International Journal of scientific research and management (IJSRM)
||Volume||xx||Issue||yy||Pages||zzzz-zzz||2018||
Website: www.ijsrm.in ISSN (e): 2321-3418

A practical strategy to improve econometric modeling–a case study for informal economy on the Republic of Macedonia

Dashamir Asani[1], Dode Prenga[2]

[1]PhD student, Department of Informatics and Statistics, Faculty of Economy, University of Tirana
ecc-da@hotmail.com
[2]Department of Physics, Faculty on Natural Sciences, University of Tirana
dode.prenga@fshn.edu.al

Abstract. Here we consider some proposal-steps in the calculation of the shadow economy as a latent or simply response variable in the models that involve high volatile observable as the money aggregates etc., and if the number of data point is small. In this case we propose to check for possible extreme behavior or self-organization regimes present in the series by testing e log-periodic fit to the data. To improve the linear regressions critical points (if found) have been excluded from series by simply truncating them. Next, the presence of more general regimes is analyzed using empirical mode decomposition techniques, and we estimate that the best truncated series to be used should exclude the edges of such regimes. In the case of short term regimes, we propose to use series in intervals that include many cycles. This technique worked for the calculation of informal economy in the Republic of Macedonia for the short period of [20o4, 2016] but it is supposed to improve calculation for other cases as well

Keywords: informal economy, linear model, log-periodic, empirical mode decomposition**.**

I. Some remarks on modeling with economic and financial time series

In a general consideration for econometric modeling we attempt to fit observed values for an economic observable to a functional of other economic variables. In this sense, from mathematical point of view it is plausible that all variables involved should be in stationary state. In special case of estimation of the unregistered economy, money models assume that it has a direct indicator the money in circulation, or money outside deposits, and therefore this variable is involved in the modeling procedure. Concrete applications are detailed largely in literature as for example in [1], [2] etc. Those models are generally based on the assumption of direct and linear relationship between response variables and indicators. More complex relationships are possible as given by structural equation approach, but linear relationships is presumed to hold again as seen in [3] for example and many other applications too. Those models involve even highly volatile quantities as money aggregates are, and some analytic strategies as removing unit root etc., are necessary. Detailed remarks on those procedures we have found in the reference [4], [1], [2] etc., However, no-linearity and non-stationary issues have been considered as problematic issues in models and regressions based on them. Here it is worth to mention the fact that when dealing with indexes and money aggregates we should take into account the self-originations behavior which has been verified and analyses discussed in [5], [6] etc. In this case it has been remarked that those structures were a specific case of fractals called discrete scale of invariance, that typically are characterized by extreme behavior which became heightened near a critical point. We underline that near those points the variable well behave in un unpredictable fashion and therefore not appropriate to be used in regressions. Note that detailed mathematical analyses as described in [4] or [7] considers and suggest effective method to deal with non-stationary series when performing linear modeling, but we want to stress specifically that element in our case study where critical behavior was present in the dynamics of at least one variable and series consist on a small number of data. In this aspect we do not want to go deeper in pure analytic procedures but we want to evoke some practical

steps that could help in the calculation. So, trying calculation of the informal economy in Republic of Macedonia for the period [2004, 2014] we noticed that the results were not so good when using CDA (currency demand approach) and MIMIC (multiple indicators, multiple causes) models. Similar calculations have been reported in many references as in [8], [9] etc., but apparently by using more data. But here we fixed this period because the methods of recording econometric and financial data have been improved significantly by end of 2010 in the country. So, trimestral and even monthly data could be easy found in official sites. In our initial work we proposed to use monthly data to improve the calculation and to overcome the problem small number of data that make regressions less credible. Remember that in models mentioned above we must use variables of the type money aggregate or their ratios based as for example the money in circulations or its ratio to the another money. In particular, the currency in circulation as many other money aggregate could present extreme behavior by nature as referred in [5], or may conduct a regime-change in a specific time. In a simple regard we can expect that a value can be measured if it's in a stationary state which in turn can be characteristic for stable economies. But In transition economies (as our case is to be), intervals of nearly stationary economic state for parameters are expected to be typically short because there exist a point of total regime change. In turn, those shortcomings could affect the calculation and therefore we considered them in an preliminary analysis.

II. Some practical proposals and approaches

We propose to identify the position of the economy itself from a more general perspective, so we can acknowledge the estimation, even by a large uncertainty. So, as a starting point, we calculate the expected position of the variable under study and tried to recognize the general trend of it. Herein we refer to the parametric distribution analyzed in [10] , [11] and [12] say

$$p(x) \sim \alpha\left(1-\beta(1-q)(x-\mu)^2\right)^{\frac{1}{1-q}} \qquad (1)$$

called q-Gaussian. They have specific advantages in the fact that q represent the distance from the Gaussian distribution, hence stability issue is discussed according to this parameter. Using this relationship we estimated the margins of informal economy for the country by using overall distribution of global economies and acknowledged the general tendencies. We were careful to not impose a trend for our variable, but to acknowledge the most possible development on it. We considered other analysis as given in [13] etc., to learn from the q-distribution fitted and to analyses it. We expect that general tendency on informal economy could be opposed form the one calculated in our case if there exist complex dynamics on the system. Self-organization and discrete scale of invariance (DSI) are typically such dynamics as seen in reference [5]. Therein, a specific group of functions called log periodic, have been proposed to catch DSI or fractal dynamics of the quantity under study. It reads

$$P = P + a(t-t_c)^m\left[1+b\cos(\ln\omega(t-t_c)+\varphi)\right] \qquad (1)$$

where t_c is the critical time that according to [5] etc., it better signify the moment where the regime change is most probable to occur. In theoretical deduction there are more terms than in (2) as introduced from the reference [5] and analyzed in [6] or in [14]. In summation, we believe that the linear model analysis will give better results if series does not contain critical in which nearby the behavior is quite complex. In following we will discuss the dynamics only for the part related to the critical time tc, leaving the cyclic frequency, the power growth (m) parameter etc., out of the discourse.

Another problem that is expected to affect the linear assumption or linearization procedures is the presence of regimes. It could be plausible that we work in one single regime and far from its edges or if analysis could include many of them if present in series. A very intriguing method based on error reduction analysis called empirical mode decomposition technique is discussed in [15] and [16], and in many applications for real systems. Shortly, the method identifies so–called modes which are not strictly orthogonal but gives principal series under e threshold condition, that at best reproduces the signal. It adaptively represents non-stationary signals as sums of zero-mean AM-FM components [15], [17]. Highly dynamical series that are difficult to be examined with analytic techniques could be investigated using EMD approach instead. Many Improvement of such calculation have been introduced successively and some of them are discussed in [17] .

III. Estimation of the boundaries from descriptive analysis

By straightforward analysis and arguments and appropriate extension as well discussed in [1], the

size of informal economy is calculated for individual countries in many works or in a general review as in [2]. Referring to them in our analysis we identify that position of the Informal Economy in the RM is found among countries with values centered around 35-40% in the period [2006, 2016] as seen on the picture of Figure 1.

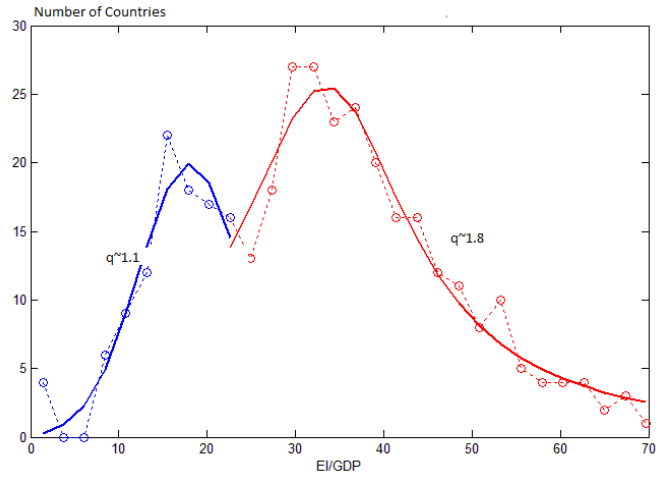

Figure 1: PDF graph for the fraction EI/GDP in the world. Data sources [2]. Author calculation

Next by analyzing the distributions using equation (2) we observe that global informality is characterized by two classes, one represented by a near to Gaussian distribution centered in 15% that belong to developed countries and another presented by high distorted Gaussian centered in 40%. The non-stationarity of the distribution is measured with q parameter in similar fashion as given in [10], or [12]. We observed that the general tendency of evolution for the group II is governed by a stabilization process because q-parameter obtained by the fit of empiric data to the equation (2) becomes slightly smaller each year. Therefore we can accept that the RM economy could have an estimated informal economy around 40% and in the most probable development it has a stabilizing trend toward smaller values. However, no significant change is expected to occur in short term, and it is possible the adverse movement too. Considering other calculations we noticed that informal economy has known different trends during [2004, 2014], so form this view we only select the boundaries as reference value. Specifically we apply CDA and MIMIC models based on different group of factor variables, and when we obtained values too different form the above findings, we changed the set of variables.

IV. The dynamics of some important variables and factors

Considering problems arising form the very small number of data points (~12) for all variables used in regressions, the use of trimestral or even monthly has been appreciated as a fruitful step. But we observe that those series were highly dynamic and according to the proposed steps above, we practiced the analysis for cycles, regimes and critical behavior for variables involved in CDA or MIMIC calculations. In CDA usually the response variable is taken the ratio of money in circulation with some other money as explained in [1], [2] etc. Here we used C/M where C is the amount of money in circulation and M is the money say base money (M_0), broad money (M_2) etc. The regression is performed for series that does not include extreme behavior or critical points so we estimated them. To realize that, we tried to find an acceptable log periodic fit of series. In this procedure we changed the start and the end date for series using an ad hoc genetic algorithm introduced firstly but us for other purpose [18]. Firstly we considered trimestral data so we have 4 times more points than yearly ones Moreover those series are plausible in CDA or MIMIC modeling because in this timing we have valuable variable as GDP, GNI, unemployment. Unfortunately monthly timing could not apply to such important variables, so the analysis for C.M variable in monthly records is performed to a batter understanding the dynamics in itself, not for direct use ion the informal economy modeling.

By reading a log-periodic fit to trimestral data of C/D variable, we obtained that possibly the system has entered a near to DSI regime which is expected to change near January 2023 (c.p). This result is not truthful in quantitative view, because it goes too far to be credible. But qualitatively we qualify this variable as appropriate in regressions so modeling with trimestral data is performed for the period [2004, 2016]. However, the results were not satisfactory and we obtained different values of informal economy using different models. Further analysis is needed. For a better understanding the dynamics of the series, we used monthly data in the successive step. As mentioned herein and according to informal economy models those data are un-useful because many economic variables are not defined for such a short period but our strategy is oriented in the knowledge of special point which should be avoided in regressions, so we proceed with them. Using monthly series of the ration C/D, we observed that a log-periodic regime appear in its trend. It starts near coordinate 33 that coincide to the year 2004, and the critical time is expected to be around 171 month later, that is around 2017 c.p,. So

far, the end data of our series (2016) seems to fall in a particular region where extreme behavior is characteristic. This "problematic" edge would better be excluded and therefore we use the period [2005, 2014] in our calculation for informal economy. In this case values of informality obtained using different models lies in the range of 35-40%

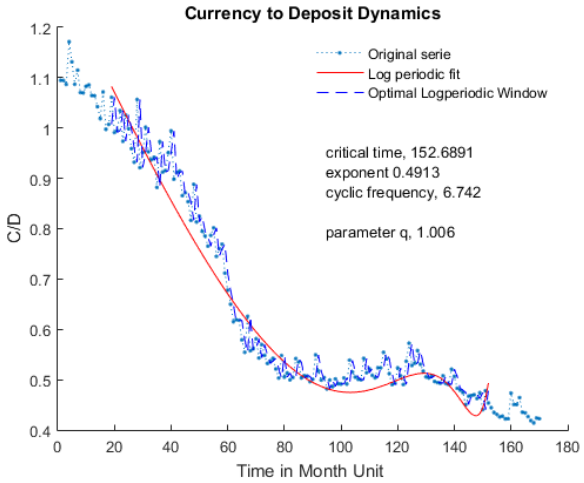

Figure 2: The log periodic approach for C/D variable

Considering rapport C/M_2 as another candidate for response variable in CDA model, we observe that a shorter regime is present in the underlining behavior. It starts at coordinate 112 that is April 2012 and finish at coordinate 170 that is January 2017. Results are presented in the Table 1. Under such condition all the period discussed [2004, 2016] contains regime changes for variable the C/M_2 and moreover, the last one correspond to a bubble-like behavior around end of 2017. It consists in an extreme event by which the trend could change drastically, so it should be avoided in linearization assumption. Practically the variable C/M_2 is considered as highly non-stationary and this property can affect the quality of regressions in linear model in this period, Figure 3. The data points around 2017 have been cut from the series used in our linear models. Next, we studied another variable of this kind seeking for a more stable behavior. By taking into account the important weight of remittances in the country, we proposed to analyses the broad money M_{22} that includes foreign currency used in the country. In this case we obtained that more than one critical–like process might underline the dynamics of variable C/M_{22} as seen in Figure 4. Those results suggest that modeling using monthly variable should avoid again the edges of the regime but we expect that the presence of short period pseudo regimes could minimize the effect of those particular points and hence the variable C/M_{22} could ore admissible to be used in models.

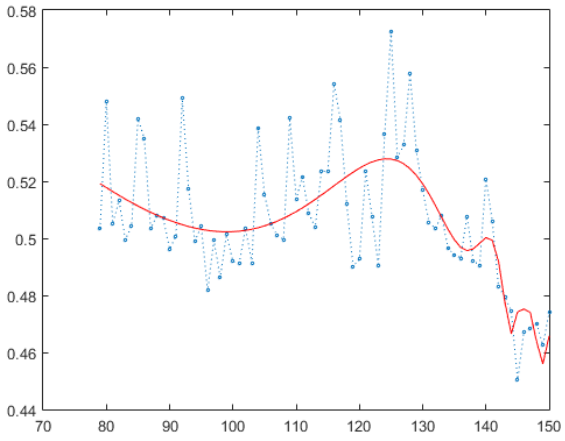

Figure 3: Extreme behavior near 2017 for C/M_2 variable

With acceptable statistical significance we identified a relatively medium term regime that start at coordinate 82 and die at 159. There are some short range processes lasting around 2 years or so as shown in figure 4.

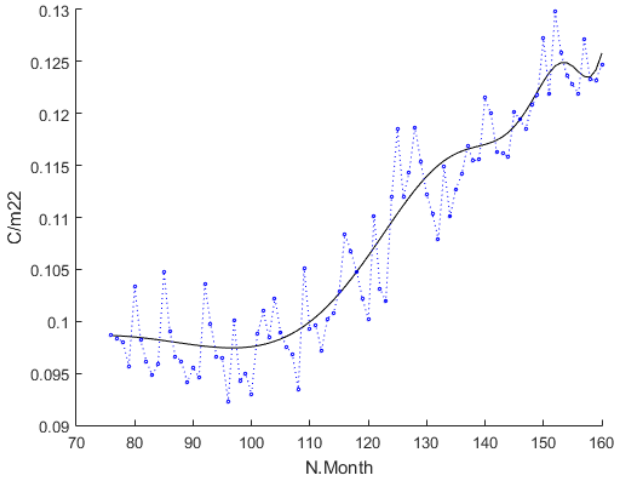

Figure 4: Log-periodic behavior of C/M_{22} variable

In this case we observe that CDA and MIMIC approach gives similar result for all the period considered, say [2004, 2016]. Up here we underline to important facts. First, findings herein give some information for precaution in the linear analysis, but are they are not sufficient for deeper knowledge and modeling. This originates from the fact that monthly data are not typical one to be used in models because many indexes and econometric variables are calculated by nature in financial year basis. Second, the dynamics observed on monthly data is not expected to be transferred in yearly data, but the presence of special point and especially in there is verified critical behavior strongly suggest that linear relationship could be destroyed nearby them. However, the verification of the log-periodic presence needs for more data and more sophisticated analysis so it exists the possibility for false alarms too. Being aware of this, we use those result as precaution measure in linear regressions as

mentioned above.

Table 1: some data from log periodic fit

	T.Start	T.End	Critical Time	Exponent	Cyclic Frequency	q_Parameter
C/M1	33.00	158.00	171.01	0.54	6.74	1.0006
C/M22	79.00	168.00	111.95	0.34	9.70	1.0016
C/M2	38.00	165.00	129.71	0.68	7.49	1.0008
C/D	13.00	159.00	165.68	0.76	7.80	1.0017

In the case where the presence of a specific DSI regime is likely to be present for a period inside the whole interval considered, the knowledge trend on theta characterize the whole series is very important. We paid attention to this point in the next paragraph.

V. Empirical regime identification

Here we used the Empirical Mode Decomposition to obtain the underlining trend for whole interval studied, not stepping on the characteristic local or self-organization dynamics as above. This method, decompose a complicated signal in so called more basic ones named "modes" which even not orthogonal, can be used as a basis for the signal. Interestingly, the last mode gives the trend that underlines the all series. Important comments for such application have been provided in [15],[16] and detailed calculation has been provided in [17] and we are not repeating them. The method is appropriate in the case of analysis for complicated signals as practically the rapport of two financial series is to be. We can read the most realistic trend and possible regimes by analyzing IMF. We notice that by construction the last IMF signify the long range trend of the series. In Figure 5 the last IMF on C/M_1 analysis shows a possible regime that is expected to end around coordinate 240 and it has possibly start near to our start point of the series. Considering the above discussion that the regime contains log-periodic behavior as well, it is supposed that empiric regime found here in coincide with self - organization one. Therefore we should choose our series as initiating later than 2004, but no need to cut terms at the other edge 2016), because the end of the regime is many points after the end of our period under study. In the same way we obtained that the variable C/M_{22} has an underlying regime that started many month before our first point in series considered (2004) and will finish around 2020, so series [2004,2016] is situated far from regime edges and hence, according to the above assumption it is appropriate to be used in linear models.

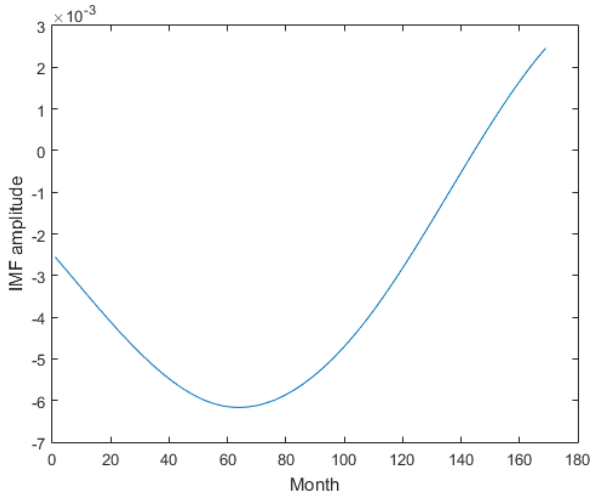

Figure 5: EMD modes for variable C/M2

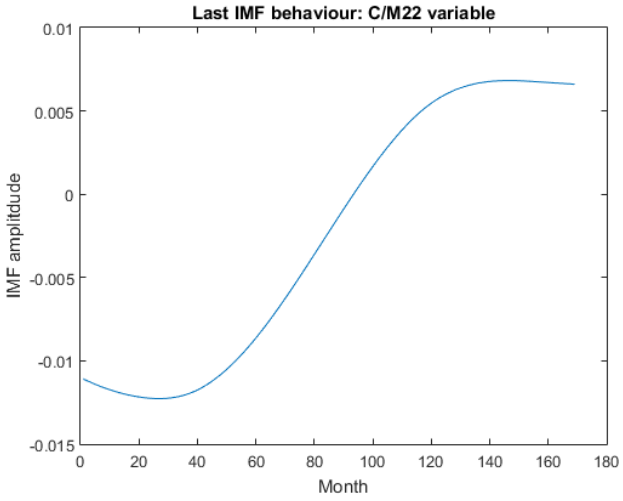

Figure 6. Last IMF behavior for C/M_{22} variable

Practically we selected this last variable to be used in the modeling for estimation of the informal economy. In this case the result obtained by different methods as CDA or MIMIC were found similar.

VI. Conclusions

In this work we propose to improve the calculation in linear modeling for econometric variables in the case of short period data series, or if complex dynamics is present on it. In modeling informal economy (for Republic of Macedonia) as a hidden variable, we obtained better result if using series that does not include critical points, edges of the regimes etc. In our case-study we avoided the critical-like point's effects by simply cutting the series before the extreme dynamics appear on the series, and this last is localized as the critical time obtained by a log-periodic fit. In the same way we identified the presence of the regime reading the last Intrinsic Mode Functions (IMF) , and in this case we propose to use series that does not include the start and the end of the regime or alternatively we the series that include many regime are qualified to

follow linear models and regions too. Another step that can guarantee the model to not be unrealistic could consider the position of the concrete economy as part of global one, but this procedure is legitimate to only obtain the boundary or envelope solution for values. It is advisable that those steps should be taken carefully and they should be based on more detailed comments according to concrete system under study. It worked in our concrete case and it supposed to be fruitful in other circumstances where the number of data in series is small and the processes underlining the dynamics of the variables are so complex that can affect directly the validity of standard modeling routines and linear regressions.

References

[1] Schneider, Friedrich and Willimas, Colin C. , The Shadow Economy (June 1, 2013). Institute of Economic Affairs, 2013.

[2] Friedrich Schneider. Estimating the size of the shadow economies of highly developed countries: selected new results. https://www.cesifo-group.de/DocDL/dice-report-2016-4-schneider.

[3] F. Shneider, R. Dell'Ano . Estimating the underground economy by using mimic models: a response to T. Breusch´s critique by dell'Ano , Working papër no. 0607. july 2006

[4] Kalr Jorskog, Arthur Goldbwerg. Estimation of a model with multiple indicator and multivple causes of ingle latent variable. Journal of the American statistical association. Volume 70, issue 351(Sep. 1975)

[5] Didier Sornette, Ryan Wodard, , Wanfeng Yan, Wei-Xing Zhou. Clarifications to Questions and Criticisms on the Johansen-Ledoit-Sornette bubble Model. ETH-RC-11-004.

[6] Zhi-Qiang Jiang, Wei-Xing Zhou, D Sornete, R. Woodard, K. Bastiaensen, p. Cauwels, "Bubble Diagnosis and Prediction of the 2005-2007 and 2008-2009 Chi nise stock market bubbles", Journal of Economic Behavior and Organization 74, 149-162 (2010

[7] Steiger, J.H. (1990), "Structural model evaluation and modification," Multivariate Behavioral Research, 25, 214-12.

[8] Vesna Garvanalieva, Vlatko Andonov, Marjan Nikolov. Shadow economy in Macedonia. Center for economic analyses (CEA). www. cea. org. mk Kadri G Yilmaz , Sedat Belbag. Prediction of Consumer Behavior Regarding Purchasing Remanufactured Products: A Logistics Regression Model. International Journal of Business and Social Research Volume 06, Issue 02, 2016

[9] Ramona Angelescu, Laura Chappell, Alex Glenie, George Mavrotas. , Dhananjayan Sriskandarajah. Development on the move: measuring and optimizing the economic and social impacts of migration in the Republic of Macedonia. EDU research. Development on the move - country study of Macedonia

[10] Constantino Tsallis Computational applications of non-extensive statistical mechanics. Journal of Computational and Applied Mathematics 227 (2009) pp 51-58.

[11] G.P. Pavlos, M.N. Xenakis,L.P. Karakatsanis, A.C. Iliopoulos, A.E.G. Pavlos D.V. Sarafopoulos. Universality of Tsallis Non-Extensive Statistics and Fractal Dynamics for Complex Systems. Chaotic Modeling and Simulation (CMSIM) 2: 395-447, 2012.

[12] S. Umarov, C. Tsallis and S. Steinberg, On a q-central limit theorem consistent with nonextensive statistical mechanics, Milan J. Math. 76 (2008) [DOI 10.1007/s00032-008-0087-y].

[13] Liza Borlnad. aTheory of Non-Gausian option Pricing: capturing the smile and the skew (2004) www. Spiedl. org/data/ conference s/SPIEP

[14] P Geraskin Everything You Always Wanted to Know about Log Periodic Power https://kr.mathworks.com/matlabcentral/answers/uploaded.../frolova-paper-15.pdf

[15] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung and H.H. Liu, "The empirical mode decomposition, and Hilbert spectrum for nonlinear and non-stationary, time series analysis," Proc. Roy. Soc. London A, Vol. 454, pp. 903–995, 1998.

[16] Zhaohua Wu and Norden E. Huang. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. Advances in Adaptive Data Analysis. Vol. 1, No. 1 (2009) 1-41.

[17] G Rilling .On empirical mode decomposition and its algorithms - citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.586.812

[18] D. Prenga, M. Ifti. Complexity Methods Used in the Study of Some Real Systems with Weak Characteristic Properties. AIP Conf. Proc. 1722, 080006 (2016)

[19]

Author(s) Profile

Dashamir Asani, has graduated in economy and has finished the research in framework of his PhD in management and marketing. His major researches were focused on modeling informal economy and some application on concrete

Dode Prenga, associate professor, has graduated in 1992 at the Faculty of Natural Sciences. He is is lecturer of nonlinear dynamics in the Faculty of Natural Sciences, University of Tirana. The main researches in are complex systems with focus on econo-physics and socio-physics