**Review of quality of the Hungarian VAT and STS data**

Zsolt Kovari, Hungarian Central Statistical Office, zsolt.kovari@ksh.hu

Abstract

Over the past decades in the data production process of the official statistics the use of administrative data plays increasingly important role. The European statistical law (Regulation (EC) no 223/2009 of the European parliament and of the council of 11 march 2009 on European statistics) says in the Article 24 “In order to reduce the burden on respondents, the NSIs and other national authorities and the Commission (Eurostat) shall have access to administrative data sources…”. In the 3th section of the ESS Vision 2020 the new data sources are mentioned as a key area. The following goal is declared in this document: “We base our statistical products and services on both traditional surveys and newer sources, including administrative data, geospatial and, where possible, big data.” In the area of business statistics the tax data (mostly VAT and corporate tax data) can be used especially effectively for various estimations, in validation processes or even to select data providers. Due to the growing interest in the service sector the stakeholders and users require monthly service turnover or price index as well as index of service volume. These data have to be produced by the lowest possible increase in data supplier’s burden and their quality must be analysed. This paper describes the efforts of the Hungarian Statistical Office tackled with the fulfilment of the new data requirements and introduces problems and solutions affected the quality.

**Keywords:** Administrative data, VAT, Short term statistics

# 1. Introduction

According to the current STS regulation, both service turnover and the service price index are to be produced quarterly. Many countries, however, report monthly turnover or price indices, and even monthly volume indices. If these are needed and significantly increase the reporting burden should be avoided, it is recommended that you use monthly administrative data. VAT data in many industries are capable of estimating turnover or its growth rate with their help. However, whatever the mathematical model we use, the quality of the data received depends heavily on the extent to which the VAT data used cover the population that supplies the STS revenue and how much the turnover from the VAT data is correspond to the STS turnover. In order to assess these two conditions, the VAT legislation must first be reviewed.

# 2. The outline of the VAT legislation

According to the Hungarian legislation the taxable persons are obliged to file a tax return in every quarter. However the enterprises can submit tax return in every month or just once a year under certain conditions. The most important conditions are related to the sum of payable VAT aggregated on annual basis from the beginning of the year. If it shows positive result and reach 1 million HUF[[1]](#footnote-1) (≈ 3,150 €) the taxable person concerned must move to monthly returns from quarterly. Other important rules:

* The enterprises that choose collective value added taxation (i.e. these form a VAT group) must submit a tax return in every month.
* The enterprises which were founded without legal predecessor are also obliged to file a tax return in every month in the first two year of its operation.

There are some other conditions which can force the change of the frequency of the taxation but they affect just few companies or apply only specific group of enterprises.

The taxable person has to submit an annual return if the sum of payable VAT aggregated on annual basis to be accounted for in the second year preceding the current year has not reached 250 thousand HUF (≈ 790 €) and the sum aggregated on an annual basis of the countervalue of his sales of products or provisions of services without tax does not exceed 50 million HUF (≈ 158,000 EUR), provided that he does not have a Community VAT number.

The deadlines for submitting tax returns are as follows:

* For taxable persons subject to monthly VAT payment, the return shall be filed by the 20th day of the month following the current month.
* For taxable persons subject to quarterly VAT payment, the return shall be filed by the 20th day of the month following the current quarter
* For taxable persons subject to annual VAT payment, the return shall be filed by 25 February of the year following the taxation year.

The Hungarian Central Statistical Office (hereinafter HCSO) receives VAT data from the National Tax and Customs Administration (hereinafter NTCA) in every month 42 days after the end of the reference period. At this moment the data of previous periods are overwritten with the updated dataset. The taxable persons can submit their corrections retroactively for five years.

# 3. Short review of the VAT database

The VAT legislation changes relatively frequently. The database itself contains more than 100 tax elements and some data that can be used to describe the reporting company. Each reporting provider has its own unique identifier, given the NACE classification of the administrative principal activity, the frequency of the data supply, an entry code that can be used to determine which one is the most recent tax return for a given period and some other information about the time of receipt of the tax return and the reference period.

If we want to evaluate the most important quality indicators based on the complete database, we receive a fairly good picture.

## 3.1 Relevance

According to the definition the relevance is “the degree to which statistical information meet current and potential needs of the users.” In this case the scope of the users is quite limited. In some respects the only real user is the NTCA, for the other authorities the VAT data are simple administrative data which will be used for some specific statistical purpose. From this point of view the most important characteristic of the administrative data is that whether it is usable for the required statistical purpose. The VAT data can be considered relevant if it allows a good estimate of the correspondent STS turnover.

The figures of data suppliers that are both in VAT and STS database were used for the compilation of this table, and although comparing the reported and estimated turnover from enterprise to enterprise sometimes quite big differences can be found, the relevance of VAT database seems to be good.

## 3.2 Accuracy

Since the tax authorities have a right to punish for the incomplete or inaccurate data provision the VAT data can be considered quite accurate. However, data can be changed by self-revisions. For handling this problem the database has to contain information about self-revision, but the necessary column has become the part of the database since 2017. The enterprises can submit self-revision five years retrospectively, which means a given record can change for five years. In case of previous years some plausible assumption can be used. For instance: the self-revision is an extra return usually submitted at the end of the year, often with negative value.

The impact of revision should be characterised calculating the well-known indicators (MAR, RMAR, MIR, etc.) and revision triangle can be built at least theoretically. Up to now the database was overwritten each month, but the internal discussion has begun at the beginning of 2018 in order to the data to be archived for at least 5 years.

## 3.3 Timeliness

As it was mentioned above the VAT database is handed over to HCSO after 42 days of the end of the reference period. The IT experts need one or two days to fulfil the checking processes and to carry out some additional calculations. (E.g.: the NTCA uses 9 digit identifier, HCSO uses 8 digit one which can be calculated from the NTCA code.) Hence the VAT data are available usually t+45 days.

## 3.4 Coherence and comparability

The assessment of comparability over time is made difficult by the changing of the legal framework. According to our plans the time series of monthly turnover begin from 2009 but since then time we had to use seven different formulas to estimate the turnover. The last formula has been used since 2015 in unaltered form though the content of some variables have been changed. For example the range of goods or services that are subject to different VAT rates can change year to year.

As for the internal coherence the following situations were checked:

* The average increase rate
* The strange (suspicious) figures (zero or negative VAT, or empty value)
* Duplicate data
* Outliers

The average increase rate of the turnover was -20 and 40% in case of 92% of the enterprises.

Duplications were not detected.

The simple methods for outlier detection revealed five enterprises from the monthly tax payers and eight enterprises from the quarterly tax payers, but these are the biggest firms in the H49 division indeed, so the reported figures can be acceptable.

# 4. Current state of play in service turnover

The service turnover data come to the statistical office using a classical data collection method. The target population consists of enterprises included in Divisions 45 and 46 as well as Sections H to J, L to N and P to S of NACE Rev. 2. according to the STS regulation. The number of possible data providers is approximately 50,000. The Business Register has online connection with the Registry Court and the data are updated every night. The frozen state of the Business Register is used as a sampling frame. All enterprises are observed if the number of persons employed is greater than 49. For enterprises employing between 5 and 49 persons the data collection is based on random stratified representative sample survey. When strata was defined the beyond the NACE division, the number of persons employed and the location of the headquarters are taken into account. There are some strata which are based on the NACE group or class.

Data collection is based on an online survey. The questionnaire contains some elementary checking rules (e.g. the turnover is greater than 0, the tax is smaller than the taxable income etc.)

## 4.1 Relevance

The concepts and methods are based on European legislation. The main international user is Eurostat. The principal domestic users are the National Central Bank of Hungary and Ministries. The relevant internal users are the National Accounts department and the Sectoral statistics department. Data about direct user satisfaction is not available, but the overall assessment of STS fulfilled by Eurostat is good.

## 4.2 Accuracy

Sampling error is calculated for the strata and for some aggregates each quarter. The relative error which is interpreted as the division of standard error and value of total was between 5.08 and 49.09 for the sample in the 4th quarter of 2016. The highest value is at H51 (Air transport). There are relatively few data providers with significant standard deviation for the turnover, which is a characteristic of the market. This is an extreme figure, the second highest value is about 25. The lowest value was measured at G46 (Wholesale trade except of motor vehicles and motorcycles).

Over-coverage was 0.16%; 0.38%; 0.43% and 0.38% in each quarter respectively. Under-coverage occurs when an enterprise that is member of the full-scope part starts to work after the update of BR. In this case this data supplier goes into the data set next quarter.

The unit non-response rate was between 3.42 (Accommodation and food service) and 10.72% (Arts, entertainment and recreation). The main reason for the higher value in arts, entertainment and recreation NACE section is that the rate of small enterprises is significant, almost 50%. The yearly average is 7.15% which is a good value. The item non-response rate was under 0.5% for the net turnover each quarter.

## 4.3 Timeliness

Data of service turnover are published in line with the STS Regulation. The online questionnaire is available for the data provider at the beginning of the quarter following the reference quarter. The deadline for the data transmission is t+20. The database is closed after the deadline, but the data sent late will be taken into account at the next period. Data are published in the dissemination database t+55.

## 4.4 Coherence and comparability

Quarterly service turnover data are available from 2008. The earliest quarterly SPPIs data are available those of 2007. Most of the data about annual performance of enterprises are accessible from 2008. Transport and tourism time series are available between 2007 and 2012. There is not known internal inconsistency.

# 5. Problems and planned actions

It is an interesting question whether the aggregated VAT turnover and grossed up STS turnover should be compared, or the turnover of individual firms listed in both databases. The problem is the enterprises in the VAT database are classified according to their main activity which was declared when they were founded. In the BR each firm is classified by its principal activity which contributes most to its total value added. These two classification are not necessary the same.

The members of VAT group do not necessary appear in the VAT database because the leader of the VAT group pays all the VAT of the transactions carried out among each other. However the members of the group can report STS turnover to the statistical office, but it may cause distortion.

Due to different reasons, certain enterprises may be exempt from tax payment, so they do not appear in the VAT database, but they still have to send data on their turnover.

It is difficult to aggregate VAT data because there may be duplicates, and there are enterprises that change the frequency of tax reporting (e.g. from quarterly to monthly).

HCSO has established a “Large Cases Unit” section under the Business Statistics Department which has two main tasks:

* observation of the effects of globalisation,
* processing and evaluation of the VAT and Corporate tax data to ensure uniform approach.

The following activities are in process:

* Recommendations for better form of the VAT database (e. g. participation in the discussion of cooperation agreement between HCSO and NTCA)
* Collection and dissemination of information among the other departments which work with VAT data (e. g. creation of own share point site on the Intranet)
* The elaboration of an effective IT system in order to visualise, find and handle to suspicious data record (concerned by the above mentioned problems)
* The examination of the inconsistent data from an accounting point of view
* Pre-processing of the data for various use (estimation of monthly turnover, examination of coherence etc.)

# 6. References

Rudi Seljak (2010), Some aspects of quality assessment of usage of administrative data in official statistics, Statistical Office of the Republic of Slovenia.

ESS (2015), Quality assurance framework of the European Statistical System

Pieter Vlag (2011), .Use of administrative and accounts data in Business Statistics WP4

1. 1 € = 317.19 HUF on 16 May 2018. [↑](#footnote-ref-1)