



Economic and Social Council

Distr.: General
24 February 2012

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on National Accounts

Eleventh session

Geneva, 30 April-4 May 2012

Item 5 of the provisional agenda

Household sector and unincorporated enterprises

Estimation of Gross Value Added Generated by Sole Proprietors – Development of Methodology

Note by the Hungarian Central Statistical Office

Summary

The estimation of gross value added (GVA) generated by sole proprietors in Hungary needed improvement, as the industrial structure and the ratios of GVA and output did not reflect the real structure, which became apparent in Supply and Use Tables. In addition, the previous methodology did not allow the recording of categories of non-observed economy (NOE). The paper presents the new methodology based on individual-level data on sole proprietors for the period 2006-2010. The results are back-casted for the period 1995-2005.

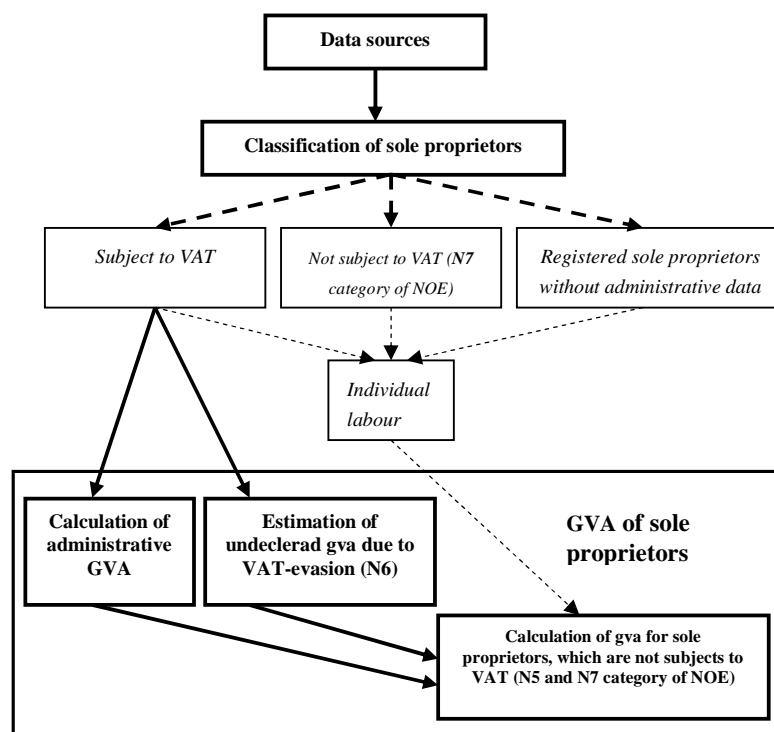
I. Introduction

1. The estimation of gross value added (GVA) generated by sole proprietors needed to improve from several reasons. The previous method assumed that the expected gross value added per labour input of sole proprietors equals to the ratio of gross value added and labour input of a reference group. The result of these calculations did not meet the requirements from some reasons. The industrial structure and the ratios of GVA and output – which are estimated by using of data of reference group – do not fit to the real structure, as became clear during the analysis in the framework of Supply and Use Tables.
2. In addition, the previous method can not record explicitly the figures of non-observed economy categories (Tabular Approach, Eurostat Guideline (2005)) and consequently the Process Table.
3. Besides, the applicability of tax audits for the estimation of non-observed economy has to be examined, according to the Commission decision (94/168/EC). In the first stage of development the applicability of tax audits' results was examined between the other, several methodological opportunities. (Murai et al (2011)). Because of the heterogeneity of sole proprietors by economic activity and size of enterprise we decided to use the results of tax audits.
4. The presented model use individual data of sole proprietors. The estimation is available for the period 2006-2010. The result of methodological change is back-casted for the period 1995-2005.

II. The Model

5. For the estimation of GVA produced by sole proprietors a new model was set up which enabled us to express the gross value added, output and intermediate consumption of sole proprietors in an exhaustive form. The model compiles the total gross value added from administrative data sources and estimates for non-observed economy (N6, N5, N7). In a first step, GVA declared form administrative sources and the undeclared gross value added for sole proprietors subject to value added tax (VAT) (N6 category of NOE) are estimated. These results allow in a next step the calculations for gross value added for those sole proprietors who do not have appropriate data sources (N5 and N7 category of NOE).
6. The model is summarized in the following figure.

Figure 1
The model



A. Data sources

7. The method demonstrates the performance calculated from administrative data sources and describes the non-observed economy's performance by NOE categories.

8. The sources of the estimation are: the Business Register; the database of tax returns of personal income tax, value added tax and simplified entrepreneurial tax.

9. In 2010 the National Tax and Customs Administration made available the final, anonymous, individual results of value added tax audits of sole proprietors for the period 2006-2008. This database contains the monitored years and the results of the audits, if the audited sole proprietor should pay additional tax, or has tax rebate or there is no problem with the tax return.

Table 1
Results of tax audits of sole proprietors (2006-2008)

<i>Industry</i>	<i>Number of audited sole proprietors</i>	<i>Number of statements with undeclared VAT</i>	<i>Number of statements with undeclared income tax</i>	<i>Total number of statements with undeclared tax</i>	<i>Sum of undeclared income tax, million HUF</i>	<i>Sum of undeclared VAT, million HUF</i>
Agr.	877	195	40	214	177	366
Manufac.	366	109	35	126	38	164
Constr.	748	333	81	370	256	1,301
Services	2,983	878	332	1,041	334	2,484
Total	4,974	1,515	488	1,751	805	4,315

Source: National Tax and Customs Administration (NAV), Control Dataset

10. In the first step the results of the tax audit were merged at individual level with databases, and were used as anonymously. From this time on those units who were obligated to pay undeclared tax are deemed to be 'tax evader'.

11. The next step was to analyze the tax audit data's utilization. The database contains only the legally binding statements; therefore most of the data are available for 2006. Unfortunately the number of observations is still not enough compared to the number and heterogeneous activities of the sole proprietors.

12. The number of the simplified entrepreneurial tax audit proved not to be enough, therefore no attempt was made to estimate the number of the simplified entrepreneurial tax evaders.

13. The analysis of the personal income tax audit revealed that the personal income tax evader behaviour can not be explained with the available variables. This problem arose partly because of the low number of tax audits results and partly because of the fact that using the available variables there were no success in finding the sensitive indicators of personal income tax evasion.

14. In the case of VAT evasion a larger database was used, therefore the estimation of the non-observed activities due to VAT evasion was set as the aim of the methodological development. Calculations were made by SPSS software package.

B. Classification of sole proprietors

15. In the first step of development we examined the available data of sole proprietors. The active sole proprietors were identified by the Business Register. Administrative data were not available for the whole population. The figures for this part of population were taken into account in the N5 category of non-observed economy (NOE), because this category includes the figures of active enterprises without administrative data.

16. The other part of population has administrative data in different level of quality and completeness. The VAT taxpayer sole proprietors provide the most complete data sets, and their personal income tax return includes data of cost composition. So we calculate GVA figures for this part of population and the hidden figures due to the tax evasion (N6 category of NOE) are estimated for this part of population, as well.

17. The last part of population includes the sole proprietors, who are not VAT taxpayers, but have any administrative data, like personal income tax or simplified entrepreneurial tax data (N7 category of NOE).

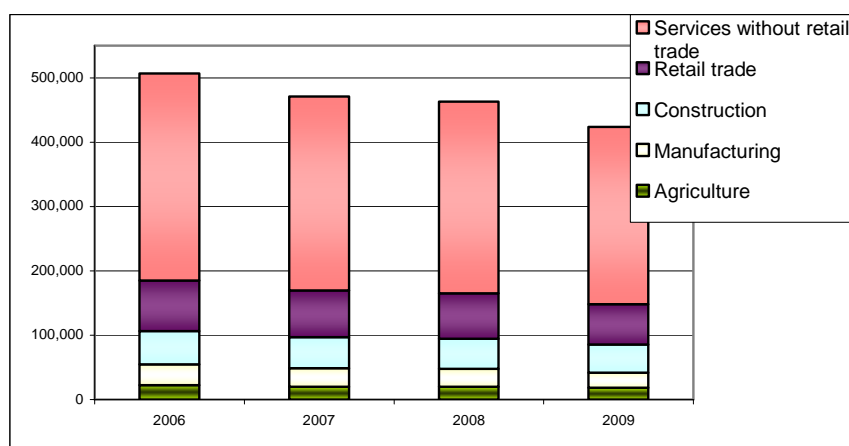
Table 2

Number of sole proprietors by categories of estimation

Categories		2006	2007	2008	2009
VAT payer		111,370	109,322	115,525	105,335
N7	Not complete data	245,516	240,250	293,421	279,577
N5	Registered entrepreneurs not included in statistics	150,070	121,311	53,991	39,055
Total		506,956	470,883	462,937	423,967

18. The shares of sole proprietors in the industries are not similar, because the economic activity can influence the categories of taxation, as well. The next figure shows the number of sole proprietors by economic activities.

Figure 2

Number of sole proprietors by industries

Source: HCSO, database of tax returns

19. The set of VAT taxpayer sole proprietors is called for 'vat', the set of sole proprietors with non-complete data is 'n7'. The last group of sole proprietors is called for 'n5' in the following description.

20. The estimation uses the NACE Rev.1.1 for the years 2008-2006, because of the estimation uses mainly the data of 2006, and for this year the NACE Rev.2 is not available. The result of 2006 and 2007 are reversed by the shares of output and intermediate consumption of year 2008 to NACE Rev.2.

C. Calculation of administrative GVA

21. In this part of the description calculation of administrative output, intermediate consumption and gross value added is explained.

22. The administrative output and intermediate consumption are determined at individual level, for VAT taxpayer sole proprietors. In general, the output is considered to be equal to the turnover of sole proprietors. The intermediate consumption is assumed to be equal to the cost of materials, goods and wrappings (this is a common category of income tax return). The structure of income tax return of sole proprietors does not allow more detailed calculation. The calculation of output and intermediate consumption differ from the above method in the case of trade, where the output equals to the trade margin.

Table 3

Administrative GVA of sole proprietors by industry

<i>Industry</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>
Agriculture	61,818	67,386	78,103	81,966
Manufacturing	72,907	78,430	81,443	67,601
Construction	77,767	84,201	86,496	75,506
Retail trade	75,953	78,294	81,009	62,747
Services without retail trade	249,311	266,030	276,006	241,691
Total	537,756	574,341	603,057	529,512

Source: Own calculation.

23. The personal income tax return of sole proprietor includes data of sold goods for retail, but these data proved to be highly unreliable during the analyses.

24. So the trade margin and ratios of GVA and output are calculated by using of the non-financial corporation sectors data in the case of trade.

25. The next formulas show the calculation of output and intermediate consumption.

$$\text{output}_{-a_{ij}} = \begin{cases} R_{ij} \cdot m_j & \text{if } 500 \leq j \leq 529 \\ R_{ij} & \text{else} \end{cases} \quad i \in \text{vat} \quad (1)$$

26. The $\text{output}_{-a_{ij}}$ means the individual output of sole proprietor i in industry j (NACE Rev.1.1), which is determined by the administrative data. The R_{ij} is turnover of the sole proprietor i , in industry j , the turnover is the data of personal income tax return. The ratio m_j signs the average margins of non-financial corporations, which is determined at three digit level of NACE Rev.1.1. The margins are calculated by the corporate income tax returns of corporations with a turnover less than five billion HUF and the ratio of trade margin less than 0.6.

$$\text{ic}_{-a_{ij}} = \begin{cases} \text{output}_{-a_{ij}} \cdot \left(1 - \left(\frac{\text{gva}}{\text{output}} \right)_j \right) & \text{if } 500 \leq j \leq 529 \\ C_{ij} & \text{else} \end{cases} \quad i \in \text{vat} \quad (2)$$

27. The ic_{aij} is the administrative intermediate consumption of sole proprietor i , in industry j (NACE Rev.1.1).

28. The C_{ij} means the costs of materials, goods and wrappings of sole proprietor i in industry j . This data is included in the income tax returns of sole proprietors. The previous

$$\left(\frac{gva}{output} \right)_j$$

formula uses also the ratio $\left(\frac{gva}{output} \right)_j$, which is the average ratio of GVA and output for non-financial corporation in the industries of trade. For the industry Trade these ratios are also determined for three digit level of NACE Rev.1.1, and the calculation of these ratios use also the data of corporations with less than five billion HUF turnover and the ratio of margin is less than 0.6.

29. This selection of corporations is needed because the composition of turnover by activities is typically available only for large corporations and the data of smaller corporations describe probably better the typical conditions of sole proprietors.

30. We had chosen the average margins and ratios of GVA and outputs of non-financial corporations at three digit level of NACE, because the number of selected corporation does not describe the distribution of sole proprietors at four digit level of NACE. The averages at two digit level of NACE would have blurred the differences of margins and ratios of GVA and output between separate trade activities.

31. The administrative GVA is calculated by the difference between the administrative output and administrative intermediate consumption.¹

D. Estimation of undeclared GVA due to VAT evasion

32. The estimation of undeclared GVA due to VAT-evasion includes three steps.² The first step is the estimation of population of VAT-evaders among VAT taxpayers. Second, we estimate the undeclared VAT. At last the undeclared GVA is determined by using the estimated, undeclared VAT. The additional, undeclared output and the needed correction of intermediate consumption due to over-reporting of costs are calculated by the estimated undeclared GVA.

1. Estimation of VAT evaders

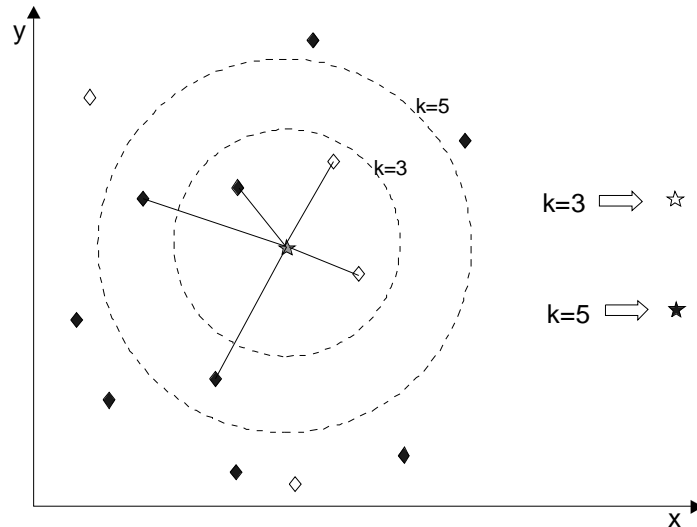
33. The estimation of population of VAT evaders is difficult, because the available data of tax audits can not be considered as random sample. The estimation uses the kNN method (k nearest neighbour algorithm), because this method does not depend on the distribution of the sample, it does not require random sample.

34. The kNN method calculates the distances between the individuals by using preselected independent variables. The median or mean of nearest individuals determine the value of the estimated, dependent variable for the individuals. The calculation is made with SPSS software, which allows the number of neighbours between 3-5 individuals.

¹ The administrative GVA is assumed as a non-negative figure, so the intermediate consumption in formula (2) is corrected by this condition, except for industry Trade.

² Giczi et al (2012).

Figure 3

Illustration of the kNN method

35. The kNN method requires relatively large sample, therefore the estimation uses the whole database of tax audits for 2006-2008.

36. The independent variables of kNN method are chosen by logistic regression, because this method allows narrowing the vast number of administrative and calculated variables, in order to get the most precise results.

37. The independent variables, the values of parameters for logistic regression and the information about the calculation of independent variables are included in the appendix. The changes of economic situation necessitate the analysis of the structure of independent variables, which can change the structure of independent variables and level of parameters.

Table 4

Estimated number of VAT evaders

<i>Industry</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>
Agriculture	1,861	1,731	1,735	687
Manufacturing	2,723	2,292	2,202	2,974
Construction	5,205	4,713	4,805	6,686
Services	20,811	18,274	18,057	19,606
Total	30,600	27,010	26,799	29,953

Source: Own calculation.

2. Estimation of undeclared VAT

38. In the second step of the estimation we calculated the undeclared VAT with linear regression. The parameters of linear regression are estimated by the database of VAT evaders of year 2006. The set of independent variables includes the administrative and calculated variables, which are joined to the database of tax audits. The tax-evaders

behaviour and the level of undeclared tax depends on the economic activity, so the linear regression is not fitted the whole population, but the groups of industries, separately.

39. The groups of industries are determined by cluster analysis on the rate of tax evaders and audited sole proprietors, and the number of audited sole proprietors by industries. We assumed that the tax office defines the frequencies of audited sole proprietors by practical experiences and models.

40. So we created four groups for industries, which are the retail trade; construction; agriculture and transportation; and the other industries. The different structure of independent variables fits to the different groups of industries. The undeclared VAT is calculated by the estimated parameters of the regressions.

41. The parameters and independent variables for linear regression are included in the appendix. The changes of economic situation can result in changes in structures of independent variables, and levels of parameters, as well. So we will examine in this respect the estimation in the future.

3. Calculation of undeclared GVA

42. In the third step the undeclared GVA due to the VAT-avoidance is calculated at individual level, by using the estimated, undeclared VAT. We assumed, that the VAT-avoidance results undeclared GVA, and the VAT evaders are interested in distorting in the items of their VAT return where they can reduce tax.

43. So the average ratios of VAT are calculated by the corresponding items of VAT returns of sole proprietors for all industries. Due to the changes of VAT rates the average ratios of VAT have to be determined for each year separately.

$$gva_n6_{im} = \frac{a_vat_{im}}{\hat{t}_m^{vat}} \quad i \in vat \quad (3)$$

44. The individual GVA of N6 category of NOE for sole proprietor i in industry m (gva_n6_{im}) equals to the quotient of the estimated, undeclared VAT of sole proprietor i (a_vat_{im}) and the average ratio of VAT (\hat{t}_m^{vat}) in the previous formula.

45. The estimated VAT evasion affects both the output and intermediate consumption of sole proprietors, because the entrepreneurs can combine the undeclared sales and the over-reporting of costs. The ratio of the undeclared output and intermediate consumption can not be determined explicitly, because the information for this calculation is not available.

46. We assumed that the ratios of undeclared output and GVA equal to the ratio of actual consumption of households sector and domestic output in all industries. The idea behind this assumption is that those sole proprietors are likely to sell without invoice who sell for the households sector.

47. The ratios of actual consumption of households are determined in the data of symmetric input-output tables (industry by industry) of 2005. The ratios equal to the quotients of actual consumption of households and domestic output for all industries, the results are used for the calculation of individual, undeclared output:

$$output_n6_{im} = m_m^{ch} \cdot gva_n6_{im} \quad (4)$$

48. The $output_n6_{im}$ in the above formula equals to the undeclared output of sole proprietor i in industry m , which arises due to the VAT evasion. The m_m^{ch} is the ratio of

actual consumption of households to domestic output in industry m. The maximum of m_m^{ch} is 100%, and the maximum of the undeclared output is the undeclared GVA. The rest of hidden GVA is explained by the over reporting of costs, so the estimated hidden intermediate consumption can be calculated by the next form:

$$ic_n6_{im} = output_n6_{im} - gva_n6_{im} \quad (5)$$

49. Because of the negative value of ic_n6_{im} the total intermediate consumption can be negative as well, when the absolute value of ic_n6_{im} exceeds the value of ic_a_{ij} (IC calculated from administrative sources). We solved this problem by assuming that the source of hidden GVA is only the hidden output in these cases. The estimated, undeclared GVA is shown in the next table.

Table 5

Estimated GVA hidden due to VAT evasion

<i>Industry</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>
Agriculture	7,865	11,256	12,439	5,046
Manufacturing	18,321	16,804	16,368	9,061
Construction	29,947	25,425	17,386	17,022
Services	97,038	93,331	92,768	120,538
Total	153,171	146,816	138,960	151,666

Source: Own calculation.

E. Calculation for enterprises which are not subject to VAT (N5 and N7 categories of NOE)

50. The non-VAT subject sole proprietors are classified into two different groups (N7 and N5 categories of NOE), but the estimation use the same assumptions. We assumed that the sole proprietors in the same industry produce the expected gross value added and labour input ratio, independent of categories (VAT taxpayer, N7 or N5 NOE categories) of sole proprietors. Labour input equals to the sole proprietors and their employees. (The employment mode of sole proprietors (full or part time) is taken into account in the labour input of sole proprietors.)

51. Therefore we calculated the ratios of GVA and labour input for VAT taxpayer sole proprietors at two digit level of NACE, the reference group for calculation is this set of sole proprietors. The GVA of VAT taxpayers equals to the sum of administrative data and the estimated, undeclared GVA.

$$gva_L_m = \frac{\sum_i (gva_a_{im} + gva_n6_{im})}{\sum_i L_{im}} \quad \forall m \quad (6)$$

52. In the previous formula the L_{im} means labour input of the VAT taxpayer sole proprietor i, who is active in the industry m (NACE Rev.1.1). The gva_a_{im} shows the administrative GVA of VAT taxpayer sole proprietors i in industry m. The gva_n6_{im} is

the estimated, undeclared GVA (due to tax evasion) of VAT taxpayer sole proprietor i in industry m .

53. The GVA of categories N7 and N5 is resulted as the multiplication of ratios from the previous formula and the individual labour input of sole proprietors in categories N7 and N5.

54. The structural analysis in the framework of supply and use tables revealed that some industries need other methods of estimation, because the conditions of production are significantly different in the different groups of sole proprietors.

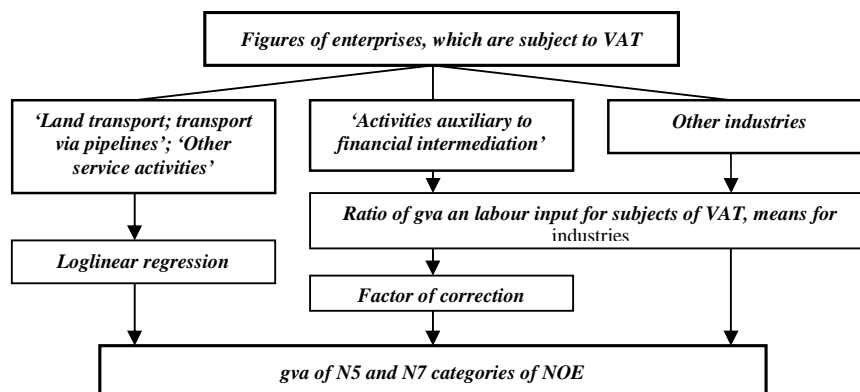
55. The model assumes that the technology of production is the same in almost all industries, except for the industries 'Land transport; transport via pipelines'; 'Other service activities' and the 'Activities auxiliary to financial intermediation'.

56. In the case of 'Transportation and 'Other services' two types of sole proprietors can be identified. The smallest sole proprietors produce without significant capital investment, because they work for larger enterprises, using their material fixed assets. The larger ones are the owner of the assets so their revenues have to cover not only for their compensation of employees but also for the consumption of fixed capital and management expenses. Our investigation showed that the larger enterprises (e.g. owner of a truck) tend to choose the VAT taxation form, while the smaller ones (e.g. truck drivers) belong to the N5-N7 categories.

57. In the 'Financial intermediation' the sole proprietors work for financial or insurance institutions as their agents, and work on commission. The system is similar to the MLM (Multi-Level Marketing) scheme. Sole proprietors at higher level of hierarchy on the multi-level scheme receive commission for sole proprietors belonging to their group. The following figure summarizes the calculations for N5 and N7 categories of NOE.

Figure 4

Scheme for estimation of GVA for N5 and N7 categories of NOE



58. In the cases of 'Land transport; transport via pipelines'; 'Other service activities' – in order to eliminate the differences between the two types of entrepreneurship described above – we use logarithmic regression for estimating the GVA of N5 and N7, which is estimated by the independent variable of natural logarithm of turnover. In the case of 'Activities auxiliary to financial intermediation' we calculate a factor of correction:

$$\gamma = \frac{\left(\frac{\sum_i R_{im}}{\sum_i L_{im}} \right)}{\left(\frac{\sum_j R_{jm}}{\sum_j L_{jm}} \right)} \quad m = 67 \quad i \in n7 \quad j \in \text{vat} \quad (7)$$

59. The γ is the factor of correction. The R_{im} and L_{im} are the turnover and labour input of sole proprietor i in N7 category of NOE. The R_{jm} and L_{jm} mean the turnover and labour input of VAT taxpayer sole proprietor i . The γ is determined only for the industry "Activities auxiliary to financial intermediation". The next formulas summarize the calculation of GVA for N5 and N7 categories of NOE.

$$gva_n7_{im} = \begin{cases} \gamma \cdot gva_L_m \cdot L_{im} & m = 67 \\ \exp(\ln_gva_{im}) & m = 60;93 \quad i \in n7 \\ gva_L_m \cdot L_{im} & else \end{cases} \quad (8)$$

$$gva_n5_{im} = \begin{cases} \exp(\ln_gva_{im}) & m = 60;93 \\ gva_L_m \cdot L_{im} & else \end{cases} \quad i \in n5 \quad (9)$$

60. The calculation of output for N5 and N7 categories of NOE based on the same method, which uses the next ratios of GVA and output of reference group:

$$r_gva_m = \frac{\sum_i (gva_a_{im} + gva_n6_{im})}{\sum_i (output_a_{im} + output_n6_{im})} \quad \forall m \quad (10)$$

61. The output of N7 and N5 categories is determined by the estimated GVA and the average ratios of GVA and output from the previous formula (r_gva_m).

$$output_n7_{im} = \frac{gva_n7_{im}}{r_gva_m} \quad i \in n7 \quad (11)$$

$$output_n5_{im} = \frac{gva_n5_{im}}{r_gva_m} \quad i \in n5 \quad (12)$$

62. The $output_n7_{im}$ is the estimated output of sole proprietors i of category N7, the $output_n5_{im}$ means the estimated output of sole proprietors i of category N5. The intermediate consumption is the difference between the output and GVA for N5 and N7 categories of NOE.

63. The next table includes the GVA of N5 and N7 categories of NOE.

Table 6

Estimated GVA of N7 and N5 categories of NOE (current prices, millions HUF)

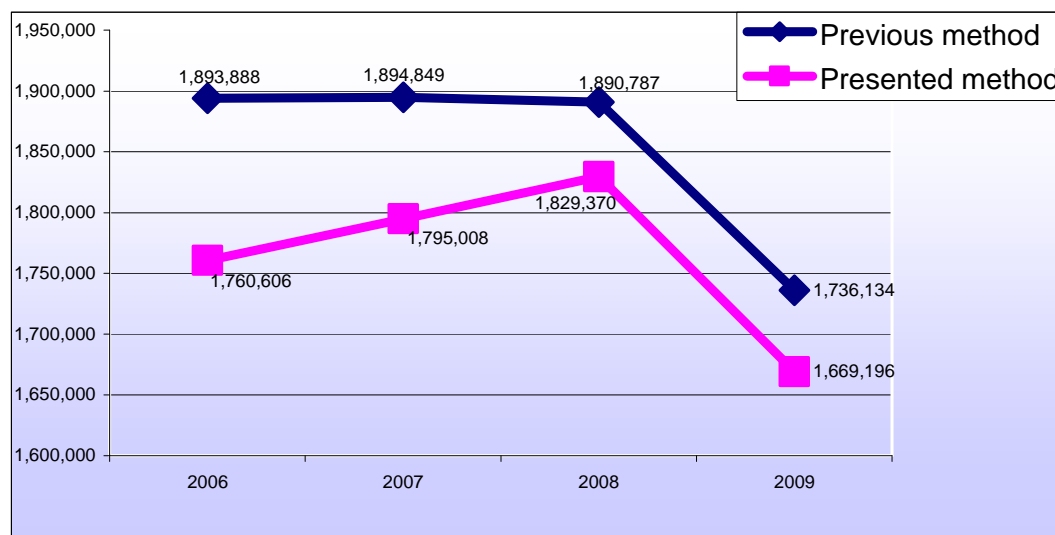
Ind.	2006		2007		2008		2009	
	GVA_N7	GVA_N5	GVA_N7	GVA_N5	GVA_N7	GVA_N5	GVA_N7	GVA_N5
Agr.	14,805	14,124	18,268	14,232	27,290	8,851	22,442	5,761
Man.	52,527	22,498	55,313	20,495	63,957	10,401	46,438	6,312
Const.	94,214	41,984	98,116	37,569	111,927	18,279	101,577	14,514
Serv.	562,634	245,984	588,901	238,572	733,398	113,252	698,301	92,675
Total	724,180	324,591	760,598	310,868	936,573	150,783	868,759	119,262

Source: Own calculation.

64. The GVA, output and intermediate consumption for the sole proprietors are the sum of the results of the different subsectors. The GVA of sole proprietors by the new model and the previous method is available for the period 2006-2009.

Figure 5

The sum of GVA by the previous and presented methods (current prices, million HUF)



65. The above graph shows, that the results of presented method are more sensitive and more volatile than the old method, due to the explicit estimation of categories of non-observed economy.

III. Back-casting of the results

66. Because of the lack of basic data the calculation could be carried out for the years 2006 and later. The results led to a sectoral structure and GVA/output ratio that is significantly different from the results of the previous calculation.

67. For this reason it was necessary to do the retrospective calculation until 1995. For the period between 2005 and 2003 the calculation was made at four digit level of NACE, between 2002 and 1995 at two digit level of NACE.

68. The retrospective calculation was made by using several methods depending on available data. Not all of the individual data could be used for the whole period, because the VAT database was not available for the years between 1995 and 2003, and the personal income tax database was not available for the years between 1995 and 2001.

69. The results of the methods were validated in the SUT framework. The method was chosen which gave the smallest difference from the expenditure side. Therefore the value indices of previous method were used for back casting.

IV. Follow-up work

70. The generated model achieved the objectives of the methodological developments. The model's continuous operation is ensured by the agreement with the National Tax and Customs Administration. According to this arrangement, from 2011 the individual results of the value added tax audits will be delivered to the Hungarian Central Statistical Office on a regular basis.

71. The tax audit's database – that is yearly renewed – ensures the estimation of the non-observed economy's level and structure being as up-to-date as possible. Based on these data the continuous analyses of tax evader's number and the non-observed activities will be possible, which can provide essential information on the comparison of these data to other macroeconomic figures.

72. The yearly renewed database may require the re-estimation of the model's parameters, because the model is only up-to-date when the estimation of tax-evasion follows the phenomena happening in the non-observed economy. This can cause the involvement of new indicators or the change of the weight of the old ones.

73. All in all, the calculations can be made in a short period of time; if the procedure is appropriately automated, and the experiences are utilized and documented.

V. References

Giczi J. – Horváth G. – Ritzlné Kazimir I. (2012): Value Added Tax Evasion of Sole Proprietors; in press

Eurostat's Tabular Approach to Exhaustiveness Guidelines, <http://www.unescap.org/stat/isie/reference-materials/National-Accounts/Eurostat-Guidelines-Tabular-Approach.pdf>

Murai B. – Ritzlné Kazimir I. (2011): A nem megfigyelt gazdaság mérésének lehetőségei
(Measuring the Non-Observed Economy), Statisztikai szemle, 89. (5). pp 501-522.

Annex 1

Estimated parameters of regressions

Categories of variables	Independent variables	Logistic regression	Parameters of linear regression ³			
			Agriculture and Transportation	Manufacturing and Services	Construction	Retail Trade
	Constant	-33,71	0	0	0	0
Economic Sociology	Year of establishment	0,02	0	0	0	0
	Sole proprietor in full time (dummy, from Business Register of HCSO) ⁴	-0,25	0	0	0	-116,81
	Region: West and Central Hungary in contrast to Nord Hungary	-0,3909	0	0	0	0
	Region: South Hungary in contrast to Nord Hungary	-0,3923	0	0	0	0
Market Characteristics	Non-registered production ⁵ or tips in households sector (dummy, from National accounts calculations)	0,38	0	82,38	0	0
	Natural logarithm of output per number of enterprises for sole proprietors and corporations by industries	-0,19	0	0	0	0
	Relative estimated output of 'missing' type of corporations from national accounts calculations in 2006, by industries	0,06	0	0	0	0
	Natural logarithm of ratio of intermediate consumption and output of corporations by industries	-0,84	0	0	0	0
	Natural logarithm of GVA per employees of corporations by industries	-0,80	0	0	0	0
Size of Enterprise	Purchases million HUF ⁶ (from VAT return)	0	0	0	152,45	0
	Turnover million HUF (from income tax return)	0	80,06	0	0	0
	Turnover under 20 million HUF (dummy, from income tax return)	0	0	0	0	271,53
	Other costs, million HUF (from income tax return)	0	-201,72	0	140,24	0
	Sales, million HUF ⁷ (from VAT return)	0	37,05	0	0	-834,14
	Total cost, million HUF, (from income tax return)	0	0	-286,75	0	0
	Basis of payable VAT, million HUF (from VAT return)	0	0	0	-57,71	0
	Basis of payable VAT, thousand HUF (from VAT return)	0	0	0	0	0,84
	Sales with VAT rate 20%, million HUF (from VAT return)	0	0	-40,81	0	0
Natural logarithm of other costs in million HUF	-0,04	0	0	0	0	

³ The parameters in the table show the results (in thousand HUF) of unit change in the dependent variable (all other things being equal, excluding the constant).

⁴ The sole proprietor is active in full time, if she or he works more than 36 hours/week for the enterprise.

⁵ The non-registered output of households includes the production in fields of agriculture, hotels and restaurants, transportation and other personal services.

⁶ The purchases equal to the sum of import and domestic purchases from the VAT return.

⁷ The sales are the sum of export and domestic sales from the VAT return.

	Natural logarithm of turnover in million HUF	0,13	0	0	0	0
	Natural logarithm of wages in HUF	-0,19	0	0	0	0
Figures of effectiveness	Ratio of individual GVA ⁸ per labour input ⁹ and mean of corporations, by activities	0	0	46,51	0	0,89
	Ratio of materials and other costs	0	0	0	-689,61	0
	Ratio of GVA and total cost	0	0	-7,34	0	0
	Ratio of GVA and other costs	0	0	745,03	0	33,04
	Ratio of GVA and labour input, million HUF	0	293,8	0	0	0
	Ratio of wages and entrepreneurial withdrawals, and labour input HUF/person	0	0	0,044	0	0
	Ratio of wages and employees million HUF/person	0	0	849,77	-4464,22	0
					20836,2	
	Ratio of wages and total cost	0	0	14905,61	9	0
	Natural logarithm of quotient of ratios of purchases and sales for individuals and mean of corporations by activities	-0,21	0	0	0	0
	Natural logarithm of ratio of intermediate consumption and output for individuals and mean of corporations by industries	-0,0019	0	0	0	0
	Natural logarithm of ratios of GVA per labour input for individuals and mean of corporations by activities	-0,03	0	0	0	0
	R square ¹⁰	0,05	0,78	0,93	0,66	0,67

⁸ The gross value added of sole proprietors is estimated by the difference between the turnover and the costs of materials, goods and wrappings.

⁹ The labour input equals to the sole proprietor and the number of his/her employees from personal income tax.

¹⁰ The table includes the adjusted R square for the linear regressions and the Cox and Snell R square for the logistic regression.