Estimation of gross value added generated by sole proprietors

...How to measure the unmeasurable?

Source of picture: http://www.tiszagazdasagert.uni-corvinus.hu/index.php/F%C3%A1jl:Civertan_rejtett_gazd_merete.jpg
Role of sole proprietors in the Hungarian economy

- Their number is ~ 424 thousand (~1 per 25 inhabitants) in 2009, of which:
  - VAT taxpayer: 105 thousand
  - Not VAT taxpayer, personal income tax payer: 280 thousand
  - Registered but do not declare any tax: 39 thousand

- Share in total GVA: ~ 8%
- Share in employment: ~ 10%
- Average nr. of employees: 0.3
The old method

- Used since the mid 90’s
- It was based on estimate of expected minimum income per labour by activity and region, for tax audit purposes
- Non of the data declared by the sole proprietors was considered reliable (except the number of persons)
- Structural problems (SUT)
- No explicit estimates for the elements of non observed economy (N1-N7)
- Presentation issue (EU GNI audits – Process Table)
Sub-sector of sole proprietors

Problems:
- Heterogeneous population
- Not enough reliable information
- Significant proportion of hidden economy
- Entrepreneurial and household costs are often mixed

Requirements:
- Estimation should fit into Process table framework
- Results should fit into the framework of Supply and Use Tables (SUT)
- Detailed results according to non-observed economy categories (N1-N7)

Using the results of tax audits seemed to be an appropriate solution from the recommended and examined methodological options.
The model

Input
- Personal income tax
- Simplified entrepreneurial tax
- Value added tax (VAT)
- Business register
- VAT audits of sole proprietors
- Corporate profit tax returns

Process

Classification of sole proprietors

- Subject to VAT
- Not subject to VAT
- Registered sole proprietors without administrative data

Individual labour input

Calculation of administrative GVA

Performance of N6

Performance of N5

Performance of N7

Final result

GVA of sole proprietors
Categories of non observed economy estimated within this method

- N6: Misreporting by the producer (= VAT tax payers with undeclared VAT)
- N7: Statistical deficiencies in the data (= Not VAT tax payers, but declaring personal income tax or simplified entrepreneurial income tax)
- N5: Registered entrepreneur is not included in statistics (= registered but no declaration provided)
Steps of the estimation

I. Estimation of undeclared GVA due to VAT evasion
   1. Estimation of the number of VAT evaders among VAT taxpayers
   2. Estimation of the value of undeclared VAT
   3. Estimation of undeclared GVA, output and IC (N6)

II. Estimation of GVA of units not subject to VAT (N7)

III. Estimation of GVA of registered units without administrative data (N5)
I. GVA of VAT taxpayer sole proprietors on the basis of administrative data

- Assumption: sole proprietors carry out activity only according to their NACE classification
- Calculations are based on VAT returns of entrepreneurs and corporate profit tax returns of other companies
- Output is considered to be equal to
  + turnover of sole proprietors,
  - Goods for resale in the case of trade (average rate from profit tax returns of smaller companies)
- Intermediate consumption
  - is assumed to be equal to material costs
  - ratios of IC is used in the case of trade (average rate from profit tax returns of smaller companies)
I. Administrative GVA of sole proprietors by industries

![Graph showing GVA by industries and years]
I. Undeclared GVA due to VAT evasion

1. Estimation of VAT evaders

Results of tax audits of sole proprietors, 2006-2008

- Number of audited sole proprietors
- Number of statements with undeclared VAT
- Number of statements with undeclared income tax
- Total number of statements with undeclared tax

Services
Construction
Industry
Agriculture
First try: Logistic regression

- Models the relative ratio of occurrence probability of two mutually exclusive categories based on the independent variable values
- Distorts the estimation, because the sample (audited units) is not randomly selected by the tax office → Not appropriate for defining VAT evaders.
- …But, still good for defining the mix of variables as inputs for the final method for defining VAT evaders.
I. Undeclared GVA due to VAT evasion

1. Estimation of VAT evaders

**Independent variables of logistic regression**

- Economic sociology
- Market characteristics
  - Year of establishment
  - Region
    - full time / part time
- Size of enterprise
- Figures of effectiveness
  - Other costs
  - Turnover
  - Wages

**Non-registered production in the households sector**

**Natural logarithm of output per number of enterprises for sole proprietors and corporations by industries**

**Relative output of „missing“ corporations**

**Natural logarithm of ratio of intermediate consumption and output of corporations by industries**

**Natural logarithm of GVA per employees of corporations by industries**

**Year of establishment**

**Region**

**full time / part time**
I. Undeclared GVA due to VAT evasion
   1. Estimation of VAT evaders

Method for defining VAT evaders:

kNN or k-Nearest Neighbour algorithm
The kNN method

- Built-in procedure of SPSS
- Method of classifying
  - classification based on \( k \) similar case
  - Similarity: based on distance
  - Distance measure:
    - City block
      \[
      CityBlock_{ih} = \sum_{p=1}^{P} w_{(p)} |x_{(p)i} - x_{(p)h}|
      \]
- Distribution free
- Nominal variables encoded as \( c \) (0,1) type
- Continuous variables
  - Without transformation
  - Built-in transformation: into (-1,1) interval
I. Undeclared GVA due to VAT evasion

1. Estimation of VAT evaders - kNN

- **Value assign**
  - Mean of neighbours (median)
  - Categorization type: that category the most neighbour falls
I. Undeclared GVA due to VAT evasion

1. Estimation of VAT evaders - kNN

- kNN method with 3 neighbours was chosen
- Calculation was made at individual level
- Dimensions of the space are the variables that were chosen by logistic regression
## I. Undeclared GVA due to VAT evasion

### 1. Estimation of VAT evaders

<table>
<thead>
<tr>
<th>Industries</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>1,861</td>
<td>1,731</td>
<td>1,735</td>
<td>0,687</td>
</tr>
<tr>
<td>Industry</td>
<td>2,723</td>
<td>2,292</td>
<td>2,202</td>
<td>2,974</td>
</tr>
<tr>
<td>Construction</td>
<td>5,205</td>
<td>4,713</td>
<td>4,805</td>
<td>6,686</td>
</tr>
<tr>
<td>Services</td>
<td>20,811</td>
<td>18,274</td>
<td>18,057</td>
<td>19,606</td>
</tr>
<tr>
<td>VAT evaders, total</td>
<td>30,600</td>
<td>27,010</td>
<td>26,799</td>
<td>29,953</td>
</tr>
<tr>
<td>Share %</td>
<td>23,8</td>
<td>23,3</td>
<td>23,2</td>
<td>28,4</td>
</tr>
</tbody>
</table>
## I. Undeclared GVA due to VAT evasion

### 2. Estimation of undeclared VAT

<table>
<thead>
<tr>
<th>Classification of independent variables – <em>linear regression</em></th>
<th>Economic sociocology</th>
<th>Market characteristics</th>
<th>Size of enterprise</th>
<th>Figures of effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agriculture and transport</strong></td>
<td>Turnover</td>
<td>Other costs</td>
<td>Sales</td>
<td>Ratio of GVA per labour input</td>
</tr>
<tr>
<td><strong>Construction</strong></td>
<td>Purchases</td>
<td>Other costs</td>
<td>Basis of payable VAT</td>
<td>Ratio of materials and other costs</td>
</tr>
<tr>
<td><strong>Retail trade</strong></td>
<td>Turnover under 20 million HUF</td>
<td>Sales</td>
<td>Basis of payable VAT</td>
<td>Ratio of GVA and other costs</td>
</tr>
<tr>
<td>Sole proprietor in full time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Industry and services</strong></td>
<td>Total costs</td>
<td>Non-registered production in households sector</td>
<td>Sales with VAT rate 20%</td>
<td>Ratio of wages per employee</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ratio of wages and total cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Natural logarithm of ratios of GVA per labour input for individuals and mean of corporations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ratio of GVA and other costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ratio of GVA and total cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ratio of wages and entrepreneurial withdrawals, and labour input HUF/person</td>
</tr>
</tbody>
</table>
I. Undeclared GVA due to VAT evasion

2. Estimation of undeclared VAT
I. Undeclared GVA due to VAT evasion

3. Estimation of undeclared GVA – N6

- VAT is a value added type tax ➔ Undeclared VAT also means undeclared GVA
- GVA:

\[ GVA_{ij} = \frac{T_{ij}^{VAT}}{\bar{t}_j} \]

\( T_{ij}^{VAT} \): estimated, undeclared VAT of sole proprietor i in industry j
\( \bar{t}_j \): means of VAT rates for industry j
I. Undeclared GVA due to VAT evasion
3. Estimation of undeclared GVA – N6

- VAT evasion affects both the output and intermediate consumption
- sales for households offers more opportunity for VAT evasion ➔ data of symmetric input-output tables (year 2005)

\[ \text{Output}_{ij} = \text{IOT \_ cons \_ ratio} \cdot \text{GVA}_{ij} \]
\[ \text{IC}_{ij} = -(1 - \text{IOT \_ cons \_ ratio}) \cdot \text{GVA}_{ij} \]
I. Undeclared GVA due to VAT evasion

3. Estimation of undeclared GVA – N6

Undeclared and administrative GVA of sole proprietors, million HUF

- Undeclared GVA (N6)
- Administrative GVA

<table>
<thead>
<tr>
<th>Year</th>
<th>Undeclared GVA</th>
<th>Administrative GVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
II.+III. Sole proprietors which are not subject to VAT – N5, N7

- Personal income tax declarations are considered to be unreliable
- Simplified entrepreneurial tax declaration does not cover costs
- The reference group for N5 and N7 estimates of sole proprietors not subject of VAT are the VAT taxpayer sole proprietors
II.+III. Sole proprietors which are not subject to VAT – N5, N7

- Assumptions:
  - Productivity for groups of sole proprietors are similar: administrative + estimated GVA per labour input of VAT taxpayers equal to that of non VAT taxpayers (at 2 digit level of NACE)
  - Employment characteristics is similar for all entrepreneurs, except:
    - Transport
    - Other service activities
    - Auxiliary financial intermediation
II+III. Sole proprietors which are not subject to VAT – N5, N7

- Other services and transport
- Financial intermediation
- All other industries

Sole proprietors which are subject to VAT

- Ratio of GVA and labour input for subjects of VAT, means for industries
- Correction factor: ratio of turnover difference by industries based on tax returns

GVA of N5 and N7 category

IC ratio

Output and IC
II.+III. Sole proprietors which are not subject to VAT – N5, N7

- Calculation was made at individual level
- Transport and other services: logarithmic regression by the independent variable of natural logarithm of turnover.
- Auxiliary financial intermediation: expected GVA per labour input + correction factor based on SUT
- All other industries: expected GVA per labour input
- All GVA estimates are broken down to output and IC based an output/IC ratios of the reference group (VAT taxpayers)
II.+III. Sole proprietors which are not subject to VAT – N5, N7

GVA of non-VAT subject sole proprietors, million HUF

- Services
- Construction
- Industry
- Agriculture

<table>
<thead>
<tr>
<th>Year</th>
<th>gva_N7</th>
<th>gva_N5</th>
<th>gva_N7</th>
<th>gva_N5</th>
<th>gva_N7</th>
<th>gva_N5</th>
<th>gva_N7</th>
<th>gva_N5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
II.+III. Sole proprietors which are not subject to VAT – N7

- Estimate for N7 covers the total GVA of sole proprietors not subject of VAT.
- In fact, a part of their GVA was declared, but these amounts are not possible to calculate from their administrative records (incomplete data). → Not the total value of N7 is due to underreporting.
Summary of results

GVA according to the previous and new methods

- Services
- Construction
- Industry
- Agriculture

<table>
<thead>
<tr>
<th>Year</th>
<th>Old</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Summary of results, 2009

<table>
<thead>
<tr>
<th>Sole proprietors GVA</th>
<th>Million HUF</th>
<th>GVA total = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative GVA</td>
<td>529 512</td>
<td>2,4</td>
</tr>
<tr>
<td>N5</td>
<td>119 262</td>
<td>0,6</td>
</tr>
<tr>
<td>N6</td>
<td>151 666</td>
<td>0,7</td>
</tr>
<tr>
<td>N7</td>
<td>868 756</td>
<td>4,0</td>
</tr>
<tr>
<td><strong>NOE total (N5+N6+N7)</strong></td>
<td><strong>1 139 684</strong></td>
<td><strong>5,3</strong></td>
</tr>
<tr>
<td>Sole proprietors, total</td>
<td>1 669 196</td>
<td>7,7</td>
</tr>
<tr>
<td>GVA, total</td>
<td>25 622 866</td>
<td>100,0</td>
</tr>
</tbody>
</table>
Summary

- The „new” method is more flexible and capable of tracking changes
- Possibility for further developments (extension to corporate sector?)
- Fits well into SUT framework
- With the „new” method, the level of performance decreased slightly, while the structure more significantly
  - IC ratios
  - Structure and ratio of services
Thank you for your attention!

peter.szabo@ksh.hu