A Thematic Discussion on the Role of Descriptive Analysis in the Study of Informal Economy–a Case Study for the Republic of North Macedonia

Dashmir Asani[1], Elmira Kushta[2], Dode Prenga[3]

Abstract: In this paper we present some thematic aspects related to the study of informal sector in the Republic of North Macedonia for the period [1998, 2016]. So, by conventional application of CDA model we perceived problematic issues in the estimation of the informal economy. To recover the model aftermaths, we assisted the calculation procedures by additional descriptive analysis aimed on avoiding the regression disturbances from critical dynamics. By evidencing the presence of self-organization regimes in some money-type variables, we identified the intervals where data series behaved highly nonlinearly. Consequently, by excluding parts of series up to a reasonable point (herein before 2004), the CDA predictions for the size of informal economy and the relationship with its factors has been improved remarkably. Next, we used factorial analysis to facilitate the design of the n-p-m MIMIC model. This investigation applied in 22 candidate factors suggested the configuration 9-2-3 as optimal structure for the model. In particular this model anticipated two terms confined in the latent variable structure, respectively in the range [0.35, 0.38] and [0.023, 0.08] part of the GDP. We assigned them as subparts of informal economy reflecting differenced influences or weights of certain cause factors. We observed that the estimated effect of the factors included in the model followed theoretical expectation.

Keywords: linear econometric model; informal economy; factor analysis; correlations.

JEL Classification: C13; C52; E25

1. Introduction

Informal economy in a country includes all economical activities that avoid government regulation and taxation or other duties. It interacts with registered economy (the GDP) by affecting its dynamics and modifying certain economic indicators, becoming therefore assessable in some way. It can be evaluated by

[1] PhD student, Faculty of Economy, University of Tirana., Albania, Address: Street Arben Broci, Tirana, Albania, E-mail: dashimrasani2@gmail.com.
[2] PhD, Department of Applied Mathematics, Faculty of Technical Sciences, University of Vlora, Albania, Address: Skelë, Rruga Kosova, Vlorë 9401, Albania, E-mail: kushtamira@gmail.com.
[3] Professor Associate, Department of Physics, Faculty of Natural Sciences, University of Tirana, Albania, Address: Street Arben Broci, Tirana, Albania, Corresponding author: dode.prenga@fshn.edu.al.

survey's analysis or by modeling. The trustworthiness of the estimation for this hidden quantity depends on methodical calculation procedures and on the characteristics of the economic system under the study. Theoretical aspects and modeling remarks have been deliberated in a large literature as for example in (Schneider & Williams & Colin, 2013), etc. In this work we consider some particularities observed in the calculation of informal economy for a concrete and specific system, the Republic of North Macedonia for the period [1998, 2016]. It is specific because the number of data points is small, the economy of the country has been under transformation processes, (Shukarov, 2012) and non-standard variables as migration and remittances impose their particular effects on the economy. We will present in following some conceivable analysis that helped to improve the measurement of informal economy when using indirect or model approach. Remember that direct approaches practice surveys, so their accurateness depends on the quality of the responded questionnaires and moreover their realization needs additional expenses. Indirect methods are mostly macroeconomic (Schneider & Buehn, 2016), so the calculations in this case use models comprising variables from professional and official databases. In this view, model approaches are practical and easy evaluation techniques. The most used indirect methods are discrepancies approach, monetary approach and physical input approach. A more general technique is based on structural equation methods (SEM) known as multiple indicator multiple causes model (MIMIC). The discrepancies approach admits that the difference between expected and real values for an economic quantity reflects the effects of informal economy, and therefore this last can be evaluated by a straightforward procedure. SCR model (simple currency ratio) asserts that informal economy is observed directly in the ratio of the currency out of deposits (C) by money in deposits (D). The simplified calculation formula has the form $\frac{InformalEconomy}{GDP}=\frac{C-kD}{(k+1)D}$ where $k=\frac{C}{D}$ in absence of informality. CDA (currency demand approach) model assumes that informal economy is generated from the fiscal evasion and other duty's avoidances. Therefore, informal activities seek to use only transactions in cash that increase the demand for currency (in circulation). Methodically, by evaluating the excess in this last, the size of informal economy is measurable using the regression of the type $\log(C/D,M)=\sum_{(i)}\alpha_i x_i$ where x are some factors and M is money aggregate. In MIMIC method, informal economy stands in-between factors or cause-variables and some macroeconomic quantities called indicators. So, by a two-step regression, this method offers the measurement of the size of informal economy, its cause factors and its indicators. From the calculation perspective, the rigorousness of linear modeling that appears in all abovementioned approaches is conditioned by the fulfillment of some requirements for variable data series. In (Dell'Ano & Schneider, 2006) it is underlined that (specific) economic mediums impose additional constraints, and from our perspective, we should consider them in modeling and data elaboration stages. In our case-study intended on evaluating informal economy in a given specific system, we got questionable results compared to the general expectation. To improve the evaluation we performed an ad-hoc stationary analysis to avoid possible causes of problems observed in the use of CDA, and the factorial analyses to assist the MIMIC model setup. The reviewed models have produced reliable results and additional information for our system. In the following, we will present those undertakings.

2. Matching the Time Interval for a Truthful Use of CDA

The measurement of the informal economy for R.N. Macedonia for various periods up to 2005 has been addressed recently in many researches as (Osmani, 2004; Willimas, 2015) and references therein. So, the level of informal sector has been reported in the range 35%-30% of the GDP. Generalized analyses given in (Schneider& Buehn and Montenegro, 2010) advocated that the size of informality for economies in transitions might take values in the range [0.25-0.55] of GDP. Knowing that the economy under study belongs to the group of economies in transition, we acknowledged those estimations as reference boundaries. In (Dietz, 2010) and (Angelescu, 2009) it is underlined that the economy of the country has known dynamical changes last years due to the remittances and migration effects. Therefore, we had to consider econometric and methodical aspects when modeling for the period considered. In the first attempt we used CDA in the Cagan form

$$\ln\left[\frac{C}{D;M}\right] = a + \alpha \ln GDP + \beta \log(1+.T) - \gamma I + ..\varepsilon \qquad (2)$$

as given in (Schneider & Williams, and Colin, 2013). In (2) M denotes a money aggregate quantity (we used M_1 and M_2), D denotes the amount of the money in deposits, *GDP* is the gross domestic production, T denotes the averaged or weighted taxes and I signify the interest rate applied on deposits, whereas ε is an uncertainty or error term. Applying standard routines of linear multivariate regressions, we observed that original *C/D* and *C/M* data series resulted co-integrated I(1). The procedure of unit root removal by first differences was not conclusive in this case. In this situation we proceeded nevertheless with the regression (2) following other standard steps of the procedure, but managing a lower significance level for statistical tests. Remember that for such a small size series (16 points), statistical analysis becomes practically incoherent. Thus made, using initial series of [1998, 2016] in the regression (2), we got abnormal low value for informal sector at around 5%-10% of GDP for the years 1998-1999. This value is characteristic for a few highly developed countries (Schneider, 2016), so we flagged it as a wrong evaluation. The estimated value for informal economy in the middle of the period was high, around 60% of the GDP. This value corresponds to the collapsing economy, which clearly was not our case. Other shortages regarding to the sign of the coefficients have been observed too. Therefore we tried the Tanzzi formulae for CDA:

$$\ln\left(\frac{C}{p}_t\right) = a_0 + a_1 \ln\left(\frac{C}{p}_{t-i}\right) + a_2 \ln(GDP_{t-i}) + a_3 \ln(I_{t-i}) + a_4 \ln(1+T_{t-i}) \quad (3)$$

where i is the time lag that considers the delayed response in the indicator variable, p is the deflator index, C is currency or normalized currency to the money M. By using (3) we attained a little improvement, but the problem of a reliable regression remained unsolved. So, for optimal time lag ($i=1$) the coefficient α_4 was found negative which is theoretically wrong. Next the abnormal low values in the edges of the period remain unresolved. Considering those findings we hypothesized that shortcomings have originated from local high non-stationary behavior of some variables included in (2) or (3). The dynamics on econometric variables and its effects on linear modeling have been addressed in (Libanio, 2005; Kwiatkowski et al, 1992). So far, in an effort to recuperate the application of the formulas (2) and (3), we assumed that the variables in l.h.s of (2) or (3) may have undergone complicated dynamics around certain time-points in the interval considered. Therefore, linear regressions (2) or (3) have lost their trustworthiness nearby those points. To localize the possible extreme events associated with such dynamics, we considered monthly data series of C/D and C/M variables. The idea is that highly non-linear dynamics of daily or monthly variable behavior would be replicated somewhat in the yearly data values, displaying local deviances from a smooth trend. Consequently we checked such series for the presence of highly non-stationary regimes known as self-organization behavior. They are common events for financial time series and typically leads to extreme behavior of the type bubble or anti-bubble as described in (Sornette et al, 2004). Based on the analysis provided in (Sornette et al, 2011; Jiang, et al, 2010), such self-organization process is characterized by a log-periodic trend

$$y(t) = A + B(t-t_c)^m \left[1 + C\cos(\ln \omega(t-t_c) + \alpha)\right] \quad (4)$$

Note that theoretically the trend (4) has a critical point $t=t_c$, but in practice its signifies strong oscillation and high amplitudes associated to the behavior of variable y. In (Sornette et al, 2004) is stated that critical time t_c denotes merely the most probable moment for a regime change to occur, so we will consider this last property of the critical points in (4). Evidently A, B, C are (real) constants, ω is the cyclic frequency and α in the initial phase. By analyzing the fit of (4) to the C/D monthly data series, we evidenced the presence of a mixed self-organization regime as seen in Figure 1. So, by spanning the time windows in the interval [1998, 2016], we observed that a near-to-log-periodic process of anti-bubble type seemed to has started around the middle of the year 2003 and would (probably) finish at June 2018 (at the critical time t_c). We referred near-to-characteristic regimes in the context of discussion in (Prenga, 2016).

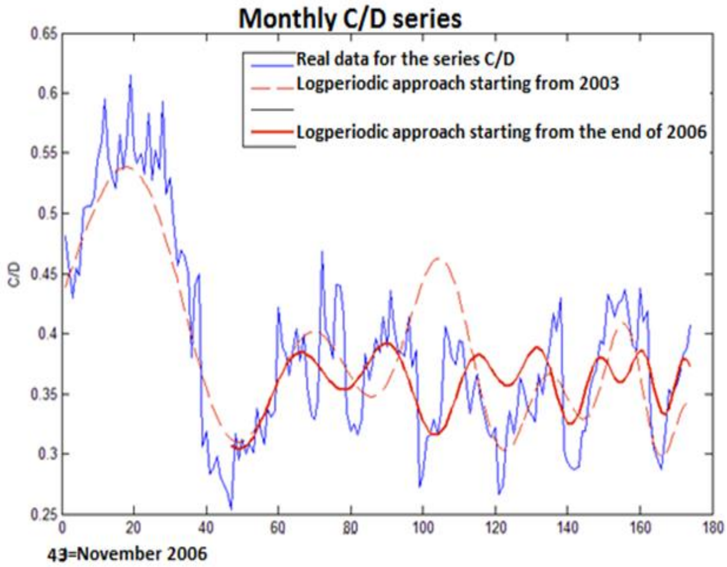

Figure 1. Mixed regime in the C/D data series

Another adjoining such process is likely to have started around November 2006 (in the monthly coordinate 43) and would be active until 2020 in condition *ceteris paribus*. Therefore a special point corresponding to a starting self-organization regime is located in-between 2003 and 2006. Similarly, the other special point corresponding to the end of this regime is found near 2018, which is out of our interval. So, we banned the data points before 2006, as hosting particular points which we assigned as undesired for regression procedures. The remaining segment [2006, 2016] was qualified as the appropriate interval for CDA regression.

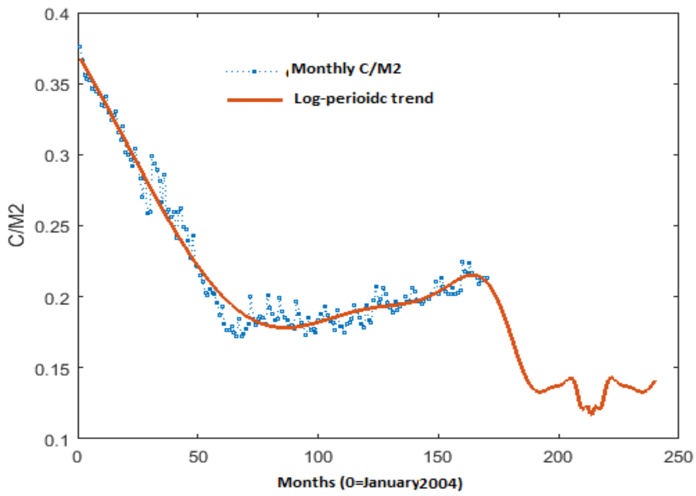

Figure 2. Near to log-periodic trend of C/M_2

Putting series C/M_2 in (4) we observed that a near to self-organization regime has originated before 2003 and will end at 2019 (around the coordinate 213), Figure 2. Therefore regarding to the formula (3) the period [2003, 2016] has been qualified as the interval where C/M_2 variable is smooth enough to be used in the regression (3). We observe that in this case, the unit root was removed in first difference of the series. Introducing above correction, we obtained remarkable improvement on the estimation of the informal economy. So, for the period [2004, 2016] this parameter is evaluated in the range [0.18-0.35] using (3). The underestimation problem remained unresolved for the beginning of the interval, but it differs significantly from the abnormal values of 5%-10% obtained when considering the [1998, 2016] time segment. Next, by putting the VAT in the role of taxes and GNP instead of GDP in (2) and (3), all the coefficients in (2) and (3) resulted with expected sign. As preliminary conclusion, we underlined that shortening the series as to exclude critical behavior from them has produced a significant improvement of the CDA predication for informal economy in our system.

3. Use of Monotonic Correlations in Modeling

A preliminary empiric view on a given system could be helpful for modeling. Note that if the number of data points is small as in our case, it is difficult to decide from the regression results which variable plays at best a certain role in the linear model. Specifically, we were interested on the role of remittances in the size of informal economy. This variable does not appear in standard models because it is not a typical economic parameter, but in our case it is significant, and is expected to affects the currency demand and other economic parameters of the country. We proposed to analyze the level of association between some variables including the remittances, and thereafter to judge which one is appropriate to be considered in CDA model. We calculated Spearman correlation coefficients for those variables in the time sequences [2003, 2008; 2009;…]. So we were able to learn about the advancement of their association by the time.

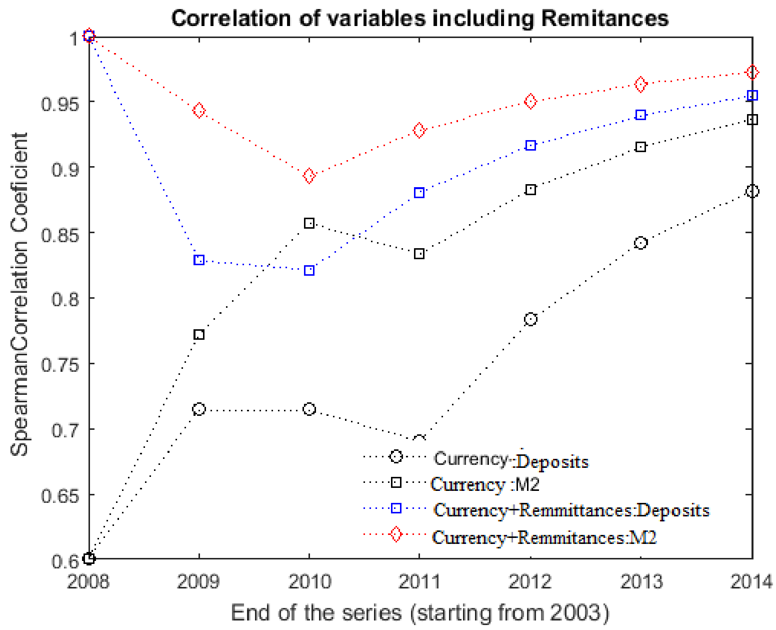

Figure 3. Monotonic Correlation coefficients for some variables

As seen in the Figure 3, the monotonic correlation coefficients between C and D or M_2 showed apparent changes for successive periods. The pair of variables C+R and D or M_2, exhibited higher monotonic correlation coefficients. Moreover, it remained high along all the core period of our interests. The monotonic correlation between {C+R, M_2} variables was found the highest among other combination and therefore we preferred them to use in CDA approach for interval [2006, 2014]. The results obtained accordingly, confirm an improvement in the regression's statistics. The size of informal economy in the problematic zone 2004-2005 is obtained at 25% of the GDP that is much better than when we using simply C in (3).

4. Using Factor Analysis to Assist MIMIC Modeling

The MIMIC model used in econometric studies is a structural equation approach. SEM models assume that latent variable *Y* are observed in the indicators set *Z*, whereas it is caused by the factors *X* where *X, Y* and *Z* in general are vectors. So the relationship between *n* factors and *m* indicators is realized by the intermediacy of *p* unknown variables. The indicators of informal economy are in general the GDP, unemployment rate, normalized currency C/M, but other socio-economic parameters may appear as indicators too. The set of factors depends on concrete economy and involve numerical and categorical variables. The generalized SEM model aims evaluating Y, having known indicators Z and factors X. It has the n-p-m matrix equation form:

$$Z = AY + u; Y = BX + v; \quad (5)$$

In (5), A and B are matrices of coefficients, u and v assign the measurement errors for each equation. They are assumed to be normally distributed with zero mean. The MIMIC calculation procedure is described in many references as in (Schneider & Williams, 2013), etc. The calculation algorithm is detailed in (Jorskog & Goldberg, 1975) and additional clarifications can be found in (Steiger, 1990). Modeling starts by fixing the structure n-p-m. In MIMIC application for informal economy, usually it is assumed p=1 but this is not compulsory. Once the n-p-m dimension is fixed, the concrete set of variables could be selected based in general econometric arguments. Specifics of the factors influencing in the informal economy depend on concrete economies as seen in (Schneider & Savaan, 2007; Schneider & Buehn; Montenegro, 2010; Davidescu, 2015) etc., therefore choosing a correct set of cause variables is very important for further analysis. In this step we performed a data-oriented empirical analysis to fix the set of cause variables. Firstly we grouped the factors in five categories, *taxes, econometric indexes, interests and rates, government performance indicators, and currency-like variables.* For each of them we inspected subsets wherein at least 80% of total variance was explained. We observed that for the group of taxes and tariffs with six elements, more than 80% of the variance was explained by a single variable and 98% in only 2 variables. For 5 variables of the type *indexes* we obtained that 82% of the variance was explained in a single variable, for 3 variables of the type *interests and rates,* 80% of the variance was explained in one variable, for 5 variables of the type *governance 85*% of the variance was explained in two variables and for *money-typ*e variables 98% of variance was explained in two variables. So far, the minimal number of factors could be reduced in 7 variables according to those findings. A more descriptive set based on 98% variances explained for all categories, should include 9 factors selected among the categories above. Finally we fixed n=9 for further calculation. The set of indicators Y is taken from general theoretical views, e.g., the effect of informality is expected to affect specifically {*GDP, Unemployment Rate, Money*} etc. Finally we used factorial analysis to fix the number of principal components that described the system of the factors and proposed to consider this last as the number of latent variables. We observed that 95% of the variance of the set of variables fixed above was described by one component whereas more than 98% by two components.

Therefore our optimal model is fixed 9-2-3 where *X= [Rural Population, VAT, Total Taxes, House Holdings, Net Wages, GDP.Capita, Interest, Remittances, Government Expenditures]; Y=[E1,E2]; Z= [Unemployment, C/ M1, GDP. Deflator]*. We obtained that the first latent variable which we identified hereto as classical informal economy had its normalized values to the GDP in the range [30%-32%] for the interval [2004, 2016], Figure 4. The other latent variable had the normalized values in the range [6%-9%], Figure 5.

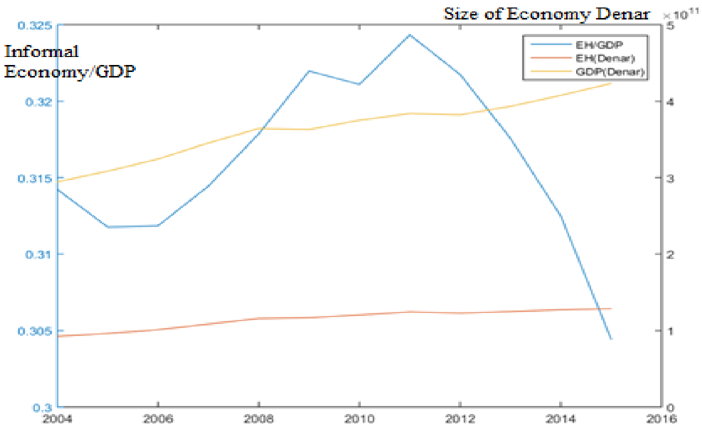

Figure 4. Recalculation of Hidden Economy recalculation in 9-2-3 MIMIC model

Thus, the informality in the R. N. Macedonia for the period [2004, 2016] would be described at best by an array encompassing two terms, which of one reflecting different relationship with the set of factors X.

Figure 5. Second latent variable using 9-2-3 MIMIC model

The total informal economy is obtained by simply summing up those two terms.

Accordingly, the level of informality in 2004 is estimated at 36.5% of the GDP, going at 39% -41% of the GDP in the period [2008, 2012] and falling to 31% of the GDP at the end of 2015 as given in Figure 5. The overall estimated informal economy for this period was found in the range 90-130 billion dinars and is keeping increasing but apparently slower than the GDP. The major contributors on informal economy have been identified the variables *rural population ratio, value added taxes* and *government expenditures*. All of those promoted it as expected in theory. Other finding have resulted statistically reliable and matching with other estimation and expectations. We qualified the model obtained as described in this paragraph as optimal and its predictions as realistic. Therefore, the factorial analysis has worked as helpful tool in the designing n-p-m structure of MIMIC model in our system. Again, it can be suggested for similar cases too.

Conclusions

The stringent use of CDA and MIMIC models in our study of informality for the R. N. Macedonia during [1998, 2016] has produced initially questionable results. By evidencing the presence of self-organization regime, we localized the zones where currency type variable showed high nonlinear dynamics and excluded them from the study. Therefore, we qualified the interval [2004, 2016] as the appropriate reference for linear regression. Using series in this interval, the CDA estimations for informal economy exhibited a remarkable improvement. Next, by using factorial analysis in designing n-p-m MIMIC structure we identified the 9-2-3 structure as optimal for this system. In particular, we observed that informal economy herein was better modeled as a two-term structure. It resulted that each of those terms encompassed differenced effects of certain factor variables, which consists in additional information learned from the model. In general, this evaluation produced an enhanced estimation for the size and the causes of the informal economy in the period [2004, 2016]. Thus, we guesstimate that descriptive analysis is a useful instrument that could help to improve the estimation and the analysis of informal economy using linear models in similar systems.

References

Schneider, F. & Williams, C. (2013). *The Shadow Economy*. Institute of Economic Affairs. Lord North Street, Westminster London.

Schneider, F. (2016). Estimating the size of the shadow economies of highly developed countries: selected new results. *Research Report*. CESifo DICE Report 4/2016 (December).

Schneider, F. & Savaan, F. (2007). DYMIMIC estimates of the size of shadow economies of Turkey and of her neighboring countries. *International Research Journal of Finance and Economics*, Issue (9), pp. 126-143.

Davidescu, A.; Strata, V. & Paul, V. (2015). Revisiting the size of Romanian informal economy using the Gutman approach. *Procedia Economics and Finance*, 23, pp. 1037 – 1045.

Schneider, F.; Buehn, A. & Montenegro, C. (2010). *Shadow economies all over the world*. New estimates for 162 countries from 1999 to 2007, Vol. 24, issue 4, pp. 443-461.

Jorskog, K. Goldbwerg, A. (1975). Estimation of a model with multiple indicator and multiple causes of ingle latent variable. *Journal of the American Statistical Association*, Volume 70, issue 351.

Dell'Ano, R. & Schneider, F. (2006). *Estimating the underground economy by using MIMIC models: a response to Breusch´s critique*. Department of Economics. Johanes Kepler University of Linz. *Working pape*r, no. 0607.

Kwiatkowski, D. & al. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, pp. 159-178.

Shneider, F. & Buehn, A. (2017). Shadow economy : estimation methods, problems, results and open questions. *Open Economics*, Vol 1(1), pp. 1-29.

Sornette, D. & al (2011). *Clarifications to Questions and Criticisms on the Johansen-Ledoit-Sornette bubble Model*. ETH-RC-11-004.

Jiang, Zh. Et al (2010). Bubble Diagnosis and Prediction of the 2005-2007 and 2008-2009 Chi nise stock market bubbles. *Journal of Economic Behavior and Organization*, 74, pp. 149-162.

Angelescu, R. et al (2009). *Measuring and Optimizing the Economic and Social Impacts of Migration in the Republic of Macedonia*. Development on the Move. IPPR-GDN. Skopje.

Steiger, J.H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research*, 25, pp. 214-12.

Davidovska-Stojanovska, B. et al (2010). Analysis of the level l of currency in circulation in the Republic of Macedonia and the reasons behind its high level. Research Department *Working Paper* No.11 . National Bank of the Republic of Macedonia.

Osmani, R. (2015). The level of the shadow economy, tax evasion and corruption: the empirical evidence for SEE countries. *The Journal of South East European University*. Vol. 11, Issue 2, pp. 6-22.

Dietz, B. (2010). *Migration and remittances in Macedonia*. A review, Arbeiten aus dem Osteuropa-Institut Regensburg, No. 281, ISBN 978-3-938980-30-9, Osteuropa-Institut Regensburg.

Libanio, G. (2005). Unit Roots in Macroeconomic Time Series: Theory, Implications, and Evidence. *Nova Econ*, Vol. 15, no. 3, Belo Horizonte.

Williams, C. et al (2015). Evaluating the Policy Approach towards the Undeclared Economy in FYR Macedonia. *International Journal of Entrepreneurship and Small Business*, Vol. 24, no. 2, pp. 268–286.

Shukarov, M. (2012). Perceptions of Transition and the Crisis in Macedonia. Croatian Economic Survey, Vol. 14, no. 1, pp. 107-131.

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