

# **The Undeclared Value Added Due to the Value Added Tax Evasion of Sole Proprietors**

Ildikó Ritzlné Kazimir

Hungarian Central Statistical Office, Budapest, Hungary

E-mail: [ildiko.ritzln@ksh.hu](mailto:ildiko.ritzln@ksh.hu)

## **Abstract**

In Hungary the estimation of undeclared gross value added due to value added tax (VAT) evasion of sole proprietors has been the part of compilation of annual national accounts since 2011. The calculations are based on a model built on anonym individual tax audits data. However the growing importance of the quarterly national accounts and the need for the improvement of data quality requires detailed quarterly estimation of households sector, as well. The paper examines the possibility of developing the model actually used in annual national accounts, to estimate the quarterly non-observed gross value added. The suggested solution is a logarithmic model that estimates the quarterly undeclared VAT by the available quarterly indicators, and so the non-observed gross value added can be calculated by the ratio of undeclared VAT and average VAT tariffs.

Keywords: estimation, modelling, national accounts, non-observed economy

## **1. Introduction**

In the non-observed economy the sole proprietors' share is significant in Hungary. The non-observed economy involves a number of different activities and so different estimation methods are required due to the different reasons for lack of the adequate data. Eurostat has developed an integrated system of tables to take into account the non-observed economic activities, and the non-observed economic activities have been divided in seven categories. The non-observed gross value added due to the VAT evasion is included in the N6 category of non-observed economy by the Eurostat. (Eurostat (2005))

The Eurostat drafted methodological recommendations for estimation of categories of non-observed economy. Eurostat recommended using the results of tax audits for the estimation of gross value added due to the tax evasion. (Eurostat (2005)) Similarly, the OECD (2002) suggests estimation based on the tax audit data. However, both papers draw the attention to the fact that the tax audit stocks from non-random sampling, so their use requires great care. Many countries have made estimates by the tax audit results to evaluate the degree of tax evasion. For example, in France the coefficients of value added and output are estimated by the use of VAT audits data (OECD (2002)), and the examination of the extent of income tax evasion have been used such samples in the U.S. (Feinstein (1990))

In Hungary a new method was introduced for estimating the annual non-observed gross value added for sole proprietors in 2011. In the new model the non-observed gross value added is calculated according to the Eurostat recommendations. (KSH (2012)) The increasing importance of the quarterly national accounts, and the demand for higher reliability of the quarterly calculations, urged the analysis of the options for the quarterly estimation of the non-observed value added due to the VAT evasion of sole proprietors. The potential extension of the annual calculation method is based on the tax audit results according to the Eurostat and the OECD's recommendations, and the quarterly system can be in harmony with the annual calculations.

## **2. The extended model for the estimation of quarterly gross value added due to VAT evasion**

The model for the estimation of gross value added of sole proprietors that has been introduced to the compilation of the national accounts in 2011, estimates separately the non-observed value added due to the VAT evasion. The calculation of non-observed value added due to the VAT evasion is based on the final, individual results of VAT audits, the data of VAT and income tax returns, and data of the business register of Hungarian Central Statistical Office (HCSO). The calculation procedure consists of three steps.

In the first step the population of VAT evaders is estimated by the “k nearest neighbour method” (kNN method). This calculation is based on the consolidated data of four years prior to the current period. The independent variables of this procedure are coming from the tax returns and the related indicators of the Business Register categories. The second step of the procedure is the estimation of the undeclared VAT. It is estimated by linear regression for the population of VAT evaders utilizing the data of sole proprietors, who have additional VAT payment obligation from the previous third-year. Finally, the non-observed gross value added is being calculated by using the average VAT tariffs. (Giczi et al. (2013)) The calculations are repeated annually using the new tax audits data.

The annual calculation model is extended to allow us the estimation of undeclared VAT and the non-observed gross value added due to the VAT evasion on quarterly basis. The available annual data is much more extensive than the quarterly data. So, this paper examines the possibility of an additional model which is set up on quarterly basis under the annual calculation.

This additional model focuses only on the estimation of the level of the undeclared tax, because tax evasion behaviour is considered stable within a longer period. The linear function in the annual calculation cannot be used in the quarterly estimation, because it does not allow the calculation of the changes between the quarters, even if quarterly indicators are being used.

The indicators for estimation the quarterly undeclared VAT are calculated from the quarterly VAT returns. The dependent variable of the extended model is the relative undeclared VAT, equal to the ratio of undeclared VAT and the difference of sales and purchases. This dependent variable eliminates the effects of the business environment changes. There is another advantage of using ratios as dependent variables. The system in this case takes into account that the entrepreneurs evaluate the chance of being detected in tax evasion before sending back the official tax returns.

The quarterly estimation model is built on indicators, which describe the relative position of the enterprise; these indicators are suitable to characterize the business environment, size, efficiency and market environment of the enterprises.

The extended model for estimation of quarterly undeclared VAT uses logarithmic regression instead of linear regression, because the logarithmic function results better fit model, and interactions are allowed between indicators of tax evasion.

However logarithmic model is not additive; so the sum of the quarterly estimates will not be equal to the annual estimates. The partial elasticities of continuous variables are constant and equal to the coefficients in this model, and the coefficients influence the percentage of change, for example the development of undeclared VAT. So the development of the undeclared VAT and thus the not-observed value added is determined by the percentage change of continuous variables and the coefficients of the model. It means that the extended model can be used only to determine the current price trends of undeclared VAT. Therefore it is necessary to adjust the quarterly undeclared VAT and value added data to the annual figures.

The parameters of extended logarithmic model are estimated by the final results of VAT audits of sole proprietors in year 2008.

For the quarterly estimation the results of tax audits, the VAT indicators on quarterly basis and the indicators of the Business Register of HCSO are combined. In the next step from this combined indicator the indicators for relative size, efficiency, and relative market position are being calculated. These new calculated indicators are the independent variables of the logarithmic model.

Four model types have been identified by the explanatory variables. The first model estimates the parameters for all sectors collectively. The other models estimate the undeclared VAT for industry groups separately. Agriculture, construction and retail trade industries are treated separately due to their characteristics, and due to the widespread tax evasion. The question is how to aggregate the rest of the industries.

In the second model different equations are being used for the industry, the services and the agriculture, the construction and the retail trade. In the third model the services is divided into two categories. The first category includes enterprises, which operate in trade, accommodation and food services, transportation and warehousing, information, communication, arts, entertainment and recreation services and other services. The second category includes activities of finance, insurance, real estate, professional, scientific, technical, administrative activities, education and human health services.

The fourth model was based on cluster analysis in which the industries were grouped by the number of tax audits and the ratio of additional tax liabilities. Thus, agriculture and transportation, construction, retail trade and all other industries have been grouped separately.

The third model was selected from our four models because this model showed the lowest root mean square error (RMSE). The parameters estimated by the third model and the values of RMSE for each industry groups are shown in the Appendix.

The explanatory variables are not independent; therefore the partial interpretation of the parameters is impossible.

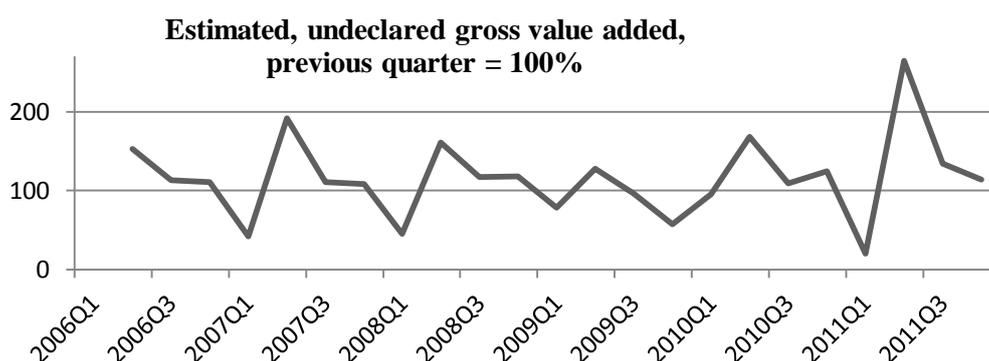
The explanatory variables can be divided into four groups. The first group includes variables from the economic sociology that characterise the operating environment of the sole proprietors. Many of these parameters are significant for the relative level of undeclared VAT in each industry group except the skilled labour-intensive service industries (second group of services). The indicators of relative size and efficiency of sole proprietors explain the level of relative undeclared VAT. The indicator, which characterized the market environment by the relative efficiency of the enterprise, is only significant in agriculture and retail trade.

The quarterly relative undeclared VAT is calculated by the parameters based on the data of year 2008. The level of undeclared VAT utilizes the quarterly data of VAT returns. The quarterly undeclared VAT is adjusted to the annual calculations to get the undeclared VAT. It is taken into account in the calculations, that the estimation by logarithmic regression is biased; the result has to be corrected by the value of mean square error.

At the quarterly calculation the undeclared value added due to the VAT evasion can be estimated by the average VAT ratios, and the price indices allow the calculations of volume indices of undeclared value added, as well.

The following illustration shows the quarterly trends of the undeclared gross value added compared to the previous quarter.

Figure 1



The strengthening seasonality of data arises due to the seasonality of agriculture, where in recent years the VAT evasion level has increased significantly.

### 3. Summary

The paper describes an option how to extend the model of annual calculation of the undeclared gross value added due to the VAT evasion and how to estimate the quarterly gross value added due to the VAT evasion. Fewer indicators are available for the quarterly calculation, therefore the extended model, which is based on quarterly VAT data uses a logarithmic regression. The relative size of VAT evasion is explained by indicators calculated from the VAT returns, and are categorized by the efficiency of the enterprise, its relative size and relative market position, and category indicators of specific conditions of sole proprietors' operation.

The selection from the examined models was based on the root mean squared error. That model was the best where the services had been divided into two categories by the skilled labour intensity.

It is necessary to adjust the output of the extended model to the results of annual calculation. Then, using the annual average tax rates the quarterly time series of the non-observed gross value added can be calculated, as well.

### References

1. Eurostat (2005) "Eurostat's Tabular Approach to Exhaustiveness Guidelines," <http://www.unescap.org/stat/isie/reference-materials/National-Accounts/Eurostat-Guidelines-Tabular-Approach.pdf>
2. Feinstein, J. S. (1990) "Detection Controlled Estimation," *Journal of Law and Economics* 33(4), 233-276.
3. Giczi, J., Horváth, G. and Ritzlné Kazimir, I. (2013) "Value Added Tax Evasion of Sole Proprietors between 2006 and 2008 in Hungary" forthcoming
4. HCSO (Hungarian Central Statistical Office) (2012) "Estimation of Gross Value Added Generated by Sole Proprietors – Development of Methodology" United Nations, Economic Commission for Europe, Conference of European Statisticians, Genf. 30.04. - 04.05. 2012 [http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2012/06\\_Hungary\\_Sole\\_Proprietors.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2012/06_Hungary_Sole_Proprietors.pdf)
5. OECD (2002) "Measuring the Non-Observed Economy, A Handbook" <http://www.oecd.org/std/na/1963116.pdf>

**Appendix table**  
**Estimated parameters of regressions**

<b>Independent Variables<sup>1</sup></b>	<b>Agriculture, forestry and fishing</b>	<b>Construction</b>	<b>Retail trade</b>	<b>Services 1<sup>2</sup></b>	<b>Services 2<sup>3</sup></b>	<b>Industry</b>
Constant	7,32 (0,53)	6,02 (0,31)	0,79 (0,57)	4,05 (0,40)	5,89 (0,32)	2,16 (1,10)
Non-registered production or tips in households sector (dummy, from National accounts calculations) <sup>4</sup>	-4,30 (0,56)					
The age of enterprise			0,52 (0,17)			1,08 (0,41)
Industry with large tax evasion (dummy)				-0,81 (0,47)		
Region: Budapest and Pest County and Transdanubia				0,62 (0,31)		
Ratio of sales and mean of sole proprietors by activities	-2,03 (0,74)	-4,25 (1,18)	-0,76 (0,11)	-1,26 (0,30)		
Ratio of payable VAT and mean of sole proprietors by activities	1,25 (0,73)	2,48 (1,11)			-0,60 (0,16)	-0,50 (0,19)
Ratio of input VAT and mean of sole proprietors by activities		1,29 (0,43)		0,44 (0,27)		
First quartile includes the sole proprietor based on sales by activities (dummy)				-2,70 (1,20)		
Second quartile includes the sole proprietor based on sales by activities (dummy)				-2,26 (0,87)		
Third quartile includes the sole proprietor based on sales by activities (dummy)				-1,18 (0,61)		
First quartile includes the sole proprietors on the basis of the purchases by activities (dummy)				2,07 (1,17)		
Second quartile includes the sole proprietors on the basis of the purchases by activities (dummy)				2,93 (0,85)		
Third quartile includes the sole proprietors on the basis of the purchases by				1,63 (0,61)		

<sup>1</sup> The number in parentheses is the standard error.

<sup>2</sup> The Services1 includes the trade, accommodation and food services, transportation and warehousing, information, communication, arts, entertainment and recreation services and other services.

<sup>3</sup> The Services2 includes activities of finance, insurance, real estate, professional, scientific, technical, administrative activities, education and human health services.

<sup>4</sup> The non-registered output of households includes the production in fields of agriculture, hotels and restaurants, transportation and other personal services.

activities (dummy)						
Ratios of purchases and sales		-10,02 (1,81)	-14,41 (3,01)	-5,25 (0,39)		-5,44 (0,50)
Ratios of payable and input VAT		-3,81 (1,69)			4,27 (0,26)	
Ratio of differences between the sales and purchases and purchases	-3,83 (0,31)	-2,55 (0,16)	-2,91 (0,10)	-2,97 (0,11)	-4,64 (0,44)	-2,56 (0,21)
Ratio of differences between the payable and input VAT and the input VAT	0,59 (0,30)				1,87 (0,43)	
Quotient of ratios of purchases and sales for individuals and mean of sole proprietors by activities	-4,64 (0,22)		8,84 (3,01)			
<b>Adjusted R Square</b>	<b>0,787</b>	<b>0,635</b>	<b>0,797</b>	<b>0,708</b>	<b>0,703</b>	<b>0,678</b>
<b>RMSE</b>	<b>2,666</b>	<b>3,190</b>	<b>2,558</b>	<b>3,011</b>	<b>2,950</b>	<b>2,794</b>
<b>N</b>	<b>199</b>	<b>224</b>	<b>282</b>	<b>395</b>	<b>151</b>	<b>86</b>