A Fuzzy Logic Approach to Modelling the Indian Underground Economy

S. Kayalvizhi[#1], T. Gunasekar[*2], S. Thenmozhi [#3]

[#1]*Department of Sciences, Mahatma Gandhi University, India*
[#2]*Department of Computer and Information Science, University Technology PETRONAS, Malaysia*
[#3]*Department of Electrical and Electronical, NPR College of Engineering & Technology, India*

Abstract— *The availability of data for the size of the "Indian Underground Economy" is important for macroeconomic policy. We use fuzzy set theory and fuzzy logic to construct an annual time-series for the Indian underground economy over the period 2010 to 2016. Two input variables are used – the effective tax rate and an index of the degree of regulation. The resulting underground economy, time-series is compared with one previous a structural "Multiple Indicators, Multiple Causes" (MIMIC) model. The two approaches each yield sensible, but somewhat different, pictures of the India underground economy over this period. The fuzzy logic approach to this measurement problem involves several subjective judgements, but our results are quite robust to these choices.*

Keywords — *Fuzzy Logics, Modelling, Time series, Multiple Indicators***.**

I. INTRODUCTION

There is renewed interest internationally in the problem of the "Indian underground economy" and its implications for the "tax-gap", the effectiveness of fiscal and monetary policy, economic growth, and for income distribution. The Indian underground economy involves activities and transactions which may themselves be legal or illegal, but which are not measured because they go unreported. The lack of reporting is generally to evade tax liabilities. Examples of these activities include unreported "cash" payments, extortion, smuggling, prostitution, narcotics sales, *etc.*.

By its very nature, the "Indian underground economy" (IUE) is not directly observable. However, various authors, to obtain measures of the UE in different countries, have used many different methods. For example, see [1] for a recent survey and some new results, and [2] for extensive international results. The empirical evidence, now available for many countries, is of varying quality. In addition, historical time series data on the UE have been constructed on a consistent basis for a few countries.

However, the available quantitative measures of the IUE point to one important fact – the size of the IUE is growing in all countries for which data have been constructed. This appears to be the case, not only in absolute (nominal) terms, but also in relative terms when we consider the ratio of the IUE to the measured GDP of each country. There is an urgent need for new and improved methods for measuring the size of the unobservable IUE. In this paper, we address this need by illustrating how the tools of fuzzy set theory and fuzzy logic can be used to generate a time-series measure of the IUE. This illustration takes the form of a rather limited, but very promising, application with India data.

The section-2 discusses some basic principles associated with fuzzy sets, and outlines our overall methodology. The step-by-step details of this methodology are provided in Section 3; and the results based on the Indian data are described in Section 4. The last section provides our conclusions, and a discussion of prospects for further research on this topic.

II. BACKGROUND PRINCIPLES

A. Historical Context

Fuzzy set theory and the associated fuzzy logic have found widespread application in many disciplines since the seminal contributions of Zadeh [3-4] and his followers. These applications are extensive in computer science, systems analysis, electrical and electronic engineering, and related fields. The construction and application of "expert systems" has touched most aspects of modern life, often without our knowledge. Examples include their inclusion in domestic appliances, vehicles, and the like.

While the use of fuzzy sets and logic has been widespread in the physical sciences (although not without criticism), the application of these tools in the social sciences appears to have been limited mainly to psychology. Applications in economics are few, some exceptions being [5-6] and others in the field of social choice.
More specifically, the use of fuzzy set theory in Econometrics is virtually unknown. To our knowledge, the only other such contributions are those of [7]. The former authors use fuzzy sets in a regression context to model non-linearity's, while Lindstrom uses fuzzy analysis to "predict" fixed investment behaviour on the basis of interest rate

levels and changes. Our own analysis here follows his methodology quite closely.

B. Fuzzy Definitions

"Fuzzy sets" deal with "concepts" and "linguistic variables". For instance, "price" is a concept, and "rather low price" is a linguistic variable. A "fuzzy set" maps from a regular set to [0, 1]. Membership of a fuzzy set is not "crisp". An example of this mapping would be: "the price of this personal computer is Rs. 1,25,000. This is one of the most expensive such computers I have ever seen, so I rate its price at 0.98." The number "0.98" is the "degree of membership", and this value should *not* be confused with a probability. For instance, degrees of membership need not sum to unity.

Examples of "fuzzy logic" are: "if the price is high, then demand will be low"; or "if taxes are high, then tax evasion will be high". The application of the inductive premise of fuzzy concepts poses some difficulties – not all of the usual set-theoretic laws are satisfied. In particular, the "law of the excluded middle" is violated, so a different group of operators must be adopted for instance, "union" is replaced by "max", "intersection" is replaced by "min", and "complement" is replaced by subtraction from unity. Then the commutative, associative, distributive, idempotency, absorption, excluded middle, involution, and De Morgan's laws are satisfied. For example, if U={a, b, c, d} and the fuzzy sets A and B are defined as A={0.3/a, 0.6,c, 1/d} and B={0.1/a, 0.5/b, 0.7/c, 0.9/d}, where the numbers are "degrees of membership", then $A\cup B$={0.3/a, 0.5/b, 0.7/c, 1/d} and $A\cap B$={0.1/a, 0/b, 0.6/c, 0.9/d}. Similarly, B^c={0.9/a, 0.5/b, 0.3/c, 0.1/d}.

The use of fuzzy sets and logic in Econometrics is an appealing possibility. For example, often our data are necessarily vague, we may have limited knowledge of the nature of the relationships between variables, and these relationships may be intrinsically non-linear.

III. OVERALL METHODOLOGY

Our task is the "measure" the size of the Indian underground economy, year by year. The following methodology is not inferential in the usual sense, and it differs from a regression-based approach using "indicators" and "causes", as none of the former variables are used. For simplicity, our application uses only two causal variables that, based on both economic theory and widespread international empirical evidence, are widely believed to be the primary determinants of underground activity [1-2], [8-13] .

These variables are the effective tax rate (the ratio of total tax revenue to GDP), TR, and an index of the degree of regulation (REG) in India. Our primary sample period is 2010 to 2016 to match that of Giles [1], who also provides data sources. Somewhat earlier data on the causal variables are available and are used in the construction of certain moving averages in our analysis. The choice of these two input variables is of course itself somewhat subjective, and work in progress explores the implications of modifying the input set. In each case, we expect a positive association between the causal variable and the size of the Indian underground economy. In fuzzy parlance, "if taxes are high and if the degree of regulation in the economy is high, then we would expect the size of the Indian underground economy to be high".

Not only is the choice of causal variables subjective, but so is the specification of the boundaries of the fuzzy sets. At what point do taxes change from "average" to "high"; and at what level does the degree of regulation change from being "low" to "very low", *etc.* Accordingly, it is important to conduct a range of sensitivity tests to determine the robustness of our results to these and other choices. It is important to note, however, that there is no need to assume anything about the functional form of the hypothesized relationship between taxes and the degree of regulation on the one hand, and the size of the underground economy on the other. The basic approach we adopt, then, is to first define fuzzy sets associated with the values of the two causal variables. Then for each variable in each year, we assign association values with the subjective levels; and then we use decision rules to establish a level for the underground economic indicator (or index), using the fuzzy operators. The details of this procedure are presented in the next section.

IV. ANALYTICAL DETAILS

A. Data Break Points

There are several possible ways to create "benchmarks" to quantify what we mean by "high", "low", *etc.* in the present context. Here, we use a moving average value for each of the TR and REG. To take account of a possible electoral cycle in the data, a minimum of six years' data has been incorporated into the moving averages. As we wish to have an IUE measure for the period 2010 to 2015. For each series, and for each year, the average in the history of data gives us a "normal" value. Therefore, in 2015, this value is the average of the data from 2010 to 2015 inclusive. Once "normal" values have been established for each of the TR and REG in each year from 2010 to 2015, we then calculate quantitative associated levels of magnitude. This is done by taking one or two sample standard deviations around the "normal" value in each period:

TR: effective tax rate = Taxes /GDP				
Very Low	**Low**	**Normal**	**High**	**Extreme**
(VL)	(L)	(N)	(H)	(EX)
-2 SD	**-1 SD**	**Mean@ time=1**	**+1 SD**	**+2 SD**

REG: level of regulation = an index				
Very Low	**Low**	**Normal**	**High**	**Extreme**
(VL)	(L)	(N)	(H)	(EX)
-2 SD	**-1 SD**	**Mean@ time=1**	**+1 SD**	**+2 SD**

In this way two sets, each of five numbers, corresponding to TR and REG, are generated for each year in question. These sets are termed "break-points" in the subsequent discussion. For example, in 2010 the following break points emerge for TR: 0.2167912, 0.2277076, ***0.2386240***, 0.2495404 and 0.2604568. The highlighted value of ***0.2386240*** is the mean of TR over the period 2010 through 2015. Similarly, the value of 0.2167912 is the above mean value minus two times the standard deviation of this particular (moving) sample.

B. Break-points and Level Association

We then associate data values with categories of magnitude. Consider the above data for 2010. The actual data value for TR in that year is 0.2400330, which places it somewhere between "normal" and "high" in that year. "Fuzzy" or "multi-valued" logic uses non-crisp sets whose members are defined by levels or degrees of association, rather than by strict "all-or-nothing" membership status. So, a particular value of TR or REG can be associated with more than one set (or relative level of magnitude in our case).

In the above example, the 2010 value of TR is both "normal" and "high", but how "normal" and how "high" it is depends on its location relative to the break points in question. In fuzzy logic, the establishment of levels of association is governed by what are termed "membership functions". These can take various forms, according to one's prior beliefs, so another element of subjectivity enters the analysis. Here we use a simple linear or distance measure to assign levels of association.

For example, the value of TR in 2010 is closer to "normal" than to "high", and a harmonic assignation is used – that is, the weights are inversely related to the distances:

VL	**L**	**N**	**H**	**EX**
0.0000	0.0000	0.8709	0.1291	0.0000

A fuzzy logic membership function of the type used here will associate observations with at most two magnitude levels, the weights for which sum to unity. Extreme observations that fall below the lowest break point, or above the highest breakpoint, are given an extreme association value equal to the relevant "outer boundary" level. A value of unity associated with any particular level indicates complete membership, while a zero value denotes no membership at all.

C. Association Level and Decision Rules

Next, we create the decision rules that will determine how particular levels of association for each of TR and REG are combined to establish the levels of association for UE itself. These rules are necessarily rather arbitrary, but the method by which they are assigned may choose in the following table:

Rule	REGS	TR	UE	Degree
1	E	E	VB	1.0
2	E	H	VB	0.8
3	E	N	S	1.0
4	E	L	S	0.8
5	E	VL	A	0.8
6	H	E	VB	1.0
7	H	H	B	1.0
8	H	N	B	0.8
9	H	L	A	1.0
10	H	VL	S	1.0
11	N	E	B	1.0
12	N	H	B	0.8
13	N	N	A	1.0
14	N	L	S	0.8
15	N	VL	S	1.0
16	L	E	B	1.0
17	L	H	A	1.0
18	L	N	S	0.8
19	L	L	S	1.0
20	L	VL	VS	1.0
21	VL	E	A	0.8
22	VL	H	S	0.8
23	VL	N	S	1.0
24	VL	L	VS	0.8
25	VL	VL	VS	1.0

E=Extreme, H=High, N=Normal, L=Low, VL=Very Low, VB=Very Big, B=Big, A=Average, S=Small, VS=Very Small

The above table is then interpreted using simple "if-then" decision criteria. For example, recall that in 2010 TR is associated with "Normal" AND with "High", so using Rule 12 above, we say the IUE is "Big". The construction of the rules in the table is rather arbitrary - the "benchmark" rules (1, 7, 13, 19, and 25) are straightforward to assign, and we then followed [11] in assigning the others symmetrically.

The column labelled "Degree" in the above table provides a quantified degree of association for the UE series. For instance, continuing with 2010 as an example, Rule 12 associates UE with "Big" at a degree of 0.8. This indicates that UE is not perfectly associated with "Big" in that year, but only associated with the extent of 8/10ths. Again, a judgement is exercised in the assigning of these degrees.

D. Derivation of the IUE Series

The last stage of the analysis involves deriving the numerical series for IUE. This is achieved by attaching the values of 0.0, 0.25, 0.5, 0.75, and 1.0 to the levels "Very Small", "Small", "Average", "Big", and "Very Big" for IUE, weighted by the relevant levels of association. Recall that for each observation on TR and REG there are at most two association values, so there are at most (2 x 2) = 4 decision rules active for each IUE value generated. Here, the fuzzy "MIN", "MAX" operators act in place of the usual "AND", "OR" operators. Therefore, in 2010, the associating values for the four different levels of magnitude are:

TR	**Normal** 0.8709	**High** 0.1291
REG	**Low** 0.7339	**Normal** 0.2661

For 2010, there are four levels of magnitude to form four possible combinations with:

TR/REG	**Rule**	**IUE level**	**IUE Association**
1. N/L	18	S: 0.8 x 0.7339	0.5871
2. N/N	13	A: 1.0 x 0.1291	0.1291
3. H/L	17	A: 1.0 x 0.2661	0.2661
4. H/N	12	B: 0.8 x 0.1291	0.1033

From the previously listed decision rules, each combination of TR and REG level is associated with a level of magnitude for IUE, along with a degree. The first combination considered in this example (1) associates "Normal" for TR **AND** "Low" for REG to produce a level of "small" with a degree of 0.8 for the IUE series. The "Normal" level for TR is 0.8709 and the "Low" level for REG is 0.7339 using the **AND** operator results in choosing the smaller value of 0.7339 to multiply against the degree value for "small" IUE level. The third column under the heading level summarizes the calculations to this point. The last column incorporates the use of the OR fuzzy operator. For 2010, decision rule 13 and 17 are activated, both resulting in a level of A, raising the question as to which "Average" should be chosen, as they both cannot be true simultaneously. The final task is to attach values for the IUE levels.

Level	**Value**	**Weight**
S	0.5871	0.25
A	0.2661	0.50
B	0.1033	0.75

WTGU derivation

$$\frac{(0.5871 \text{ x } 0.25) + (0.2661 \text{ x } 0.5) + (0.1033 \text{ x } 0.75)}{(0.5871 + 0.2661 + 0.1033)}$$

= 0.3737 = IUE index value of 2010.

For the index value for IUE to lie in the interval [0, 1] the sum of the weights must equal 1.0, which is accomplished by dividing by their sum. The IUE index value of 0.3737 indicates that for 2010 in India the willingness of agents to "go underground" was less than neutral. An average agent, on balance, would tend towards working openly and above board.

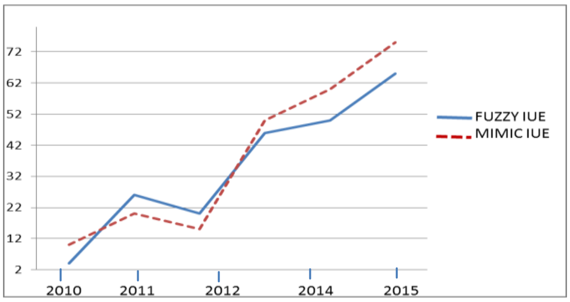
V. FINAL RESULTS

The resulting index values for each year have been scaled so that the "Fuzzy UE" series is comparable to that generated by Giles [1] - he used MIMIC model analysis, and levelled the resulting index by using a currency-demand model. The two different time-series of the Indian underground economy for 2010 to 2015 appear in Figure-1. We see there that although the two series follow a similar upward trend over time, their cyclical movements differ quite sharply. Of course, the true series of values for IUE is unknown, so which of these two measures is the more accurate cannot be determined.

We have examined the robustness of our "Fuzzy IUE" series to changes in the various subjective assumptions that have been made in its construction. We have found the results to be quite insensitive to the choice of the decision-rule "degrees"; to the use of the mean or the median as the "benchmark" for the break-points; and to the number of standard deviations used about these "benchmarks". We have also constructed corresponding "Fuzzy IUE" series using other causal variables that have been adopted in other analyses of the underground economy. For instance, the use of the inflation rate in conjunction with TR or REG yields strikingly similar results.

The most likely explanation for the different cyclical patterns of the MIMIC and FUZZY series is that the former is based on ten causal variables, and not just two. Extending the above fuzzy logic analysis to incorporate more than two causal variables is not straightforward in terms of the subjective judgements that need to be made.

Figure-1: Two measures of the Indian Underground Economy



As an approximation to a full such analysis, we have experimented with "hierarchical" structures of two-variable models. For instance, the resulting series for the IUE here can be taken as one new "composite" causal variable. A second such composite causal variable can be obtained in an analogous manner by using two quite different basic causal variables to generate a separate time path for the UE. Then, the methodology outlined here for the two-variable case can be applied to these composite inputs. This assumes, of course, an inherent "separability" of the effects of the two pairs of basic input variables, and this may be unrealistic. Our work to date along these lines has not yielded significantly different results to those reported in Figure-1.

VI. CONCLUSIONS

Clearly, much remains to be done to refine the procedures outlined in this paper. However, the preliminary results reported here are extremely encouraging, and do not appear to be especially sensitive to the various subjective prior judgements that have to be made in applying this methodology.

The size of the Indian underground economy is unobservable, but it is important for policy-makers to have reliable measures of its magnitude, trend, and cyclical characteristics. The recent resurgence of interest among policy-makers in Europe, the U.S., Canada, the U.K. and New Zealand on this topic makes it all the more timely to explore alternative methods for measuring the underground economy internationally.

Our use of fuzzy set theory and fuzzy logic in the novel in this context, and among other things it provides useful crosschecks on other measures that are available. Work in progress is extending this analysis in various ways, notably to incorporate a more comprehensive array of causal variables, and to consider alternative "membership functions". Finally, this same type of analysis can be used to measure other interesting, but intrinsically unobservable, economic variables. Examples include capacity utilization and price (and other) expectations.

REFERENCES

[1] Giles, D.E.A., Taxes and the Canadian Underground Economy, Canadian Tax Foundation, Toronto, Forthcoming, 2000.

[2] Data Engineering: Fuzzy Mathematics in Systems Theory ad Data Analysis. Olaf Wolkenhauer by John Wiley & Sons, Inc. 2001.

[3] Zadeh, L.A., *Fuzzy sets*, Information and Control, Vol. 8, 1965.

[4] Zadeh, L.A., *Fuzzy Sets and Applications*, Wiley, New York, 1987.

[5] Richardson, G., *The structure of Fuzzy Preference: Social Choice Implications,* Social Choice and Welfare, Vol. 15, 2008.

[6] Sengupta, K., *Choice Rules with Fuzzy Preferences: Some Characterizations,* Social Choice and Welfare, Vol. 16, 2009.

[7] Shepherd, D., and K. F. C. Shi, *Economic Modelling with Fuzzy Logic*, Paper Presented at CEFES'98, Society for Computational Economics, Cambridge, U.K., 1998.

[8] Shiangi Sengar, Sakshipriya and Urvashikhatuja, *Fuzzy set Approaches to Data Mining of Association Rule,* International Journal of Computer Science and Informatics, Vol.2, No. 4, 2012.

[9] Lindstrom, T., A Fuzzy design of the Willingness to invest in Sweden, Journal of Economic Behaviour and Organization, Vol.36, No.1, 1998.

[10] M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[11] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[12] (2016) The Incometaxindia website. [Online]. Available: http://www.incometaxindia.gov.in/

[13] Priyanka Kaushal, Neeraj Mohan and Parvinder S. Sandhu, Relevancy of Fuzzy Concept in Mathematics, International Journal of Innovation, Management and Technology, Vol.6, No.3, 2010.