ESSnet

Use of Administrative and Accounts Data in Business Statistics

WP4

Timeliness of Administrative Sources for Monthly and Quarterly Estimates


The Use of Admin Data for Monthly and Quarterly Estimates:

Common Issues and Challenges in Estonia, Finland, Germany, Italy, Lithuania, The Netherlands and the United Kingdom

Pieter Vlag*, Shirley Ortega Azurduy, Epp Karus

* p.vlag@cbs.nl

February, 28th 2011
1. Introduction

Several National Statistical Offices within the European Statistical System (ESS) use Administrative Data (Admin Data) as information source for monthly and quarterly turnover and employment estimates.

The use of the Value Added Tax (VAT)-registers to estimate the turnover for Short-term Statistics (STS), is allowed when two conditions are met.

a. The definition of the so-called (VAT-)turnover corresponds with the definition of the turnover in the STS-regulation in such extent that turnover growth rates can be derived from VAT-registers when this administrative data source is complete.

b. Sufficient VAT-data are available in the VAT-registers when the STS-estimates for a certain period have to be made.

Although some exceptions do exist, for example in branches with intermediate services (travel agencies) and trade in second-hand goods, it can be stated that VAT-registers of Finland, Germany and the Netherlands fulfil the first condition for a number of economical sector covered by the STS- and SBS (Structural Business Statistics)-regulation. This means that VAT-data can be used to estimate turnover growth rates when both turnover definitions agree and the VAT-register is complete. This issue won’t be discussed further in this paper.

This document deals with the challenges to meet the second condition. In many European countries, when the estimates have to be made the completeness of the VAT-register gets more complicated. Two factors determine the complexity of the problem:

1. The VAT reporting period might differ from the time period for which the statistical estimates have to be made. In practice, this means that for the shortest statistical reporting period (= month) the VAT-data are by definition incomplete because enterprises may report their VAT for a longer period like quarter and year. Formally, this is a periodicity issue.

2. The VAT-data might be reported later than the STS-estimates have to be made. Formally, this is a timeliness issue.

In practice, most European countries deal with both challenges. Therefore, both issues are included within workpackage (WP) 4 of the ESSnet Admin Data. Finland, Germany, Lithuania, the Netherlands and the United Kingdom compared their practices and challenges in the use of incomplete VAT-data for monthly and quarterly turnover estimates as part of WP4. The results are described in previous deliverables of this project (Sirviö, 2011a,b; Lorenz, 2011, Kavaliauskiene, 2011; Vlag et al., 2011; Orchard et al., 2011). In this document, the main similarities, differences and challenges that were encountered across WP 4 countries’ current practices and R&D studies (and observed during a study visit at Statistics Sweden) are summarised. Moreover, this summary is used to define the WP4’s future work in the next two years.

Notice that current practices on the use of (incomplete) Social Security Administrative data to produce quarterly estimates on Employment for STS-statistics was thoroughly described and discussed by Italy (Baldi et al., 2011) and Lithuania (Šličkutė-Šeštokienė, 2011). From this work, it was concluded that ‘periodicity’ and ‘timeliness’ issues observed in Social Security Admin Data are quite similar to those found in VAT-registers. Consequently, the conclusions concerning the use of incomplete VAT-data for statistical purposes are also applicable for incomplete Social Security data, unless otherwise stated.
2. General conclusions about the use of (incomplete) VAT-data for short-term estimates

The availability of Admin Data for the monthly and quarterly estimates are summarised in an overview table, which is published as deliverable 4.8 for SGA-2010 (Ortega Azurduy, 2011). Despite the differences in the availability of Admin Data, six general conclusions can be drawn from the country reports, namely:

a. All countries use a mixed-mode model to estimate turnover, implying that
   - only survey-data are used for the large(st) enterprises.
   - only VAT-data or VAT-data in combination with survey-data are used for the small and medium size (SME) enterprises.

Note that the survey’s coverage differs across countries.

b. Incompleteness of VAT-data differs per period (Q/M), estimate (1st, 2nd of definitive) and country. This is caused by two factors:
   - depending on the size of the enterprises, VAT might be declared on a monthly, quarterly or yearly base in all participating countries. However, the thresholds for monthly, quarterly and yearly declarations differ per country (Ortega Azurduy, 2011).
     The most important exceptions are:
     - monthly VAT-reporting is on a voluntary base only in the Netherlands.
     - quarterly VAT-reporters are allowed to declare in three ‘staggers’ in the United Kingdom.
     - the deadline for reporting VAT to the tax office differs per country.

Although the survey’s coverage differs across countries.

c. The contribution of the yearly VAT-reporters to the total turnover is very small (generally < 1%). Hence, it is assumed that monthly and quarterly VAT-reporters are representative for yearly reporters when estimating the first monthly and quarterly growth rates.

d. Challenges on timeliness and periodicity are focussed on monthly estimates, because
   - for quarterly estimates both monthly and quarterly VAT-data are available and the timeliness of Admin Data is a less important factor. The latter is because the deadlines for statistical output are generally later than 45 days after the end of the quarter.
   - for monthly estimates, only (a limited group of) monthly VAT-data are (is) available. Furthermore, the timeliness issue becomes much more important since some publication deadlines for monthly turnover estimates are scheduled early. An example of an early publication is the “t+ 30 days estimate” for the retail trade according to the STS-regulation.
   - several countries (e.g. the Netherlands, Finland, Sweden) faced legislation changes which resulted in a sharp decrease of the number of monthly VAT-reporters.

e. When describing practices about the use of VAT-data for turnover growth rate estimations for SMEs, two statistical approaches can be detected:
   - estimations based on VAT-data alone.
   - estimations using VAT-data as auxiliary information.

Both practices are applied in case of a good coverage of the available VAT-data for the reporting period. Moreover, the latter can be subdivided into two types.

---

1 Reporters may report for quarters ending March, June, September and December (stagger 1); for quarters ending April, July, October and January (stagger 2) and for quarters ending in May, August, November and February (stagger 3). Monthly reporters are referred to as “stagger 0”.
1. Using VAT-data as auxiliary information when weighting (SME) Survey-data for current period $t$. These are the so-called regression estimators.

2. Using VAT-data from previous periods as auxiliary information (in combination with data for current period $t$ from other sources). These are the so-called inter- and extrapolation techniques.

f. The mean average error and the mean average bias (see formulas 1 and 2) between the first and definitive (= when the Admin Data are complete) estimate for period $t$ are generally considered as proxies for the quality of the 1st estimates. It should, however, be noted that not all participating countries measure their revision or revise their results when more Admin Data become available.

\[
B = \frac{\sum_{i=1}^{T} e_i}{T}
\]

\[
E = \frac{\sum_{i=1}^{T} |e_i|}{T}
\]

with $e_i$ revision error between the first and definitive estimate for $t$, $T$ number of analysed periods

Points a., b., c. and d. are related to the used (input) data on which the estimates are based. These points will be clarified in chapter 3. Points e. and f. are related to estimation procedures and they will be described in chapter 4.

Note that the ESSnet WP 4 is only halfway of a 4-year period. Hence, this deliverables presents, in contrast to the description of the national practices (Baldi et al., 2011; Kavaliauskiene, 2011; Lorenz, 2011; Orchard et al., 2011, Sirviö, 2011a,b; Šličkutė-Šeštokienė, 2011; Vlag et al., 2011), only a compilation of the preliminary results. As a result, chapters 3 and 4 provide only a superficial description of the results obtained until the end of February 2011. Despite this superficial description, this document links the results obtained so far with the work to be done in the next period.


3.1 Survey for the largest enterprises

As mentioned in chapter 2, all participating countries use (or intend to use) a mixed-mode model. This means that for the largest enterprises (LEs) Survey-data (complete enumeration) are used as data source, while VAT-data are either the main or the auxiliary data source for SMEs. The implicit assumption for this workpackage is that all Survey-data are available in time (although this it not the case for all participating countries). The cut-off point used to identify the survey part differs per country. For example, enterprises with at least 15 mln € annual turnover or at least 250 persons employed are surveyed in Germany. In Italy, only enterprises with more than 500 persons employed are subject to survey to support STS-employment estimates. The Netherlands uses a complete enumeration approach of enterprises with more than 100 persons employed (but it tries to reduce it to the largest 1900 enterprise groups). Finland has a survey for the largest 2000 enterprises only (plus a survey for 1000 enterprises to measure industrial production).

Ideally, the coverage of the large enterprises survey should be related to the availability of the Admin Data and the desired quality of the output. In practice, this relationship seems to be more complex.
Therefore, one of the challenges of ESSnet Admin Data WP 4 is to provide some recommendations about survey size for the largest enterprises.

3.2 Incompleteness of VAT due to yearly VAT-reporters are almost negligible. Challenge must focus on monthly-reporters

In most European countries, VAT may be declared on a monthly, quarterly and yearly base (Ortega Azurduy, 2011). The large enterprises have to declare their VAT on a monthly base, while the other enterprises are quarterly VAT-reporters. Only very small enterprises may declare their VAT on a yearly base. The contribution of yearly reporters to total turnover is usually very small.

The threshold between quarterly and monthly VAT-reporting differs per country. As a consequence, the share of the total turnover of enterprises reporting monthly varies from more than 90 % (Finland, Germany, Lithuania) to less than 10 % (United Kingdom). In the Netherlands, this share of the turnover is difficult to determine since monthly VAT-reporting is only on a voluntary base and monthly VAT-reporters are not representative for the entire population anymore.

A general conclusion is that – apart from the timeliness issue – all countries do have sufficient VAT-data for quarterly estimates, but this is not the case for monthly turnover estimates.

In the UK, VAT is reported by three month periods, implying that VAT-data has to be interpolated into (artificial) monthly data for monthly estimates. To produce quarterly estimates, these (artificial) monthly data have to be ‘added’ to calendar quarterly data. This results in 'smoothing' of the VAT-data which can be addressed by a survey component for estimation purposes, as is the case in Canada and New Zealand.

At the National Statistical Offices participating in workpackage 4, most VAT-data (> 95%) are available between \( t + 27 \) days (Lithuania) and \( t+60 \) days (United Kingdom) after the reporting period. It means that quarterly turnover estimates can be based on almost complete VAT-datasets since,

a) data from monthly and quarterly VAT-reporters are available for quarterly estimates. Together these groups cover almost the total turnover of the population (as the share of yearly reporters can be neglected).

b) by far most quarterly turnover growth rates are published between 45 to 60 days after the end of the quarter when nearly all VAT-data are at the statistical office.

Hence, it is not surprising that WP 4 detected three practices for quarterly turnover estimates (Finland, Germany and the Netherlands) based on an almost complete dataset. Main similarities and differences between these practices are described in chapter 4. Statistics Finland uses the same practice for regular monthly estimates, because the share of monthly VAT-reporters to the total turnover is > 90 %. It should also be noted that Italy uses a quite similar practice for Employment estimates (based on Social Security data).

For monthly estimates, the situation is more complicated. The major two issues are:

a) The publication deadlines for monthly turnover estimates are earlier (for example: \( t+25 \) days in Lithuania, \( t+30 \) days for the retail trade according to the STS-regulation, \( t+40 \) days for industrial production – the latter is often derived from turnover). Hence, the monthly VAT-data are not available yet when the estimates have to be made.

b) Monthly VAT-data are not representative for the entire population, because it is a (relatively) small group dominated by enterprises belonging to a certain activity.

ESSnet AdminData WP 4 has also described some – more experimental – practices about the use of Admin Data for monthly turnover estimates under these complicated conditions (Kavaliauskiene, 2011; Sirviö, 2011b; Vlag et al., 2011). Chapter 4 describes the challenges of these methods.
It should be noted that the use of VAT-data for monthly turnover is an actual topic in many countries, because
- the reduction of administrative burden can be substantial by reducing the number of monthly surveys. Especially, in the case of surveys which are only intended for early turnover estimators (for instance in the case that VAT-data become available a few days later).
- some countries have faced changes in tax legislation which has resulted in less monthly VAT-data, i.e. a decrease on the number of monthly VAT-reporters.


4.1 Estimates based on an almost complete VAT-dataset

As mentioned in chapter 3, workpackage 4 detected three practices for turnover estimates based on almost complete VAT-dataset. In the current situation, these practices are used with a turnover coverage of the available VAT-data of more than 90 %. All practices are used for quarterly estimates only, with the exception of Statistics Finland (which use the same approach for monthly turnover estimates).

These practices are characterised by the fact that the estimates for small and medium sized enterprises are based on VAT-data alone. The practices from Germany and Finland seem to be focussed on estimations of growth rates (this is also the case for similar practices from Italy for Employment estimates), while the Dutch approach – which has not proven itself yet in regular production – is focussed on both turnover growth rates and turnover levels.

Some remarkable differences are observed between practices. The most important ones are:

a. Estimation of missing values
   - Germany estimates missing values by \( \frac{O_t}{O_{t-1}} \) (=current period over previous period) growth rates calculated at “NACE2-digit level x 2 size classes” level.
   - The Netherlands estimates missing values by \( \frac{O_t}{O_{t-12}} \) (=current period over same period previous year) growth rates, calculated at “NACE3/4-digit level x 9 size classes”.
   - Finland uses five different imputation rules calculated at microlevel (See Figure 1). These five imputation rules are tested by comparing them with data for the five previous periods. The imputation rule closest to these ‘realisations’ is selected for imputation the ‘missing’ values for current reporting period \( t \).

b. The role of the Business Register

The Business Register has also a different position in the several practices. For example, the Netherlands and the United Kingdom use the Business Register as population frame for each period \( t \), while Italy (employment) matches the Admin Data to the Business Register once a year. Germany matches the Business Register to replace the code of economic activities allocated by fiscal authorities with the code of economic activity of the Business Register. Italy and Germany derive population changes for current period \( t \) (starting/stopping enterprises) from the Admin Data. These different positions of the Business Register across countries is partly related to the actuality of the Business Register.

It is in connection to these differences that the aim of next period is to detect why different choices in practices have been made. Although some of these differences are connected to Admin Data availability, like the role of the Business Register. the hypothesis is that these different
methodological choices might also be related to a) different cost-benefit considerations and b) practical implementation (Lorenz, 2011).

As the VAT is currently nearly complete in these countries (when the estimates have to be made), these differences might have few impact on the results. Therefore, all these practices can be recommended with a turnover coverage of at least 90%. However, these practices might produce different results with a lower share of VAT-data to the total turnover. To determine this threshold, some testing work is planned. Determination of this threshold might be useful for countries which plan to use these methods but, have a smaller share of VAT-data to the total turnover. Furthermore, Italy plans to test the German imputation method on their Social Security Administrative Data, because this method seems quite straightforward and produce good results.

**Figure 1.** The situation in Finland (Sirviö, 2011a) concerning the five imputation rules for ‘missing’ data.

Five different imputation rules are used for enterprises with missing data (i = month to be estimated):

\[ \hat{x}_t = \frac{x_{t-1} + x_{t-2} + x_{t-3}}{x_{t-1} + x_{t-2} + x_{t-3}} \times x_{t-12} \]  
(Mean annual change)

\[ \hat{x}_t = \sqrt[3]{\frac{x_{t-1} \cdot x_{t-2} \cdot x_{t-3}}{x_{t-1} \cdot x_{t-2} \cdot x_{t-3}}} \]  
(Geometric mean of monthly changes)

\[ \hat{x}_t = x_{t-1} \]  
(Previous turnover)

\[ \hat{x}_t = \text{avg}(x_{t-1}, x_{t-2}, x_{t-3}) \]  
(Mean turnover)

\[ \hat{x}_t = x_{t-12} \]  
(Turnover of comparison month)

The methods are tested with data concerning the five latest months.

When testing these practices the next year (and comparing the advantages-disadvantages of the different methods) the question about the desired quality raises in relation to the complexity (=costs) of the system. It is proposed to use the test results to provide recommendations about this question in the next two years. The question about measuring desired quality will be discussed in chapter 4.2.

### 4.2 Quality of the estimates

As mentioned in chapter 3 the quality of the first estimates (when the Admin Data are incomplete) is often determined by the average error and the average bias. It was, however, concluded by the WP 4 participants that the average error and average bias are only measured for the total estimate, including the Survey part for the largest enterprises. This is an obvious choice from the publication point of view but, this is not a practical approach to measure the quality – and provide recommendations on cost-benefit ratios - for the ‘missing data’ estimations in the VAT-part (usually connected to small and medium sized enterprises). This is because the average error and bias might be dominated by revisions in the Survey part.

Hence, in the next period, recommendations will be provided to distinct the average error and bias due to the accumulation effect in the “VAT” part from the average error and bias in the total estimate (Röstel, 2011). This is illustrated in Schema 1.

The ‘decomposed’ average error and bias cannot only be used to relate the practices in case of a good coverage of the Admin Data to the desired quality, but also – or even especially – to relate the more experimental practices in case of low completeness of the available Admin Data to the desired quality. The challenges of the latter practices will be described in chapter 4.3.
### 4.3 Challenges in turnover estimations based on incomplete VAT datasets

Basically, two types of practices were detected for estimating turnover growth rates in case of bad (and not representative) coverage of the VAT-data for current period $t$. Both approaches are characterised by using VAT from previous periods as auxiliary information for the estimates in stead of using the available VAT-data for current period $t$. Hence, these methods can also be used in case of no VAT-data are available yet when the estimates have to be made. All these approaches are used for monthly turnover estimates only (!)

The two types of estimates encountered are:

- **Using a mini-Survey for small and medium sized enterprises for current period $t$ and weighting this Survey by using VAT-data from the previous period. These so-called regression estimators are used in Lithuania.** Advantages of using regression estimators are that they
  - correct for (limited) effects connected to definition differences in VAT-turnover and turnover according to the STS-definition
  - enable to reduce the needed Survey-data to the minimum.

Disadvantages of these methods might be (1) their inaccuracy in case of strong seasonal changes (for example from December to January) and (2) lack of robustness against strong changes in turnover due to sudden events or changes in the business cycle. This should be further examined by some testing. Based on the work of Šličkutė-Šeštokienė (2011), the applicability of using regression estimators to produce estimates for low aggregation levels should also be explored as
well as the approach of Statistics Sweden to weigh a mini-Survey for monthly turnover with annual VAT-data from the previous year.

b. Extrapolating VAT-data from previous periods to estimate growth rates for current periods \( t \).

Up to now, three practices were worked out in this field.

- extrapolating VAT-data with a time-series model using linear trends and seasonal factors only (Orchard et al., 2011).
- extrapolating VAT-data with a time-series model by using the trend and an auxiliary variable (for example wages) for current period \( t \) (Sirvio, 2011b).
- extrapolating VAT-data by using the growth rates for the large enterprises for current period \( t \) (derived from a Survey) and the differences in growth rate between the large enterprises and the small/medium sized enterprises in previous periods (Vlag et al., 2011).

Similarly to the case of regressions, the accuracy of these extrapolating methods should be tested in case of sudden events or changes in the business cycle. Recommendations for these methods might include their possible applicability for the \( t+30 \) days estimate for the retail trade and the required size of the large enterprises Survey.

4.4 Challenges in harmonisation in the use of timeliness-related terms

Another worth to mention challenge of the ESSnet Admin Data WP4 is the harmonisation of timeliness-related terms, concepts and definition to create a common ground of discussion and the efficient implementation and communication of the results achieved so far by this workpackage. Estonia has contributed to this topic by the compilation of a Glossary of used statistical terms in the deliverables 4.1 until 4.7 of SGA-2011 (Karus, 2011). This glossary is published at the Information Centre of Essnet Admin and will be used to harmonise terms further.

5. Conclusions.

Basically, two types of practices were detected within the ESSnet Admin Data WP 4. Type 1 practices comprehend ‘traditional’ practices, for which the estimate is based on a Survey-data for the large enterprises (complete enumeration) and only VAT-data for small and medium sized enterprises. These practices can be recommended but do require a good coverage of the VAT-data when the estimates have to be made (in practice > 90 % turnover coverage at NACE 2-digit level in the participating countries). They are generally used for quarterly turnover estimates. Within the ESSnet Admin Data, differences between these practices were observed. These differences might be related to different cost-benefit considerations. Hence, it is planned to determine by testing the location of transition points, i.e. below which turnover coverage of the available VAT-data these practices will provide different results. This might be useful information for countries considering to use this method but not having a 90 % turnover estimate. The desired quality of the estimates will be quantified by average bias and average error, the most widely used quality indicators for early estimates, for the VAT-part.

Type 2 practices include more ‘experimental’ practices and is used for monthly turnover estimates in case of bad coverage of the VAT-data when the estimates have to be made. Within these practices, the use of regression estimators can be distinguished from extrapolation techniques. Further research and recommendations should be focussed on the applicability of these methods in case of strong seasonal patterns and changes in the business cycle. Recommendations should be focussed on its use for early monthly turnover estimates, like the \( t+30 \) days estimates for the retail trade.
Acknowledgements

The authors would like to thank all the countries participating within ESSnet Admin Data WP4 who share their ideas, interest and comments with us over the last year. Their commitment motivated and supported the work across the Work Package 4: Timeliness of Administrative Sources for Monthly and Quarterly Estimates. We would also like to acknowledge the engagement and assistance provided by Eurostat and the programme ESSnet: Use of Administrative and Accounts Data in Business Statistics, Grant number: SGA-2010.

References


D. Kavaliauskiene, 2011, USE OF REGRESSION TYPE ESTIMATION TECHNIQUE AS A TOOL FOR SOLVING TIMELINESS PROBLEM IN SHORT-TERM STATISTICS, deliverable ESSnet WP4, January 2011

R. Lorenz, 2011, IMPUTATION OF MISSING VALUES IN THE VAT DATA IN GERMANY, deliverable ESSnet WP4, January 2011

C. Orchard, A. Langford, K. Moore, NATIONAL PRACTICES OF THE USE OF ADMINISTRATIVE AND ACCOUNTS DATA IN UK SHORT-TERM BUSINESS STATISTICS, deliverable ESSnet WP4, January 2011

S. Ortega Azurduy, 2011, OVERVIEW ON COUNTRY-RELATED ISSUES CONNECTED TO THE USE OF INCOMPLETE ADMIN-DATA FOR STS-ESTIMATES, deliverable ESSnet WP4, February 2011

D. Röstel, ATTEMPTS TO IMPROVE METHODS OF PLAUSIBILITY CHECKS OF COMBINED TURNOVER DATA IN GERMAN SERVICE STATISTICS (STS), deliverable ESSnet WP4, January 2011

M. Sirviö, 2011a, TURNOVER INDICES (INCL. RETAIL TRADE) AND VALUE ADDED TAX (VAT) DATA, deliverable ESSnet WP4, January 2011

M. Sirviö, 2011b, INDUSTRIAL PRODUCTION INDEX AND VALUE ADDED TAX (VAT) DATA, deliverable ESSnet WP4, January 2011

M. Šličkutė-Šeštokienė, 2011, APPLICATION OF RATIO AND GENERALIZED REGRESSION ESTIMATORS FOR LITHUANIAN LABOUR STATISTICS BASED ON ADMINISTRATIVE DATA, deliverable ESSnet WP4, January 2011