WHAT IS THE SIZE OF THE SHADOW ECONOMY IN ROMANIA?
A COMPARISON OF DIFFERENT ESTIMATION APPROACHES

Roberto Dell’Anno † and Adriana AnaMaria Davidescu‡

Abstract
The paper aims to estimate the shadow economy (SE) in Romania using quarterly data covering the period 2000–2014 by some most utilized macro econometric approaches: the currency demand approach, the MIMIC model and the labor approach. Principal component analysis and dynamic factor models are applied to construct a composite index of the estimated series. Taking into account that all macro econometric methods produce inaccurate outcomes in terms of absolute estimates of the SE, we find that the three estimated series showed evident similarities as concern trends. Accordingly, a composite index of SE can be considered a useful tool to grasp the development of the informality.

Keywords: shadow economy, MIMIC model, labor approach, dynamic factor models, principal component analysis, Romania.

JEL codes: E26, O17, C22, C5

† Roberto Dell’Anno, Department of Economics and Statistics and CELPE, University of Salerno, Via Giovanni Paolo II, 1, 84084 Fisciano (Sa), Italy; e-mail: rdellanno@unisa.it
‡ Adriana AnaMaria Davidescu, Department of Statistics and Econometrics, Bucharest University of Economic Studies and Romanian Scientific Research Institute for Labour and Social Protection, Bucharest, Romania, e-mail: adrianaalexandru@yahoo.com. (Corresponding Author).
1. Introduction

There is no precise measure of the dimension of the shadow economy (SE). According to the national accounting community (e.g., OECD, 2002) macro-model estimates of SE are unreliable. A heated debate exists in economic literature on the reliability and precision of the existing estimates (e.g. Slemrod and Weber, 2012; Feige, 2016). Following Dell’Anno and Schneider (2009), we consider all the estimates of SE as “vulnerable” and no one can guarantee the accuracy of her estimates.

In this paper, we aim to contribute to this issue by showing how, the method of estimation, influences the estimated size of SE with reference to a single country analysis.

The first issue in this literature on SE is about its definition. Despite some attempts to find a definition of SE able to harmonize the national accounting system (SNA) approach with economic research (e.g., Dell’Anno, 2016), a conventional and internationally accepted classification of SE is still missing. Following Smith (1984), Alm and Embaye (2013, p. 512) proposed a definition of the SE containing “all market-based goods and services (legal or illegal) that escape inclusion in official accounts”. This definition is a wider definition of the SE proposed by Schneider et al. (2010, p.444) that only include “all market-based legal production”. In this paper, we consider our estimates of the SE as an approximation of all market-based legal production that escape inclusion in official accounts.

The measurement of SE is very difficult because of the nature of shadow economic activity. In literature, three main methods of assessment of the size of the SE are used: direct approaches, indirect approaches and structural equation models (Garcia-Verdu, 2007).

In the category of direct methods are included survey and tax auditing methods, which have the particularity to highlight the structure of informal activities at a certain moment of time, instead of offering information about the pattern over time (Schneider and Williams, 2013).

In the category of indirect methods are included “the discrepancy between national expenditure and income statistics, the discrepancy between the official and actual labor force, the monetary approaches and the physical input (electricity consumption) methods” (Schneider and Enste, 2000).
Out of all these methods, the currency demand approach and structural equation models are the most frequently used in the literature for estimating the magnitude of SE (Schneider and Williams, 2013). The first one has the particularity of considering the cash as the main medium of exchange in the unofficial sector and if the amount of cash used in the under covered sector can be estimated, we can obtain an estimation of SE using the information provided by the money speed of rotation.

Instead of the first one, who takes into account a single potential indicator of SE (the cash), structural equation models start from the hypothesis according to which the informal economy is considered to be a “latent variable” linked with multiple causes and having multiple effects(Schneider et al, 2010).

According to Dell’Anno(2003), SE is “linked to one hand with a set of observable indicators (reflecting changes in the size of SE) and on the other to a set of observable causal variables considered to be the most important determinants of unreported economic activity”. This particular SEM specification is labeled as the MIMIC (Multiple Indicators Multiple Causes) model.

It is important to mention that until now, we cannot claim that there is an optimal method of estimation for this phenomenon of shadow economy. Each method has its own shortcomings and although the empirical estimates in the literature are not exempted from biases, the lack of national accounting estimates of SE across countries and for an adequate period, made the macroeconomic methods the only alternative methods to get an approximation of the SE for positive and normative analysis in this field of economic research.

In this paper, we focus on the Romanian SE. Looking at some of the existing estimates: the Romanian Institute of Statistics (RIS) estimated the non-observed economy (NOE) as representing between 14.5% and 23.5% of the official GDP; Schneider et al.(2010) determined the dimension of the shadow economy as being close to about 30% of the official GDP in 2010. Furthermore, it followed a descendant trend in the last years, attaining about 28% of the GDP at the end of 2014. Table 1 summarizes the existing macroeconomic estimates of the SE for this country.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Period</th>
<th>Size of SE (% of official GDP)</th>
</tr>
</thead>
</table>

The differing results show how the relevance of different definitions, methods, hypotheses affect the estimation of SE. The central idea of this paper is that, although collecting several biased estimates doesn’t generate unbiased estimates, they can shape an overview of the SE from different points of view; therefore, a combination of these estimates may provide a better approximation of the “true” range and the trend of informality in the economy.

Accordingly, we aim to minimize the measurement errors by constructing two composite indexes – by applying the principal component analysis and dynamic factor models – of the Romanian SE determined using three of the most common methods in the literature adapted to the Romanian economy specificities—the currency demand approach, the MIMIC model and the labor approach covering the period 2000-2014.
Baptista (2014, p. 5), states that Maggino and Zumbo (2012) argued that a “potential advantage of developing composite indices is that they can help to overcome problems concerning precision, reliability, accuracy and validity that are associated with using individual indicators”. Also, Kaufmann and Kraay (2007) stated that building a composite indicator can reduce the measurement error of an individual indicator.

On the other hand, Molle and Mollinga (2003) and Lohani and Todino (1984) drown attention about the fact that the measurement error can be propagated by aggregation of individual indicators into a composite index (Baptista, 2014). In the process of constructing a synthetic indicator, besides the methodology, the dataset used and the framework are very important (OECD, 2008).

The paper is organized as follows: after the introduction, Section 2 presents the methodology, data issues and the main empirical results of the size of Romanian SE using the currency demand approach, the MIMIC model and the labor approach. Section 3 compares the results of the estimation methods. Section 4 shows the composite indices of SE. The paper ends with main conclusions.

2. Estimating the size of the Romanian shadow economy

2.1. The currency demand approach

2.1.1. Data and Methodology

The method, initially proposed by Cagan (1958) and later developed by Gutmann (1977) and Tanzi (1983), considered shadow economy as a response to the tax burden mainly based on cash. The method has been used in the studies of Cagan (1958), Gutmann (1977), Tanzi (1980, 1983), Johnson et al. (1998), Bhattacharyya (1990), Spiro (1996), Hernandez (2009), Gadea and Serrano-Sanz (2002), Macias and Cazzavillan (2009) and also Alm and Embaye (2013).

The method supposed two main steps. According to Macias and Cazzavillan (2009) and also Alm and Embaye (2013), in the first step, the observed demand for currency (tax evasion and no tax evasion) and demand for formal transactions were econometrically estimated. The currency demand for formal
transactions was computed assuming no tax evasion by using the equation of the observed demand in which the taxes and also other fiscal variables have been set just below to their lowest historical levels. The difference between the observed demand for currency and the estimated amount used for formal transactions can be considered to be the currency demand associated with tax evasion.

In the second step, an estimation of the total amount spent in informal activities can be determined using the estimated currency for informal transactions multiplied by the velocity of money, assuming this velocity is the same in both sectors and finally this amount can be reported to official GDP.

Some of the main limitations of the method are: not all transactions in SE are paid in cash; tax burden is not the only cause of SE; the hypothesis that the velocity of money is the same in both sectors is questionable; the assumption of no SE in a base year is improbable.

According to Ahumada et al. (2000, p. 7) and Hernandez (2009, p. 88), “the multiplicative demand function for currency is:

\[
C_0 = a(1 + F)^\alpha Y_0^\beta e^{\gamma i}
\]

where: \(C_0\) is the observed currency in real terms, \(F\) is a fiscal variable related to the evolution of informal activities (e.g. total taxes over GDP or government expenditure over GDP); \(Y_0\) is a variable associated with the level of transactions in the economy such as observed real GDP; \(i\) represents the opportunity cost of holding currency (e.g. nominal interest rate or inflation rate); \(a\) is the intercept”. The expected sign for \(\alpha\) and \(\beta\) is positive while for \(\gamma\) it is expected a negative impact. Following Hernandez (2009), an increase in taxation or in government consumption will push individuals more and more to the informal sector in which they used the cash.

Table 1 from the appendix offers detailed information on the variables. According to Spiro (1996), the currency which refers mainly to transactions is recommended to be deflated using M1 and not M2, considered as being inappropriate since it highlighted the long-term accumulation. Schneider and Enste (2000), Ögunç and Yilmaz (2000) and Macias and Cazzavillan (2009) suggest the usage of currency and currency per capita in real terms. In our analysis, the series had been deflated by GDP deflator with the
exception of the nominal interest rate, inflation rate and real currency outside banks. The data has been seasonally adjusted by means of Census X-13 method.

The model has been linearized by natural logarithm. The sample covered the period 2000-2014, quarterly data. Following the work of Alm and Embaye (2013), Hernandez (2009) and Macias and Cazzavillan (2009), the currency demand model specified for the case of Romania is:

$$\ln C_t = a + \alpha \ln(1 + F_t) + \beta \ln Y_t + \gamma_1 \ln(1 + \pi_t) + \gamma_2 R_t + \gamma_3 Urb\_empl + \gamma_4 dummy\_crisis + \epsilon_t$$

(2)

Where: $\ln C_t$ is the natural logarithm of currency in circulation outside the banks (at the end of the period in millions national currency (RON) normalized by the GDP deflator; $\ln(1 + F_t)$ is the natural logarithm of (1 + total tax revenues normalized by GDP); $\ln Y_t$ is the natural logarithm of real GDP; $\ln(1 + \pi_t)$ is the natural logarithm of (1 + inflation rate); $R_t$ is the nominal deposit interest rate; $Urb\_empl$ is the employment rate from urban area; $\epsilon_t$ is the error term; $a$ is the intercept of the model.

The monetary approach starts from the main idea that transactions in SE are only made by currency enabling individuals to avoid authorities; SE is a direct result of high taxes; the opportunity cost of holding currency, i.e. measured by the inflation rate and interest rate affects the currency demand; the velocity of circulation for currency is the same in both economies. As far as the most important determinant of SE for currency approach is concerned, i.e. tax rate, it is expected to have a direct influence on currency and a negative impact on demand deposits, taking into account the anonymity of cash.

Inflation is another relevant factor in the analysis of the determinants of the shadow economy. Alm and Embaye (2013, p. 516) stated that “when tax systems are not indexed, higher inflation creates tax bracket creep and increases the tax liabilities of taxpayers, increasing incentives for greater tax evasion and also affecting the decision to evade taxes by eroding the real value of a given level of nominal disposable income, which gives taxpayers the incentive to evade more taxes to restore their purchasing power”.

The potential explanation for the inclusion of inflation in the model specification was that it could be regarded as a taxation tool used to counterbalance the fiscal revenues not collected having as the main
cause tax evasion (Alm and Embaye, 2013) and in such a way offering more details about the effective tax rate\(^1\).

Another explanatory variable of the currency approach is the interest rate, seen more as an opportunity cost of holding currency and the expected impact should be a negative one; an increasing interest rate leading to a lower demand for currency.

Real income is the most important control variable of the currency demand approach. So, if the income will increase, more money is needed to make transactions and there is a positive relationship between income and the demand for money.

Another particular factor that has been mentioned in this context of estimating the currency used for informal transactions was the degree of urbanization. The explanation offered by Cagan (1958) was that urbanization constrained individuals to make commerce in unknown places so the use of currency is encouraged but we must not overlook that the checks are more used in urban areas comparative with rural areas, so the impact could be doubtful. Given that the urbanization degree is not a quarterly series, we used as proxy the employment rate in urban area.

Finally, we also control the effects of the economic crisis using the dummy variable, taking the value of 1 for the period 2008Q3-2011Q4 and 0 otherwise.

In the process of estimating the currency demand function the main steps were the analysis of stationarity tested by unit root tests (ADF and PP), the Johansen multivariate cointegration test, the estimation of a VAR/VECM models. If the variables are not cointegrated, then the vector error correction (VAR) model in the first difference will be estimated, otherwise will we estimate the vector error correction (VECM) model.

---

\(^1\) Their explanation was that “many developing countries have a lower tax to GDP ratio not because they impose lower tax rates but because they exhibit higher tax evasion. The response is to print money to make up for the shortfall, which results in higher inflation as an implicit tax on individuals and firms” (Alm and Embaye 2013, pp. 8-9).
2.1.2. Analysis of Cointegration

The analysis of non-stationarity of variables verified by unit root tests pointed out that the variables are non-stationary and need to be differenced once in order to achieve stationarity. Knowing this, a potential long-run relationship has been tested using Johansen cointegration approach, estimating a VAR model in level, determining the optimal number of lags checking the main hypotheses on the residuals and also the stability condition. According to informational criterions Akaike (AIC) and Schwarz (SBC), the optimal lag is found to be 3. The Pantula principle, used to determine the appropriate restrictions on the intercept and trends in the short-run and long-run models specified as the appropriate specification the model with intercept and no trend in cointegrating equation, and a model without intercept in VAR. The empirical results of Johansen cointegration test revealed the existence of at least one cointegrating vector at the level of 5%. Furthermore, a VECM model has been estimated, allowing for interactions between variables both on long-run and short-run, instead of VAR models who allow only for short-run relationships between variables. The empirical results of the VECM model are presented in Table 2.

### Table 2. The empirical results of the VECM model

<table>
<thead>
<tr>
<th>Trace Statistic</th>
<th>5% level</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. eigenvalue</td>
<td>5% level</td>
<td>1</td>
</tr>
<tr>
<td>$C_{t-1}$</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$Y_{t-1}$</td>
<td>3.08* (0.30)</td>
<td></td>
</tr>
<tr>
<td>$\pi_{t-1}$</td>
<td>35.83* (7.00)</td>
<td></td>
</tr>
<tr>
<td>$Tax_{t-1}$</td>
<td>3.33** (1.65)</td>
<td></td>
</tr>
<tr>
<td>$R_{t-1}$</td>
<td>-2.78* (0.48)</td>
<td></td>
</tr>
<tr>
<td>$Urb_empl_t$</td>
<td>1.41*** (1.28)</td>
<td></td>
</tr>
<tr>
<td>Dummy crisis</td>
<td>-0.07** (0.03)</td>
<td></td>
</tr>
<tr>
<td>Cons.</td>
<td>-27.11* (2.70)</td>
<td></td>
</tr>
<tr>
<td>$ECT_{t-1}$</td>
<td>-0.45* (0.13)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>F-stat</td>
<td>5.97*</td>
<td>Normality test</td>
</tr>
<tr>
<td>Autocorr. LM test</td>
<td>53.83 [0.29]</td>
<td>White test</td>
</tr>
</tbody>
</table>

Notes: *, **, *** statistically significant at 1%, 5% and 10%; standard errors are given in parentheses.
Foreseeable, the tax burden, the inflation rate, urban employment and also the real GDP manifested a statistically significant and positive long-run effect, while the interest rate had a negative impact on currency demand. The dummy variable used to capture the effect of economic crisis reveal a negative impact on long-run. The value of GDP coefficient confirmed an elastic effect on currency demand suggesting that an increase in the tax burden most likely will push individuals to the informal sector in order to complement their earnings. The error correction term coefficient is highly significant at 1% level and negative suggesting the presence of causality in at least one direction. Its value of -0.45 indicated that the deviation from the long-term equilibrium is corrected by 45% over each quarter.

The overall degree of determination in the model was high enough; the simultaneous action of all exogenous variables explained almost 53% of the variation of the currency demand. The analysis of VECM residuals pointed out the lack of autocorrelation of residuals and confirmed the homoscedasticity hypothesis but infirmed the last hypothesis of residuals’ normality.

2.1.3. The size of the shadow economy

Using the log-run coefficients for VECM estimation, the size of SE was estimated using:

$$\ln(C^*) = -27.11 + 3.33\ln(1 + F) + 3.08\ln Y + 35.83\ln(1 + \pi) - 2.78R + 1.4Urb\_empl, - 0.07\text{dummy\_crisis},$$

Following the assumption that the amount of currency demanded used for the official transactions is calculated supposing no tax evasion, i.e., by setting the tax burden just below its lowest historical level\(^2\) and the inflation rate to its minimum level (i.e. zero), \(\ln(C^*)\) is calculated by preserving all the other coefficients fixed (Macias and Cazzavillan, 2009). The gap between these two variables can be considered the total informal money of the economy. According to Tanzi (1983), the main assumption was that the money velocity is equal in both sectors:

\(^2\ln(1+F_{\text{min}}) = \ln(1+0.307) = 0.268. The tax rate is set to the value of 0.265. Because the minimum value of the inflation rate approaches zero, we set the inflation rate to 0 and this explains why \(\ln(1 + \pi)\) is used in the estimation since the logarithmic term also approaches zero.
\[
\frac{Y}{M_1 - EC} = \nu
\]  

(4)

where: \( Y \) is the real GDP, \( M_1 \) is the real narrow money supply formed by currency, deposits in circulation and informal money. The difference between \( M_1 \) and \( EC \) can be interpreted as the amount of legal money used in the economy (Macias and Cazzavillan, 2009). Using the velocity of money computed by the equation (4), the dimension of SE is determined by multiplying the informal money with this velocity of money as follows:

\[
EC * \nu = Y_{shadow}
\]

(5)

Having the amount of money spent in SE, we can obtain an estimation reported to official GDP. Starting from the observation of Ahumada et al. (2007, 2008) the estimates of SE were misplaced and contradicted the hypothesis of equal income velocity in both sectors (formal and informal) if income elasticity is not unitary. As in our case, the income elasticity is not unitary; we corrected our estimates using Ahumada et al. (2007) method:

\[
\frac{Y_{shadow}}{Y_{official}} = \left( \frac{C_{shadow}}{C_{official}} \right)^{\beta^{-1}} = \left( \frac{Y_{shadow}}{Y_{official}} \right)^{\beta^{-1}}
\]

(6)

where: \( Y \) is the real GDP, \( C \) is the real currency, and \( \beta \) is the income elasticity.

Finally, the Romanian SE recorded the value of 35% of official GDP in the first quarter of 2000 and followed a downward trend registering about 21% in the third quarter of 2009. From this point forward, the size of the SE begins to increase, reaching the value of 34% of official GDP in the third quarter of 2010. For the last years, there is a slightly declining tendency in the size of the representing 25% at the end of 2014.
2.2. The structural equation modelling

2.2.1. Data and methodology

The Romanian shadow economy was modelled using the MIMIC model based on latent variable theory, having the particularity of considering the shadow economy as an ‘unobserved’ variable, taking into account multiple causes and multiple effects of the “this latent variable” (Dell'Anno, 2003; Dell’Anno, 2007; Dell’Anno et al., 2007).

According to Schneider et al. (2010, p. 9-10) “this model presents the statistical relationships between the latent variable and the observed variables divided into causes and indicators, using the covariance matrix of them”.

Dell’Anno and Solomon (2007, p. 6-7) mentioned that “the model has two kinds of equations, the structural model- who revealed the relationships among the latent variable (η) and the causes (X_q) and – the measurement model- who links indicators (Y_p) with the latent variable (non-observed economy)”.

In the estimation of the Portugal shadow economy, Dell’Anno (2007) mentioned the following the mathematical specification of the MIMIC model:

\[ Y = \lambda \eta + \varepsilon \]  
\[ \eta = \gamma' X + \xi \]

where:

\( \eta \) is the scalar latent variable (the size of the shadow economy); \( Y' = (Y_1, ..., Y_p) \) is the vector of indicators; \( X' = (X_1, ..., X_q) \) is the vector of causes; \( \lambda_{(p \times 1)} \) and \( \gamma_{(q \times 1)} \) vectors of parameters; \( \varepsilon_{(p \times 1)} \) and \( \xi_{(p \times 1)} \) vectors of scalar random errors. Giles and Tedds (2002) mentioned that in order to estimate the model, it is fundamental to impose a restriction on one element of \( \lambda \).

As Schneider and Buehn (2016) already mentioned the most used method of estimation is the Maximum Likelihood (ML). An important attention must be paid to the multivariate normality of data. ML estimators are known to be consistent and asymptotically efficient in a large sample (Bollen, 1989). Several statistical methods for structural equation modelling such as the ML, Generalized Least Squares...
(GLS) and Full Information Maximum Likelihood (FIML) methods were applied in the case in which the data is multivariate normally distributed. When the multivariate normality is violated, it is recommended to use other estimator such as Weighted Least Squares (WLS) or Robust Maximum Likelihood (RML) who correct the bias in standard errors.

The MIMIC approach has its disadvantages or limitations presented by Dell’Anno (2003) and Dell’Anno and Schneider (2009). Among the most important we can mention: the sense of the latent variable, the problem of indefinite matrix, concerns related to the benchmarking procedure, the sensitivity to the exogenous value of the SE, the ambiguity related to the choice of the observed variables (causes vs. indicators).

In the estimation process of the Romanian SE, we consider a MIMIC model specification with 10 “causes” and 3 “indicators” (MIMIC 10-1-3). The causal variables considered are: tax burden decomposed into indirect taxes, direct taxes and social security contributions; unemployment rate; self-employment; government employment; employed persons having a second job; government consumption; real interest deposit rate and regulatory quality index. As indicators of latent variable we include: the index of real gross domestic product index; currency ratio and the labor force participation rate. A detailed presentation of the variables and their source is offered in Table A3 in the appendix.

Using the coefficients of the structural equation (eq.8), the dimension of SE can be computed over the sample period.

As mentioned earlier by Giles and Tedds (2002), in order to estimate the dimension of the shadow economy it is mandatory to set a scale and therefore the index of real gross domestic product (base year 2005=100) $Y_1$ was chosen as a reference or scale variable in our model and we set the coefficient to be equal to -1, making the assumption that formal and informal sectors are most probably substitutes and an indirect relationship exist between them (Dell’Anno, 2008). A detailed discussion about this topic is provided by Dell’Anno (2003). One important remark about the interactions between these two sectors is
offered by Schneider (2005) who considered that these interactions are related to the degree of economic development.

The source of the data for this estimation is the same as the currency demand approach, i.e. quarterly data from 2000:Q1 to 2014:Q4. With the exception of the real interest rate and the ratio C/M1, the data was seasonally adjusted by means of the Census X-13 method.

The series in levels or differences have been tested for unit roots using the ADF test. We have taken the first difference in all variables in order to have stationary data and given this, also the latent variable will be estimated in the first difference.

In order to identify the estimation method, the series has been tested for multivariate normality using Mardia test. The empirical results pointed out a clear violation of the assumption of multivariate normality, therefore it is not recommended to apply ML but rather the RML\textsuperscript{3} method.

\textsuperscript{3} This method allows the use of ML estimator despite the violation of normality hypothesis, because it corrects the bias in standard errors. Hoyle (1995, p. 368) suggests the usage of ML estimator with robust standard errors in case of continuous non-normal variables, calculates the same parameters of the ML, but both statistics on model reliability (e.g., Chi-square) and standard errors are corrected in large samples.
2.2.2. Empirical results

As Dell’Anno (2003, p.8) stated “the identification procedure of the best model starts from the most general model specification and continues removing the variables which have not structural parameters statistically significant”. The estimated output of different MIMIC models is presented in Table 3.

The positive signs of causal variables confirm the assumption regarding the scale variable. The empirical evidence pointed out that tax burden and its components (direct taxes, indirect taxes and social contributions) unexpectedly do not have any significant impact on SE.

The general government final consumption expenditure is included as a proxy of the size of government. The assumption is that governments that provide high-quality and/or quantity of public services and
goods reduce the agents’ incentive to operate in the unofficial economy; therefore, higher government final consumption is associated with lower levels of informality.

The government employment is a proxy of the regulations’ intensity. It has a positive impact on SE. Also variables such as unemployment rate, self-employed persons and second job employees have a positive impact on SE. The fluctuations of real interest rate on deposits do not impact significantly the informal activity. According to World Bank Worldwide Governance Indicators (2015), the regulatory quality “reflects perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector”. It has a negative impact on SE as expected because poor institutions discourage people to work in official economy.

Regarding the model indicators, we find empirical support of currency ratio approach because the SE has a positive impact on currency. Concerning the labor force participation rate, we don’t find a statistically significant effect of SE on this indicator. Indirectly, this result does not confirm the crucial idea behind the labor approach to estimate the SE, namely, if the activity rate decrease this can be a sign of increasing activity in the informal sector.
### Table 3. Estimated Coefficients of the MIMIC Models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MIMIC 10-1-3</td>
<td>-</td>
<td>0.39</td>
<td>0.24</td>
<td>0.08</td>
<td>0.23 ***</td>
<td>1.34 ^*</td>
<td>0.57</td>
<td>1.91</td>
<td>0.96 ***</td>
<td>-9.70 ^*</td>
<td>-0.006</td>
<td>0.30</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.18)</td>
<td>(0.25)</td>
<td>(0.12)</td>
<td>(0.59)</td>
<td>(0.49)</td>
<td>(1.19)</td>
<td>(0.50)</td>
<td>(3.72)</td>
<td>(0.127)</td>
<td></td>
<td>(0.20)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>MIMIC 9-1-3</td>
<td>-</td>
<td>0.40</td>
<td>0.24</td>
<td>0.08</td>
<td>0.23 **</td>
<td>1.34 ^*</td>
<td>0.57</td>
<td>1.91</td>
<td>0.69</td>
<td>-9.73 ^*</td>
<td>-</td>
<td>0.30</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.18)</td>
<td>(0.26)</td>
<td>(0.11)</td>
<td>(0.59)</td>
<td>(0.47)</td>
<td>(1.19)</td>
<td>(0.49)</td>
<td>(3.52)</td>
<td>-</td>
<td></td>
<td>(0.20)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>MIMIC 7-1-3</td>
<td>-</td>
<td>0.19</td>
<td>-</td>
<td>-</td>
<td>0.26 **</td>
<td>1.39 ^*</td>
<td>0.81 ***</td>
<td>2.03 ***</td>
<td>0.92 **</td>
<td>-8.77 ^*</td>
<td>-</td>
<td>0.30</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.12)</td>
<td>(0.69)</td>
<td>(1.04)</td>
<td>(1.04)</td>
<td>(1.02)</td>
<td>(3.02)</td>
<td>-</td>
<td></td>
<td>(0.20)</td>
</tr>
<tr>
<td>MIMIC 6-1-3a</td>
<td>0.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.30 ^*</td>
<td>1.30 ^*</td>
<td>0.62 ***</td>
<td>1.00 **</td>
<td>-7.92 ^*</td>
<td>-</td>
<td>-</td>
<td>0.32</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.12)</td>
<td>(0.60)</td>
<td>(0.39)</td>
<td>-</td>
<td>(3.09)</td>
<td>-</td>
<td></td>
<td>(0.22)</td>
<td>(0.47)</td>
</tr>
<tr>
<td>MIMIC 6-1-3b</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.25 ^*</td>
<td>1.43 ^*</td>
<td>0.79 ***</td>
<td>1.90 ***</td>
<td>0.92 **</td>
<td>-8.82 ^*</td>
<td>-</td>
<td>0.30</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.11)</td>
<td>(0.71)</td>
<td>(0.43)</td>
<td>(1.02)</td>
<td>(0.45)</td>
<td>(3.08)</td>
<td>-</td>
<td></td>
<td>(0.20)</td>
</tr>
<tr>
<td>MIMIC 5-1-2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.25 **</td>
<td>1.10</td>
<td>1.07 ^*</td>
<td>1.92</td>
<td>0.86 **</td>
<td>-</td>
<td>-</td>
<td>0.31</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.11)</td>
<td>(0.73)</td>
<td>(0.44)</td>
<td>(1.21)</td>
<td>(0.44)</td>
<td>-</td>
<td></td>
<td>(0.28)</td>
<td></td>
</tr>
<tr>
<td>MIMIC 4-1-2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.27 **</td>
<td>1.15 ^*</td>
<td>1.09 ^*</td>
<td>0.78 **</td>
<td>-</td>
<td></td>
<td>0.45 ***</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### Global Goodness of Fit Statistics

<table>
<thead>
<tr>
<th>Models</th>
<th>SRMR</th>
<th>CD</th>
<th>AIC</th>
<th>SBC</th>
<th>Degrees of freedom</th>
<th>LL(model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIMIC 10-1-3</td>
<td>0.05</td>
<td>0.35</td>
<td>1506.41</td>
<td>1545.89</td>
<td>19</td>
<td>-734.20</td>
</tr>
<tr>
<td>MIMIC 9-1-3</td>
<td>0.05</td>
<td>0.34</td>
<td>1274.97</td>
<td>1312.37</td>
<td>18</td>
<td>-619.48</td>
</tr>
<tr>
<td>MIMIC 7-1-3</td>
<td>0.06</td>
<td>0.29</td>
<td>986.05</td>
<td>1019.29</td>
<td>16</td>
<td>-477.02</td>
</tr>
<tr>
<td>MIMIC 6-1-3a</td>
<td>0.05</td>
<td>0.31</td>
<td>1097.73</td>
<td>1128.89</td>
<td>15</td>
<td>-533.86</td>
</tr>
<tr>
<td>MIMIC 6-1-3b</td>
<td>0.06</td>
<td>0.29</td>
<td>841.44</td>
<td>872.61</td>
<td>15</td>
<td>-405.72</td>
</tr>
<tr>
<td>MIMIC 5-1-2</td>
<td>0.03</td>
<td>0.24</td>
<td>911.56</td>
<td>934.41</td>
<td>11</td>
<td>-444.78</td>
</tr>
<tr>
<td>MIMIC 4-1-2</td>
<td><strong>0.02</strong></td>
<td><strong>0.29</strong></td>
<td><strong>950.58</strong></td>
<td><strong>971.35</strong></td>
<td><strong>10</strong></td>
<td><strong>-465.29</strong></td>
</tr>
</tbody>
</table>

Notes: *, **, *** represents statistical significance at 1%, 5% and 10%.
Following the MIMIC identification procedure, the best model is the MIMIC 4-1-2, chosen based on the following criteria (Hoyle, 1995): the statistical significance of parameters, the value of standardized root mean squared residual (SRMR), the value of coefficient of determination (CD), the values of AIC and SBC\(^4\).

The main drivers of SE are unemployment rate, self-employment, persons employed having a second job and government consumption and the presence of SE is reflecting best in the increase of currency reported to monetary aggregate M1.

As respects the unemployment, some “official” unemployed spend a part of their time working in the SE (Giles and Tedds, 2002). As part of the “official” unemployed go to the informal sector in order to supplement their earnings, Dell’Anno et al. (2007, pp. 8-9) already stated that “there is a flow of resources from the official sector to SE in recession cycles”.

Also, Dell’Anno and Solomon (2007) highlighted that shadow economy was acting as a buffer, absorbing part of the officially unemployed workers into the shadow economy. Saafi and Farhat (2015) and Saafi et al. (2015) revealed the existence of a causal relationship running from unemployment to unofficial sector.

Self-employment, measured as a percentage of labor force, is confirmed to be one of the main determinants of Romanian SE, due to the fact that these workers have more opportunities to evade (Bordignon and Zanardi, 1997).

A potential explanation for the relationship between those employed who have second jobs and SE could be the fact they have lower incentive to declare a second official job (e.g. in terms of social security contribution, higher marginal direct tax rate, loss of subsidies, etc.).

Government consumption expenditure can be seen also as a proxy for regulation burden. Higher government consumption increases the size of the public sector and decreases the economic freedom leading to a rise in the size of SE.

---

\(^4\) A perfect fitting corresponds to SRMR equal to 0. A good fitting corresponds to a low value, but limited by the threshold 0.08. CD is the coefficient of determination and it should be closed to 1. Lower is the value of AIC and SBC better is the model specification.
In order to compute the size of SE as percentage of GDP, an exogenous value of the SE for the reference period, i.e., the 2005 is needed in the calibration procedure. Taking into account the reference variable \( Y_1 : \text{Real\_GDP}/\text{Real\_GDP}_{2005Q3} \), we scale up the SE to a value in 2005 – base year – by using the most recent estimate of Romanian SE calculated by Schneider (2013), i.e., 32.2%.

In order to calculate the ordinal time series index for latent variable, we used the coefficients of the structural equation (eq. 10):

\[
\frac{\Delta \tilde{\eta}_t}{\text{GDP}_{2005Q3}} = 0.27 \Delta X_{5t} + 1.15 \Delta X_{6t} + 1.09 \Delta X_{7t} + 0.78 \Delta X_{9t}
\]  

(10)

Following Dell’Anno (2007), the index is scaled to take up to a value of 32.2 % in 2005Q3 and then converted as a ratio of current GDP using the benchmark equation described in Dell’Anno and Schneider (2009):

\[
\frac{\tilde{\eta}_t}{\text{GDP}_{2005Q3}} \times \frac{\tilde{\eta}^*_{2005Q3}}{\text{GDP}_{2005Q3}} \times \frac{\text{GDP}_{2005Q3}}{\text{GDP}_t} = \frac{\hat{\eta}_t}{\text{GDP}_t}
\]  

(11)

where: \( \tilde{\eta}_t/\text{GDP}_{2005Q3} \) is the index of SE calculated by eq. (10); \( \tilde{\eta}^*_{2005Q3}/\text{GDP}_{2005Q3} = 32.2\% \) is the exogenous estimate of SE; \( \tilde{\eta}_{2005Q3}/\text{GDP}_{2005Q3} \) is the value of the index estimated by eq.(10) in 2005Q3; \( \text{GDP}_{2005Q3}/\text{GDP}_t \) is to convert the index of changes respect to base year in values of SE respect to current GDP ; \( \hat{\eta}_t/\text{GDP}_t \) is the estimated SE as a percentage of official GDP.

The SE reached in the first quarter of 2000 almost 34% of GDP and followed a descending trend attaining the value of 30% in the start moment of economic crisis in Romania (the third quarter of 2008). From this point forward, the size of the SE begins to slowly increase, oscillating to about 33% of official GDP in the third quarter of 2010. For the last four years, the magnitude of the informal sector measured about 28% - 29% of official GDP.
2.3. Labor approach

2.3.1. Data and methodology

According to Schneider and Buehn (2016, p. 12), the labour approach relies on the idea that, “if total labour force participation is assumed to be constant, a decline in participation of the labour force in the official economy can be seen as a signal of increased activity in the SE”.

Schneider and Buehn (2016) also mentioned the fact that these fluctuations in participation rate can have other reasons and also individuals can work both in formal and informal economy thus this method provides relatively unreliable estimates of SE. The research of Crnkovic-Pozaic (1999) and Svec (2009) for Croatia and Nastav and Bojnec (2007) for Slovenia constituted the base for the estimation of SE using the labor approach. Following the approach of Crnkovic-Pozaic (1999), started from the very restricted assumption according to which any fluctuation in the participation rate can be putted on behalf of informal activities (which is not very credible).

The method involves two active population-hypothetical and de-facto (actual) and the difference between both of them could be considered as the number of people who are push in SE.

In order to compute the hypothetical activity rate it is necessary to set a reference period in which it is supposed that the informal activities are negligible. For obtaining the size of SE as % of official GDP using this method, it is mandatory to take into account information related to the productivity of these people and the most common assumption was that the productivity is the same in both sectors due to the lack of such an information in SE. The source of data is provided in Table A.2 of the appendix.

2.3.2. Empirical Results

As the reference period has been considered, the third quarter of the first year 2000, in which the facto activity rate reaches its maximum, 70.2%. The activity rate decreases until the end of 2005, while furthermore the trend is inverted until the end of 2014. Activity rates had an oscillating trend ranging from 60% to 66%. It was observed that when the activity rate decreased, the number of persons involved
in SE increased, revealing that most probably there was a flow between both sectors, but with the mention that fact is only partially true; not all these people are going to the informal sector, part of them becomes inactive.

Adjusting for seasonality the series using the Census X-13 and making the assumption that both sectors have the same labor productivity, we can multiply the number of individuals that operate in the SE with the real labor productivity; we will obtain the size of SE in terms of GDP.

The size of the SE registered the value of 7% at the beginning of 2000 and followed an upward trend until the end of 2005, reaching the value of 15%. During the period of 2006–2010, the size of the SE decreased, registering the value of 7%. For the year 2011, the dimension of SE increased to 11.4% in terms of GDP. For last periods, SE followed a continuous downward trend attaining about 6.3% reported to GDP. The empirical results need to be seen with due reserve, taking into consideration the underestimation of the phenomenon.

3. Comparisons between the main estimates of the Romanian shadow economy

Starting from the observation of Schneider and Williams (2013), an ideal or optimal method of estimation of SE doesn’t exist; each of the methods in literature have pluses and minuses and none of them does not offer a precise figure of this phenomenon. Nevertheless, it becomes plausible to use several methods to capture the big picture of the phenomenon regarded from different points of view, coming somewhat closer to the actual value of the SE.

Comparing the three estimates of the SE using different approaches, we can conclude that the currency demand approach and MIMIC led to relatively close values and to a common trend.

Table 4. Descriptive statistics of SE estimates

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currency ratio approach</td>
<td>60</td>
<td>30.48</td>
<td>4.48</td>
<td>19.10</td>
<td>41.05</td>
</tr>
<tr>
<td>MIMIC approach</td>
<td>60</td>
<td>32.00</td>
<td>1.75</td>
<td>29.40</td>
<td>35.60</td>
</tr>
<tr>
<td>Labor approach</td>
<td>60</td>
<td>10.11</td>
<td>3.52</td>
<td>0.03</td>
<td>19.00</td>
</tr>
</tbody>
</table>
All the macro econometric approaches estimate a decreasing trend starting from 2003 to 2014, with the exception of a two slight increases in the 2004–05 and 2009–10. By comparing our own estimates with Schneider et al. (2010) and Schneider (2015), we find strong similarities in terms of both trend and magnitude. Schneider (2013) mentioned that the decline in the shadow economy in 2012/2013 comparative with 2008/2009 can be attributed to the recovery from the economic crisis. Making the official economy more attractive could represent a disincentive for going in the informal sector.

Finally, we compare our estimates with the estimates of the size of NOE estimated by the RIS using the labor input method. Although differences can be also due to diverse definitions of SE compared to NOE, Figure 2 shows as the RIS calculates a value of NOE in terms of GDP of 18.1% in 2000, and it slowly increased to about 19.8% at the end of 2009.
4. Building a composite index of the Romanian shadow economy

4.1. The empirical results of the principal component analysis (PCA)

Using the estimates obtained by the currency demand approach, MIMIC model and labor approach, we calculate a composite index of the Romanian SE. Following Gerlach and Yiu’s (2005) approach, we first apply Principal Components Analysis (PCA) to extract the first principal component from the data to obtain an impression of the unobservable co-movement component of the three SE series. The quarterly series of SE have been seasonally adjusted using Census X-13. Taking into account that all the three estimates of SE tend to approximate the same latent variable, we use the Cronbach’s Alpha\(^5\), to measure the internal consistency of the three estimates of SE as a group. The value of Cronbach’s Alpha coefficient of 0.67 highlighted the consistency of the three estimates of the SE and thus the aggregate index may be computed.

The PCA indicates the existence a first component that recovers 62.04% of the variance of original variables (Table 5).

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comp1</td>
<td>1.86112</td>
<td>0.6204</td>
<td>0.6204</td>
</tr>
<tr>
<td>Comp2</td>
<td>0.87169</td>
<td>0.2090</td>
<td>0.9109</td>
</tr>
<tr>
<td>Comp3</td>
<td>0.26718</td>
<td>0.0891</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

In order to evaluate the statistical validity of PCA results, the measure of the sampling adequacy Kaiser–Meyer–Olkin (KMO) was applied and its value greater than 0.5 is indicative of the underlying common factors.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comp1</th>
<th>Unexplained</th>
<th>KMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currency demand approach</td>
<td>0.6556</td>
<td>0.2002</td>
<td>0.5306</td>
</tr>
<tr>
<td>MIMIC approach</td>
<td>0.6632</td>
<td>0.1815</td>
<td>0.5294</td>
</tr>
<tr>
<td>Labor Approach</td>
<td>-0.3612</td>
<td>0.7572</td>
<td>0.8360</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.5473</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^5\) It is considered to be a measure of scale reliability.
Accordingly, taking into account a unique principal component, we use the squares of the loadings to construct the composite index of SE.

\[ SE_{PCA} = (0.656)^2 SE_{CD} + (0.663)^2 SE_{MMIC} + (-0.361)^2 SE_{LA} \]  

(12)

**Figure 3. The evolution of the composite index of the Romanian SE**

According to the evolution of the composite index, SE followed a downward trend during the period 2000–2009. For the period 2009–2010, the magnitude of the SE slowly increased, while during the last years has been decreasing. The downward trend in SE in the first period of analysis may be primarily motivated by the preparations for EU membership and by the rapid spread of electronic payment services (Antohi et al., 2007). Accordingly, we can conclude that in fighting SE, it is recommended to encourage electronic payments. The introduction of compulsoriness of electronic payment for the amounts that pass a certain level may be a suitable method to increase tax compliance. Once Romania joined EU, at the beginning of 2007, SE decreased, registering the value of 25.6% of official GDP.

By matching the trend of the estimated SE with Romanian reforms, we can examine the effects of these reforms on the SE. In particular, we report in Figure 3, the coming into force of: an amendment of Labor
Code that eliminated the collective labor agreement and introduced the criminalization of undeclared work (in the second quarter of 2011); the introduction of the authorized physical person (PFA) (in the last quarter of 2008); the decreasing of 25% of salaries in the public sector from the third quarter of 2010 to the first quarter of 2013; the VAT increasing from 19% to 24% in the third quarter of 2010. The introduction of the authorized physical person (PFA) (in the last quarter of 2008) has contributed to the increase of SE, allowing some employees to contribute with less money. Few contribute to such systems and therefore have great financial shortcomings. There is a gap between taxpayers depending on the form of income and this differentiation favorite PFAs. The new trend is to work on PFA instead of an employee because they are much smaller contributions. There are from 300,000 to 400,000 of active PFAs in Romania.

4.2. The empirical results of dynamic factor models (DFM)

This aim of this section is to apply the single-index DFM, advanced by Stock and Watson (1991) to produce a composite index for the Romanian SE. PCA was used to highlight the common component of the three indicator series. From the three estimates of SE, the labor approach series has been seasonally adjusted by the X-13 method.

Analyzing the order of integration of all three series, using Augmented Dickey Fuller test, all the series have a unit root, being I(1).

According to Duarte (2014), the main hypothesis of these kinds of models is that each informality estimation can be written as the sum of two components: the “true” unobserved variable \( f_t \), which is common for all three estimates and an idiosyncratic component \( u_t \), which stores the specific characteristics of each series. Accordingly, the model can be written as:

\[
Z_t = L f_t + u_t
\]

where \( Z_t \) is a (3×1) vector of variables, \( f_t \) a vector of common factors in terms of the first differences and \( L \) a (3×1) vector of factor loadings’’.
The main advantage of DFM comparative with usual factor models is the fact that it allows the autocorrelation of the factors and idiosyncratic components (Breitung and Eickmeier, 2006). Following Duarte (2014), we specify the equations for the common factor $f_t$ and the idiosyncratic component $u_t$ as:

\[ f_t = A_1 f_{t-1} + \ldots + A_p f_{t-p} + w_t \]  \hspace{1cm} (14)

\[ u_t = C_1 u_{t-1} + \ldots + C_q u_{t-q} + \varepsilon_t \]  \hspace{1cm} (15)

where $w_t$ and $\varepsilon_t$ are white noise processes. The model can be estimated using the Kalman filter.

The optimal specification for the DFM was determined at one lag for $f_t (p = 1)$ and zero for $u_t (q = 0)$ based on the lowest value of AIC and SCH. The factor loadings for MIMIC approach and currency approach were statistically significant at 5% and 10%, respectively, while the factor loading of labor approach is no statistically significant.

<table>
<thead>
<tr>
<th>Table 7. The empirical results of dynamic factor models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>$F_{t,1}$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Mimic_std</td>
</tr>
<tr>
<td>Currency_std</td>
</tr>
<tr>
<td>Labor_std</td>
</tr>
<tr>
<td>Var (e.mimic_std)</td>
</tr>
<tr>
<td>Var (e.currency_std)</td>
</tr>
<tr>
<td>Var (e.labor_std)</td>
</tr>
</tbody>
</table>

The empirical results revealed that the unobserved factor was a significant predictor for the observed variables. Figure 4 shows the estimated unobservable common component, ($\hat{F}$) in first differences, against the first differences in each of the three SE series.
We calculate the co-movement component $f_i$ as a weighted average of the first difference of the indicator series. Following Gerlach and Yiu (2005, p. 7), “the weights are those used to construct $f_i$ from the indicator variables and has been estimated from the Kalman Filter algorithm”. Figure 5 shows the first difference of the estimated factor ($\hat{f}$) of SE obtained using DFM against the first difference of SE using static PCA.

Analyzing both the empirical results, we can highlight the similarities in the evolution of absolute changes in the Romanian shadow economy.
5. Conclusions

This paper aims to contribute to existing literature from both methodological and positive viewpoints.

From a methodological perspective, since the literature unambiguously states that it does not exist an ideal or dominating macro-econometric method to estimate the dimension and development of the SE, we consider worthwhile use different estimation methods to come a somewhat closer to the actual size and development of the SE.

From a positive side, we estimate the size of the Romanian SE in terms of GDP using quarterly data for the period 2000–2014. In particular, we apply three of the main commonly used approaches of the estimation of SE, that is, currency demand approach, MIMIC model and labor approach.

Given the methodological premise, the results obtained in this study should always be seen as an approximation and treated with a certain reserve, given the limitations of the methods.

Following the results of the currency demand approach and the MIMIC model, the size of SE registered about 35%-38% of the official GDP in 2000 and engaged on a downward trend until 2009 and increased to almost 32% in the period 2010/2011. In the last period, there was a slow decline in the size of the shadow economy reaching the value of 26%-30%. Using the labor approach, the magnitude of SE quantified almost 18% of the official GDP in 2000 and declined about 11% in the last period.

Using the previously estimates of the Romanian SE, we made a first attempt by computing a composite index of the Romanian SE using both the PCA and the DFM, assuming that all the three informality estimates aim at measuring the same unobservable variable.

Taking into account that all macro econometric methods produce inaccurate results in terms of absolute estimates of SE – e.g. because all of them need an exogenous estimate of SE in a reference year to calibrate the estimated index – we find that the estimates of SE reveal evident similarities in terms of the trend from 2003 to 2014. Accordingly, a composite index of SE, both based on PCA and DFM can be considered a useful instrument to grasp the dynamics of informality.
## APPENDIX

### Table A.1. The description and source of the data - Currency demand approach

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>Natural logarithm of real currency holdings measured as nominal currency in circulation normalized by GDP deflator (205=100). 2000Q1-2014Q4 in national currency (mil.RON). The values of currency, all notes and coin outside the banking system are those at the end of the quarter.</td>
<td>The series of nominal currency is available in National Bank of Romania Monthly Bulletins 2000-2015. The series of GDP deflator (2005=100) is available in Eurostat, Quarterly National Accounts database.</td>
</tr>
<tr>
<td>$M_1$</td>
<td>Natural logarithm of real monetary aggregate $M_1$, normalized by GDP deflator (205=100). 2000Q1-2014Q4 in national currency (mil. RON). The definition of monetary aggregate according to the methodology of European Central Bank beginning with 2007 produces some changes in monetary aggregate $M_1$ who includes from 2007 in addition to the structure employed until December 2006, the demand deposits of household savings expressed in RON and sight foreign currency deposits of residents previously included in quasi-money, being considered as having the same degree of liquidity as the demand deposits of economic agents. The series has been recalculated for the period 2000-2006 using data from National Bank of Romania, Monthly Bulletins 2000-2014.</td>
<td>Monthly bulletins of National Bank of Romania. M1 comprises currency outside banks and demand deposits.</td>
</tr>
<tr>
<td>Tax</td>
<td>Natural logarithm of 1+ total of tax revenues (taxes on production and imports, current taxes on income, wealth, social contributions) over GDP. 2000Q1-2014Q4 in %.</td>
<td>Quarterly Government Finance Statistics, Eurostat.</td>
</tr>
<tr>
<td>$R$</td>
<td>Natural logarithm of nominal interest deposit rate. 2000Q1-2014Q4 in %. International Monetary Fund defines the interest rate as the average rate offered by credit institutions on outstanding time deposits of nonfinancial corporations and households in national currency.</td>
<td>International Financial Statistics (IFS) database of International Monetary Fund, Financial Indicators section.</td>
</tr>
<tr>
<td>Urb_empl</td>
<td>Employment rate from urban area, 15-64 years, 2000Q1-2014Q4, %.</td>
<td>Tempo database of National Institute of Statistics.</td>
</tr>
<tr>
<td>$\pi$</td>
<td>Natural logarithm of (1+inflation rate). 2000Q1-2014Q4%. The series of quarterly inflation rate is calculated as a arithmetic mean of the chain-based monthly harmonized indices of consumer prices, from which the comparison base equaling 100 is subtracted. In the raw data the lowest value for inflation was -0.3667%. Since we use a logarithmic transformation for the inflation rate in our regressions, we add 1 to this variable to make the lowest observation 0.</td>
<td>Inflation Database, Eurostat.</td>
</tr>
</tbody>
</table>

### Table A.2. The description and source of the data - Labor approach

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working Age population (15-64 years)</td>
<td>2000Q1-2014Q4</td>
<td>Employment and Unemployment database, Eurostat.</td>
</tr>
<tr>
<td>Employment (15-64 years)</td>
<td>2000Q1-2014Q4</td>
<td>Employment and Unemployment database, Eurostat.</td>
</tr>
<tr>
<td>ILO unemployed (15-74 years)</td>
<td>2000Q1-2014Q4</td>
<td>Employment and Unemployment database, Eurostat.</td>
</tr>
<tr>
<td>Labor productivity</td>
<td>Real labor productivity per employed. 2000Q1-2014Q4</td>
<td>Employment and Unemployment database, Eurostat.</td>
</tr>
</tbody>
</table>
### Table A.3. The description and source of the data - The MIMIC model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate</td>
<td>Unemployment rate represents the ratio of the unemployed, according to international definition (ILO*), in economically active population, 2000Q1-2014Q4, %. Unemployed persons are persons aged 15-74 who were without work during the reference week, were currently available for work and were either actively seeking work in the past four weeks or had already found a job to start within the next three months.</td>
<td>Labor Force Survey, Eurostat.</td>
</tr>
<tr>
<td>Employed persons having a second job</td>
<td>Population in employment having a second job refers only to persons with more than one job at the same time. Consequently, persons having changed job during the reference week are not covered. 2000Q1-2014Q4, %.</td>
<td>Labour Force Survey, Eurostat.</td>
</tr>
<tr>
<td>Real interest deposit rate</td>
<td>Real interest deposit rate, 2000Q1-2014Q4, %. International Monetary Fund defines the deposit rate as the average rate offered by credit institutions on outstanding time deposits of nonfinancial corporations and households in national currency. The series is computed using nominal interest rate and inflation rate by the formula: ( 1 + R_r = (1 + R)/\left(1 + \pi\right) ), with: ( R_r = ) real interest rate; ( R = ) nominal interest rate; ( \pi = ) inflation rate.</td>
<td>International Financial Statistics (IFS) database of IMF, Financial Indicators section.</td>
</tr>
<tr>
<td>Regulatory quality index</td>
<td>Reflects perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development. The scores of this index lie between -2.5 and 2.5, with higher scores corresponding to better outcomes. 2000Q1-2014Q4</td>
<td>Worldwide Governance Indicators database, World Bank.</td>
</tr>
<tr>
<td>Index of real GDP</td>
<td>Index of real GDP, 2005=100, %</td>
<td>Quarterly Government Finance Statistics of Eurostat.</td>
</tr>
</tbody>
</table>

---

30
References
Alexandru, Adriana, Dobre, Ion, 2011. The impact of unemployment rate on Romanian shadow economy: An empirical investigation using Granger causality analysis, Economic Computation and Economic Cybernetics Studies and Research, 45, 55-76.


Data sources:

*** Employment and Unemployment database, Eurostat.


*** Quarterly Labor Force Survey database, Eurostat.

*** Quarterly Monetary and Financial Statistics database, Eurostat.

*** Quarterly National Accounts database, Eurostat.

*** Tempo database, National Institute of Statistics
