

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

Dashamir Asani¹, Dode Prenga²

¹PhD student, Department of Informatics and Statistics, Faculty of Economy, University of Tirana

²Department of Physics, Faculty on Natural Sciences, University of Tirana

Abstract:

Here we consider some proposals in the calculation of the shadow economy using linear models that involve if the number of data point is small and observables are highly dynamical. We firstly suggest checking for possible critical points in the series by testing a log-periodic fit to the data. To improve the results from the model, we used intervals that do not contain critical points. Next, the presence of regimes is analyzed by using empirical mode decomposition technique, 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 [2004, 2016] but it is supposed to improve calculation for other similar cases as well.

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

1. 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 ones. From mathematical point of view it is plausible that variables involved in the model could be stationary. If data series are not stationary the remove of unit root by some additional operation is needed. In some models for estimation of unregistered economy, a key variable that is money aggregate [1],[2], by nature is non-stationary and usually volatile. Procedures of data elaboration for this case are detailed largely in literature as for example in [2], [3] etc. More detailed algebraic analysis and remarks on those procedures have been described in the reference [4], etc. However, indexes and money aggregates could be characterized from a special non-linear dynamics called self-organization behavior [5], [6] that could not be elaborated easily to be used in linear models. When dealing with a model of estimation for a latent quantity as the informal economy for example, it worth to consider the additional effects of critical behavior in the dynamics model's variables. Those effects are expected to become important if the series consist on a small number of data. In the calculation of the informal economy in Republic of Macedonia for the

period [2004, 2014] using standard models CDA (currency demand approach) and MIMIC (multiple indicators, multiple causes) as proposed in [2] etc., we noticed that the results were not so good. Calculation for other periods have been reported in many references as in [8], [9] etc., but apparently by using more data. In our initial work we proposed to use monthly data to improve the calculation and to overcome the problem of small number of data that make regressions less suitable. But the key variable of the model, the currency in circulation, in monthly series showed selforganization-like behavior. So we must consider this problematical behavior in the modeling process. Next, in a brief regard we can expect that an observable can be measured or known accurately only if its (eigen) state is stationary. But in transitive economies (as our case is to be), there exist at least one point when a total regime change occur. Other disturbances could be present too. In short, specific systems are not as good as mathematical models want them to be. Therefore we assumed that removing those shortcomings could have improved the result of linear modeling and therefore we considered them in a preliminary analysis. In our case study we considered the analysis of the state itself from a general point of view and next we explored about specifics of the dynamics on the series used for the calculation.

2. Some proposals and approaches

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

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

called q-Gaussian. They have specific advantages in the fact that q-parameter therein represents the distance from the Gaussian distribution [10]. Using this relationship we estimated the margins of informal economy for the country by using the distribution of global economies and acknowledged the general tendencies considering other analysis as in [12] or [13]. Next, we expect to improve the result of the model by addressing complex dynamics issues for money aggregates variables used in regressions of CDA or MIMIC models. In reference [5] the self-organization and discrete scale of invariance (DSI) have been analyzed theoretically. Therein, specific functions called log periodic have been proposed to catch DSI or fractal dynamics of the quantity under the study. It reads

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

where t_c is the critical time. According to [5] it signifies the moment where a regime change is most probable to occur. In other consideration there are more terms than in (2) as analyzed in [6] or in [14]. We assume that linear models are expected to fit better if series do not contain critical points. In following we will discuss the dynamics related to the critical time t_c , leaving the cyclic frequency (ω), the power exponent (m) out of the comments. Another problem that is supposed to affect the linearization procedure on models mentioned above is the presence of regimes. It could be reasonable to use series belonging to one single regime and far from its edges, so we propose to analyses the regimes on the series used. 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 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]. In following we will describe the

procedures used in our concrete calculation.

3. Estimation of the boundaries from descriptive analysis

By straightforward analysis discussed in [1], the size of informal economy has been calculated for individual countries in many works [2]. International organization as IMF or World Bank updated every time their own estimation of this parameter.. Referring to them we estimate that the Informal Economy of the country for the period [2006, 2016] could be found in the zone centered around 35%-40% as seen on the picture of Figure 1. Remember that this parameter is usually represented as fraction of the GDP.

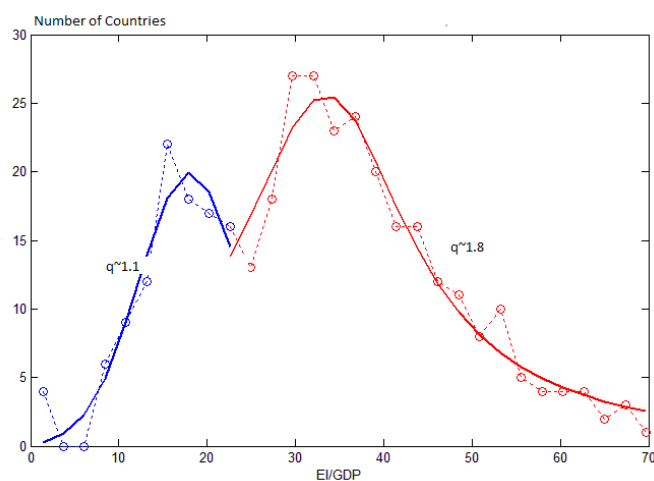


Figure 1: PDF graph for the fraction EI/GDP in the world. Data sources [2]. Author calculation

Using equation (2) we observe that the informality in world is characterized by two distinct classes grouped in two separate distributions. The first one is characterized by a near to Gaussian shape centered at 15% and include developed economies. The other shows a distorted Gaussian distribution centered at 40%. Examining the q parameter in different periods, we obtained that the distribution of informal economies on the second group shows a tendency for stabilization. This is read from the value of q parameter that became marginally smaller by time. Therefore we can accept that the RM economy might have an estimated informal economy around 40% in this period and probably has a stabilizing trend toward smaller values. As an immediate consequence, the estimations for informal economy of the country that result outlying this reference values must be reviewed in details or even rejected. Considering other calculations as in [8] or [9] we noticed that informal economy has been reported in different trends during [2004, 2014]. Therefore, from the above picture we fixed

the boundaries. Particularly the threshold is considered at 15% that belong to the developed countries as seen in the Figure 1, because we are sure the economy under analysis does not belong to this group. By empiric application of linear models we obtained values of the country informal economy outlying those mentioned above, therefore we proposed more careful analysis based on the dynamics of variables used.

4. The dynamics of some important variables and factors

In the CDA method, the indicator of informal economy is assumed to be the ratio of money in circulation with some other money aggregates or component as explained in [1], [2] etc. We used the ratio C/M where C is the amount of money in circulation and M were taken the base money (M_0), narrow money (M_1), broad money (M_2) etc. according to above remarks, the regression should be performed for series that does not include extreme behavior or critical points. To realize that, we checked the log periodic fit of the data series. In this procedure we changed the start and the end date for partial series under test, using an ad hoc genetic algorithm as in [18]. Firstly we considered trimestral data so we had 4 times more points than yearly ones, but still the number is smaller for such quantitative analysis. But those series are acceptable in CDA or MIMIC modeling because the values for important model variables as GDP, GNI, were still available in this format. So we learned qualitatively the behavior in this case. Monthly series could be better in this view, even they cannot be used directly in the CDA model for example because some important variables doesn't appear in monthly records. But we can still analyze them for a better acknowledgment of the dynamics characterizing the variable under study. For sure some properties are expected to disappear in longer period series but 'unwanted' zones could have been localized so far.

By reading a log-periodic fit to trimestral data of C/D variable in [2002,2016], we obtained that possibly the system has entered a near to a DSI regime which is expected to change near January 2023 (c.p). This result is not truthful in quantitative view, because it outlays the period of the study but can be considered as an argument that critical zone is not in the adjacent of the end of the period considered. Using monthly series of the ratio C/D , we observed that a good log-periodic fit has been found and a regime is expected to be present in its trend. It starts near the coordinate 33 that coincide to

the September 2004, and the critical time was 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 28%-43% and do not differ remarkably. We used this approach to detect which variable was the most suitable to be used as first indicator of informal economy according to the CDA models. But theoretically it other C/M ratios are more important for such models.

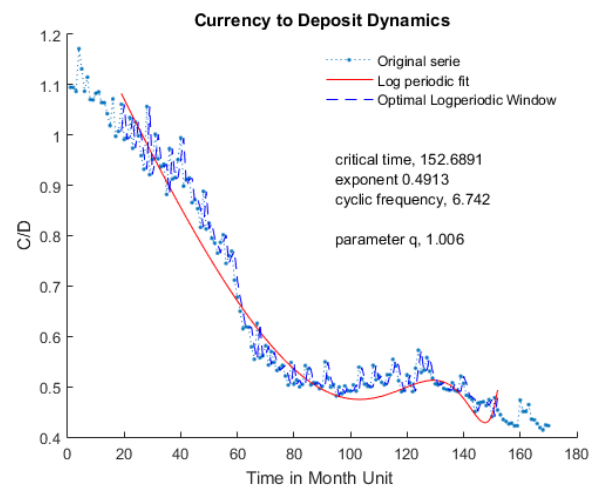


Figure 2: The log periodic approach for C/D variable

Considering rapport C/M_2 as another candidate for response variable in CDA model, we observe that a shorter regime has been found in its underlying behavior. It probably started at coordinate 112 that is April 2012 and finished at coordinate 170 that is January 2017. Results are presented in the Table 1. Under such condition all the period discussed [2004, 2016] contains a regime change point for variable the C/M_2 and moreover, it correspond to a bubble-like behavior around end of 2017, Figure 3. The data points around 2017 impose high deviations when used in our linear models as CDA or MIMIC. Next, taking into account the important weight of remittances in the country, we proposed to analyses the ratio of C to the broad money M_{22} . 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. We expect that the presence of more than one critical point would minimize the deviation effect of each other. With acceptable statistical significance we identified a relatively medium term regime that start at coordinate 82 and

die at 159.

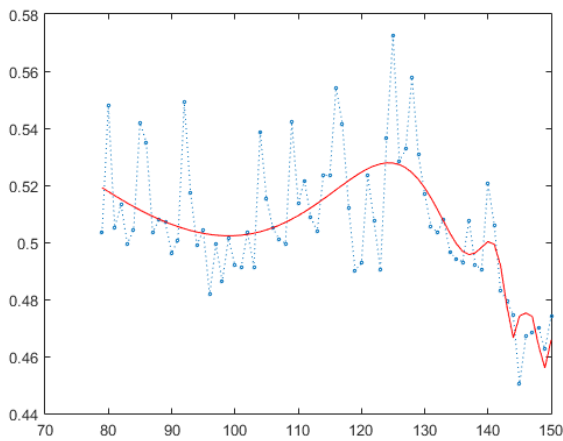


Figure 3: Extreme behavior near 2017 for C/M_2 variable

There are some short range processes lasting around 2 years or so as shown in figure 4. Hence the variable C/M_{22} spanning all target intervals [2002, 2016] is found admissible to be used in models.

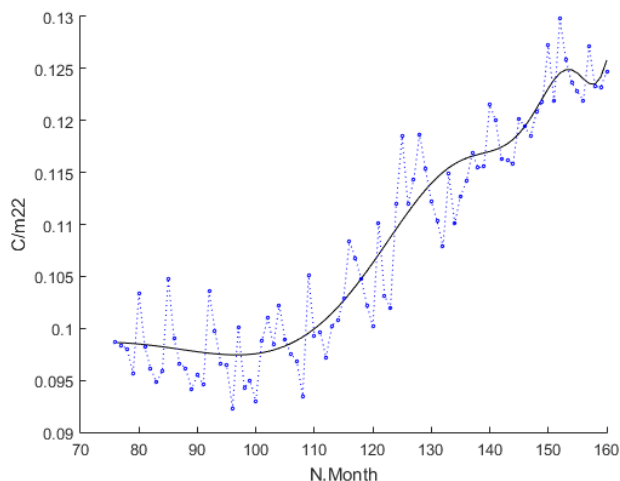


Figure 4: Log-periodic behavior of C/M_{22} variable

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 they are not considered sufficient for a trustworthy modeling. The dynamics observed in monthly data is not expected to be transferred in yearly data, but the presence of special points strongly suggested 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 false alarms are possible. 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

As long as the presence of regimes is deduced but its nature is not verified, the best strategy is to identify the real trend and regimes no matter what type they really are. In the case where the presence of a DSI regime is likely to be present we identify a critical point that behaves differently from the others. But other processes could be present and the resulting regime is complex, hence unknown. In this case we suggested the analysis of empirical regime using empirical mode decomposition.

5. Empirical regime identification

We used the Empirical Mode Decomposition (EEMD algorithm) [17] to obtain the underlining trend for whole interval studied, neglecting characteristic local or self-organization dynamics as above. By this technique a complicated nonlinear and non-stationary signal was decomposed in so called “intrinsic modes” which are not rigorously orthogonal. Interestingly, the last mode gives the trend that underlines the data. Important comments for such application have been provided in [15],[16] and detailed calculation in [17]. We noticed that by construction the last IMF signifies 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 started near to our start point of the series (January 2002). 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 if we choose our series in the interval [2004, 2016], the edges effect of regimes whatever type they are, is removed. For better analysis we should consider other IMF and not surprisingly 2-10 year cycles have been observed, but those findings has not been used in our work. In the same way we obtained that the variable C/M_{22} has an underlying regime (last IMF cycle) that started many month before our first point in series and will finish around 2020, Figure 6. So in the period [2004, 2016] contains no edges of medium term regime and hence, according to the above assumption it is appropriate to be used in linear models.

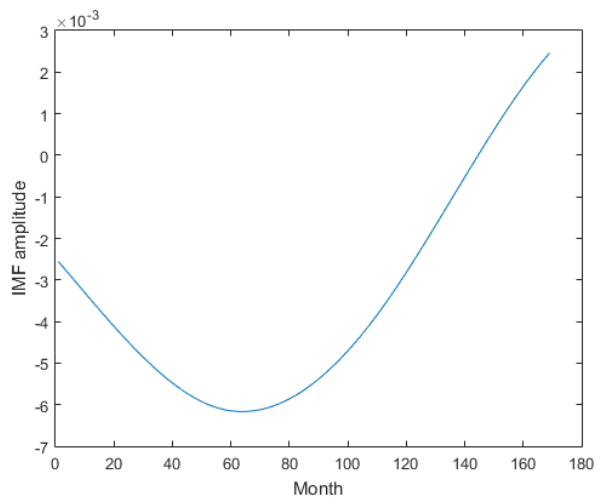


Figure 5: EMD modes for variable C/M2

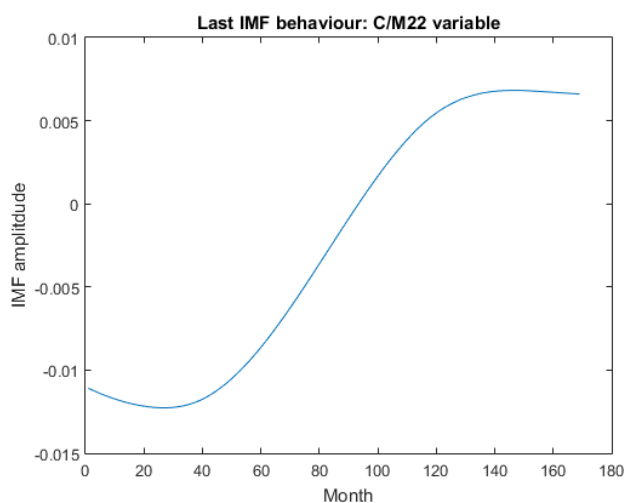


Figure 6. Last IMF behavior for C/M₂₂ variable

We performed such corrected calculation and obtained that informal economy in the country has reached a maximum of 38%-41% around the years 2010-2011 and at 2016 has decreased to the value 32%-34%. By now is slowly going down.

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. It worked in our concrete case and it supposed to be fruitful in other similar circumstances

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 paper no. 0607. july 2006
- [4] Kalr Jorskog, Arthur Goldbweg. 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 Garvanaliev, 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 Srisikandarajah. 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-Gaussian option Pricing: capturing the smile and the skew (2004) www.spiedl.org/data/conferences/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)

Author(s) Profile

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

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