy methods are:

- Fuzzy models in the literature consist of 3 steps — fuzzification, fuzzy inference and defuzzification —, while the current model applies a structural equation modeling concept with fuzzy data to construct the principal model;
- The indicators — in addition to causes — are included in this method to estimate the underground economy, which improves much the estimation results, while other fuzzy methods only use the causes;
- In previous fuzzy methods, the final result is one crisp number while the result here is an interval of crisp numbers for the size of the underground economy in each year, with a central value as the most possible size of the underground economy.

We used this fuzzy model to determine two key specifications of the underground economy: its size and trend. The resulting time series provide information about the maximum and minimum possible size of the underground economy as a fuzzy number in each year. The central time series represents the more probable size of the underground economy, and its upper and lower bounds (the spreads of the mode) in each year demonstrate the appropriate and inappropriate conditions.

The method developed in this work can be applied to different countries. Moreover, the fuzziness or spreads of the results – the divergence of the left and right bounds from the central time series – indicates the magnitude of the

uncertainty of the raw data. In other words, it tells us how much the assumed initial conditions for decision policies are relevant to the real initial conditions. There are also grounds for claiming that fuzziness is inversely related to achievement of policy goals. So increasing fuzziness reflects decreasing achievement of the various policy goals and vice versa. The results of the proposed model (three time series) may therefore be crucial to decision-makers as they can obtain a size for the underground economy by calculating the weighted average of the three time series. The greater the fuzziness of results, the more advisable it is to increase the weighting percentage of the right and left bounds.

## References

Albu, Lucian-Liviu (1995). Underground economy and fiscal policies modeling. Working Paper, Bucharest, National Institute of Economic Research, Romanian Academy.

Arabacioglu, Burcin Cem (2010). Using fuzzy inference system for architectural space analysis. *Applied Soft Computing* 10: 926–937.

Arabpour, Alireza, and M. Tata (2008). Estimating the parameters of a fuzzy linear regression model. *Iranian Journal of Fuzzy Systems* 5:1-19.

Bordignon, Massimo, and Alberto Zanardi (1997). Tax evasion in Italy. *Giornale degli economisti e annali di economia* 56: 121–169.

Breusch, Trevor (2005). Estimating the underground economy using MIMIC models. Working Paper, Canberra, National University of Australia.

Buehn, Andreas, and Matthias Goethel (2010). The shadow economy and implications for money demand in Germany. Dresden.

Buehn, Andreas, and Friedrich G. Schneider (2008). MIMIC models, cointegration and error correction: An application to the French shadow economy. Working Paper 2200, Bonn, IZA CESifo.

Chang, Yun-Hsi, and Bilal M. Ayyub (2001). Fuzzy regression method - a comparative assessement. *Fuzzy sets and systems* 119: 187-203.

Dell'Anno, Roberto, and Friedrich Schneider (2009). *A complex approach to estimate the shadow economy: The structural equation modelling*. *Coping with the complexity of economics, new economic windows*. Milan, Springer.

Dell'Anno, Roberto, Miguel Gómez-Antonio, and Angel Pardo (2007). Shadow economy in three very different Mediterranean countries: France, Spain and Greece. A MIMIC approach. *Empirical Economics* 33: 51-84.

Dell'Anno, Roberto (2007b). The shadow economy in Portugal: An analysis with the MIMIC approach. *Journal of Applied Economics* 10: 253-277.

Diamond, Phil (1988). Fuzzy least squares. *Information Sciences* 46: 141-157.

Draeseke, Robert, and David EA Giles (2002). A fuzzy logic approach to modeling the New Zealand underground economy. *Mathematics and Computers in Simulation* 59: 115–123.

Elgin, Ceyhun, and Oguz Oztunali (2012). Shadow economies around the world: model based estimates. Working Paper 5, Bogazici University.

Ene, Corina-Maria (2010). Romanian underground economy–a fuzzy approach. *International Audit Risk Management* 2: 29-38.

Ene, Corina-Maria, and Andrei ğefğnescu (2011). Size and Implication of underground economy in Romania - A MIMIC Approach. *Annales Universitatis Apulensis Series Oeconomica* 13: 77-86.

Frey, Bruno S., and Friedrich Schneider (2000). Informal and underground economy, Working Paper 4, Department of Economics, Johannes Kepler University of Linz.

Gerxhani, Klarita (2004). Informal sector in developed and less developed countries: A literature survey. *Public choice* 120: 267-300.

Giles, David EA., and Lindsay M. Tedds (2002). *Taxes and the Canadian underground*. Toronto, Canadian Tax Foundation.

Gylys, Povilas (2005). Economy, anti-economy, underground economy: Conceptual and terminological problems. *Ekonomika* 72: 1-11.

Ihrig, Jane, and Karine S. Moe (2004). Lurking in the shadows: The informal sector and government policy. *Journal of Development Economics* 73: 541-557.

Jie, Sim Wan, Huam Hon Tat, Amran Rasli, and Lee Thean Chye (2011). Underground economy: Definition and causes. *Business and Management Review* 12: 14-24.

Kazemi, A., H.G. Shakouri, M.B. Menhaj, M.R. Mehregan, and N. Neshat (2011). A hierarchical fuzzy linear regression model for forecasting agricult ure energy demand: A case study of Iran. 3rd International Conference on Information and Financial Engineering, Singapore.

Lindström, Tomas (1998). A fuzzy design of the willingness to invest in Sweden. *Journal of Economic Behavior & Organization* 36: 1-17.

Marler, R. Timothy, and Jasbir S. Arora (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization* 26: 369–395.

Öğünç, Fethi, and Gökhan Yılmaz (2000). Estimating the underground economy in Turkey. Discussion Paper 15, Central Bank of the Republic of Turkey.

OECD (2014). Link: http://stats.oecd.org/

Risteski, Hristijan (2009). Assesing handicraft shadow economy in Macedonia. Center for Economic Analyses, Skopje.

Schneider, Friedrich (2005). Shadow economies around the world: What do we really know? *European Journal of Political Economy* 21: 598-642.

Schneider, Friedrich, and Dominik Enste (2000). Shadow economies: Size, causes, and consequences. *Journal of Economic Literature* 38: 77–114.

Schneider, Friedrich (2009). Size and development of the shadow economy in Germany, Austria and other OECD-countries. *Revue économique* 60: 1079-1116.

Schneider, Friedrich, and Andreas Buehn (2012). Shadow economies in highly developed OECD countries: What are the driving forces? Discussion Paper 6891, Bonn, IZA.

Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro (2010). New estimates for the shadow economies all over the World. *International Economic Journal* 24: 443-461.

Shapiro, Arnold F. (2005). Fuzzy regression models. Penn State University.

Smith, Philip (1997). Assessing the size of the underground economy: The Canadian statistical perspectives. In: *The underground economy: Global evidence of its size and impact*. Vancouver, The Fraser Institute.

Tanaka, H., S. Uejima, and K. Asai (1982). Linear regression model with fuzzy model. *IEEE Transactions on Systems, Man and Cybernet* 12: 903-907.

Tanzi, Vito (2009). The shadow economy , its causes and its consequences. International Seminar on the Shadow Economy Index in Brazil.

Tanzi, Vito (1999). Uses and abuses of estimates of the underground economy. *Economic Journal* 109: 338-347.

Wang, Hsiao-Fan, and Ruey-Chyn Tsaur (2000). Insight of a fuzzy regression model. *Fuzzy Sets and Systems* 112: 471-475.

Yu, Tiffany Hui-Kuang, David Han-Min Wang, and Su-Jane Chen (2006). A fuzzy logic approach to modeling the underground economy in Taiwan. *Physica A: Statistical Mechanics and its Applications* 362: 471-479.