Benchmarking monthly indicator series with quarterly information derived from administrative data

Pieter Vlag

Statistics Netherlands, P.O.Box 24500, 2490 HA Den Haag, The Netherlands

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Conference

Abstract

Several countries in Europe use the Value Added Tax (VAT) for short-term statistics (STS). An important challenge when using VAT for STS is that the VAT is incomplete when the estimates have to be made. This is because not all VAT-data are timely available or reported for the same period as the statistical output. An European project has focused on the challenge of incomplete sets of VAT used in the production of quarterly and monthly turnover estimates.

Two situations can be distinguished when producing STS in a VAT based system with the largest enterprises being surveyed. Situation I is that VAT provides good coverage when the estimates have to be made. Situation II is that no or only few VAT is available.

Situation I is in general applicable for quarterly estimates. The main conclusion of the project is that good quality level and growth rate estimates can be produced with VAT under these circumstances. These level and growth rate estimates are called the structural series.

Provided that sufficient VAT-data are available to produce structural series for the quarters, the challenge is to find estimation methods for (early) month when no or few VAT is available. The project suggests that monthly estimates should be considered as an 'indicator', which are benchmarked with the structural series. The estimate for the current month is based on a benchmark-nowcast. The latter means that the estimate is determined by the indicator value for the current month corrected with the average difference between the original indicator values and benchmarked indicator values in the previous months.

A small survey under the larger enterprises is sufficient to determine the inidcator value, if benchmarking demonstrates that the short-term trend is well measured by these enterprises. In this case, monthly estimates can be provided by either

- the growth rates of the surveyed enterprises only (in case the average differences between the original indicator values and benchmarked indicator values in the previous months are close to zero).
- the growth rates of the larger enterprises plus a correction for the difference in the long-term trend between surveyed larger and non-surveyed smaller enterprises (in case the average differences between the original indicator values and benchmarked indicator values in the previous months are systematically positive of negative).

This paper presents a case study on the retail trade and commercial services enterprises in the Netherlands.

Introduction

The use of administrative data for statistical purposes has increased considerably during the last decade in Europe. National Statistical Institutes (NSIs) in Europe use administrative data in two ways; as auxiliary information or to replace survey. Two examples of using administrative data as auxiliary information are that administrative data are used 1) to check the validity of outlying survey values and 2) to check the accuracy of estimates based on surveys. However, an increasing number of countries in especially northern and north-western Europa use administrative data

to replace survey. A common example is using the VAT (Value Added Tax) registration of the tax office for turnover estimates. Note that the VAT in Europe is comparative with sales taxes in Northern Armerica. Another widespread example of replacing surveys with administrative data in Europe is the use of social security data for employment estimates. Statistics Netherlands uses both the VAT-registrations for turnover estimates as well as social security data for employment estimates.

A problem is that administrative data are incomplete at the time when they are needed for enterprise statistics. This is particularly the case when using Value Added Tax (VAT) for turnover estimates, because these administrative data are still incomplete when the monthly or quarterly turnover estimates have to be produced. This incompleteness might be temporal (e.g. due to late responses of enterprises) or structural (e.g. because enterprises below a fixed income threshold may report to the admin data holder for a different periodicity).

In the Netherlands, commercial enterprises have to report their VAT on a monthly or quarterly basis within 30 after the reporting period. Only very small enterprises may declare annually. Taking into account these reporting periods and reporting deadlines, three situations exist with regard to the completeness of VAT-data for statistical purposes.

1. The VAT-data are complete for annual statistics. Taking into account the VAT-regulation, this implies that we do have for approximately all enterprises an annual VAT-turnover.
2. The VAT-data are almost complete for quarterly statistics. This is because commercial enterprises have declared their monthly or quartely VAT already 45 days after the end of quarter and quarterly estimates have to be produced later (45 or 60 days after the end of the quarter). As the number of late reporters to the tax office is very limited, the only missing data are those of the annual VAT-reporters. However, the share of the annual reporter to the turnover totals is very limited.
3. Selective and incomplete VAT-data are available when the monthly estimates have to be produced. Firstly, VAT-reporting is late for estimates to be produced 30 days after the month. More importantly, monthly VAT-reporting is voluntary in the Netherlands. Hence monthly VAT-reporters are a selective group dominated by enterprises expecting a tax return.

Despite the use of VAT for turnover estimates, the larger enterprises (> 100 persons employed) are still surveyed by Statistics Netherlands. This is because large enterprises often have a complex structure and their impact on the estimates is high. Statistics Netherlands therefore considers surveyed observations from those large enterprises as crucial for producing reliable STS figures. The large enterprises are completely enumerated in the survey, with VAT being used for the remaining smaller enterprises.

Two types of estimation

Looking at the available VAT-data, two situations can be distinguished when estimating turnover with VAT. Situation I is that almost all of by far most admin data are available when the estimates have to be made. It can be therefore be assumed that the available data are representative. Situation I applies to annual estimates and quarterly estimates to be produced 45 or 60 days after the end of the quarter in the Netherlands. Situation II is that no or few admin data are available when the estimates have to be made. If few admin data are available, they are not representative. Situation II applies to monthly estimates in the Netherlands.

Turnover estimations with almost complete VAT: annual and quarter

As a rule of thumb, Statistics Netherlands considers like most European countries a coverage of about 80% of the total turnover by the available VAT and the LE-survey as a good coverage. This is generally the case for annual and quarterly estimates.

Statistics Netherlands produces like most NSIs in Europe estimates for levels and growth rates in this situation. The estimates are generally produced with micro-level data. Therefore, missing data are imputed. Most NSIs in Europe impute missing values by using Y_t/Y_{t-1} (current period – previous period) or Y_t/Y_{t-12} (current period – same period last year) ratios of available VAT-data. Statistics Netherlands uses Y_t/Y_{t-12} ratios. However, findings of an European project on the use of administrative data, the so-called ESSnet AdminData project, for short term statistics (STS) indicate that the choice for using either Y_t/Y_{t-1} ratios or Y_t/Y_{t-12} ratios for imputation has negligible impact on the

results. This is because the coverage of the available data is very high. Therefore, this project has recommended at one should aim for an optimal trade-off between benefits and costs when choosing an imputation method, rather than aiming for the “best” theoretical quality.

The most important issue for the STS estimates is uncertainty about which enterprises are active and which are not. As a consequence, the question arises in statistical production whether VAT-data are missing due to late reporting or because the enterprise is not active (any more). In the latter case the missing unit does not need to be imputed. Analyses in Italy, Germany and Estonia by the ESSnet AdminData project show that the main revisions to STS-VAT estimates are caused by uncertainty in the active enterprise population. Therefore, it is recommended that, when developing a statistical production system for STS-estimates using VAT, research and development should be focussed on the choice of the best method for determining the active population (i.e. which enterprises with missing VAT-data are to be imputed).

To deal with this challenge, a two step imputation model is recommended:

1. determine a provisional target population: Define a list of missing enterprises to be imputed. The most generally applicable approach for the determination of the provisional active population is that NSIs simply check whether the enterprise has reported any VAT-data to the tax office for the last few months.
2. impute the values of the defined list: For this part one has roughly speaking the choice between imputing missing VAT with average Y_t/Y_{t-1} (current period / previous period) or Yt/Yt-12 (current period / same period last year) growth rates of available VAT. These average growth rates can be calculated at high or low aggregation level.

The findings of the ESSnet AdminData project confirm the conclusions of Statistics Netherlands and other northern European countries that it is possible to produce high-quality turnover growth rate and turnover level estimates by using VAT as replacement for a survey for medium-sized and small enterprises using micro-imputation, if

- the data transfer of VAT from tax office to NSIs is guaranteed
- >> 80 % turnover covered by VAT
- a (provisional) target population cen de defined

For further details, we refer to de Waal, 2012, Vlag et al., 2011, Maasing et al., 2013 en references herein. In the remaining part of this paper, we assume that high quality VAT-based turnover estimates can be made for the quarter and that these quarterly estimates can be used as benchmarks for the monthly estimates.

Turnover estimations when few or no VAT available yet: month

Provided that turnover levels and growth rates can be estimated with VAT for the quarters, the challenge is to find estimation methods for month when no or only a few VAT-data are available. This as alternative for a relatively large standard monthly survey among all enterprises within a branch, which is not connected to the later available quarterly VAT-estimates.

The work of the ESSnet AdminData project has demonstrated that if few VAT is available, this VAT cannot be used to replace a monthly survey. Main reason is that, as confirmed by analyses in Finland, the Netherlands and the United Kingdom, the available VAT is selective. In other words, the available VAT covers a specific group of enterprises within a branch. Therefore, the turnover growth rates derived from the available VAT do not necessarily correspond with the growth rates of the entire branch. It is not straightforward to develop estimation techniques correcting for this selectivity, because the selectivity of the available VAT may vary over time and cannot be determined beforehand, e.g. when the STS-estimates for month *t* have to be made. Therefore, it was concluded that few available VAT-data cannot be used to replace a survey.

However, VAT can be used to check whether large enterprises and small enterprises record similar long-term trends and short-term movements of the economic growth. Depending on the outcome, this information can be used to decide whether

- a survey under the largest enterprises (a LE-survey) only is sufficient for the monthly estimates, knowing that VAT covering the entire population becomes available for the quarters. This option is called alternative I in the remainder of this paper.

- the LE-survey should be combined with a separate estimate for the smallest enterprises based on extrapolation of the VAT-series. This option is called alternative II in the remainder of this paper.

VAT of previous quarters is in both alternatives used as auxiliary information for the monthly estimates in both alternatives. The European ESSnet AdminData has explored both alternatives. Statistics Netherlands has explored alterrnative I as it considers this alternative as a promising approach for a monthly estimate for the retail trade and the service sector without using extensive surveying. In the remainder of this paper we'll present the findings of Statistics Netherlands for alternative I. Alternative II is not discussed further in this paper. For more information about alternative II, we refer to Sirvio, 2010 and references herein.

Monthly estimates: indicator series and temporal estimations

With VAT being available for the quarter and high-quality level and the finding that level and growth rates estimates can be based on almost complete VAT, the monthly estimate could be seen as an indicator, which is benchmarked (= made consistent from a time-series perspective) to the quarterly series. The latter are called structural series, as both detailed turnover levels and turnover growth rates can be derived from these series.

Therefore, we propose that the general set-up of an indicator series is

- a surveyed part of the population. These are the largest enterprises which are completely enumerated.
- a part of the population without data for current period, which might be temporary estimated. These are the medium-sized and small enterprises.

Ideally the surveyed part for the monthly indicator series is equal to the surveyed part of the structural quarterly series with are mainly VAT-based. The survey among the largest enterprises is called the large-enterprise (LE) survey in the remainder of this paper.

The idea of benchmarking is that the indicator is made consistent with the higher-quality structural series. When the indicator is benchmarked, we speak about the benchmarked indicator series. Several approaches for benchmarking do exist. Dagum en Chollette (2006) generalised the several benchmark-approaches used at Statistics Canada in the following formula:

$$f(Y_1, \ldots\ldots, Y_T) = (1-\rho^2) * \left(\frac{Z_1 - Y_1}{Z_1^{\gamma}}\right) + \sum_{t=2}^{T} \left[\left(\frac{Z_t - Y_t}{|Z_t^{\gamma}|}\right) - \rho\left(\frac{Z_{t-1} - Y_{t-1}}{|Z_{t-1}^{\gamma}|}\right)\right]^2$$

with

Z_t	: original bias-corrected indicator series
Y_t	: benchmarked indicator series
ρ, γ	: parameters

We refer to Dagum and Cholette (2006) for more information about the several benchmarking techniques. Statistics Netherland has used the parameterisations $\rho = 1$, $\gamma = 1$ as default. These parameterisations corresponds with the so-called proportional Denton-method.

A benchmark nowcast is applied for the current month with no structural VAT-data available yet. Basic idea behind the benchmark nowcasting is that the adjustments between the original and the benchmarked indicator series over the last x periods are used to improve the estimate of the indicator series for current month *t* not covered by quarters. This can be done by summarising the adjustments of the previous periods into a factor C_1.

In formulas:

$$Y_{\text{indicator, benchmark nowcast}} = Y_{\text{indicator original}} * C_1$$

Several methods do exist to determine C_1. These methods can be subdivided into so-called implicit and explicit benchmark nowcasts. When applying implicit benchmark nowcast, the standard benchmark formulas are used for the extremities (the current months not covered by the quarters). When using an explicit benchmark nowcast a specific value is assigned to C_1. An explicit benchmark nowcast is used in this study. More specifically the value C_1

is determined by the mean of the adjustment between the original and the benchmarked indicator series over the last 6 months. For an overview of exiting periods to calculate C1, we refer to Fortier et al. (2007).
Statistics Netherlands has tested a method of monthly estimates at which

- a LE-survey is used to determine to original monthly indicator-series, i.e. $Y_{\text{indicator original}}$
- the quarterly turnover series are used as benchmarks. These series are based on a LE-survey and VAT
- benchmarking and implicit benchmark nowcast-techniques are used to calculate C_1
- the estimate for month t is based on $Y_{\text{indicator original},t} * C_1$

In this testing the LE-survey is similar for the monthly as quarterly estimates. Hence, the factor C_1 implictly measures the differences in growth between large enterprises and medium-sized plus small enterprises which are covered by VAT for the quarters and are not ‘ measured’ for month. Factor C_1 can therefore be considered as a temporal estimation for medium-sized plus small enterprises.

Conditions for applying benchmarking techniques

With the largest enterprises being surveyed, two situations may exist for the temporary estimation of small enterprises.

The first situation is that:

- the <u>short-term</u> movement of the growth of the non-surveyed small enterprises is similar to the short-term movement of the surveyed large enterprises; and
- changes in the business cycle and sudden events are simultaneously registered in the surveyed and non-surveyed parts.

The second situation is that:

- the <u>short-term</u> movement of the growth of the non-surveyed small enterprises may differ from the short-term movement of the surveyed large enterprises; and/or
- time-lags may exist between these groups; and/or
- changes in the business cycle are differently recorded in the surveyed and non-surveyed parts.

The first approach is valid for enterprises related to consumption (like retail trade and part of the services) , because it is unlikely that consumers suddenly start consuming in larger or smaller shops when the economy changes. The first approach might also be valid for homogeneous enterprise activities without demand-supply or subcontractor-contractor relationships within the branch. Nevertheless, it is strongly recommended to test the validity of these assumptions on historical data, such as VAT-data on a quarterly base.

The second approach is valid for enterprise activities with demand-supply, subcontractor-contractor relationships or for smaller enterprises which are more concentrated on sub-activities (e.g. the construction sector). Within manufacturing, it is possible that small enterprises supply parts to the larger enterprises, which are more focussed on assembly. In this case, changes in the business cycle may be differently recorded in small and large enterprises. Again, it is strongly recommended to test the validity of these assumptions on historical data.

The threshold between large and small enterprises is a matter of debate. For this study large enterprises are defined by enterprises with 20 or more persons employed. Formally, these are medium-sized and large enterprises (MLE), a term which is used in the remainder of this paper. Analyses with large enterprises being defined as enterprises with 100 or more persons employed provide similar results for the high aggregation levels. These results are not discussed in this paper.

If the growth of the medium-sized and large enterprises is related to the growth of the small enterprises, the total estimation can be determined by:

$$G_{t,t-1} = \frac{Y_{LE}.G_{t.t-1;MLE} + Y_{SE}.G_{t,t-1;MLE}.C}{Y_{MLE} + Y_{SE}}$$

with

$G_{t,t-1}$, $G_{t,t-1;MLE}$ the growth rates of the entire target population and the surveyed medium-sized and large enterprises (LE) respectively.

Y_{LE}, Y_{ME} the (extrapolated) turnover level for the MLE and the small enterprises (SE), respectively

C a factor to correct for possible systematic differences in growth between MLE and SE. As the factor C multiplies the growth, the presented approach is a so-called multiplicative model (which assumes that the growth of the non-surveyed part may differ by a constant factor from the growth of the surveyed part)[1]

The most simple model arises from the assumption that C = 1. In this case, the growth rate of the MLE equals the growth of the entire population. In more sophisticated models C may be determined by benchmark nowcasting. A necessary condition for benchmark nowcasting is that growth rates are related. Hence, before applying benchmarking methods it should be checked that

- the short-term movement of the growth of the non-surveyed small enterprises is similar to the short-term movement of the surveyed enterprises; and
- changes in the business cycle and sudden events are simultaneously registered in the surveyed and non-surveyed parts.

Benchmarking techniques tested by Statistics Netherlands

The current monthly survey for the retail trade covers all enterprises. This survey is under pressure in its present form as it is considered as too costly. Therefore Statistics Netherlands continues looking for alternatives to reduce the monthly STS-survey for turnover estimates, knowing that good-quality VAT-based turnover estimates are available for the quarter. Therefore, it investigates

- whether monthly estimates can be based on surveying LE only
- how these monthly indicators can be benchmarked to the structural quarterly "VAT-based" series.

Previous analyses suggest that such approach is promising for especially the retail trade. Time-series for the period 2005-2010 suggest that LE record the same short-term movement of growth than the SE (fig 2 in chapter 2.2).

Therefore, the two simulations have been done. Simulation 1 covers the period 2005 – 2011. In simulation I is

- the indicator series a monthly series for enterprises with 20 or more persons employed. The data are derived from the current monthly survey for the retail trade
- the structural series a quarterly series covering all enterprises. Like the indicator series, the data are derived from the current monthly survey for the retail trade
- the indicator benchmarked to the structural series. A benchmark nowcast has been made for current month.

In simulation II is

- the indicator series a monthly series for enterprises with 20 or more persons employed. The data are derived from the current monthly survey for the retail trade. Hence, this indicator series is the same as in simulation I
- the structural series a quarterly series covering all enterprises. The data are derived from VAT and that LE-survey for enterprises with 100 or more persons employed. Hence, this series differs from simulation I
- the indicator is benchmarked to the structural series and a benchmark nowcast has been made for current month.

Simulation II covers the period 2010 – 2011 only, because high-quality VAT-estimate are not available before 2010.

The estimation process is largely based on Fortier et al., 2007 for simulation I. More precisely, it consists of four steps:

[1] Additive models also exist. These models assume that the growth of the non-surveyed part differs by a constant value from the growth of the surveyed part. Additive models are not discussed for the sake of clarity, because previous analyses demonstrated that additive models are needed for these series.

Step 1: <u>calculating the turnover level of the LE for current month</u>:

$$Y_{t,MLE} = Y_{t,LE} + Y_{t,ME} = \sum (y_{LE}) + \frac{N}{n} \sum y_{ME}$$

with $Y_{t,LE}$ the estimate for the large enterprises which are completely enumerated
$Y_{t,ME}$ the estimate for the medium enterprises which are randomly sampled and weighted with the Horwitz-Thompson estimator.

Step 2: <u>calculating the indicator series by adding the small enterprise estimate</u>

$$Y_{t,indicator} = Y_{t,LE} + Y_{t-1,SME} * G_{t;t-1,LE} * C_{bias} = Y_{t,LE} + (Y_{t-1,indicator} - Y_{t-1,LE}) * \frac{Y_{t,MLE}}{Y_{t-1,MLE}} * C_{bias}$$

$$C_{bias} = \frac{\sum_{q=1}^{4} \frac{1}{q} * G_{t;t-1,q,MLE}}{\sum_{q=1}^{4} \frac{1}{q} * G_{t;t-1,q,SE}}$$

with $Y_{t,indicator}$ the indicator for the target population (=all enterprises)
$Y_{t-1,SE}$ the turnover of the small enterprises in previous period. As this turnover is not observed, it is derived from the indicator of previous period minus the estimate derived from the observed large and medium sized enterprises
$G_{t;t-1}$ the growth rate between current periods and previous period
C_{bias} the bias correction. It is derived from the average of the last four quarterly growth rates for MLE and SE, respectively. This long period has been chosen to reduce the impact of noise in the individual quarter-to-quarter growth rates on the bias correction (note that small enterprises are observed on a quarterly base in these simulations)

With respect to the calculations of the turnover estimate of small enterprises ($Y_{t-1,\ SE}$); benchmarked indicator series values have been used until the last 'covered' quarter with estimations for the entire target population.

A bias correction has been applied, because the growth-rates of the MLE are slightly but structurally slightly higher than the growth rates of SE, despite similar recordings of the short-term movement of the growth (see fig. 2 – chapter 2.2). The higher growth rates for MLE are caused by a long-term trend; more consumers are shopping at larger general stores instead of small shops, explaining that on average growth rates of large enterprises are systematically slightly higher than those of small enterprises .

Step 3: <u>the benchmark nowcast</u>

The indicator series is benchmarked to the structural series using the 'tempdisagg' package within R software of EuroStat. Within this software package the "proportional Denton method" (Denton, 1971) has been selected.

For the current month, with no quarterly observations over the entire target population available yet, a benchmark nowcast has been applied.

$$Y_{t,indicatorbenchmarked} = Y_{t,indicator} * \frac{\frac{1}{p} \sum_{p=1}^{6} Y_{p,indicatorbenchmarked}}{\frac{1}{p} \sum_{p=1}^{6} Y_{p,indicator}}$$

where p denotes the previous periods.

The benchmark nowcast implies that the indicator is corrected with the average difference between the benchmarked indicator series and the original indicator series over the last 6 months. The long period has been chosen to reduce the possible impact of noise, but is shorter than the covered period for the determination of the bias correction (12 month) to take into account changes in the trend. If the previous month is also benchmark nowcast (because no quarter available yet), it is considered as a benchmarked indicator in these calculations.

Step 4: revising the benchmark nowcasts with traditional benchmarks

Successively, an indicator for month t is calculated at a first stage. Then the indicator series is adjusted with a benchmark nowcast. In a last step, the benchmark nowcast is replaced by the traditional benchmark after the quarter is covered by quarterly data.

This procedure has been followed for 26 aggregation level of the retail trade, including retail trade total, retail trade food, retail trade non-food. The main results are summarised in figures 1 and 2.

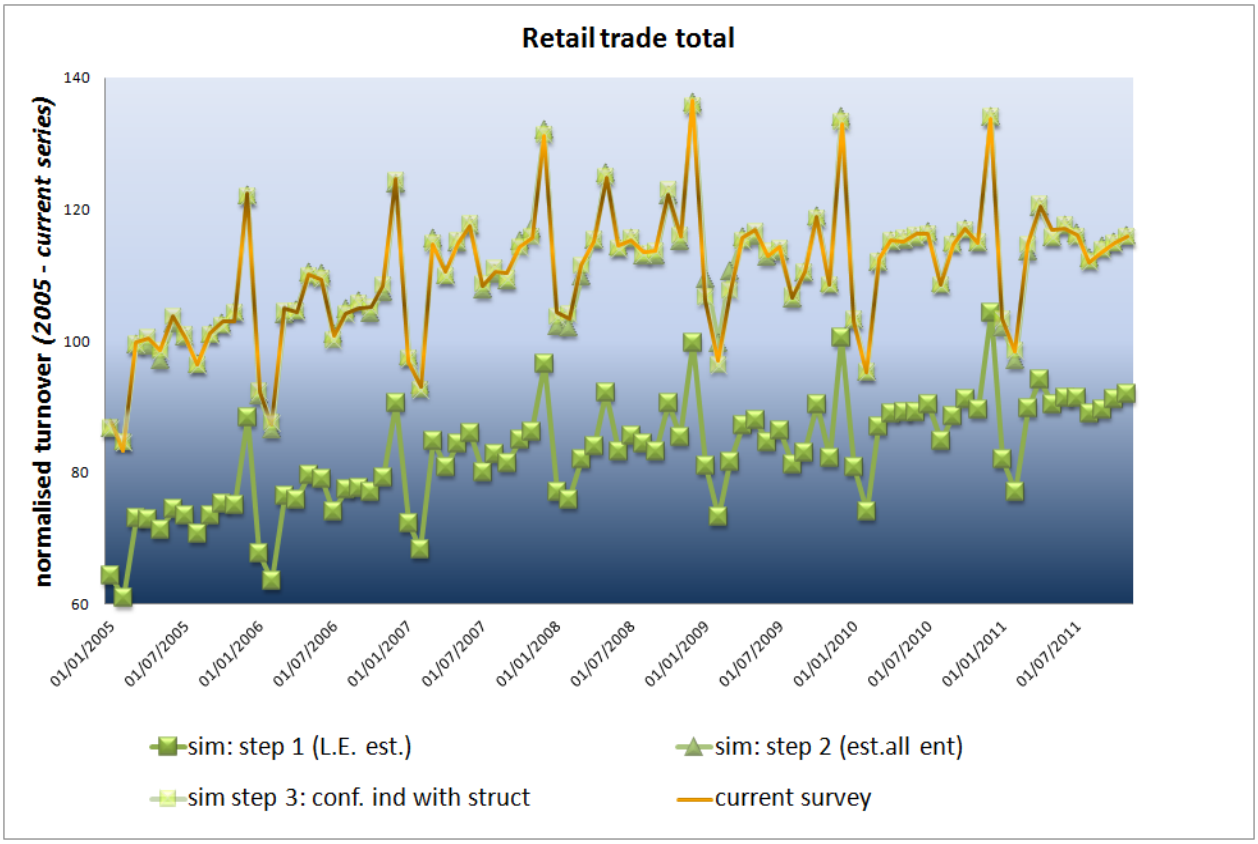

Figure 1 *Benchmarking an indicator series for the retail trade total in the Netherlands (simulations I – see text) The lines denote the growth rates for the "medium and large enterprise estimate" (step 1 in simulation I), the "indicator series" (step 2 in simulations I) and the "benchmark nowcast" (step 3 in simulations I). The results after traditional*

benchmarking are not shown as almost identical to the benchmark nowcast. Instead the results of the current survey are shown as reference.

The results reveal that if

- the share of the surveyed part the population is more than 65 %
- the large and medium enterprises record the short-term movement of the growth well (the latter can be checked by the quarterly data which cover the entire target population)
- and the bias correction is well determined (which is possible if the long-term trend, slightly but structural higher growth for larger than for smaller enterprises, is stable in time).

the indicator value approaches the final estimate. Adding benchmark nowcasts and traditional benchmarking when the quarterly data become available do not structurally change the results. Therefore, the benchmark nowcast and traditional benchmark could be considered as a quality check for the indicator. This hypothesis is confirmed by further analyses on these data, because these analyses revealed that incidental higher revisions after application of benchmarking and benchmark nowcasting on the indicator were in almost all cases related to "artefacts" in either the indicator series or the structural series. Analysing the results also revealed that the estimates after benchmark nowcasting are close to the results of the current survey. Hence, they look plausible from this point of view.

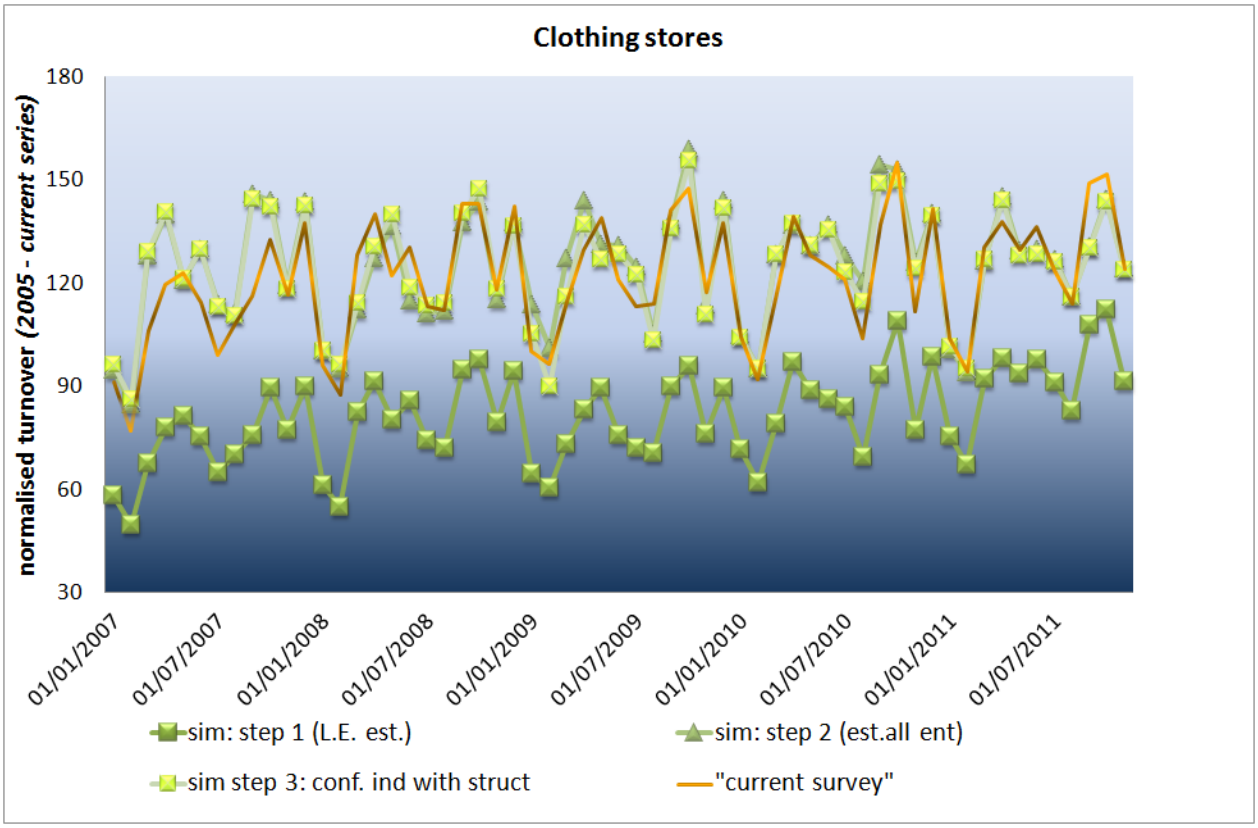

Figure 2 *Benchmarking an indicator series for the clothing stores in the Netherlands, a subgroup of the retail trade. Calculations are according to simulations I (see text for a full description) The lines denote the growth rates for the "medium and large enterprise estimate" (step 1 in simulation I), the "indicator series" (step 2 in simulations I) and the "benchmark nowcast" (step 3 in simulations I). The results after traditional benchmarking are not shown as almost identical to the benchmark nowcast. Instead the results of the current survey are shown as reference.*

The situation described above appeared to be valid for the highest aggregation levels within the retail trade. This situation is also valid for 14 of the 20 subgroups (for example, clothing stores). The share of these 14 subgroups to

total retail trade is, however, more than 80 % in terms of turnover. The remaining 6 subgroups are dominated by small enterprises (like bicycle shops, flower shops and street market trade).

Increasing the threshold for the MLE to enterprises with 100 or more persons employed provide similar results for the highest aggregation levels, but more subgroups with not plausible due to domination by small enterprises.

It should be noted that the conditions for simulations are almost ideal. The datasources and the weighting were identical for the monthly and quarterly series. The data have been thoroughly analyzed before using them for these simulations. Earlier tests on less analyzed data for this period showed much more erratic results. Hence, these simulations only proofs that benchmarking is a useful tool to resolve the timeliness problem of VAT, if the underlying data are ideal.

Benchmarking: level based series versus index series

Unlike simulations I, the source data are not ideal for simulations II of the Netherlands. As the structural quarterly series is a VAT-based estimate, differences between month and quarter might be caused by

- differences in microdata (VAT-declarations versus survey results)
- different impact of weighting

Furthermore, the used time-series are much shorter (2008-2011) and – more importantly – not stable in time, because

- the VAT-based quarterly estimates of 2010 have slightly less quality than those of 2011 (because it was the first year using VAT and the results were not published yet).
- The quarterly estimates of 2008-2009 are still based on a survey. These years have been added to extend the time-series and check how the benchmarking methodology acts in case of a moderate level-shift in the time-series.

Simulations II are described in Bikker (2013), this paper provides only a short summary of the results. Bikker (2013) used the proportional Denton method, like the simulations I. The level shift in January 2010, due to a change from survey based quarterly estimates to VAT-based quarterly estimates, has an impact on the results. This confirms the hypothesis that benchmarking may only be applied be applied if the structural series is continuous and of high-precision. A more practical consequence is that a structural bias in growth between MLE on one hand and SE on the other hand is impossible, because this structural bias differs for the 2008-2009 series and the 2010-2011 series. Therefore the results of Bikker (2013) are hard to interpret in absolute terms. The study shows, however, that the revisions between

- the original indicator series,
- the benchmark nowcast
- the benchmarking after the quarterly information becomes available

is more erratic if a survey based indicator series is benchmarked to a quarterly VAT-based structural series. This effect is more pronounced in the 1st quarter of the year. The larger revision in the 1st quarter can be explained by two factors, both related to the imprecision of the monthly series

- the panel of the sample survey under the medium sized enterprises is renewed at the beginning of the year. As the sample is small this renewal has an impact on the levels of the monthly estimates, because the 'new' enterprises in the panel may have different turnovers and are weighted with high weighting factors.
- the impact of merging/splitting enterprises and outlying values can be much higher on the 'survey based' monthly estimates than on the 'VAT-based' quarterly estimates, due to the high weighting factors of the first.

Hence, the second simulations confirm which was already concluded from theoretical point of view; the indicator series needs to have a high signal to noise ratio, i.e. the impact of the sampling error of the indicator series needs to be limited, when benchmarking it to a structural quarterly series.

To illustrate this effect, two sets of monthly indicators were prepared by Bikker (2013). For the first set all available records of companies with 20 or more 20 people employed were weighted to the total population for each month separately. This gives the best estimate for the level of turnover in each individual month. The second set of monthly

indicators was prepared by forming a panel from all the records of companies with 20 or more persons employed available in two subsequent months. These records were weighted to the total population, leading to an estimate of the growth rate from month $m - 1$ to m. These growth rates were then chained to form index series. This gives the most stable month-to-month growth rates. The first set is called a level-based series. The second set is called an index series.

Tables 1 and 2 summarize the differences between the initial nowcast and the final benchmarked annual growth rates. Mean square error (MSE), the mean difference (= average bias) and the variance of the differences have been calculated for the period 2010-2011. The t-values (mean divided by standard deviation), which are also calculated for 2011-2012, show that the average bias are not significantly different from zero. The much lower variance of the index series does show, however, that it produces more stable results. As the created bias is insignificant and cannot accumulate because the series are benchmarked after the quarter, a conclusion of this work is that an index series should be strongly considered when benchmarking a survey based monthly indicator series with high weight factors to a structural quarterly series based on another data-source like VAT.

	MSE	Av.bias	Variance	t-value
Total retail trade	1.9	0.0	1.9	0.00
Food	0.9	-0.4	0.8	-0.46
Non food	3.8	0.3	3.7	0.17

Table 1. *Level based series. The difference between the annual growth rates of the benchmark nowcast and the benchmarked (in %-point). Results are for the period 2010-2011.*

	MSE	Av-bias	Variance	t-value
Total retail trade	0.9	-0.3	0.8	-0.30
Food	0.3	-0.4	0.1	-1.09
Non food	2.4	-0.2	2.4	-0.12

Table 2. *Index series. The difference between the annual growth rates of the benchmark nowcast and the benchmark (in %-point). Results are for the period 2010-2011*

Conclusions

Two simulations have been described about calculating an indicator series in case the surveyed large enterprises record similar short-term movements of the growth than the non-surveyed small enterprises. These simulations show promising results, indicating that it should be possible to produce monthly estimates without surveying the smallest enterprises. This under the condition that VAT-data become available soon after the end of the quarter. However, the examples do also show some points to take into account:

- as the results of a small dataset are 'extrapolated' to the target population when calculating the indicator series, small artefacts in the dataset may have a large impact on the quality of the indicator (and revisions when benchmarked or replaced by VAT-data at a later stage). Hence, thorough data-analyses are recommended when applying such methods.

- if the bias correction is well determined and the indicator series is of high-quality, benchmarking and the benchmark nowcast might be seen as a quality indicator. In case of a good quality the revisions after applying a benchmark nowcast and benchmarking after the end of the quarter should be small.

- if the indicator series is based on a small survey, a panel-based index series should be used.

- benchmarking only produces good results if the quarterly structural series is of good quality and not affected by level shifts.

Literature

Brown I (2012), An Empirical Comparison of Benchmarking Methods for Economic Stock Time Series, Proceedings ICES-IV conference- June 18-21, 2012, Montreal, Quebec, Canada, pp. 399-412

Dagum E B & Cholette P (2006), Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series Data. New York: Springer-Verlag, Lecture Notes in Statistics #186.

Denton F (1971), "Adjustment of Monthly or Quarterly Series to Annual Totals: An approach Based on Quadratic Minimization," Journal of theAmerican Statistical Association, 82, 99–102.

Fortier S & Quenneville B (2007), Theory and application of benchmarking in Business Surveys, Proceedings ICES-III conference- June 18-21, 2007, Montreal, Quebec, Canada, pp. 422-434

Maasing, E., Remes, T., Baldi, C., Vlag, P.A., (2013), STS estimates based solely on administrative data: final esults and recommendations, Deliverable 4.1 of the ESSnet on AdminData, http://essnet.admindata.eu

Sirvio M (2010), Turnover indices (incl. Retail trade) and Value Added Tax (VAT), *Report 1 Statistics Finland of ESSnet AdminData – SGA2010*, http://essnet.admindata.eu

Sax C , Steiner P (2012) Temporal disaggregation of time series, The R journal, vol., 5/2, 80-87

Vlag, P A (2012), Imputing missing values when using administrative data for short-term enterprise statistics, Paper for UNECE work session on Statistical Data Editing

Vlag, P.A , Bikker, R., de Waal, T., Toivanen, E., Teneva, M., (2013), Extrapolating admin data for early estimation: some findings and recommendations for the ESS, deliverable 4.3 of the ESSnet Admin data – SGA2011, http://essnet.admindata.eu

De Waal, A.G., Vlag, P.A., Baldi, C. Tuzi, D (2012), The use of administrative data for STS. Situation I: Good coverage provided by administrative data, Milestone of work package 4. http://essnet.admindata.eu