

USING MIMIC MODELS TO EXAMINE DETERMINANTS OF VAT GAP IN LITHUANIA

Gindra Kasnauskienė*

Vilnius University

Jolita Krimisieraitė

Vilnius University

Abstract: *In recent years analysis of economic loss attributed to different aspects of shadow economy has attracted much attention of both academics and policy makers. Recent statistical data shows that new member states have on average a 9 percent higher VAT gap than the older members of the European Union. Knowing that economies of emerging markets rely on the VAT for a substantially higher percentage of their government revenues, it is very important to understand the determinants limiting revenue mobilization in those countries. In Lithuania, the VAT gap increased dramatically after the crisis of 2008, and now is one of the largest in the EU. However, few studies have empirically tested some hypotheses about the VAT as a revenue-raising instrument in the country. The purpose of this study is to identify the determinants significantly influencing the size of the VAT gap in Lithuania using the MIMIC method for quarterly data of the period 2000-2013. The applied MIMIC model indicated that two factors (General government consumption expenditure and inflation) have a statistically significant impact on the VAT gap in the long-run. The results of the eMIMIC model show that two determinants (inflation and household deposits) have a statistically significant influence on the gap in the short-run. The authors believe that the key findings of the study can be used as one of the supporting tools in adjusting Lithuanian pro-growth tax policy and improving administration of VAT taxes.*

Key words: *VAT gap, VAT revenue, tax evasion, MIMIC.*

1. Introduction

Value Added Tax (VAT) revenue is one of the most important sources of the state's income. According to the latest annual data on the implementation of the National Budget presented by the Ministry of Finance of the Republic of Lithuania, it amounted to 40.76 percent of the state budget income. Analyses of economic trends in the various phases of the economic cycle highlight the problem of the shadow economy,

* Corresponding author: Faculty of Economics, Department of Quantitative Methods and Modeling, Vilnius University, Saulėtekio Ave. 9, LT – 10222, Vilnius, Lithuania; e-mail: gindra.kasnauskienė@ef.vu.lt

particularly the VAT compliance gap, which is henceforth simply called the “VAT gap”. The most commonly used definition describes the VAT gap as the “difference between the theoretical tax liability according to the tax law and the actual revenue collected” (CASE, 2013, p.18). Tax evasion and tax avoidance are widely believed to be important causes of inefficiency in the functioning of the market, limiting the country’s capacity to mobilize money and implement their economic and social policies.

VAT gaps have a very wide dispersion range across countries. One of the findings presented in the report of the “Study to quantify and analyse the VAT gap in the EU-27 Member States” commissioned by the EC is that in 2011 the largest VAT gap as a share of GDP was in Romania, Latvia, Greece and Lithuania (7.9, 4.7, 4.7, and 4.4 percent, respectively) (CASE, 2013). Naturally, better understanding of determinants of the VAT gap as a response to economic events or policy decisions in Lithuania is required.

There is a cause - effect relationship between the VAT gap and many macroeconomic variables, for example, a shortfall of the National Budget revenue, decline in the GDP, decrease in additional wages and other indicators that reflect negative trends in the economy. Since causes determining tax revenue and the strength of their impact are different for each type of the tax, it is not appropriate to assess all the revenues as a single variable. Consequently, in this study we decided to analyse separate economic factors which may have statistically significant influence on the VAT gap. It should be mentioned that there is no single methodology for precise estimating of the VAT gap; its results could not be safely used as a basis for practical application. Having conducted a critical review of the methods in this field we decided to apply the most appreciated method for our study – Multiple Indicators and Multiple Causes (MIMIC) model, where the VAT gap is treated as a latent variable.

The main purpose of this study is to review the various methods estimating the VAT gap and, using MIMIC models for quarterly 2000-2013 period data, to identify the determinants significantly influencing the VAT gap in Lithuania. The paper is organized as follows:

The following part discusses the methods for quantifying the VAT gap and existing estimates. Further, the analysis of the variables influencing either the VAT gap or VAT performance used in the MIMIC models is reviewed. In the next part of the paper, the construction and analysis of the long run eMIMIC models is presented. The fifth part analyses the results of the short run eMIMIC models, including comments on the economic interpretation. The sixth section describes some limitations of the study and suggestions for the future research. In conclusions section the results are summarized and discussed.

2. Review of Literature

VAT non-compliance problems are widely studied in scientific literature. Historically, M. G.Allingham and A.Sandmo (1972) were among the first authors who published the paper on tax evasion. The study mainly focused on the individual tax-payers’

decision to avoid taxes. Later on, it became more important to concentrate on the VAT gap measurement and identifying factors having influence on such a gap. Basically, there are two main approaches in the field to estimate the tax gap: macro (top-down) and micro (bottom – up). Macro methods allow researchers to calculate the VAT gap at the macroeconomic level and are preferred when analysing indirect taxes. Micro methods focus on direct taxes in the sense that they measure missing tax income based on samples of individual tax-payers (Gemmell & Hasseldine, 2012).

The macro or *top – down* method was used to estimate the VAT gap in the „Study to quantify and analyse the VAT gap in the EU – 25“ over the period from 2000 to 2006 carried out by a London-based economic consultancy company Reckon (2009). Total accrued tax receipts were compared with a theoretical tax liability calculated from general economic data. However, when using this method the structural information about the most fraudulent companies or economic sectors is lost. Additionally, the use of this method in order to calculate the VAT gap involves more disadvantages: firstly, particularity of available data, and, secondly, part of taxable activities, such as construction of own dwelling or exceptions to small business, is outside the National accounts; thirdly, it is difficult to determine whether the National accounts estimates are accurate. The study used the ESA 95 standard, which requires that influence of the shadow economy should be included in the National accounts data, so the accuracy and the quality of the data depends on how logically and properly calculations of the shadow economy are performed. However, total estimated liabilities of the VAT tax are calculated as a sum of total household consumption, gross fixed capital formation, other consumption, which contains government intermediate and final consumption expenditure, intermediate consumption of other sectors and adjustments that are related to small business exemption, company cars, business environment and changes in worth. It should be noted that top-down estimation method of the compliance gap was also used in the recent CASE (2013) study.

One of the *bottom – up* (micro method) analyses was performed by the specialists of the Swedish National Tax Agency (2008). An extensive survey, which was intended to help determination of the behaviour of the taxpayers under different circumstances, tax rates and so on, was conducted to identify the difference between the official sources of household consumption expenditure and those obtained by the survey results. Hanousek and Palda (2002) used the bottom – up method to describe evolution of tax evasion over time and to get a detailed view on the structure of it. The major disadvantage of such (micro) methods is that the main source of data is a survey, which is difficult to be controlled. Moreover, the respondents can interpret the same terms differently or can be simply afraid to tell the truth. On the other hand, if the same parameters are estimated using different sources of data, they can produce imprecise and sometimes conflicting results.

The VAT gap, the calculation of which is based on the above discussed methods, is widely used in the econometric models. However, as it was mentioned earlier, these

methods may measure the VAT gap incorrectly. Therefore, it was decided to use a different model in this study – Multiple Indicators and Multiple Causes (MIMIC) model. The MIMIC model is a special type of the structural equation model that specifies the causal relationships among observable variables and an unobservable variable. The major advantage of the MIMIC model is that the main attention is paid to the latent or unobserved variable, for example, such a hardly measurable variable as the VAT gap, which we can investigate and evaluate only indirectly, and assess its effects on directly monitored variables. All the methods described above designed to estimate the size of the VAT gap consider just one indicator that is supposed to include all the effects of the VAT gap. But it is obvious that it is very difficult to evaluate all these effects correctly and accurately. By contrast, the MIMIC method is quite different from those and is based on the statistical theory of “unobserved variables, which considers multiple causes and multiple indicators of the phenomenon to be measured” (Schneider, 2006, p.48).

The MIMIC model has become very popular to be applied in measuring hidden economy and the tax gap in recent years. It is well appreciated among the scientists. D.Giles and L.Tedds (2002) have used the MIMIC method in order to evaluate underground economy in Canada, A. Bühn and F.Schneider (2008) in France. R.Dell’Anno (2006) applied it in order to find plausible answers to the following questions: What are the dynamics and size of the Portuguese shadow economy (as percentage of the official GDP) in the last thirty years (from 1977 to 2004)? What are the main causes of shadow economy? What kind of economic policies could be effective in reducing shadow economy? C.- M. Ene and A. Ștefănescu (2011) have also used the MIMIC approach in their work „Size and implication of Underground Economy in Romania – a MIMIC approach“.

The general structure of the MIMIC model is presented in Figure 1:

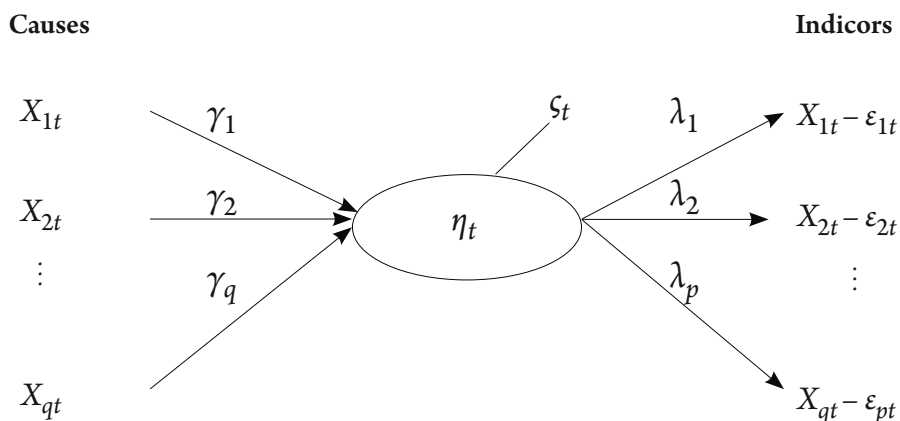


FIGURE 1. General Structure of the MIMIC model

Source: A. Bühn & F Schneider, 2008, p. 6.

As we see from Figure 1, formally, the MIMIC model consists of two parts: the structural equation model and the measurement model. The structural equation is specified by this equation:

$$\eta_t = \gamma^T \chi_t + \zeta_t \quad (1)$$

Where $\chi_t^T = (\chi_{1t} \chi_{2t} \dots \chi_{qt})$ is a $(1 \times q)$ vector of a set of observable time series variables (causes), $\gamma^T = (\gamma_{1t} \gamma_{2t} \dots \gamma_{qt}) - (1 \times q)$ vector of coefficients in the structural model, which describes the causal relationship between the latent variable η_t and its causes; ζ_t is the error term, which represents the unexplained component. Symbol T henceforth means that an appropriate vector is transposed.

Latent variable (η) describes the endogenous variables that depend on the measurement errors (the measurement model):

$$y_t = \lambda \eta_t + \varepsilon_t, \quad (2)$$

where $y_t^T = (y_{1t} y_{2t} \dots y_{pt})$ is a $(1 \times p)$ vector of individual endogenous time series variables (indicators) $y_{jt} = (1, \dots, p)$, $\varepsilon_t = (\varepsilon_{1t} \varepsilon_{2t} \dots \varepsilon_{pt})$ is a $(p \times 1)$ vector of disturbances where every ε_{jt} ($j = 1, \dots, p$) is a white noise term the covariance matrix of which $(p \times p)$ is given by Θ_ε . λ_j ($j = 1, \dots, p$) is a $(p \times 1)$ vector of regression coefficients λ . Both structural noise ζ , as well as measurement error ε are represented by normal distribution and are linearly independent of each other. In order to create the model, the following assumptions were taken:

- 1) $E(\eta_t) = E(\chi_t) = E(\zeta_t) = E(y_t) = E(\varepsilon_t) = 0$ – means are equal to 0;
- 2) $E(\chi_t \zeta_t^T) = E(\zeta_t \chi_t^T) = 0$ – the error terms in the structural model do not correlate with the causes;
- 3) $E(\chi_t \varepsilon_t^T) = E(\varepsilon_t \chi_t^T) = 0$ – the error terms in the measurement model do not correlate with the causes;
- 4) $E(\eta_t \varepsilon_t^T) = E(\varepsilon_t \eta_t^T) = 0$ – the error terms in the measurement model do not correlate with the latent variable;
- 5) $E(\zeta_t \varepsilon_t^T) = E(\varepsilon_t \zeta_t^T) = 0$ – the error terms in the structural model do not correlate with the error terms in the measurement model;
- 6) $E(\zeta_t^2) = \sigma^2$ – the variance of the error terms in the structural model is a constant.

By using equations (1) and (2) the MIMIC model covariance matrix Σ (3) can be defined. It reveals the structure between the observable variables and the latent variable:

$$\Sigma = \begin{pmatrix} \lambda(\gamma^T \Phi \gamma + \Psi) + \Theta_\varepsilon & \lambda \gamma^T \Phi \\ \Phi \gamma \lambda^T & \Phi \end{pmatrix} \quad (3)$$

As the latent variable is unobserved, its size is unknown, so the model parameters must be evaluated using the relationships between variance and covariance of the

observed variables. The aim of the assessment procedures is to find values for the parameters γ and λ and also the values of covariance matrix, ψ and Φ (where Φ – is the $(q \times q)$ covariance matrix of the causes χ_t , ψ – is the variance of ζ_t and ε – is $(p \times p)$ covariance matrix of the white noise term) which are compiled to matrix, which is as close as possible to the covariance matrix of observed variables and indicators, i.e., all χ_t and all y_t .

The following MIMIC model was considered as a long-run MIMIC model in this study. The aim of such a model was to find the trends and mutual associations between the variables during all the period (from 2000 to 2013). However, it is also very important to examine an instantaneous relationship between variables. Therefore, an error correction MIMIC (eMIMIC) model was used, where the long run equilibrium errors together with causal variables were evaluated (Bühn & Schneider, 2008). Thus, after adding the long run equilibrium errors (the number of the errors must be equal to the number of indicators) to the long run MIMIC model, the short run eMIMIC model was constructed. The long run equilibrium errors are found by cointegrating the variables. If a stationary linear combination of non-stationary variables exists, the variables are co-integrated (Enders, 1995). While creating the model, due to the fact that only non-stationary variables can be co-integrated, a stationary variable (if it exists) must be used as an exogenous variable.

In summary, analysis of scientific literature reveals that there are almost no econometric models which would be suitable in the analysis of the VAT gap. Most of them are based on the principle that primarily the theoretical revenue of VAT is calculated and only after that the econometric models are constructed and, according to the results of the models, appropriate conclusions are drawn. However, such techniques of the VAT gap calculation are not always sufficient and do not determine the situation when incorrect and inaccurate conclusions are being made. Therefore, the most appropriate method is the MIMIC model, in which the VAT gap is considered to be a latent variable. On the one hand, it is associated with a set of observable variables and, on the other, – with the causal variables, which affect the VAT gap activity.

3. Selection of variables used in the MIMIC model

The level of tax evasion, or the VAT gap, depends on a number of different factors. There are numerous studies describing the variables likely to affect non-compliance. Sometimes the study results may differ meaningfully depending not only on the use of the VAT gap estimation method, but also on the selected indicators of the economic variables. Therefore, it is very important to choose variables very precisely in order to get the most accurate results. Based on the analysis of scientific literature (Allingham & Sandmo, 1972; Christie & Holzner, 2006; Schneider, 2006; Schneider & Bühn, 2007; Nerre, 2008; Bühn & Schneider, 2008; Fathi & Esmailian, 2012), the variables believed to determine the VAT gap were selected for empirical work:

- Variables affecting the VAT gap – General Government consumption expenditure, inflation, gross fixed capital formation, additional amount of wage and money on deposit;
- Variables affected by the VAT gap – real GDP per capita, VAT revenue, cash and cash equivalents.

As stated in the literature (Tedds, 2002), development of the public sector, i.e., increasing government expenditure or economic regulation is often one of the incentives to take informal or illegal activities. That is why the first variable which is likely to affect the VAT gap, *General Government consumption expenditure* (GGCE), is chosen.

Inflation (INFL) is defined as a sustained increase in the general level of prices of goods and services. Along with the increase in the inflation, consumer purchasing power decreases. Such a situation for both consumers and producers, who want to avoid negative effects of the inflation, may encourage them to look for the alternatives such as purchasing goods and / or services illegally at a relatively low price. The inflation is calculated according to the EU methodology and using statistical data on consumer price changes in the consumption pattern of households (the harmonized consumer price index) (Statistics Lithuania, 2014).

The subject's decision to carry out the tax obligations is to a great extent determined by the mood of the taxpayers, expectation for the future and the total economic situation in the country. As a result, the selected variable that should influence the mood of the taxpayers is an *additional amount of wages* (AAW), which was calculated as a difference between the average and the minimum wages. This variable indicates how much money above the minimum wage workers receive. The growth of this amount may be associated with the higher quality of life – individuals who receive higher wages can spend more money on consumption or saving, i.e., investments, *ceteris paribus*. On the other hand, in terms of the employers, decline in the wages can be associated with decision to conceal the income, respectively, to pay the salaries in an “envelope” and thus to evade tax obligations.

Another variable of the model is *household deposits* (HD). The growth of wages does not necessarily mean that consumption will grow, too. Households decide what to do with their free money: release it for consumption or for saving. If they decide to save (for example, to put money on a deposit), then the country's economy will not grow to the extent that it could at the given period.

The last determinant is *gross fixed capital formation* (GFCF), which otherwise is considered as investments and represents physical capital replenishment. It is logical to assume that when the national economy is growing, the business must ensure a sufficient supply of goods and services in order to survive and to be competitive in the market. This increase of supply is inseparable from the greater need of funds for investment. Therefore, it is appropriate to analyse whether the investments tend to be financed from the “saved” money, i.e., concealed income, or whether growing investment leads to higher revenue and reduces the incentives to evade tax obligations.

The first indicator – *real GDP per capita* (rGDP) – is one of the most important variables in the model. The real GDP (measured at constant prices (the chain-linking method)) rather than the nominal (nGDP) is used in the analysis in order to avoid the influence of the inflation. The studies suggest that income which is earned in the informal economy increases GDP growth, since some of it is released in the formal economy (Krumplytė, 2010). So, national recession could lead to a tendency of individuals to save money instead of paying VAT. On the other hand, there are more opportunities to circumvent the law, and tax evasion may become easier when the economy grows. In addition, part of the „shadow“ money is released officially, which leads to economic growth.

The second indicator is *VAT revenue* managed by the State Tax Inspectorate (VATR). It is evident that not fulfilling tax liabilities on time or not fulfilling at all (both intentionally (for example, fraudulently avoiding to pay VAT) and unintentionally (for example, due to a lack of knowledge, taxpayers complete the VAT declaration incorrectly and pay less tax accordingly)), directly affects the VAT revenue. According to the Republic of Lithuania Law on Tax Administration, there are two institutions that are responsible for the VAT revenue collection: the State Tax Inspectorate and the Customs of the Republic of Lithuania (Chapter IV, Art. 15). However, there are some differences in the structure and the nature of work between these two institutions. Because of these differences and in order to make the study more accurate, VAT revenue which is managed by the State Tax Inspectorate was selected for our empirical work.

The third indicator is *cash and cash equivalents* (MONEY) (the money that is in the company's accounts and short – term (up to 3 months), i.e. liquid, investments, which can be quickly converted to cash and the risk of the change value is negligible) (S BAS of Business accounting standards, 2008). This indicator was selected because money is used not only in formal economy, but also in informal – money is also needed in order to carry out the illegal transactions on which the relevant taxes are not paid. Thus, the growth of the informal economy should increase the demand for money.

Quarterly data for the period of 2000-2013 were collected from various sources: General Government consumption expenditure, inflation, additional amount of wage, real GDP per capita and cash and cash equivalents were obtained from the Official Statistics Portal of Statistics Lithuania; money on deposit data was extracted from the “Report on financial market state and execution of functions” that is prepared by the Central Bank of the Republic of Lithuania; gross fixed capital formation was taken from the Eurostat Statistics Database; VAT revenue from the “Data on the Implementation of the National Budget” prepared by the Ministry of Finance of the Republic of Lithuania.

Prior to initiating the econometric model, the data needs to be prepared - seasonally adjusted and made stationary. These procedures are necessary in order to get the correct results and reasonable conclusions. It should be noted that this model formation is not mentioned in most of the articles, which could mean that not all the results of these models are reliable. In this study, first of all, part of the variables, i.e., those which have multiplicative decomposition, were logarithmized (for example, the case of *Cobb* –

Douglas production function). These include: General Government consumption expenditure, the gross fixed capital formation, the additional amount of wages, the households deposits, VAT revenue managed by the STI, real GDP per capita and cash and cash equivalents, i.e. variables, which are calculated as a percentage value of other variables or indicators and cannot gain negative values. The latter condition is relevant due to the fact that this procedure cannot be performed for negative or equal to 0 values. Since quarterly data are used in the model, it is possible that the seasonal component can be statistically significant. In order to avoid that influence which could distort the results, the variables were seasonally adjusted. This procedure was performed using the command “*decompose*” of the software R. Depending on the mathematical expressions of the variables, the suitable method of the seasonal adjustment was selected: multiplicative model for the logarithmized variables (here the seasonal variations are roughly constant through the series) and additive for the other variables (here the seasonal variations are changing proportionally to the level of the series).

Then, to overcome the problem of spurious regressions, all the variables were transformed into stationary ones. It is often omitted or not emphasized that this step is performed in the econometric models. In order to find the order of integration, several unit root tests: *Augmented Dickey – Fuller (ADF)*, *Elliot Rothenberg Stock (ERS)*, *Phillips – Perron (P – P)*, *Kwiatkowski, Phillips, Schmidt & Shin (KPSS)* and *Zivot – Andrews (Z – A)* were run. All of these tests were performed at a 95 percent confidence level, i.e., at a 5 percent probability of error.

TABLE 1. Results of the unit root tests

Variables	ADF		ERS		P – Z		KPSS		Z – A	
GGCE	-3.70 (1)*	3**	-0.27 (1)	3	-13.54 (1)	-	0.21 (1)	-	-2.49 (0)	4
INFL	-13.82 (1)	0	-10.21 (1)	0	-4.69 (0)	-	0.25 (0)	-	-3.60 (0)	1
AAW	-2.26 (1)	1	-2.54 (1)	1	-4.60 (1)	-	0.27 (1)	-	-2.93 (0)	1
HD	-9.56 (1)	0	-2.72 (1)	0	-10.15 (2)	-	0.06 (2)	-	-2.84 (0)	1
GFCF	-6.74 (1)	0	-6.74 (1)	0	-6.81 (1)	-	0.15 (1)	-	-5.13 (0)	0
rGDP	-10.86 (2)	2	-1.33 (1)	2	-9.53 (1)	-	0.12 (1)	-	-3.48 (0)	3
VATR	-7.49 (1)	0	-7.58 (1)	0	-7.54 (1)	-	0.07 (1)	-	-5.36 (0)	0
MONEY	-6.93 (1)	0	-6.47 (1)	0	-8.05 (1)	-	0.14 (1)	-	-3.41 (0)	0
Critical values (5%)	-1.95		-1.95		-2.92		0.46		-4.8	

Notes:

* The number in parenthesis shows the order in which the process is integrated. For example, we say that the process is integrated of order 1 if it is calculated according to the formula $\Delta Y_t = Y_t - Y_{t-1}$.

** Number of lags

Source: calculated by the authors with the use of R software.

Based on the results of these tests, it can be concluded that the variables are not stationary and it is necessary to differentiate them, moreover, part of the variables have

from one to three periods to be lagged. In addition, Z – A test showed that GFCF has a unit root with a structural break in the intercept which occurred in 2008 Q3. For this reason, GFCF was used as an exogenous variable to find the number of cointegrating vectors. Z – A test showed that VATR also has a unit root with a structural break at the 18th observation, but including this break in the model is not considered as appropriate for the following reasons: the critical value of this test is -5.34 (1 % level) and it is slightly different from the calculated value -5.36, so hypothesis H_0 cannot be rejected. In addition, calculated break point is quite close to the beginning of the observations – therefore, it is probable that the structural break occurred at the beginning of the investigation period, i.e., when Lithuania became a member of the European Union.

Prepared-logarithmized and seasonally adjusted data was used to estimate the MIMIC and eMIMIC models with the use of R software. The following packages were used to estimate the models: *lavaan*, *urca*, *vars*, *semPlot*, *Hmisc*, *tseries*, *lmtest* and *fArma*.

4. Construction and analysis of the long run MIMIC models

It was estimated that the three variables (AAW, rGDP, GGCE) are stationary with 1 to 3 period lags, so in order to find the best structure of the MIMIC model, the variables were included both with and without lags. To compare the models the following criteria were used:

1. χ^2 – *Chi-square* – a compatibility criterion which is used to verify if observed distribution is compatible with the theoretical one (p-value must be more than 0.05);
2. DF – *Degrees of freedom* – a criterion which is calculated using the following formula: $0.5 \times (p + q) \times (p + q + 1) - t$, where p is a number of indicators, q is a number of causes and t is a number of free parameters;
3. RMSEA – *Root mean square error of approximation*; if RMSEA is less than 0.05, it means that the model is appropriate. RMSEA values can vary from 0.0 to 1.0;
4. AIC – *Akaike information criterion* is a criterion used for selecting among nested econometric models. The preferred model is the one with the minimum AIC value.

Firstly, the long run MIMIC model with five causal variables and three indicators was constructed. After evaluating the statistical significance of the causal variables, the least statistically significant causal variables were eliminated one by one until all remaining variables were statistically significant (estimated values of the information criteria are summarized in Table 2). The statistical significance was assessed by using z statistics. Since the critical values of z statistics vary depending on the degrees of freedom, *p-value* was used: the variable was considered as statistically significant when *p-value* < 0.05.

The MIMIC model presented in Figure 2 shows that both statistically significant causal variables have positive effect on the VAT gap: GGCE – 0.156 (at significance level of 0.043) and INFL – 4.713 (at significance level of 0.010). It should be noted

that because each of these variables was used at a stationary state, interpretation of the calculated estimates would be: 1 percent change in GGCE increases the VAT gap by 0.156 notional units, *ceteris paribus*; 1 percent change in INFL increases the VAT gap by 4.713 notional units, *ceteris paribus*.

TABLE 2. Analysis of the information criteria

	With all variables	Without HD	Without HD, AAW	Without HD, AAW, GFCF
X ² (p-value)	0.802	0.991	0.972	0.977
DF	10	8	6	4
RMSEA (p-value)	0.000 (0.862)	0.000 (0.994)	0.000 (0.980)	0.000 (0.982)
AIC	2491.60	2257.09	1993.35	1599.93

Source: calculated by the authors with the use of R software.

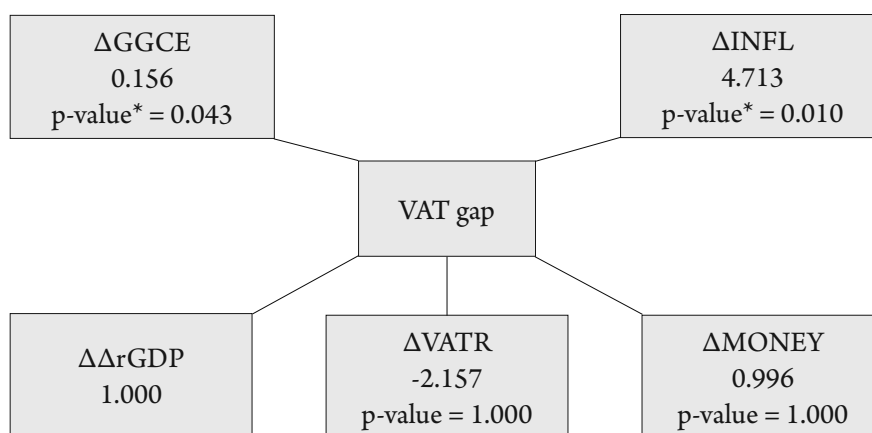


FIGURE 2. The estimates of the long run MIMIC model parameters

Notes:

* p-value – if it is greater than 0.05, it means that factorial weight which participates in the calculation of the VAT gap is not statistically significantly different from zero

Source: calculated by the authors with the use of R software.

According to the latest data provided by Statistics Lithuania (2014), one of the highest growths of the Government spending in absolute terms was on social protection (from 5.77 billion Litas in 2001 to 13.5 billion Litas in 2012). Although it is only a part of the Government spending, partial observations can be made. Rising expenditure on social protection means that people who are on low income are less interested to work, because the benefit from work is not sufficient enough for them to become active labour market participants. Although people have less revenue, there is more free time for other activities. On the other hand, lower expenses do not allow the business

to develop. In order to operate in a competitive environment, it has to find financial resources by other means. One of the alternatives is a failure to comply with the tax obligations, i.e., instead of paying taxes, including VAT, business may spend the money on the technological capacity building or increasing workers' salaries, etc.

The obtained results show that the growth of the VAT gap can be explained by the rising inflation, *ceteris paribus*. The individuals are very sensitive about the rising price level because it has a direct effect: purchasing power is declining. When inflation is low, overall price level changes slightly, but prices of particular goods and / or services that are relevant to certain individuals may actually have increased, and it reduces the purchasing power of customers nevertheless. In order to maintain the current level of consumption, individuals are forced to look for additional sources of income – both legal (e.g., looking for a new job) and illegal (e.g., engaging in illegal activity, hiding income and thereby avoiding tax compliance) ways. If individuals do not intend to search for additional sources of income and reduce the consumption of less necessary goods and / or services, the revenue of relevant companies decreases as well. The result is a similar situation as in the case of GGCE increase – rising inflation increases the VAT gap.

According to the results, the remaining variables (HD, AAW and GFCF) were not statistically significant. The inclusion of HD in the model was based on the consideration that it would have a positive effect on the VAT gap, *ceteris paribus*. However, it was found that this variable was not statistically significant and the direction of the impact cannot be defined. On the other hand, the estimated coefficient of HD was negative in the primary model, so it tentatively suggests that positive growth in the changes of HD (since the variable was integrated of order 2) determines the higher VAT gap. It implies that higher household savings are not conducive to business as demand for goods and services decreases, consequently, the National Budget income is reduced because the tax revenues are not being collected and the VAT gap tends to grow.

Another not statistically significant causal variable was AAW. Although increasing additional amount of wages should reduce the VAT gap, the estimate of DU was negative, moreover, this variable was not statistically significant. The reason for this result can be the fact that individuals who receive fixed salary are able to better plan their budgets, so the concealment of income and / or wage payment in the “envelopes” in order to evade tax obligations is less likely.

The study results show that the changes of gross fixed capital formation do not affect the VAT gap. One of the reasons can be that the changes of these investments during the analysed period from 2000 to 2013 are not considerable enough to significantly affect the businesses decision whether to carry out the tax obligations or not.

There were no statistically significant indicators in the long run MIMIC models: *p-value* of both the indicators VATR and MONEY was close to 1. It is evident that the VAT gap should negatively affect the VAT revenue. However, based on our results, it cannot be stated that the trend is either positive or negative. MONEY was not statistically significant in the first models, still it is considered to be important to the

assessment of the VAT gap, because transactions aimed to reduce or avoid the VAT liability are performed in cash. The close evaluation of the models indicates that inclusion of the lagged variables in the model was not expedient because none of them were statistically significant. It can be stated that the taxpayer's decisions are determined not by the previous but by the current period information – it is likely that decisions are made impulsively and based on the latest information. After all, if decisions were made prudently, it would take time to analyse all the available information and find the most effective and appropriate solutions.

It should be noted that the MIMIC method has not been modelled yet to analyse the factors of the VAT gap in the case of Lithuania. Some of our results are in line with other researchers' findings. F. Schneider and A. Bühn in their study "Shadow Economies and Corruption All Over the World: Revised Estimates for 120 Countries" (2007) have analysed the shadow economy and the VAT gap in three groups of countries: developing countries over 1999 to 2006, Eastern European and Central Asian countries over 1999 to 2006 and High income OECD countries over 1995 to 2006. Lithuania belongs to the second group, and as the MIMIC estimation results show, the three most statistically significant causal variables in this model were business freedom (t statistics was -7.85), fiscal freedom (-3.95) and inflation rate (2.88). As we can see from the results, inflation rate has statistically significant positive effect just as it was estimated in our research. However, this comparison should be evaluated carefully first of all because of the different time periods analysed. F. Schneider and A. Bühn's study covered a period from 1999 to 2006; our research involved a period from 2000 to 2013, so in the first study influence of the 2008 crisis was not accounted for. Moreover, other causal variables were different in these two models, that is why there is no further reason to compare them.

5. Construction and analysis of the short run eMIMIC models

In order to derive a short run eMIMIC model it is necessary to include the long run equilibrium errors (the number of which must be equal to that of indicators) in the long run MIMIC model. When designing the eMIMIC model it was discovered that inclusion of all the variables in the model did not yield any statistically reliable data of the existence of 3 cointegrating vectors. Therefore, after exclusion of one of the indicators (rGDP) two co-integrating vectors were found. In order to get reliable, fair and accurate results, rGDP was not included in the estimation of the short run eMIMIC model.

To determine the number of the co-integrating vectors the Johansen procedure was used. The procedure consists of two tests: the Trace test and the Maximum eigenvalues test. During these tests vector autoregression (VAR (p)) model is composed (Lütkepohl & Krätzig, 2004). Optimal number of lags (the errors are uncorrelated at lag 1 ($p = 1$)) was determined by using the autocorrelation function (ACF). The variable stationarity tests identified that there can be structural breaks in two variables: GFCF and VATR (this structural break is likely to be statistically insignificant), therefore additional Lütkepohl –

Saikkonen – Trenkler test was performed in order to test the co-integration rank. This test implies a possibility of the structural breaks when testing for unit root processes. Due to the large number of variables and a relatively short time series it was not possible to calculate the theoretical values of λ_{trace} . Since GFCF is stationary with a structural break, it was decided to take GFCF as an exogenous variable in these tests, i.e., we searched for a stationary combination of non – stationary variables, considering that there is a stationary variable with a structural break. The numbers of characteristic roots that are statistically different from unity are conducted by using the following test statistic in the Trace test:

$$LR(r) = -T * \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i). \quad (4)$$

where r – is the number of co-integrating vectors, T – the number of usable observations, $\hat{\lambda}_i$ – the eigenvalues of matrix π . The following sequence of null hypothesis is tested during this procedure:

$H_0: r = 0$, there are no co-integrating vectors;

$H_1: r > 0$, there is at least one co-integrating vector.

The null hypothesis is rejected when $LR(r) > theoretical \lambda_{trace}$. Then the following hypothesis is tested:

$H_0: r \leq 1$ – there is one or less than one co-integrating vector;

$H_1: r > 1$ – there is at least one co-integrating vector.

The procedure of the rejection of the null hypothesis is similar to the former case. The test is completed when the null hypothesis is not rejected at 5 percent probability of error. The obtained results of the Trace test are presented in Table 3 (the results of the Maximum eigenvalue test were similar, so they were not included in this paper). Table 3 shows that the null hypothesis, which claims that there are 2 or less co-integrating vectors, cannot be rejected at a 5 percent level of significance because in this case the theoretical test statistics was lower than the estimated one ($51.70 < 53.12$).

TABLE 3. Results of the Trace test

	Calculated values	Critical values			Decision
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	
$r \leq 5$	2.60	7.52	9.24	12.97	-
$r \leq 4$	9.26	17.85	19.96	24.60	-
$r \leq 3$	28.86	32.00	34.91	41.07	-
$r \leq 2$	51.70	49.65	53.12	60.16	51.70 < 53.12, the null hypothesis is not rejected
$r \leq 1$	87.07	71.86	76.07	84.45	87.07 > 76.07, the null hypothesis is rejected
$r = 0$	141.92	97.18	102.14	111.01	141.92 > 102.14, the null hypothesis is rejected

Source: calculated by the authors with the use of R software.

According to the *p-value* of the model compatibility chi – square (χ^2) criteria, the most relevant model is the one which includes all the causal variables; *p-value* of the last model (which was without GFCE, AAW and GGCE) was over 0.05 as well. It means that empirical distribution of all the models is compatible with the theoretical one. Other criteria used to compare the models were RMSEA and AIC. Results of both the criteria suggested that the most suitable is the last model, where RMSEA *p-value* and meaning of AIC were respectively 0.210 and 1568.16 (see Table 4). Based on all these results it could be stated that the last model is the best for further analysis.

TABLE 4. Analysis of the information criteria

	With all variables	Without GFCE	Without GFCE, AAW	Without GFCE, AAW, GGCE
χ^2 (p-value)	0.383	0.278	0.247	0.162
DF	6	5	4	3
RMSEA (p-value)	0.034 (0.470)	0.069 (0.353)	0.081 (0.312)	0.115 (0.210)
AIC	2537.42	2149.68	1910.71	1568.16

Source: calculated by the authors with the use of R software.

As it can be seen from Figure 3, there were no statistically significant causal variables in this evaluated eMIMIC model – *p-value* for both INFL and HD variables exceeds 0.05. However, in the course of this work, instead of GFCE, another variable – index of economic freedom – was included in the model both with HD and without HD (i.e., with the assumption that consumption is directly affected by growing AAW); it was calculated that both INFL and HD were statistically significant. When the model with three causal variables GGCE, INFL and HD was evaluated, no statistically significant variables were identified, *p-value* of the least significant variable (GGCE) was not very high – it did not exceed 0.15. It would be inappropriate to say that none of the used causal variables have no statistically significant effect on the VAT gap in the short run, therefore INFL and HD influence should be evaluated.

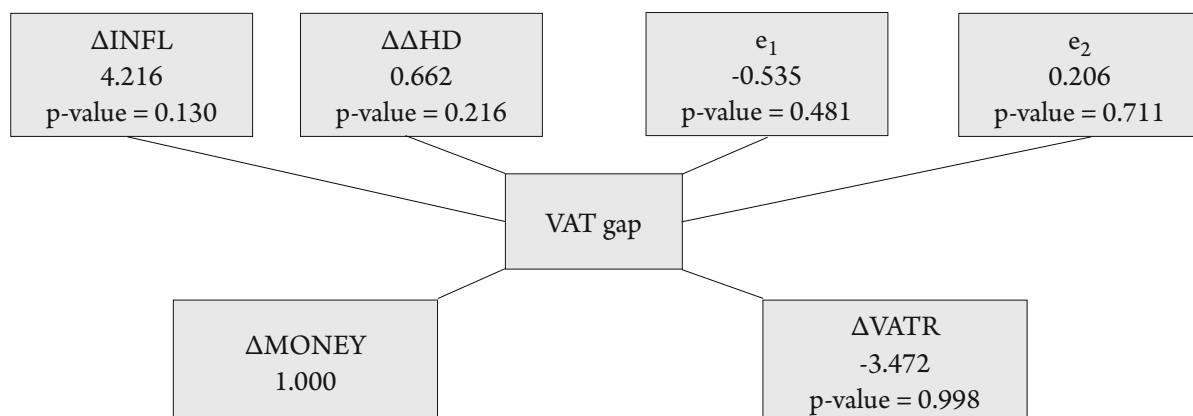


Figure 3. The estimates of the short run MIMIC model parameters

Source: calculated by the authors with the use of R software.

The estimated coefficient of INFL was the same as in the long run MIMIC model. It means that a positive change of inflation increases the VAT gap, *ceteris paribus*. It follows that in order to prevent an increase in the VAT gap, a stable price level should be maintained. As it is mentioned in the strategy of the monetary policy of ECB, price stability is defined as a year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below 2% (ECB, 2015). In order to pursue the stable price level, inflation rate should be maintained below but close to 2 percent over the medium term.

The estimated coefficient of another variable – HD was also positive: growing propensity to save, i.e., the allocation of available resources not to consumption but to deposits may be the reason for the growth of the VAT gap in the short run, *ceteris paribus*. This variable was not statistically significant in the long run, which means that business takes into account the changes in the preferences of household savings in the short run, but adapts to them in the long run and copes with these challenges in other ways than avoiding tax obligations.

Similarly to the long run MIMIC model, there were no statistically significant lagged variables and indicators in the short run eMIMIC model. These results may be explained by the fact that the impacts of the causal variables eliminate each other and the amount of the VAT gap remains constant over the long term, regardless of the outside effects (the variables which were used in the research), whereas taxpayers systematically attempt to avoid their tax obligations. It was found that the VAT gap does not have statistically significant influence on the indicators of the short run eMIMIC model.

6. Limitations and suggestions for further research

Some limitations of our study should be also noted. First of all, there are doubts about the clear theoretical definition of the “VAT gap”: this concept can include some other possible elements such as the level of socio – economic development or social welfare etc. Moreover, all the estimates depend on the limited availability of data and the subjective decisions of the authors. Although the inclusion of all the variables in our models was based on the scientific literature in the field, due to the fact that not all the data was available during all the investigation period (for example, consumer confidence) or some of the data was not available at all (for example, the weighted average VAT rate), it was not possible to estimate the influence of some of the meaningful variables on the VAT gap.

In addition to the above considerations, all the results were obtained making certain assumptions, e.g., that the error terms in the measurement model do not correlate with the latent variable. That specification of the MIMIC model simplifies description of the functioning of the economic system. Finally, the MIMIC model itself has the limitations as there may be other unidentified variables potentially correlated with

the VAT gap. The study intended to determine the significant factors, therefore the coefficients obtained should be understood as approximations rather than exact figures.

Nevertheless, the limitations mentioned above do not affect the general appropriateness of the MIMIC method in estimating determinants of the VAT gap.

In the future studies, in order to better understand the nature and “composition” of the VAT gap, it would be important to deepen the knowledge on a set of other observable macroeconomic factors, such as turnover, labour market activity rate, the weighted average VAT rate, efficiency of the State Tax Inspectorate performance, assurance of measures on executing tax obligations, etc. Moreover, in real life, relationships between economic variables are not exactly linear. Therefore further work is required to analyse non-linear relationships between variables both in the MIMIC and eMIMIC models.

7. Conclusions

The VAT gap in Lithuania being one of the largest in the EU definitely inflicts a significant loss on public finance. Therefore it is very important to understand what determinants limit the country’s capacity to mobilize revenue. The literature review suggests that the MIMIC model is one of the most reliable methods of measuring the VAT gaps.

The results of the estimated long run MIMIC model obtained show that only two out of five causal variables have a statistically significant influence on the VAT gap: the changes in General Government consumption expenditure and the changes in inflation. The growth of changes of the General Government consumption expenditure determines the growth of the VAT gap, *ceteris paribus* (1 percent changes in GGCE increases the VAT gap by 0.156 notional units). Since the core part of these expenditures in absolute terms was allocated to social protection, it may be assumed that growing sponsorship to this sector determines the fact that people getting low income are less interested to work, because the benefit from work is not sufficient enough for them to become active labour market participants. Such people spend less money on consumption, which reduces the opportunity for business development.

The VAT gap growth in the long run can be explained by changes in inflation growth, *ceteris paribus* (1 percent change in INFL increases the VAT gap by 4.713 notional units); i.e., individuals, in order to maintain at least equal level of consumption, are forced to look for additional sources of income by other means, for example, engaging in illegal activity, hiding income and thereby avoiding tax compliance.

Due to the fact that there were no statistically significant indicators in the long run MIMIC model, it can be concluded that the impact of the causal variables eliminates each other, i.e., the amount of the VAT gap is permanent in the long run – regardless of the outside effects, and the taxpayers systematically attempt to avoid their tax obligations.

The results of the short run MIMIC model obtained show that only two out of five causal variables have statistically significant influence on the VAT gap: inflation and

household deposits. Positive changes of the inflation increase the VAT gap, *ceteris paribus*. Thus, in order to maintain at least the same level of the VAT gap, it is important to ensure price stability in the country. It decreases not only the fluctuation of consumer purchasing power, but also the VAT gap. The growth of the changes in household deposits (HD) may be the reason why the VAT gap increases in the short run, *ceteris paribus*. In the long run this causal variable was not statistically significant, so it can be concluded that business takes into account the changes in the preferences of household savings in the short run, but adapts to these changes in the long run.

The evaluation of both MIMIC and eMIMIC models indicated that there are no statistically significant indicators, which means that the VAT avoidance related business decisions cannot be taken spontaneously, i.e., in compliance only with the changes in causal variables used in the model. Decisions to carry out the tax obligations or not may be determined by many microeconomic and macroeconomic factors which were not included in this research or are hard to be measured.

It is important to mention that involvement of the lagged variables in the long run MIMIC model was not expedient, because none of them were statistically significant. Therefore, it is most probable that both in the long and the short run the taxpayers make decisions based on the latest information, rather than depending on the long run trends.

Finally, we believe that the present study delivers reliable estimates and expands previous research on VAT evasion. The results of the models applied can be used not only as a basis for further discussion in this area, but also as one of the potential supporting tools in adjusting Lithuanian tax policy and VAT administration.

References

Allingham, M. G., Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323 – 338. Retrieved from <http://eml.berkeley.edu/~saez/course/Allingham&SandmoJPubE%281972%29.pdf>

Bühn, A., Schneider, F. (2008). MIMIC models, co-integration and error correction: an application to the French shadow economy. *CESIFO Working Paper* No. 2200 Category 1: Public Finance, 28 p. Retrieved from http://www.econ.jku.at/members/Schneider/files/publications/cesifo1_wp2200.pdf

Business accounting standards (2008). 5 BAS “Cash flow statement”. Retrieved from <http://www.aat.lt/index.php?id=85>

Central Bank of the Republic of Lithuania (2014). *Report on financial market state and execution of functions*. Retrieved from https://www.lb.lt/pranesimas_seimui_2014_m_spalis

Christie, E., Holzner, M. (2006). What Explains Tax Evasion? An Empirical Assessment based on European Data. *Working papers*. The Vienna institute for International Economic Studies. Retrieved from <https://ideas.repec.org/p/wii/wpaper/40.html>

CASE (2013). *Study to quantify and analyse the VAT Gap in the EU-27 Member States*. Final Report. Retrieved from http://ec.europa.eu/taxation_customs/resources/documents/common/publications/studies/vat-gap.pdf

Dell'Anno R. (2006). The Shadow Economy in Portugal: an analysis with the MIMIC approach. *Journal of Applied Economics*, Vol X, No. 2 (Nov 2007), 253-277. Retrieved from <http://core.ac.uk/download/pdf/7004117.pdf>

Dell'Anno R., Schneider F. (2005). Estimating the Underground Economy by using MIMIC Models: A Response to T. Breusch's critique. *Working Paper, Department of Economics, Johannes Kepler University of Linz*, No. 0607. Retrieved from <http://www.econstor.eu/bitstream/10419/73283/1/wp0607.pdf>

Enders, W. (1995). *Applied Econometric Time Series*, 1st edition. Iowa State University. John Wiley & Sons, inc. Retrieved from <http://www.allbookez.com/applied-econometric-time-series-walter-enders/>

Ene C. M., Ștefănescu A. (2011). Size and implication of Underground Economy in Romania – a MIMIC approach. *Annales Universitatis Apulensis Series Oeconomica*, 13(1). Retrieved from <http://www.oeconomica.uab.ro/upload/lucrari/1320111/08.pdf>

European Central Bank (2015). *The definition of price stability*. Retrieved from <https://www.ecb.europa.eu/mopo/strategy/pricestab/html/index.en.html>

Eurostat (2014). *Eurostat Statistics Database*. Retrieved from <http://epp.eurostat.ec.europa.eu>

Fathi, B., Esmailian, M (2012). Evaluation of Value Added Tax (VAT) and Tax Evasion. *Current Research Journal of Economic Theory*, 4 (1), 1–5. Retrieved from <http://maxwellsci.com/print/crjet/v4-1-5.pdf>

Gemmell N., Hasseldine J. (2012). The Tax Gap: A Methodological Review. *Working Papers in Public Finance*, September. Retrieved from http://www.victoria.ac.nz/sacl/about/cpf/publications/pdfs/WP09_TaxGap_14092012.pdf

Giles D.E.A., Tedds L. M. (2002). Taxes and the Canadian Underground Economy. *Canadian Tax Paper*, No. 106. Retrieved from <http://www.ctf.ca/ctfweb/Documents/PDF/2005ctj/05ctj2-breusch.pdf>

Hanousek, J., Palda F. (2002). Why People Evade Taxes in the Czech and Slovak Republics: A Tale of Twins. *CERGE-EI Discussion Paper*, Series No. 85 (home.cerge-ei.cz/hanousek/)

Krumplytė, J. (2010). *Šešėlinės ekonomikos masto vertinimas*. The doctoral dissertation, social science, economics (04S) Retrieved from http://vddb.library.lt/obj/LT-eLABa-0001:E.02~2011~D_20110224_160950-30036.

Lietuvos Respublikos Mokesčių Administravimo Įstatymas (2014). Actual from 2014–04-04. *Valstybės žinios*, 2004, Nr. 63 – 2243. Retrieved from http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=468559

Lütkepohl H., Krätzig M. (2004). *Applied time series econometrics*. Cambridge University Press.

Ministry of Finance of the Republic of Lithuania (2013). *Data on the Implementation of National Budget*. Retrieved from

http://www.finmin.lt/web/finmin//auktualus_duomenys/biudzeto_pajamos/nacionalinis_biudzetas

Nerré B. (2008). Tax Culture: A Basic Concept for Tax Politics. *Economic Analysis and Policy*, 38, No. 1, 153–161. Retrieved from http://www.eap-journal.com/archive/v38_i1_11.pdf

Reckon LLP (2009). *Study to quantify and analyse the VAT gap in the EU-25 Member States*. Retrieved from http://ec.europa.eu/taxation_customs/resources/documents/taxation/tax_cooperation/combating_tax_fraud/reckon_report_sep2009.pdf

Schneider F. (2006). Shadow Economies and Corruption All Over the World: What Do We Really Know? *Discussion Paper Series*. IZA DP No. 2315. Retrieved from <http://ftp.iza.org/dp2315.pdf>

Schneider F., Bühn A. (2007). Shadow Economies and Corruption All Over the World: Revised Estimates for 120 Countries. *Economics*, the Open – Access, Open – Assessment E-Journal, Vol. 1, 2007-9, July 24, 2007. Retrieved from <http://www.economics-ejournal.org/economics/journalarticles/2007-9>

Statistics Lithuania (2014). *The Official Statistics Portal*. Retrieved from- <http://osp.stat.gov.lt/temines-lenteles39>

Swedish National Tax Agency (2008). *Tax Gap Map for Sweden. How was it created and how can it be used?* Retrieved from http://www.skatteverket.se/download/18.71004e4c133e23bf6db8000109879/Report_2008_1B.pdf