

## **The use of VAT for short term statistics: some quality aspects**

Pieter Vlag and Dirk van den Bergen  
*Statistics Netherlands*

**Abstract:** This paper describes a method to benchmark the input for the flash estimate of the GDP with VAT-turnover. This so-called benchmark method uses the VAT-registration for early estimates of quarterly (VAT-)turnover growth rates. For this estimation the VAT-declarations are divided into eight clusters: 1) four clusters for existing enterprises sub-divided by declaration period (month or quarter) in the current quarter and the corresponding quarter of previous year, 2) two clusters for starting enterprises sub-divided by declaration period in the current quarter, 3) two clusters for stopping enterprises sub-divided by declaration period in the corresponding quarter of previous year. For each group missing data due to the timeliness problem are estimated with the available VAT-turnover data for the current quarter. This approach was tested for the period 2006-2008 by comparing turnover growth rates based on an incomplete VAT dataset and growth rates derived from a complete VAT-dataset. The incomplete VAT dataset is constructed 27 days after the end of the quarter when due to the timeliness problem not all VAT is declared yet. The results do show that this approach can be used for early estimates at high aggregation levels. For this reason, it was decided to use this method from 2009 to benchmark the regular input for the GDP flash estimate, which based on preliminary results of the STS turnover survey. To use as many VAT-data as possible for the early estimates, VAT turnover growth rates are directly based on a copy of the VAT-register of the tax office. As a result the NACE-classification is not perfect for low aggregation levels and this approach cannot be used for estimations at low aggregation levels and regular STS-output.

Currently Statistics Netherlands develops a production system for STS-estimates which is based on the experiences with this benchmarking system.

### **1. Introduction**

Aim of this research was to investigate whether an incomplete VAT-declaration dataset can be used for estimating quarterly turnover growth rates at an early stage. The results of these early “VAT” estimates will be used to check the quality of regular, survey derived input, for the flash estimate of the GDP (gross domestic product)<sup>1</sup>. In other words, aim of this research was to develop a method for benchmarking the early turnover growth rates derived from

incomplete survey data with VAT. It is important to have an independent check about the quality of these survey data. Even under ‘normal’ circumstances the uncertainty around these early quarterly turnover growth rates (and input for the GDP flash estimate) is relatively large, because STS-survey data are still incomplete and possibly even biased. As it could not be excluded that this uncertainty was even larger during the economic turmoil at the end of 2008 and beginning of 2009, a benchmarking system using VAT-declarations was developed.

For the benchmark methodology a prototyping system was built to calculate quarterly turnover growth rates using VAT-turnover data. This prototyping system is based on the results of Vlag, 2008. This prototyping system also is the starting point for a (currently being developed) “STS” production system for estimating quarterly turnover growth rates with VAT-data. One reason is that although initially meant for benchmarking the quality of turnover growth rates derived from surveys for high aggregation levels, this prototyping system also provided useful information about the possibilities for the use of VAT-data for quarterly turnover growth rates.

<sup>1</sup> For the flash GDP estimates (40 days after the end of the quarter), quarterly turnover growth rates have to be provided to National Account 28 days after the end of the quarter.

## **2. Background information**

Some background about availability of VAT-data for statistical purposes in the Netherlands is provided in this chapter.

- a. VAT can be declared on a monthly or quarterly base in the Netherlands.
- b. Extremely small enterprises may declare their VAT on a yearly base. However, taking into account their low contribution to total turnover levels ( $< 0,3$  % for all commercial enterprises in 2007); their contribution to early turnover growth rates can be neglected.
- c. VAT has to be declared electronically within 30 days after the reference period (month or quarter).
- d. In practice, most enterprises declare their VAT 20-30 days after the reference period. Generally, one may assume that 27 days after the end of the reference periods 60-70 % of total VAT-turnover has been declared. Exact percentages depend on enterprise group and the position of weekends/holidays when the declaration have to be made.
- e. VAT data are electronically transferred from the tax office to Statistics Netherlands on a daily base. Hence, we may assume that for most enterprises only a small time shift exists between the VAT-declaration and the availability of VAT-data for Statistics Netherlands.
- f. Fiscal (VAT) enterprises units do differ from statistical (enterprise) units as derived from the business register. Generally we could only match about 70 % of the fiscals unit to the statistical units in 2006-2007. On the other hand we may assume that all commercial enterprises declare VAT (this assumption was also checked by data). Hence, the “STS-SBS” population is completely covered by VAT-registration. Having the aim to use the maximum number of VAT-data for the benchmarking system, it was decided to use fiscal (VAT) units for this system. The business register was only used for attributing NACE-codes to the fiscal units, if possible (see chapter 3.1).

To improve the match between fiscal units and statistical (enterprise) units, Statistics Netherlands implemented a redesign for the business register in January 2010.

- g. In contrast to regular STS-production in future, VAT-declarations are also used for the largest enterprises.

### 3. Methodology: general description

This chapter describes the methodology used for estimating quarterly turnover growth rates with an incomplete dataset of VAT-declarations. Starting point of the method is that an estimate is made for the VAT-turnover level for quarter  $t$ . To determine turnover growth rates, this turnover estimate is divided by the turnover level for the corresponding quarter in last year ( $t-4$ ).

Assuming that the contribution of the yearly VAT-declarations can be neglected, a quarterly VAT-turnover level can be easily calculated by adding VAT-turnovers of all enterprises within an enterprise group. However, an estimation procedure had to be developed for estimates when the VAT data are not complete yet. In this paper, we'll describe such an estimation method for turnover with an incomplete VAT dataset.

For estimating the turnover level for current quarter  $t$ , we decided to impute for 'missing' VAT-turnover data (due to the timeliness problem). For this estimating method, the VAT- declaration period (month or quarter) for an enterprise in the corresponding quarter of last year ( $t-4$ ), was determined in a first step. Then the declaration period of the enterprise in the current quarter  $t$  was predicted. By matching this prediction for quarter  $t$  with the declaration for quarter  $t-4$ , eight independent clusters per enterprise group are created. Estimation for 'missing' VAT-turnover data is based on the following assumption: the turnover growth rate for an enterprise which didn't declare its VAT yet equals to the average VAT-turnover growth rate quarter  $t-4$  - quarter  $t$  within the cluster for enterprises which did already declare their VAT for current quarter  $t$ .

Following this approach a 'complete' VAT-turnover dataset for current quarter is created. This complete dataset is based on 1) available VAT-turnover data and 2) estimated VAT-turnover values. A turnover level for quarter  $t$  is estimated by adding the available turnover data with these estimations. Turnover growth rates are calculated by dividing these estimated turnover levels for quarter  $t$  with the turnover levels for quarter  $t-4$ . The latter turnover levels are only based on 'real' VAT-data because all VAT is already declared for  $t-4$ .

Quality of the estimated turnover level (and growth rate) for quarter  $t$  is, is tested by recalculating the turnover level for quarter  $t$  when all VAT-data are available.

This method was tested for the period 2006-2008. It is used for benchmarking the regular "STS turnover" input for the GDP flash estimate from the first quarter of 2009.

For the benchmarking the regular survey STS-input for the flash estimate of the GDP, we use monthly and quarterly VAT-data which are available 27 days after the end of quarter. This is because Statistics Netherlands starts calculating the GDP flash estimate 28 days after the end of the quarter and the VAT-data processing takes one day. Note that 27 days after the quarter the VAT-data are still incomplete, because monthly quarterly VAT-declarations and VAT-declarations of the 3<sup>rd</sup> month of

the quarter might be done until the end of the month (30-31 days after the end of the quarter). Monthly VAT-data of the 1<sup>st</sup> and 2<sup>nd</sup> month of the quarter are complete when the ‘benchmark’ estimates are made.

In the next chapters of this paper, we’ll describe some general characteristics about VAT. Then we’ll describe some the methodology in more detail. In the last part we discuss some results.

#### 4. Methodology (detailed): the VAT-declaration

In the Netherlands three possible VAT rates exist: the 0% rate, the 6% rate or the basic 19% rate. Exempted goods and services do also exist. Enterprises have to declare their turnover for all three VAT-rates (and exempted goods+services). The total VAT-turnover, which is used for statistical purposes, is determined by adding these groups.

After deduction of the input tax (for purchased goods and services by the enterprise), the tax office determines the VAT to paid by the enterprise. Statistics Netherlands receives the complete tax declaration. For this research the only following variables from the VAT declaration are used:

**Table 1** Used variables from the VAT declaration.

Fiscaal nummer (Finr)	Fiscal unit
Volgnummer (VolgNr)	Subdivision of the fiscal unit. The combination Finr + Volgnr is unique
Maandnummer (MaandNr)	Month for which the declaration has been made.
Tijdvaktype	Denotes whether the VAT-declaration covers a monthly (M) or a quarterly (Q) period. Quarterly declaration are administrated as declarations for months 3,6,9,12. Enterprises may change their declaration periods within a calendar year.
Branche	Enterprise activity determined by the tax office (NACE <sub>74</sub> -code).
Datum van aangifte (Reg_datum)	Exact date of VAT-declaration
Eenheid	Department of tax office. Department 136 implicitly denotes that VAT is declared by a foreign company.
Fiscale vergunning (Fv_vergunning)	Fiscal enterprise code. Fv_vergunning = 2 or Fv_vergunning = 3 denotes that VAT has been declared by a foreign company.
BTW omzet (netto_omzet_BTW)	VAT_turnover (in €). Determined by adding the turnovers for the separate VAT rates.

## 5. Methodology (detailed): population and enterprise activity

### 5.1 Foreign enterprises without production in the Netherlands

Foreign enterprises without production in the Netherlands do not belong to the “STS/SBS” enterprise population and should be excluded from the VAT-data. This is implemented by excluding fiscal (enterprise) unit having an *Fv\_vergunning* 2 or 3 (foreign enterprises: VAT-declaration has been made by a fiscal representative in the Netherlands) or having an *Eenheid* 136 identification (foreign enterprises which declared their VAT to a special department of the VAT-office) from the estimates.

To obtain consistency with other input sources for the GDP flash estimate, fiscal (enterprises) units with only transit trade in the Netherlands were also excluded from the estimates. This was implemented by using a list of the 1000 largest ‘transit trade enterprises’ in the Netherlands (from the International Trade Statistics Department).

In practice, excluding foreign enterprises and enterprises with only transit trade appeared to be a complicated step with some arbitrary decisions. For example, foreign enterprises having *Eenheid* 136 appeared to have a total VAT-turnover of more than 1 milliard euro per quarter. However, this group is not stable in time. Hence, excluding the up-to-date group of “*Eenheid* 136” enterprises for each quarter led to scattered quarterly turnover growth rates. To avoid this problem, we arbitrarily decided to construct a list of “*Eenheid* 136” enterprises which is based on average results for the period 2006-2008. Concerning fiscal (enterprises) units with only transit trade, we arbitrarily decided to include enterprises below the “TOP-1000” in the estimates because 1) no stable procedure for detecting these enterprises could be developed and 2) the contribution of smaller – formally Dutch – ‘transit trade companies’ to total transit trade in only 10 % in terms of turnover.

Summarizing arbitrary decisions had to be taken when excluding foreign enterprises without production and ‘transit trade’ enterprises from the statistical population frame. It was decided that this point is subject for further research when developing a regular production system for STS-estimates with VAT.

### 5.2 Enterprise activity (NACE-classification)

National Accounts still used the NACE<sub>93</sub> classification in 2009. Hence, to use VAT-data for benchmarking the regular STS-input for the flash GDP estimate, one has to attribute a NACE<sub>93</sub> code to each fiscal unit. This attribution (or matching procedure) can be divided into three steps.

**STEP 1:** All enterprises (and VAT-declarations) do have a NACE<sub>74</sub> code attributed by the tax office. By using a NACE<sub>74</sub> – NACE<sub>93</sub> transformation table, a NACE<sub>93</sub> code is added to each (enterprise with a) VAT-declaration. This NACE<sub>93</sub> code is used in the statistical process.

Advantages of this approach are that 1) the tax office checks the enterprise activity regularly and 2) a NACE<sub>93</sub> code can be assigned to all fiscal units.

Disadvantages of this approach are that the NACE<sub>74</sub> – NACE<sub>93</sub> transformation may be ambiguous. Especially the case for the services sector the NACE<sub>74</sub> classification might be outdated and matches in

some cases to several NACE<sub>93</sub> codes <sup>1</sup>. A second disadvantage is that – for several reasons – the tax office may classify an enterprise activity differently than the business register from Statistics Netherlands. As the business register is used for the STS-survey, these classification differences lead to different results between the STS-survey results and the VAT-results for especially the lower aggregation levels.

Taking into account these advantages and disadvantages, we decided to use the NACE<sub>93</sub> code derived from the enterprise activity determined by the tax office apart from the following cases:

**STEP 2:** A NACE<sub>93</sub> code has been added manually for the 1000 fiscal enterprise units with the highest turnover in 2008 (the so-called TOP-1000) <sup>2</sup>. This NACE<sub>93</sub> code overwrite the results of the first step (=STEP 1).

**STEP 3:** If the fiscal enterprise unit matches unambiguously with a statistical unit from the business register, the NACE<sub>93</sub> code is attributed to the fiscal unit (and overwrites the results of STEP 1).

## 6. Methodology (detailed): clusters

### *Estimation procedure*

After determining the enterprise activity and excluding the foreign plus ‘transit trade’ companies, turnover estimates for current quarter t are made by imputing turnover values for ‘missing’ enterprises. The general characteristics of the estimation procedure have been described in chapter 3. In this chapter we’ll describe the first step of this estimation procedure in more detail.

When benchmarking the input for the GDP flash estimate at day 27 after the end of the quarter, the VAT-dataset is still incomplete for a) the quarterly declarations and b) the 3<sup>rd</sup> month of the monthly declarations. For example, table 2 shows the available VAT-turnover data for the early estimation for the first quarter of 2009.

**Table 2** Available VAT-data at April 27, 2009 as percentage of total turnover (determined 15 May 2009 when all VAT-data for the first quarter of 2009 were available). Total VAT-turnover covered by monthly declaration was about 280 mld € for the 1<sup>st</sup> quarter of 2009. Quarterly VAT-turnover declarations covered only 30 mld €.

	<b>January 2009</b>	<b>February 2009</b>	<b>March 2009</b>
Quarterly VAT-declaration	-	-	59,4 %
Monthly VAT-declaration	99,5 %	98,0 %	72,9 %

<sup>1</sup> When this was the case we arbitrarily decided to match the NACE<sub>74</sub> to one NACE<sub>93</sub> code. The one with the likely matching in 1993.

<sup>2</sup> As fiscal units may differ from statistical units, this manually attributed NACE<sub>93</sub> code may still differ from “corresponding” NACE<sub>93</sub> code in the business register.

On the other hand one may conclude that (up to 2009) monthly VAT-declaration cover most turnover and monthly VAT-data are approximately complete for the first two months of the quarter.

An estimation procedure was developed for the ‘missing’ VAT-turnover data for the third month of the quarter and the quarterly declarations. For this estimation period the fiscal units are in a first step divided by enterprises activity. As aggregation level the so-called “regkols” levels of the National Account system is used. These levels generally correspond with NACE 2-digit levels). In a next step these so-called enterprise groups are subdivided into eight clusters. For these clusters, the declaration period of an enterprise for current quarter  $t$  is predicted into a first step. These four declaration periods for current quarter  $t$  are:

- a monthly VAT-declaration for an existing enterprise
- a quarterly VAT-declaration for an existing enterprise
- a monthly VAT-declaration for a new enterprise
- a quarterly VAT-declaration for a new existing enterprise

In a next step these predicted VAT-declaration period for quarter  $t$  are combined with the known VAT-declaration period for quarter  $t-4$ . Eight clusters are formed after combining the declaration periods for quarters  $t-4$  and  $t$  (tab. 3). These eight clusters can be distinguished into:

- four clusters with existing enterprises in quarter  $t-4$  and quarter  $t$ .
- two clusters with new enterprises in quarter  $t$  (compared to quarter  $t-4$ ).
- two clusters with stopped enterprises in quarter  $t$  (compared to quarter  $t-4$ ).

VAT-turnover for missing enterprises in the clusters with existing enterprises is estimated by assuming that the turnover growth rate for a ‘missing’ enterprise equals to the average VAT-turnover growth rate  $t-4 - t$  for enterprises which did already declare their VAT. VAT-turnover for missing enterprises in the clusters with new enterprises is estimated under the assumption that the turnover level for missing enterprises equals to the average VAT-turnover level for enterprises which did already declare their VAT in quarter  $t$ . VAT-turnover for missing enterprises in the clusters with stopped enterprises equals to 0. The estimation procedure is described more precisely in chapter 7.

Note that enterprises which changed their activity between quarter  $t-4$  and quarter  $t$  are considered as new enterprises for quarter  $t$  (tab. 3).

Also note that if an enterprise has already declared its VAT when the estimate is made, its assignment to one of clusters is obvious (tab. 3). If the VAT-declaration is missing, its assignment is still predicted (tab. 3).

The number of stopped enterprises (clusters  $M_x - Q_x$ , tab. 3) is slightly underestimated, because it is assumed that

- no enterprises with a quarterly VAT-declaration stop during current quarter  $t$ .
- no enterprises with a monthly VAT-declaration stop in the last two months of current quarter  $t$ .

This slight underestimation for the number of stopped enterprises leads to slight overestimation for turnover (because of missing are incorrectly clusters  $M_x - Q_x$ , tab. 3) is slightly underestimated

**Table 3.** Eight clusters used for estimating of ‘missing’ VAT-turnover.

<b>Cluster</b>	<b>VAT-declaration quarter T-4</b>	<b>VAT-declaration quarter T</b>	<b>Prediction VAT-declaration period (in case of ‘missing’ declaration quarter T)</b>
MM	Month	Month	Monthly declaration: month 3, quarter T-4 Monthly declaration: month 1 or 2, quarter T
MQ	Quarter	Quarter	Monthly declaration: month 3, quarter T-4 Quarterly declaration, previous quarter No monthly declaration: month 1 or 2, quarter T
QM	Quarter	Month	Quarterly declaration: quarter T-4 Monthly declaration: month 1 or 2, quarter T
QQ	Quarter	Quarter	Quarterly declaration: quarter T-4 Quarterly declaration, previous quarter No monthly declaration: month 1 or 2, quarter T
Mx	Month	not-existing	Monthly declaration: month 3, quarter T-4 No quarterly declaration, previous quarter No monthly declaration: month 1 or 2, quarter T
Qx	Quarter	not-existing	Quarterly declaration: quarter T-4 No quarterly declaration, previous quarter No monthly declaration: month 1 or 2, quarter T
xM	not-existing	Month	No declaration: quarter T-4 Monthly declaration: month 1 or 2, quarter T
xQ	not-existing	Quarter	No declaration: quarter T-4 Quarterly declaration, previous quarter No monthly declaration: month 1 or 2, quarter T



On the other hand, we also underestimate the number of starting enterprises (clusters xM – xQ, tab. 3), because it is assumed that

- no enterprises with a quarterly VAT-declaration start during current quarter  $t$ .
- no enterprises with a monthly VAT-declaration start during the last month of current quarter  $t$ .
- starting enterprises from the 2<sup>nd</sup> month of current quarter  $t$  don't have turnover (implicit assumption).

In chapter 7, it is discussed whether underestimations for starting and stopping enterprises average out.

## 7. Methodology (detailed): estimation

After determining the clusters, the missing VAT-turnover for month 2 and 3 (and for the quarter in case of quarterly VAT-declarations) is estimated<sup>3</sup>. This estimating method is described in this chapter.

### 7.1 Cluster MM

The VAT-turnover for 'missing' enterprises in cluster MM is determined by the VAT-turnover of these enterprises in quarter  $t-4$  multiplied by the average VAT-turnover growth rate of the enterprises within the same cluster which did already report their VAT. In equations:

$$O_{m,h,MM(t)}^{*t} = g_{m,h,MM(t)}^{t,t-1} \sum_{i \notin v_t} O_{m,h,MM(t),i}^{t-1} \quad (1)$$

$$g_{m,h,MM(t)}^{t,t-1} = \frac{\sum_{i \in v_t} O_{m,h,MM(t),i}^t}{\sum_{i \in v_t} O_{m,h,MM(t),i}^{t-1}} \quad (2)$$

$O_{m,h,c(t)}^{*t}$  denotes estimated turnover for month  $m$ , current year  $t$  for 'missing' enterprises in cluster  $c(t)$  of enterprise group  $h$ ,

$O_{m,h,c(t),i}^t$  denotes reported VAT-turnover for month  $m$ , current year  $t$  for enterprise  $i$  in cluster  $c(t)$  of enterprise group  $h$ ,

$v_t$  denotes the enterprises which declared already their VAT.

### 7.2 Cluster QQ

---

<sup>1</sup> Monthly VAT-declarations for month 1 of current quarter  $t$  are almost complete when the flash estimates have to be made. Analyses for the period 2006-2008 revealed that on average 0,5 % of total VAT-turnover for the 1<sup>st</sup> month is missed. This 'missed' turnover is fairly constant for all quarters. Therefore, it was decided to correct for this 'missing' turnover by multiplying the reported VAT-turnover for the 1<sup>st</sup> month of the quarter with a factor 1,05 (=5 %). As the assignment to the clusters is more complex for the first month of a quarter, it was decided not to use the estimating procedure for the 2<sup>nd</sup>, 3<sup>rd</sup> month and quarterly declarations.

The VAT-turnover for ‘missing’ enterprises in cluster QQ is determined similarly. Compared to equations (1) and (2) only months  $m$  need to be changed quarters  $k$ :

$$O_{k,h,KK(t)}^{*t} = g_{k,h,KK(t)}^{t,t-1} \sum_{i \notin v_t} O_{k,h,KK(t),i}^{t-1} \quad (3)$$

and

$$g_{k,h,KK}^{t,t-1} = \frac{\sum_{i \in v_t} O_{k,h,KK,i}^t}{\sum_{i \in v_t} O_{k,h,KK,i}^{t-1}} \quad (4)$$

### 7.3 Cluster MQ

Theoretically, the VAT-turnover for ‘missing’ enterprises in cluster MQ can be estimated by equations (3) en (4). When implementing this system we chose, however, for a simplification to save computing time. The (simplified) assumption is that the turnover coverage of early and late VAT-declarations is similar for quarter  $t-4$  and quarter  $t$ :

$$\frac{\sum_{i \in v_t} O_{n,h,MK(t),i}^{t-1}}{\sum_{i \in v_t} O_{k,h,MK(t),i}^{t-1}} \approx \frac{\sum_{i \notin v_t} O_{n,h,MK(t),i}^{t-1}}{\sum_{i \notin v_t} O_{k,h,MK(t),i}^{t-1}} \quad (5)$$

with  $n$  the third month of the corresponding quarter:

$$n = \max(m : m \in k) \quad (6)$$

Then VAT-turnover for ‘missing’ enterprises in cluster MQ can be estimated by

$$O_{k,h,MK(t)}^{*t} = g_{k,n,h,MK(t)}^{t,t-1} \sum_{i \notin v_t} O_{n,h,MK(t),i}^{t-1} \quad (7)$$

$$g_{k,n,h,MK(t)}^{t,t-1} = \frac{\sum_{i \in v_t} O_{k,h,MK(t),i}^t}{\sum_{i \in v_t} O_{n,h,MK(t),i}^{t-1}} \quad (8)$$

As quarterly VAT-turnover can be simply calculated by adding the reported monthly VAT-turnover data for quarter  $t-4$ , one may improve the methodology by using equations (3) and (4) instead of (7) and (8). This improvement is currently necessary, because the importance of cluster MQ considerably increased due to a recent legislation change.

### 7.4 Cluster QM

The VAT-turnover for missing enterprises in cluster MQ is also estimated by equations (9) and (10) and similarly to clusters MM and QQ:

$$O_{m,h,KM(t)}^{*t} = g_{m,k,h,KM(t)}^{t,t-1} \sum_{i \notin v_t} O_{k,h,KM(t),i}^{t-1} \quad m \in k \quad (9)$$

$$g_{m,k,h,KM(t)}^{t,t-1} = \frac{\sum_{i \in V_t} O_{m,h,KM(t),i}^t}{\sum_{i \in V_t} O_{k,h,KM(t),i}^{t-1}} \quad (10)$$

### 7.5 Clusters Mx and Qx

Assumption is that enterprises in this cluster do not exist anymore. Hence, the estimated turnover is 0.

### 7.6 Cluster xM

The VAT-turnover for missing enterprises in cluster xM is estimated by using the reported VAT-turnover for the 1<sup>st</sup> month in current quarter  $t$ . Main assumption is that the contribution to the total VAT-turnover of missing enterprises in the 3<sup>rd</sup> month of the quarter is equal to contribution to the total VAT-turnover for these enterprises to the total VAT-turnover in the 1<sup>st</sup> month of this quarter. The estimation is

$$O_{m,h,xM(t)}^{*t} = \left( O_{m,h,MM}^{*t} + O_{m,h,MM}^t + O_{m,h,KM}^{*t} + O_{m,h,KM}^t \right) \frac{O_{p,h,xM(t)}^t}{\left( O_{p,h,MM}^t + O_{p,h,KM}^t \right)} \quad (11)$$

$p$  denotes the 1st month of the current quarter  $t$ .

The same assumption (and estimation) is made for missing enterprises in the 2<sup>nd</sup> month of the quarter. Note that this assumption implicitly implies that starting enterprises from the 2<sup>nd</sup> month of current quarter  $t$  don't have turnover (see chapter 6).

### 7.7 Cluster xQ

The VAT-turnover for missing enterprises in cluster xQ is estimated by using the reported quarterly VAT-turnover in previous quarter for enterprises which did not have a VAT-declaration in quarter  $t-4$  (in the same enterprise group). Basically, the same estimation procedure is used as for missing enterprises in cluster xM. However a correction factor for underestimation is added to this estimation.

As stated in chapter 6 the number of enterprises are underestimated in clusters xQ and Qx. These underestimations for starting and stopping enterprises may average out for VAT-turnover estimates if two conditions are fulfilled:

- The underestimation for the number of starting (xQ) and stopping enterprises (Qx) should be similar. This depends on the business cycle.
- The turnover is similar for the 'underestimated' starting (xQ) and the 'underestimated' stopping (Qx) enterprises. For example, if the average turnover of starting enterprises is generally higher than the average turnover of stopping enterprises then the underestimating for starting enterprises is higher than for stopping enterprises (even in hypothetic the case that the underestimating in terms of number of enterprises is similar for both groups). The effect of the business cycle on this condition is uncertain.

Analyses on complete VAT-data for the quarter in 2007 and 2008 revealed that the average VAT-turnover for the starting enterprises in cluster xQ is a factor 2 higher than the average VAT-turnover

for the stopping enterprises in cluster Qx. Although one should take into account that the turnover for the Qx cluster belongs to quarter  $t-4$  (and turnover growth rates were relative high in 2007-2008), this factor 2 is too high to be related to annual growth. Therefore, it suggests a general higher turnover for ‘starting’ enterprises. This implies that underestimations for starting and stopping enterprises do not average out for VAT-turnover estimates, but leads to an underestimation of the growth rate<sup>4</sup>. It was decided to correct for this underestimation when estimating the ‘missing’ enterprises in cluster xQ.

Based on comparisons between estimates for the aggregated Qx and xQ clusters on incomplete VAT-data and estimates for the aggregated Qx and xQ clusters for complete VAT-data in 2007-2008, it was decided to correct for the underestimation with a factor 7/6. As a result the turnover of missing enterprises in cluster xQ is estimated by:

$$O_{k,h,XX}^{*t} = \frac{7}{6} \left( O_{k,h,KK}^{*t} + O_{k,h,KK}^t + O_{k,h,MK}^{*t} + O_{k,h,MK}^t \right) \frac{O_{k-1,h,XX}^t}{\left( O_{k-1,h,KK}^t + O_{k-1,h,MK}^t \right)} \quad (12)$$

A similar underestimation exists for the xM and Mx clusters. However, as this underestimation is much smaller and assumed to be less than 4 % for these small groups, no arbitrary correction factor for this underestimation was made.

The estimation for starting and stopping Enterprises should be improved. A relatively easy improvement is to use the turnover ratio of starting and stopping enterprises from the previous quarter (based on complete VAT-data) for the estimation of the current quarter. This point is subject for further research when developing a regular production system for STS-estimates with VAT.

### 7.8 Summary

The estimations for each cluster are summarized in table 4.

**Table 4.** Estimation of VAT-turnover for missing enterprises.

Cluster	turnover estimate
MM	Month 1: turnover + 0,5 %. Month 2 + 3: turnover growth rate available VAT-data
MQ	turnover growth rate available VAT-data.
QM	Month 1: turnover + 0,5 %. Month 2 + 3: turnover growth rate available VAT-data
QQ	turnover growth rate available VAT-data
Mx	0

<sup>4</sup> As the underestimation is larger for starting enterprises than for stopping enterprises.

Qx	0
xM	contribution turnover coverage in month 1.
xQ	contribution turnover coverage in previous quarter, multiplied by 7/6

## 8. Results

### 8.1 Comparison early estimates with regular estimates

A comparison between the early estimates (on incomplete VAT-data) and regular quarterly estimates (on complete VAT-data) could be made from the 3<sup>rd</sup> quarter of 2007. Before this period, this comparison is not possible because the VAT-data were not transferred to Statistics Netherlands on a daily base.

On average, the growth rate of the early estimate is approximately 0,5 % lower than the growth rate of the regular estimate (tab. 5,6). This underestimation is mainly caused by the quarterly declarations (tab. 6) and can be explained by several factors:

1. Enterprises which were not active (anymore) during in the most recent quarter tend to declare their VAT earlier. For example in every month or quarter a few enterprises with extremely negative turnover growth declare their VAT before (or just after) the end of the reference period. The number of enterprises which declare VAT-turnover=0 is also relatively in the first two weeks after the end of the declaration period. As a result early VAT-declarations do have a non average lower turnover growth rate than the complete enterprise population.
2. VAT-turnover of new enterprises is underestimated (see chapter 7.7). The correction factor of 7/6 is too low for 2007/2008. However, we have chosen for a 'conservative' correction factor because it could not be excluded that the underestimation for new enterprises might be smaller from the first quarter of 2009 (due to the economic crisis).
3. The effects are probably larger for quarterly VAT-declarations than for monthly declarations due to a more dynamic (fiscal) enterprise population.

**Table 5.** Early (t+27 days) minus regular (t+45 days) turnover growth rate per enterprise sector.

2008 – quarters:	I	II	III	IV
	<i>procentpunt</i>			
Manufacturing industry	0,8	0,2	0,9	-1,8
Energy	0,3	1,1	-1,9	-0,7
Construction	0,8	-0,1	-0,3	0,9
Motor vehicles trade	-4,3	3,7	-1,3	-3,9
Wholesale trade	0,4	0,3	-0,6	-0,5
Retail trade	-0,9	-0,2	-0,3	-0,8
Hotels and restaurants	-0,7	0,1	-1,0	0,0
Transport and communication	-1,9	-1,1	-2,3	0,3
IT-services ; commercial services	0,2	-0,8	-3,8	-2,0

**Table 6.** Early (t+ 27 days) minus regular (t+45 days) turnover growth rates, SBS/STS population

	Quarterly decl	Monthly decl	Total
	%		
2007 III	-2,7	-0,3	-0,5
2007 IV	-2,4	-0,6	-0,7
2008 I	-4,1	0,1	-0,2
2008 II	-0,7	0,3	0,2
2008 III	-4,3	-0,3	-0,6
2008 IV	0,1	-0,9	-0,8

## 8.2 Comparison VAT turnover growth rates with regular input (=STS-survey) for flash estimate

In a last step differences in VAT-turnover growth rates and the regular input for the GDP-estimate, derived from (in some cases) incomplete STS-survey date, were compared. Most important finding is that differences between the early and regular VAT-turnover growth rates (tab. 5) are considerably smaller than differences between VAT-turnover growth rates and survey data (tab. 7,8). Hence, differences between VAT-turnover growth rates and survey data depend only for a small part on the time of estimation but are more likely caused by the fact that VAT-turnover growth rates do differ from survey derived growth rates. Likely explanations for this observation are:

1. A good comparison between survey results and VAT-data is (even for the highest aggregation levels) impossible: Due to the matching problem, the NACE-codes for the fiscal units differ too much from the NACE-codes for statistical units in the business register. This is likely for some small enterprise sections like “mining and quarrying” and “energy”. However for large enterprise sections this suggestion is weakened by the observation that VAT-turnover growth rates for the manufacturing industry correspond well with STS-turnover growth rates.
2. The VAT-declarations are not realistic, because they contain many errors. This suggestion is, however, weakened by the observation that the VAT-remittance by the enterprises as derived from the VAT-declarations corresponds with the actual received VAT by the Dutch government. The fact that growth rates in VAT-turnover and VAT-remittance are higher than turnover and production growth rates as determined by Statistics Netherlands (tab. 9) weakens the suggestion that VAT-declarations are affected by widespread VAT-fraud.
3. The STS-results are not perfect. It is a matter of debate whether the current STS-estimates (base done a chain index) deal perfectly with changes in enterprise populations. The current STS-system is basically a panel approach and only a few survey data are available for starting enterprises. It is known that VAT-turnover estimates are not perfect for new enterprises, but the same might be true for the current STS-survey. This may lead to underestimated turnover growth rates by the current STS-system. Higher growth rates in VAT-turnover and VAT-remittance enhance this suggestion. This suggestion is also enhanced by the observation that the growth rate in VAT-turnover is

- considerably higher for enterprise sectors with a large number of small enterprises like trade, hotel and restaurants and services (tab. 8)
- comparable with the STS-turnover for enterprise sectors with few small enterprises like industry.

4. The STS-results are not perfect when the flash GDP estimate is made due to incomplete STS-survey data (for example for the services section).

Table 7.VAT-turnover growth rate (early) – regular input for flash estimate GDP

	2007 III	2007 IV	2008 I	2008 II	2008 III	2008 IV
	%point					
-	-28,0	-19,4	-7,9	-19,1	-5,8	-5,2
Manufacturing industry	-0,8	2,2	0,0	-0,1	1,6	-1,4
-	-4,2	-8,5	-0,9	-2,3	0,9	13,0
-	1,4	1,7	-0,7	-0,5	-1,1	-2,2
-	3,5	5,5	-3,0	-1,3	-6,7	-16,2
-	-2,1	0,6	5,8	7,3	4,5	-0,1
--	2,6	5,0	7,2	6,3	3,6	-1,8
-	3,9	3,0	6,1	4,3	2,8	1,8
-	1,4	0,9	2,3	1,8	4,3	2,4
-	-5,8	1,9	-4,4	3,9	5,1	8,5
Commercial services	3,4	0,2	4,2	3,1	3,9	3,4

Table 8.VAT-turnover growth rate (regular) – regular input for flash estimate GDP

	2007 III	2007 IV	2008 I	2008 II	2008 III	2008 IV
	%point					
-	-18,2	-2,2	-1,7	-19,1	-8,9	-5,3
Manufacturing industry	-2,1	1,1	-0,9	-0,4	0,5	0,2
-	-1,2	-15,6	-1,2	-3,6	3,2	13,9
-	2,0	4,0	-1,6	-0,4	-0,7	-3,2
-I	3,1	3,1	1,6	-4,9	-5,4	-12,9
-	0,5	2,9	5,4	6,9	5,2	0,4
-	3,1	5,3	8,1	6,5	4,0	-1,0
-	3,2	4,0	6,9	4,2	3,8	1,8
-	0,0	3,1	4,5	2,7	6,9	2,5
-	-9,1	-2,7	-2,4	4,8	3,4	3,3
Coomercial services	3,5	4,0	4,0	4,1	8,3	5,6

Tabel 9. Ratio VAT-turnover and production in Demand-Supply table (Nat.Accounts), STS/SBS domain, excluding trade.

	2006	2007	2008
	%		
Jaar	112,2	113,5	114,9

## 9. Results 2009

This system was used for benchmarking the input for the GDP flash estimate from the 1<sup>st</sup> quarter of 2009. Most important findings are:

- Early VAT-turnover growth rates are slightly lower (~0,5 %) than regular VAT-turnover growth rates, confirming the results from 2006-2009.
- Especially for manufacturing industry (but also for transportation and wholesale trade) VAT-turnover growth rates correspond with regular STS-input for the flash estimates.
- In spite of the rapid economic decline, the timing of huge changes in STS-turnover growth rates and VAT-turnover growth rates corresponds remarkably well (suggesting that both VAT and survey data are actual).
- The use of VAT-turnover growth rates led to some revisions for services sector.
- In contrast to the period 2004-2008 (economic growth) for which higher VAT- than STS-turnover growth rates were observed, the VAT-growth tend to be lower from the last quarter of 2008 (economic decline). This confirms that Statistics Netherlands did not underestimate the economic decline.

## 10. Conclusion

This report presents a method to estimate quarterly VAT-turnover growth rates using an incomplete VAT-dataset. The incomplete VAT-dataset consists of all monthly and quarterly VAT-declaration up to 27 days after the end of the quarter. To estimate VAT-turnover growth rates the declaration period for the current quarter  $t$  is predicted for each enterprise. Combined with the declaration period in the corresponding quarter of last year, eight clusters are formed. The turnover growth rate for enterprises which did not declare their VAT yet is assumed to be similar to VAT-turnover growth rate for enterprises in the same cluster which declared already their VAT.

For the STS-SBS domain the VAT-turnover growth rates based on incomplete VAT-data are on average 0,5 % lower than VAT-turnover growth rates based on a complete VAT-dataset. This difference is partly related to the estimation for starting and stopping enterprises. The current method needs to be improved for these enterprises. This point is subject for further research.

Another major problem when using VAT-data for turnover estimates is the selection of foreign enterprises without production in the Netherlands and 'transit trade' companies. This point is also subject for further research. The use of VAT-data for STS-estimates can be considerably improved when a better matching between fiscal and statistical units is established in 2010. Without this improvement the quality of the NACE-code for fiscal units is a matter of debate.

Due to uncertainties around the NACE-classification, the current results can only be used for high aggregation levels. Especially for manufacturing industry, VAT turnover growth rates seems plausible and correspond with the regular input for the GDP flash estimate derived from surveys. For other enterprise sectors larger differences between VAT-turnover growth rates and turnover growth rates derived from surveys are observed, but this may also be related to an imperfect quality of the STS-estimates.



Growth rates in VAT-turnover and VAT-remittance are clearly higher than growth rates in production in the demand- and supply table for 2007 en 2008. These differences tend to be highest for enterprise section with many small enterprises suggestion that the current STS-system (survey) underestimates the changes in enterprise population when the economy is growing. Remarkably, VAT turnover growth rates tend to be lower from the last quarter of 2008 when the economic crisis started.

This method is used for benchmarking the input for the GDP flash estimate from the first quarter of 2009. Currently Statistics Netherlands develops a production system for STS-turnover estimates which is based on the experiences with this benchmarking system.