Discussion Paper

Correspondence between survey and admin data on quarterly turnover

The views expressed in this paper are those of the author(s) and do not necessarily reflect the policies of Statistics Netherlands

2017 | 03

Arnout van Delden
Sander Scholtus
February

Content

Summary

Statistics Netherlands (SN) publishes monthly turnover growth rates based on sample survey data and quarterly growth rates based on a combination of Value Added Tax (VAT) and survey data, referred to as census data. Those data are produced for a range of economic sectors and underlying industries. In the near future, outcomes of the monthly growth rates will be benchmarked on the growth rates of the quarterly census data, from 2015 onwards. Preliminary results over 2015 for the industries within the sector Retail trade showed that monthly growth rates were adjusted downwards in the first quarter of the year while they were adjusted upwards in the fourth quarter. In the current paper we investigate to what extent this effect was due to differences in seasonal patterns in the VAT data relative to those in the survey data. The VAT data are known to have reporting patterns that might lead to under reporting in the beginning of the year and over reporting near the end of the year. We used robust linear regression as well as a linear mixture model to describe the relationship between VAT and survey data in dependence of the quarter of the year, for the years 2014 and 2015. We indeed found some evidence for over reporting near the end of the year, but the effects differed between economic sectors and they were stronger in 2015 than in 2014. Correcting the census data for these effects slightly reduced the differences between the survey and census outcomes before benchmarking. Unfortunately, we could not yet explain all the effects from the reporting behaviour. We recommend that two topics should be addressed in the near future. Firstly, we should repeat the research with 2016 data and analyse which groups of units cause these effects. Secondly we should extend the mixture model to account for multiple groups in the population and allow the regression parameters to vary with those groups.

Keywords

Administrative data, businesses, mixture models, robust regression, statistical methods

1. Introduction

Statistics Netherlands (SN) publishes monthly turnover growth rates based on sample survey data, for different economic sectors (Mining and quarrying, Manufacturing, Construction, Retail trade and Import of new cars) and for different industries within those economic sectors (see Table 1). These monthly figures are part of the Short-term statistics (STS). The basic publication of the STS is in the form of a monthly index, where the average monthly index of a reference year is set to 100. Each month the current index is computed as the product of the previous index times an estimated month-on-month growth rate, see Eurostat (2016).

In addition to these monthly STS growth rate estimates, SN also computes quarterly level and growth rate estimates based on census data. This is done within a production system which is referred to as DRT (van Delden and de Wolf, 2013). These census data are based on a combination of administrative and survey data. The administrative data concern Value Added Tax data (VAT) that are used for the small and simple units, the so-called non-topX units. For the larger and more complex units, the so-called topX units, the monthly survey data are re-used and added to a quarterly total.

The quarterly census data are used for two purposes. Firstly, for most of the industries, four quarterly census values are combined into a yearly turnover level estimate. These yearly values are used to benchmark estimates of the so-called structural business statistics. In turn these structural business statistics are used as input into the yearly supply and use tables of National Accounts. Secondly, for some of the industries also quarterly growth rates are computed, in the form of index values, and these values are input for the quarterly supply and use tables.

Table 1 Economic sectors used in the present study

Sector	Abbreviation	Nace code	Number of underlying industries
Mining and quarrying	E(*)	06 - 09; 35, 36 and 38	13
Manufacturing	M	10 - 33	107
Construction	C	41 - 43	19
Retail trade	R	47	55
Import of new cars	I	45111	1

(*) E from "Energy companies"

Quarterly growth rate estimates are not only available from the census data, they can also be derived very simply from the monthly survey estimates. As a consequence of this, two different estimates for the quarterly growth rates are published for those monthly economic sectors: one based on sample survey data and one based on the census data. Large differences between those two types of estimates are undesirable since users may wonder which of those two estimates is the correct one. In the near future, SN therefore aims to benchmark the monthly growth values into the quarterly ones, using a Denton method (Bikker et al., 2013; Denton, 1971). SN intends to do this once a year, after the values of the four quarters of the year are available. The Denton method tries to preserve the original monthly growth rates, with the restriction that the quarterly indices from the sample survey data and those from DRT (the census data) are identical.

Table 2 Original, adjusted figures after reconciliation for Retail trade

Year	Period	Index	Year-on-year growth rates [1]			Period-on-period growth rates		
		diff	orig	adj	diff	orig	adj	diff
2015	M1	-0.5	1.1	0.5	-0.5	-17.5	-17.9	-0.4
2015	M2	-0.5	1.3	0.7	-0.6	-9.3	-9.4	-0.1
2015	M3	-0.3	-0.6	-0.9	-0.3	12.5	12.8	0.3
2015	Q1	-0.4	0.6	0.1	-0.5	-12.7	-13.1	-0.4
2015	M4	0.4	1.4	1.8	0.4	4.9	5.7	0.8
2015	M5	0.6	-1.5	-1.0	0.6	2.6	2.8	0.2
2015	M6	0.6	4.9	5.5	0.6	-0.1	-0.1	0.0
2015	Q2	0.5	1.5	2.1	0.5	11.6	12.7	1.1
2015	M7	0.3	3.1	3.4	0.3	-0.8	-1.1	-0.3
2015	M8	0.1	-1.2	-1.0	0.1	-7.1	-7.3	-0.2
2015	M9	0.3	4.5	4.8	0.3	3.1	3.3	0.2
2015	Q3	0.2	2.2	2.4	0.2	-3.8	-4.0	-0.3
2015	M10	0.8	2.2	3.0	0.8	6.5	7.0	0.5
2015	M11	0.9	-1.6	-0.7	0.9	-6.7	-6.5	0.1
2015	M12	1.3	3.2	4.3	1.1	20.7	20.9	0.2
2015	Q4	1.0	1.3	2.3	1.0	8.1	8.8	0.8
2016	M1	0.3	-1.5	-0.6	0.9	-21.2	-21.8	-0.6
2016	M2	0.5	2.8	4.0	1.2	-5.3	-5.1	0.2
2016	M3	0.7	4.1	5.2	1.1	13.9	14.0	0.1
2016	Q1	0.5	1.8	2.8	1.0	-12.3	-12.6	-0.4
2016	M4	1.0	0.2	0.8	0.6	1.0	1.3	0.3
2016	M5	1.6	-0.4	0.6	1.0	2.0	2.6	0.6
2016	M6	1.9	1.8	3.0	1.3	2.1	2.3	0.3
2016	Q2	1.5	0.5	1.5	0.9	10.2	11.2	1.0

[1] original (orig), adjusted (adj) and difference (diff = adj - orig).

The intention is that 2015 is the first year where the monthly survey data are to be benchmarked to the quarterly census data. Surprisingly, preliminary results of benchmarking the 2015 data showed that, for the majority of the industries, the year-on-year (yoy) growth rates of quarterly turnover from the survey were adjusted downwards in the first quarter of the year and upwards in the fourth quarter of the year. For Retail trade adjustments of yoy growth rates of quarterly turnover for Q1 2015 , …, Q2 2016 were {-0.5, 0.5, 0,2, 1.0, 1.0, 0.9} per cent points, with similar values for the adjustments of the yoy growth rates for monthly turnover (Table 2). The adjustments in Q4 2015 and Q1 2016 are larger than the 95 per cent margins for the yoy growth rate of the monthly turnover for Retail trade (sector total), which are about 0.7 per cent points (Van Delden, 2012; Scholtus and de Wolf, 2011). This raised the question whether there are some systematic effects in the adjusted figures. We suggested two possible explanations for these large reconciliation differences:

- the Denton method might lead to high or low growth rates near the end of the benchmarked time series;
- the seasonal pattern of the census data differs from that of the survey data.

The first possible explanation can be understood as follows. The Denton method aims to maintain the original growth figures while benchmarking to the census data to the four quarters of the year. For many industries in Retail trade in the Netherlands, turnover will be larger in the final quarter than in the third one, due to sales for Christmas and "Sinterklaas", then in the first quarter sales drop again. What might happen is that the positive growth rate from the third to the fourth quarter is overestimated by Denton since data of the next first quarter are not available for benchmarking yet. This property of the Denton method has been described previously by Boot et al. (1967). We believe that it is useful to investigate the impact of this end-effect of the Denton method on the reconciled outcomes, but the present study focuses on the second possible explanation.

This second explanation can be understood as follows. The turnover in the census data for the non-topX units is based on VAT data. Within the VAT data, there are specific reporting patterns. For instance, some of the businesses that report VAT at a monthly frequency in fact deliver eleven four-week values and one eight-week value (Scholtus, 2012). Also, there are specific quarterly reporting patterns, for instance only one of the quarters of the year has a non-zero value ("x,0,0,0", "0,x,0,0", "0,0,x,0" or "0,0,0,x"). Other examples are that one quartely value differs from the other three ("y,x,x,x", "x,y,x,x", "x,x,y,x" or "x,x,x,y"), one quarter with a negative value (e.g. "w,x, -z, y") and four equal values ("x,x,x,x"), Ouwehand, 2010). According to Ouwehand (2010) about ten per cent of the units report VAT according to one of those patterns. At least part of these patterns assume that the yearly value is correct, but not necessarily also the values of the individual quarters.
Likewise, it may well be that a group of units report different non-zero values in all four quarters of the year, with a tendency to underreport in the first quarters of the year and over report in the final quarter to ensure that the yearly value is correct.

We decided to first investigate the possible difference in seasonal pattern of VAT versus survey data, since it suggests that if these seasonal patterns differ and if they are left uncorrected, that may lead to biased outcomes. The aim of the present paper is to investigate whether the relation between VAT and survey turnover (for the non-topX units) depends on the quarter of the year. Although reconciliation effects were prominent for the sector Retail trade, we will also look into the other economic sectors for which monthly STS figures are published, since the reporting patterns in Ouwehand (2010) were found in all economic sectors.

The remainder of the paper is organised as follows. In section 2 we describe the data and methodology we used. In section 3 we present results on the quarterly relation between VAT and survey turnover. Section 4 investigates whether we can explain the results that we found in section 3. Next in section 5 we discuss results, we conclude and we give some recommendations for future research. Appendix A gives formulas of the mixture model that we used and in Appendix B we documented which industries were excluded from our analysis.

2. Methodology

2.1 Data sources

We wish to compare monthly sample survey data with quarterly VAT data for the non-topX units, using 2014 and 2015 data of the economic sectors in Table 1. Both data sets are available at the level of the statistical units, the enterprises. Quarterly turnover data based on the sample survey were linked to the VAT turnover data by using a unique enterprise identification number. Below we will describe the two data sources.

VAT turnover data. We obtained data from the DRT production system (Van Delden and De Wolf, 2013) that is used to process the VAT data. VAT is reported on a monthly, quarterly or yearly basis by fiscal units. For the quarterly DRT estimates the values of the yearly reporters and estimates for non-responding fiscal units are obtained by imputation. In the current study we only included microdata that are fully based on response.

Within DRT, the fiscal units are linked to the enterprises. Not all fiscal units that report VAT can be linked to a unique enterprise. Those enterprises to which the fiscal units cannot be linked uniquely are added to the topX units. In other words, 1-n linkages between enterprises and fiscal units are included unless one of the underlying fiscal units also links to one or more other enterprises. Therefore, VAT of the fiscal units can normally be uniquely linked to all non-topX units. An exception may occur when the relation between an enterprise and a legal unit changes during the months within a quarter.
. The turnover of topX units is a considerable part of the total turnover. The relative turnover contribution of the topX units varied from 88 per cent in Mining and quarrying to 36 per cent in Construction (Table 3).

In the production system, VAT data are automatically corrected for four-week reporters (see section 4.3 for more explanation). Next, turnover at enterprise level is derived from the VAT declarations. Then, the data are cleaned by automatic and manual data editing. So the data should not contain too many errors anymore. At SN, we also have VAT data from before 2014, but we did not use these data in the present study since they have not been processed by the DRT system.

Sample survey data. The monthly survey data are based on a random panel sample of enterprises, stratified by NACE code and size class (based on number of employees). To limit the response burden, not all size classes are included in the samples: a cut-off value is used that varies with economic sector. For the sectors Mining and quarrying and Manufacturing only enterprises with ≥ 20 employees are included, for Construction and for Car import only enterprises with ≥ 10 employees are included, and for Retail trade the sample is taken from enterprises with ≥ 2 employees.

The monthly data are processed in a production system. Within this system missing data are imputed. In our study we only used data that are fully based on response. A preliminary study showed that results on quarterly effects were different when imputed data were included. A first reason for this finding is that imputation values differ from respondent values. A second reason is that imputation values in the DRT production system differ from those in the survey production system, which is due to the availability of more background information (more units) in DRT than in the sample survey and due to differences in the imputation methods.

Likewise to the VAT data, the sample survey data are cleaned by automatic and manual data editing. The production system to process sample survey data has been revised in 2015.

Table 3 Some basic figures concerning the data based on the census data.

Sector	Year	TopX	Turnover (euro)		Number of enterprises	
			(abs $\times 10^9$)	(rel.)	total	imputed
E	2014	Yes	18.6	0.88	202	108
	2014	No	2.6	0.12	2048	124
	2015	Yes	16.0	0.87	216	114
	2015	No	2.4	0.13	2113	98
M	2014	Yes	59.7	0.74	1554	434
	2014	No	21.5	0.26	56572	3262
	2015	Yes	58.6	0.72	1495	352
	2015	No	22.4	0.28	58501	3056
C	2014	Yes	7.2	0.36	534	194
	2014	No	12.6	0.64	143339	5478
	2015	Yes	7.3	0.35	515	177
	2015	No	13.4	0.65	149662	5267
R	2014	Yes	16.2	0.56	322	49
	2014	No	12.9	0.44	110440	6658
	2015	Yes	17.0	0.56	316	40
	2015	No	13.5	0.44	115138	6070
I	2014	Yes	2.3	0.88	15	0
	2014	No	0.3	0.12	128	5
	2015	Yes	2.7	0.86	14	1
	2015	No	0.4	0.14	154	6

We derived the micro data on quarterly turnover of the sample data by adding up the three monthly values. We only included quarterly values for which we had three months of response. That means that the enterprises must exist all three months of the quarter, they must be in the sample for all three months of the quarter and they must have responded all these three months.

Enterprises in the survey that were excluded from the estimation of the growth rates in the production system were also excluded from our analyses. This concerns units with compositional changes (splits or mergers etc.) that have a large impact on the turnover estimates of the domain in which they take part. For instance, when a large enterprise that sells both books and toys with main activity toys is split into a book enterprise and a toy enterprise, turnover in toys may suddenly drop and that in books may increase.

2.2 Motivation for the data selection steps

In the current paper we did not use all data, but we applied a number of selection steps because we had to deal with a number of issues:

1. imputations and compositional changes;
2. highly implausible values in the combined data;
3. base cells where the VAT data cannot reliably be used to estimate target turnover, due to definitional differences;
4. outlying values in the combined data;
5. systematic differences between VAT and survey turnover.

We aimed to use as many data as possible, but data that might lead to incorrect results were left out. Data concerning Issue 1 (see section 2.1), 2 (see section 2.3) and 3 (see section 2.4) might lead to incorrect regressions, we therefore left out those units. The units concerning issue 1 and 2 were selected automatically. Concerning issue 3, we asked staff from the statistical production to appoint for which base cells they considered VAT to be unsuitable for turnover estimation. This selection was done "blinded", i.e. without knowing the outcomes of the regression analyses. Units of issue 4 and 5 were left in; we choose the regression models such as to deal with them, see sections 2.6 and 4.1.

2.3 Treatment of highly implausible values in the combined data set

Let y_i^q denote the survey turnover of quarter q $(q = 1, \cdots, 4)$ for unit i and let x_i^q denote the quarterly VAT turnover for unit i. In the combined data we found some extreme data pairs (x_i^q, y_i^q) like (1.5, 300) and (500, 0.0) (in 1000 euros) that are likely to be erroneous. Analysts found that there were still some so-called `thousand errors´ in the data. Apparently, not all errors have been removed by the data processing steps. Besides these thousand errors there may also be other kinds of errors. For instance a business may report a zero turnover value in the survey when it has ceased to exist, whereas it may still need to pay some money to the tax office.

The tax office may ask this business to use the VAT form for this, although it does not concern turnover of the actual month.

We removed these highly implausible values as follows. We computed the ratio $\max(y_i^q, 1) \,/\, \max(x_i^q, 1)$. When this ratio or its reciprocal was ≤ 0.01 we removed the pairs from the data. Note that the maximum value of 1 ensures that points like (0.0, 500) are removed from the data (otherwise the ratio is not defined). A value of 1 was chosen because in the survey data this was the smallest possible positive value (turnover is provided in thousands of euros).

Note that in Van Delden et al. (2016) those erroneous data pairs were also found but then for comparisons of yearly sample versus VAT turnover. They restricted their analyses to enterprises with a yearly (survey or VAT) turnover of at least 10 000 euro, and all smaller values were removed. We compared results from removing all enterprises with a quarterly turnover of less than 2 500 euro versus including those results, in both cases we applied our ratio function $\max(y_i^q, 1) \,/\, \max(x_i^q, 1)$ to remove erroneous values. When including units with a non-negative turnover smaller than 2 500 the slope estimates and their standard errors were very close to those where those small units were excluded (not shown). The intercepts, when including the small units, were closer to zero which corresponds with what we expected them to be (when there is no special VAT regulation). We therefore decided to include all turnover values ≥ 0.

2.4 Overview of the data selections

Table 4 Average quarterly turnover (in 10^9 euro) and average quarterly number of units in different stages of data selection.

Sec-tor	Year	Total survey data	Total survey data	Valid survey data	Valid survey data	Comb. valid data	Comb. data (valid T & w)	Comb. data (valid T & w & no extr.)
		Turn	Number	Turn	Number	Number	Number	Number
E	2014	19.0	238	1.2	118	113	113	109
	2015	16.6	248	1.3	124	118	118	116
M	2014	73.7	4905	14.0	3359	3190	3189	3163
	2015	73.2	4880	14.2	3329	3166	3164	3135
C	2014	13.3	1776	6.0	1234	1145	1145	1141
	2015	14.1	1698	6.5	1150	1072	1071	1065
R	2014	25.8	6689	8.5	5214	5109	5108	5083
	2015	27.3	6688	8.9	4958	4865	4863	4831
I	2014	2.6	25	0.3	9	9	9	9
	2015	3.1	23	0.4	10	10	10	10

Table 4 gives the average quarterly values of the data after the different stages of the selections described in the section 2.1 and 2.3. The third and fourth column describe the full survey data (turnover and units) that were available, so the industries for which the level and/or growth rate estimates are considered to be implausible (see section 3.1) are included.

The column "valid survey data" stands for non-topX units, for which the quarterly turnover is fully based on observations (no imputations) for units that are in the survey all three months of the quarter and that are used all three months of the quarter in estimations for the short-term statistics (Dutch: conjunctureel). The column "combined valid data" gives the number of units for which both valid survey and valid census data are available. The valid census data concern non-topX units without imputations. The column "combined data (valid T & w)" concerns the number of units with combined valid data and units for which the survey design weights are non-negative. The final column concerns the previous selection after removal of the highly implausible values. We found that selecting non-topX units yielded the greatest reduction in available turnover and in number of units. The removal of the highly implausible values concerned only a limited number of units.

2.5 Regression equations

For most industries within each economic sector, the relationship between VAT turnover x_i^q and survey turnover y_i^q can be described well by a linear regression equation (see Van Delden et al., 2016, and Figure 2 and Figure 3 in this paper). In production, some of the industries were judged not to deliver plausible outcomes for either the level estimates or the change estimates (see Appendix B). Their impact on the estimated regression coefficients is analysed in section 3.1.

All regressions were done separately for each year t and for each sector h. All units from different industries within a sector are pooled. The subscripts t and h are therefore omitted from the notation unless we need to express differences between years and sectors. The simplest linear relationship between y_i^q and x_i^q that we consider, is described by

$$y_i^q = \alpha + \beta x_i^q + \varepsilon_i^q \qquad (1)$$

where the regression coefficients α (the intercept) and β (the slope) do not vary with the quarter of the year (but they do vary by sector and year). We assume (for now) that the ε_i^q are independently and identically normally distributed with mean 0 and variance σ^2. We use the term residuals for the estimates of the errors, $\hat{\varepsilon}_i^q$, that follow from the regression equation as estimated from the observed data. Throughout the paper, a ‘hat’ denotes an estimate. We allow the regression coefficients to vary by year, mainly because the extent of measurement errors may vary from year to year (see later). We included an intercept in the regression, although we would expect the intercept to be (close to) zero. We prefer to estimate this from the data. We refer to equation (1) as model A.

In addition to model A we are interested to test whether the slope of the regression varies with the quarter of the year. Let $\delta_{q*}^q \in \{0,1\}$ be a variable that indicates whether $q = q^*$, with $q^* \in \{2,3,4\}$. We used the following model:

$$y_i^q = \alpha + \left(\beta_1 + d\beta_2\delta_2^q + d\beta_3\delta_3^q + d\beta_4\delta_4^q\right)x_i^q + \varepsilon_i^q \qquad (2)$$

In equation (2) the coefficient α stands for the common intercept, β_1 stands for the slope in quarter 1 and $d\beta_{q*}$ stands for the difference in the slope between quarter q^* and quarter 1. We refer to equation (2) as model B.

Finally we want to verify whether the regression fit would be further improved when not only the slopes vary by quarter of the year, but also the intercepts. This leads us to the model:

$$\begin{aligned} y_i^q = &\left(\alpha_1 + d\alpha_2\delta_2^q + d\alpha_3\delta_3^q + d\alpha_4\delta_4^q\right) \\ &+ \left(\beta_1 + d\beta_2\delta_2^q + d\beta_3\delta_3^q + d\beta_4\delta_4^q\right)x_i^q + \varepsilon_i^q \qquad (3) \end{aligned}$$

In equation (3) the coefficient α_1 stands for the intercept in quarter 1 and $d\alpha_{q*}$ stands for the difference in the intercept between quarter q^* and quarter 1. We refer to equation (3) as model C. Note that model C is almost equivalent to fitting separate regressions for each quarter, the only difference being that in model C we assume that $\text{var}\left(\hat{\varepsilon}_i^q\right) = \sigma^2$ is constant across the quarters within a year.

Let $\boldsymbol{\theta}$ be the vector of regression coefficients to be estimated in equation (1), (2) or (3) and let s^q be the set of observations that is available for the regression of quarter q (in year t for sector h). As in Van Delden et al. (2016), each regression is estimated by a robust version of weighted least squares:

$$\text{Min}_{\boldsymbol{\theta}}\, z = \sum_{q=1}^{4} \sum_{i \in s^q} \hat{g}_i^q (\hat{u}_i^q)^2 \qquad (4)$$

where $\hat{u}_i^q$ is a standardised residual and $\hat{g}_i^q \in [0,1]$ is a weight that accounts for the presence of outliers. Both will be explained further in the next section.

2.6 Standardised residuals

We use weighted least squares to account for three issues: heteroscedasticity, sampling weights and the presence of outliers (see section 2.7).

- *Heteroscedasticity*. Large enterprises tend to have larger residuals than smaller ones. We correct for this by applying the weighting factor $1/\max\left(x_i^q, 1\right)$. The resulting weighted least squares estimates are optimal if we assume that $\text{var}\left(\hat{\varepsilon}_i^q\right) = c^2 x_i^q$.
- *Calibration weights.* In the present analyses we wish to account for the fact that we have a stratified sample of the population and we are interested to describe the relationship at population level. If there are (minor) deviations from a linear relationship, this can be accounted for by including calibration weights, where the smaller units represent more units in the population than the larger ones. Let k stand for the sampling stratum[1], which was determined by the size class and NACE code of the enterprises. Also, let n_k^q be the number of available (non-topX) data pairs (x_i^q, y_i^q) in stratum k for the regression and N_k^q be the total population

[1] The stratum indicator that was available in this study was for most of the cases equal to the actual stratum that was used for the sampling, but to in some cases it was more detailed.

size of non-topX units for quarter q in stratum k. The calibration weights are given by $d_i^q = N_k^q / n_k^q$. Note that these weights are an approximation since n_k^q concerns only units that exist a whole quarter whereas N_k^q also includes units that are born within the quarter.

– *Standardised residuals*. Let $w_i^q = d_i^q / \max(x_i^q, 1)$ be the weight that accounts for both heteroscedasticity and the calibration weight. We now define the weighted residual as $\hat{\tilde{\varepsilon}}_i^q = \sqrt{w_i^q}\,\hat{\varepsilon}_i^q$, where $\hat{\tilde{\varepsilon}}_i^q$ is i.i.d. $N(0, \tilde{\sigma}^2)$. The standardised residual is now given by

$$\hat{u}_i^q = \frac{\hat{\tilde{\varepsilon}}_i^q}{\sqrt{\widehat{\mathrm{var}}\left(\hat{\tilde{\varepsilon}}_i^q\right)}} = \frac{\hat{\tilde{\varepsilon}}_i^q}{\sqrt{\tilde{\sigma}^2(1 - \hbar_{ii})}} \tag{5}$$

– where $\hbar_{ii}$ stands for stands for the leverage, which is ith diagonal element of the hat matrix (see, e.g. Draper and Smith, 1998). The estimation of the variance $\tilde{\sigma}^2$ in equation (5) depended on the method that we used to deal with outliers (see section 2.7). In case of the mixture model, the estimation of $\tilde{\sigma}^2$ is given in Appendix A. In case of the Huber estimator, $\tilde{\sigma}^2$ is estimated in a robust manner as:

$$\hat{\tilde{\sigma}}^2 = \frac{1}{.6745} \text{median} \mid \hat{\tilde{\varepsilon}}_i^q - \text{median}\left(\hat{\tilde{\varepsilon}}_i^q\right) \mid \tag{6}$$

– Note that the constant '0.6745' is needed to provide an unbiased estimate of the standard deviation of the residuals of independent observations from a normal distribution (see Kutner et al., 2005).

2.7 Accounting for outliers

When plotting the data it was clear that there are outliers (see Figure 2 and Figure 3). In terms of the modelling approach, outliers are values that are considered to be correct but their residuals $\hat{\varepsilon}_i^q$ are large. In fact one would expect the turnover value derived from VAT and the turnover measured in the sample survey to be identical when there are no errors. Examples of explanations for those large residuals are (minor) differences in the definition of VAT turnover versus survey turnover, administrative time delays, errors in the delineation of the statistical units, reporting errors and linkage errors. The production staff from which we obtained the data did not have the time to analyse the cause of those large residuals. Therefore, in the remainder of this paper, we handle those outliers by selecting an appropriate model to deal with them.

Those outliers may seriously hamper accurate estimation of the regression coefficients. We compared two methods to deal with those outliers: using a *Huber* estimator and using a mixture model. In both approaches units that are outlying are given a reduced weight $0 < \hat{g}_i^q < 1$ and non-outlying units have a weight $\hat{g}_i^q = 1$.

The Huber estimator (Fox and Weisberg, 2013; Kutner et al., 2005) uses the weight function:

$$\hat{g}_i^q = \begin{cases} 1 \text{ if } |\hat{u}_i^q| \leq \gamma \\ \gamma / |\hat{u}_i^q| \text{ else} \end{cases} \tag{7}$$

where γ is a the tuning constant. We used the value γ = 1.345 which yields a 95 per cent efficiency for data generated by a normal error regression model. Equation (7) means that if the absolute value of the standardised residual is larger than γ the reduced weight is given by $\gamma/|\hat{u}_i^q|$. The Huber weight function is estimated by iteratively reweighted least squares (IRLS).

An alternative to the Huber estimator is to assume that the data are a mixture of two distributions: one for the bulk of the data and one for the outlying values. The idea behind this is that the bulk of the data shows a close correspondence between survey and VAT data and outlying units concern data with (considerable) errors, e.g. linkage or measurement errors. We assumed that the residuals of the two distributions are different while the regression coefficients of both distributions are the same. This leads to a model which is close to that described in Di Zio and Guarnera (2013). In the mixture estimator the residual term, "ε_i^q", in equations (1), (2) and (3) is replaced by "$\varepsilon_i^q + z_i^q e_i^q$", where $z_i^q \in \{0,1\}$ is an indicator with $P\left(z_i^q = 1\right) = \pi$ and $e_i^q \sim N\left(0, (\vartheta - 1)\tilde{\sigma}^2/w_i^q\right)$ is an additional residual that occurs when $z_i^q = 1$.

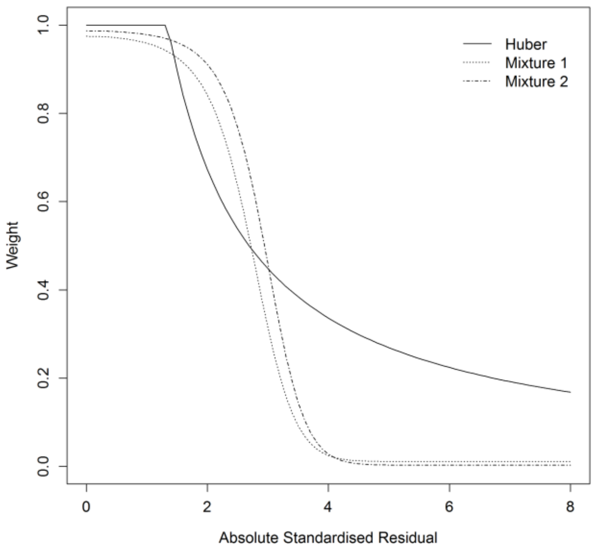

Figure 1. Value of the weight ($\hat{g}_i^q$) as a function of the absolute standardised residual ($|\hat{u}_i^q|$) for the Huber estimator (γ = 1.345) and the mixture model (1: $\pi = 0.2$, ϑ=90, 2: $\pi = 0.2$, ϑ=350).

Appendix A explains that weights $\left(1 - \tau_i^q\right)w_i^q + \tau_i^q w_i^q/\vartheta$ are used in the mixture model. In terms of the standardised residuals $\hat{u}_i^q$, the mixture model can be written as a weighted least squares problem of the form (4) using a weight function

$$\hat{g}_i^q = 1 - \tau_i^q(1 - \frac{1}{\vartheta}) \tag{8}$$

where τ_i^q is the expectation of z_i^q given the covariates for unit i (which includes at least a constant and x_i^q), y_i^q and $\boldsymbol{\theta}$. τ_i^q follows from a mixture of two normal distributions and can be expressed in terms of the standardised residual $\hat{u}_i^q$. The mixture model is estimated by an ECM algorithm using different starting values (see Appendix A). Note that, like $\tilde{\sigma}^2$, the parameters π and ϑ are supposed to be equal for all quarters in a given year and sector, but they can differ between years and sectors.

Figure 1 displays the relation between the weight $\hat{g}_i^q$ and the absolute value of $\hat{u}_i^q$ according to the Huber estimator and the mixture model. For the mixture model we used: $\pi = 0.20$ which was a commonly estimated value for our data and for ϑ we used lower and upper estimates for our data set (see Table 18). Figure 1 illustrates that the drop in weight for increasing values of $\hat{u}_i^q$ is much sharper for the mixture model than for the Huber estimator. In other words: the division of residuals into two groups is more distinct for the mixture model than for the Huber estimator.

2.8 Testing for significance of regression coefficients

The first test that we would like to do is to compare the fit of the different model complexities (A, B,C). We start by testing whether model B is significantly better than model A, thus whether the use of quarterly slopes gives a significant improvement in fit compared to a common slope. Likewise, we want to test whether the use of quarterly intercepts gives a significant improvement in fit compared to a common intercept, given that we account for quarterly slopes (model C versus B).

Let $\boldsymbol{\theta}_C$ denote a sub-vector of length c of regression coefficients for which we would like to test whether their value is zero. For instance, when comparing model B with A we are interested to know whether $d\beta_2 = d\beta_3 = d\beta_4 = 0$ (so $c = 3$), because if so then model B reduces to A. The commonly used F-test cannot be used in this case, because the assumption that the residuals ε_i^q are i.i.d. $N(0, \sigma^2)$ does not hold. Instead, we can used the Wald test, which is in our case based on the statistic

$$W = (\boldsymbol{\theta}_C - \mathbf{0})'\mathbf{V}_C^{-1}(\boldsymbol{\theta}_C - \mathbf{0}) = \boldsymbol{\theta}_C'\mathbf{V}_C^{-1}\boldsymbol{\theta}_C, \tag{9}$$

which follows a chi-squared distribution with c degrees of freedom. Here, $\mathbf{V}_C$ denotes the variance-covariance matrix of the estimated parameters in $\boldsymbol{\theta}_C$.

For the selection between the models A, B and C we used a p-value of the Wald test of 0.01 and smaller. We used this rather small p-value because we only wish to correct the data for quarterly effects when we have strong evidence that these effects exist. In this application, we find incorrectly rejecting the null hypothesis (type 1 error) to be more severe than incorrectly not rejecting the null hypothesis (type 2 error).

After we selected the most appropriate regression model, we tested the significance of individual regression coefficients according to $t = \hat{\theta}_j / \sqrt{\widehat{\mathrm{var}}(\hat{\theta}_j)}$ which follows a Student t-distribution.

2.9 Computations

All computations were performed in R, using the packages lm (linear regression), rlm (robust linear regression) and svyglm (for survey designs). The latter was used to compute the correct standard errors, accounting for the calibration weights. For the estimation of the mixture model, we implemented the ECM algorithm from Appendix A in R.

3. Results

Throughout this section we use model B as the default, since we have no reason to believe that the intercept is affected by seasonal reporting patterns. Furthermore, we also use the Huber estimator as the default, for reasons explained in section 3.2. In the description of the results we use the term mixture estimator rather than mixture model, to avoid confusion with the terms model A, B and C.

3.1 Effect of implausible industries

We investigated the effect of implausible industries with respect to level and change estimates on the estimates of the regression coefficients. We estimated the regression coefficients with and without the implausible industries (Table 5). For the sectors Manufacturing, Construction and Retail trade, the estimated slopes and their standard errors were not greatly affected by implausible industries (compare columns L1C1 with L0C0). The standard errors were slightly larger when implausible industries were removed, probably because there were fewer units left for the analyses. The estimated intercepts were affected somewhat more than the slopes by removing the implausible industries, but the effect was not large.

Two exceptions to these findings were the sectors Mining and quarrying and Car Import. The sector Car Import, which consists of one industry, was judged to have an implausible level estimate. Therefore, regression analysis was not possible when industries with implausible level estimates are removed from the analysis. For Mining and quarrying, regression estimates became unreliable when industries with implausible outcomes for both level and change estimates were removed. The reason is that for Mining and quarrying only the industry codes 06100, 06200 and 3600 were considered to be plausible for both level and change estimates. For these industries only 3 (2014) and 4 (2015) enterprises were available for the analyses which is far too few.

Table 5 Regression coefficients and standard errors for implausible industries on level estimates (L) or for changes (C) included (1) or excluded (0) in the analysis (model B, Huber estimator). (2014)

Sr	Coef	L1C1 Estim	L1C1 se	L1C0 estim	L1C0 se	L0C1 estim	L0C1 se	L0C0 estim	L0C0 se
E	α	1.642	0.566	2.957	1.045	2.616	2.720	22813	2410
	β_1	0.943	0.016	0.916	0.039	0.925	0.020	0.105	0.033
	$d\beta_2$	0.003	0.023	-0.035	0.058	0.025	0.024	-0.045	0.046
	$d\beta_3$	0.016	0.023	-0.012	0.059	0.029	0.027	-0.089	0.056
	$d\beta_4$	-0.032	0.024	-0.038	0.052	0.003	0.031	-0.024	0.042
M	α	0.571	0.568	-1.021	0.593	0.076	0.541	-1.066	0.597
	β_1	0.974	0.002	0.976	0.002	0.975	0.002	0.976	0.002
	$d\beta_2$	-0.004	0.002	-0.005	0.002	-0.004	0.002	-0.005	0.002
	$d\beta_3$	-0.003	0.002	-0.005	0.002	-0.004	0.002	-0.005	0.002
	$d\beta_4$	-0.007	0.002	-0.009	0.002	-0.007	0.002	-0.008	0.002
C	α	10.122	1.962	10.848	2.069	9.776	1.949	10.848	2.069
	β_1	0.949	0.003	0.946	0.004	0.950	0.003	0.946	0.004
	$d\beta_2$	0.006	0.004	0.007	0.004	0.006	0.004	0.007	0.004
	$d\beta_3$	0.001	0.004	0.003	0.004	0.001	0.004	0.003	0.004
	$d\beta_4$	-0.006	0.004	-0.005	0.004	-0.006	0.004	-0.005	0.004
R	α	0.129	0.022	0.013	0.017	0.394	0.052	0.169	0.055
	β_1	0.956	0.002	0.962	0.002	0.948	0.002	0.951	0.002
	$d\beta_2$	0.001	0.002	0.002	0.002	0.001	0.003	0.002	0.003
	$d\beta_3$	-0.001	0.002	-0.001	0.002	-0.002	0.003	-0.001	0.003
	$d\beta_4$	-0.007	0.002	-0.005	0.002	-0.008	0.003	-0.006	0.003
I	α	-648.082	154.968	-648.082	154.968	(*)	(*)	(*)	(*)
	β_1	1.149	0.048	1.149	0.048				
	$d\beta_2$	-0.070	0.050	-0.070	0.050				
	$d\beta_3$	0.034	0.052	0.034	0.052				
	$d\beta_4$	-0.050	0.051	-0.050	0.051				

(*) The sector Car Import consists of one industry. This industry is considered to be implausible for level estimates so no regression coefficients are available for this part of the table.

In the remainder of the paper we will limit the analyses of the sectors to the industries with plausible levels and changes, to avoid any disturbing effects due to implausibility of the VAT data. Only outcomes for the sectors Manufacturing, Construction and Retail trade are available.

Table 5. (cont.) (2015)

Sr	Coef	L1C1 Estim	L1C1 se	L1C0 estim	L1C0 se	L0C1 estim	L0C1 se	L0C0 estim	L0C0 se
E	α	11.447	5.691	19.054	8.488	10.905	17.638	-115.9	97.7
	β_1	1.003	0.011	1.036	0.029	0.981	0.011	0.930	0.057
	$d\beta_2$	-0.034	0.016	-0.075	0.036	-0.011	0.015	-0.010	0.067
	$d\beta_3$	-0.048	0.018	-0.071	0.039	-0.077	0.018	0.003	0.099
	$d\beta_4$	-0.010	0.016	-0.039	0.039	-0.009	0.017	0.010	0.064
M	α	1.115	0.602	-0.165	0.705	0.420	0.721	-0.209	0.706
	β_1	0.971	0.002	0.972	0.002	0.971	0.002	0.972	0.002
	$d\beta_2$	0.002	0.002	0.001	0.002	0.002	0.002	0.001	0.002
	$d\beta_3$	-0.003	0.002	-0.003	0.002	-0.003	0.002	-0.003	0.002
	$d\beta_4$	-0.009	0.002	-0.008	0.002	-0.009	0.002	-0.008	0.002
C	α	10.052	2.207	13.084	2.519	12.631	2.548	13.084	2.519
	β_1	0.959	0.003	0.956	0.004	0.957	0.004	0.956	0.004
	$d\beta_2$	-0.007	0.004	-0.006	0.004	-0.006	0.004	-0.006	0.004
	$d\beta_3$	-0.011	0.004	-0.012	0.004	-0.010	0.004	-0.012	0.004
	$d\beta_4$	-0.015	0.004	-0.015	0.004	-0.014	0.004	-0.015	0.004
R	α	0.179	0.024	0.077	0.020	0.485	0.059	0.340	0.067
	β_1	0.956	0.002	0.965	0.002	0.948	0.002	0.954	0.003
	$d\beta_2$	-0.009	0.002	-0.008	0.002	-0.009	0.003	-0.008	0.003
	$d\beta_3$	-0.006	0.002	-0.008	0.002	-0.006	0.003	-0.008	0.003
	$d\beta_4$	-0.017	0.003	-0.018	0.003	-0.017	0.003	-0.017	0.004
I	α	-267.175	125.000	-267.175	125.000				
	β_1	1.101	0.024	1.101	0.024				
	$d\beta_2$	-0.017	0.032	-0.017	0.032				
	$d\beta_3$	-0.033	0.055	-0.033	0.055				
	$d\beta_4$	-0.084	0.031	-0.084	0.031				

3.2 Selection among model complexities

Results of the Wald test for the Huber estimator revealed that for most of the sectors quarterly effects on the regression coefficients were stronger in 2015 than in 2014 (Table 6). The sector Construction was the only sector where model C (quarterly effects on both the slopes as well as for the intercepts) gave a significant ($p < 0.01$) improvement of fit compared to model B and likewise model B compared to model A. In 2014 model B also tended to be better than model A ($p = 0.012$), and model C was better than model B ($p < 0.01$).

The Wald tests for the mixture estimator and the Huber estimator were in line with each other for Construction and Retail trade, but they differed for Manufacturing. The mixture estimator tended to have smaller estimated standard errors (see Table 7 - 9) than the Huber estimator. It may well be that the estimated standard errors of the regression coefficients for the mixture estimator are an underestimation of the true ones, since in its computation we assumed that the parameters π and ϑ are fixed. In fact we should also account for uncertainty in those parameters. Likewise, we expect that the Wald test values for the mixture estimator are too large (and p values too small). We therefore based the selection among the model forms A, B and C on the Huber estimator.

Table 6 Wald test on models.

Sector	Year	Test	DF	Huber estimator			mixture estimator		
				Wald	p	Selected model	Wald	p	Selected model
M	2014	A→B	10835	15.134	0.002	B	21.480	0.000	C
	2014	B→C	10832	2.928	0.403		79.750	0.000	
	2015	A→B	10738	22.373	0.000	B	7.183	0.066	A
	2015	B→C	10735	2.333	0.506		0.651	0.885	
C	2014	A→B	4296	11.042	0.012	A	11.858	0.008	B
	2014	B→C	4293	14.015	0.003		4.963	0.175	
	2015	A→B	3991	16.211	0.001	C	20.652	0.000	C
	2015	B→C	3988	34.696	0.000		15.663	0.001	
R	2014	A→B	10757	6.369	0.095	A	6.352	0.096	A
	2014	B→C	10754	3.090	0.378		3.450	0.327	
	2015	A→B	10138	24.138	0.000	B	25.544	0.000	B
	2015	B→C	10135	0.153	0.985		1.134	0.769	

3.3 Quarterly effects

The estimates of the regression coefficients for model B and the p-values for their significance confirm that effects in 2015 (Table 8) were clearly larger than in 2014 (Table 7). Furthermore, quarterly effects on the regression coefficients differed per sector:

- *Manufacturing.* In 2014, we found a reduced slope in the second ($p < 0.01$), third ($p < 0.05$) and fourth ($p < 0.01$) quarter for Manufacturing. In 2015, we found a reduced slope in the fourth quarter ($p < 0.01$) only (model B).
- *Construction*. In the sector Construction, we found a reduced slope ($p < 0.01$) in the third and fourth quarter of 2015 in model B. In addition, the intercept in the second and third quarter ($p < 0.05$) tended to differ from that of the first quarter in 2015 (model C). In 2014, the slopes of the second, third and fourth quarter did not differ significantly from that of the first quarter. Turning to the results for model C, we found that the intercept of the third quarter tended to differ from that of the first quarter ($p < 0.05$).
- *Retail trade*. For Retail trade, only quarterly effects on the slope were found in 2015 where the slope in the second, ($p < 0.05$), third ($p < 0.05$) and fourth quarter ($p < 0.01$) were significantly smaller than that of the first quarter.

Table 7 Estimates of regression coefficients (2014, model B).

Sector	coef	Huber estimator			mixture estimator		
		estim	se	p	estim	se	p
M	α	-1.066	0.597	0.074	-4.710	1.145	0.000
	β_1	0.976	0.002	0.000	0.984	0.001	0.000
	$d\beta_2$	-0.005	0.002	0.009	-0.005	0.002	0.002
	$d\beta_3$	-0.005	0.002	0.013	-0.006	0.002	0.000
	$d\beta_4$	-0.008	0.002	0.000	-0.007	0.002	0.000
C	α	10.848	2.069	0.000	1.635	0.800	0.041
	β_1	0.946	0.004	0.000	0.965	0.002	0.000
	$d\beta_2$	0.007	0.004	0.067	0.005	0.003	0.090
	$d\beta_3$	0.003	0.004	0.510	0.001	0.003	0.868
	$d\beta_4$	-0.005	0.004	0.235	-0.005	0.003	0.130
R	α	0.169	0.055	0.002	-0.070	0.046	0.124
	β_1	0.951	0.002	0.000	0.963	0.002	0.000
	$d\beta_2$	0.002	0.003	0.544	0.002	0.003	0.467
	$d\beta_3$	-0.001	0.003	0.754	-0.004	0.003	0.160
	$d\beta_4$	-0.006	0.003	0.070	-0.003	0.003	0.209

The estimated values for π and ϑ for model B with the mixture estimator are given in Table 18 (Appendix A). The value of π of about 0.2 for most of the sector by year estimates indicates that about that about 20 per cent of the data pairs were outlying. The variance of those outlying data pairs was at least a factor ϑ > 90 larger than that of the non-outlying data pairs.

Regressions with the mixture estimator are illustrated in Figures 2 and 3 for Construction and Retail trade in 2015. We selected those two sectors since in those sectors the non-topX units represent a relatively large proportion of the total turnover, so a quarterly effect for the non-topX population has a relatively large impact on the outcomes of the total population. The red points in those figures stand for units with $\tau_i^q > 0.8$, ("clearly outlier"), black points for $\tau_i^q < 0.2$ ("no outlier") and the blue points for units with $0.2 \leq \tau_i^q \leq 0.8$ ("intermediate"). First of all, the figures illustrate that there is "a lot of noise" in the data, from which it is difficult to estimate the correct slope. We further see that the mixture estimator nicely separates outlying points from non-outlying points. Note that there are still extreme points in the figures, such as (202,5831) in Figure 2 but they are within the threshold that was used to remove erroneous values.

Table 8 Estimates of regression coefficients (2015, model B).

Sector	coef	Huber estimator			mixture estimator		
		estim	se	p	estim	se	p
M	α	-0.209	0.706	0.768	-4.715	1.097	0.000
	β_1	0.972	0.002	0.000	0.979	0.001	0.000
	$d\beta_2$	0.001	0.002	0.677	0.002	0.002	0.293
	$d\beta_3$	-0.003	0.002	0.129	-0.002	0.002	0.359
	$d\beta_4$	-0.008	0.002	0.000	-0.002	0.002	0.171
C	α	13.084	2.519	0.000	4.300	1.041	0.000
	β_1	0.956	0.004	0.000	0.975	0.002	0.000
	$d\beta_2$	-0.006	0.004	0.147	-0.006	0.003	0.022
	$d\beta_3$	-0.012	0.004	0.003	-0.008	0.003	0.002
	$d\beta_4$	-0.015	0.004	0.000	-0.012	0.003	0.000
R	α	0.340	0.067	0.000	0.050	0.056	0.375
	β_1	0.954	0.003	0.000	0.961	0.002	0.000
	$d\beta_2$	-0.008	0.003	0.015	-0.008	0.003	0.011
	$d\beta_3$	-0.008	0.003	0.023	-0.007	0.003	0.027
	$d\beta_4$	-0.017	0.004	0.000	-0.016	0.003	0.000

Table 9 Estimates of regression coefficients (Construction, model C).

Sector	Coef	Huber estimator			mixture estimator		
		estim	se	p	estim	se	p
2014	α_1	12.810	3.532	0.000	3.255	1.757	0.064
	$d\alpha_2$	-0.186	6.700	0.978	-3.003	2.593	0.247
	$d\alpha_3$	-9.412	3.969	0.018	-2.475	1.982	0.212
	$d\alpha_4$	5.043	5.557	0.364	2.444	3.019	0.418
	β_1	0.944	0.004	0.000	0.963	0.003	0.000
	$d\beta_2$	0.008	0.007	0.254	0.008	0.004	0.047
	$d\beta_3$	0.010	0.005	0.061	0.003	0.004	0.467
	$d\beta_4$	-0.007	0.006	0.217	-0.006	0.004	0.140
2015	α_1	10.835	3.207	0.001	0.681	1.524	0.655
	$d\alpha_2$	-8.935	3.589	0.013	-0.167	1.620	0.918
	$d\alpha_3$	12.225	5.797	0.035	7.459	2.598	0.004
	$d\alpha_4$	7.841	4.632	0.091	4.630	2.782	0.096
	β_1	0.958	0.004	0.000	0.978	0.002	0.000
	$d\beta_2$	-0.001	0.005	0.923	-0.007	0.003	0.023
	$d\beta_3$	-0.021	0.006	0.001	-0.014	0.004	0.000
	$d\beta_4$	-0.020	0.006	0.000	-0.016	0.003	0.000

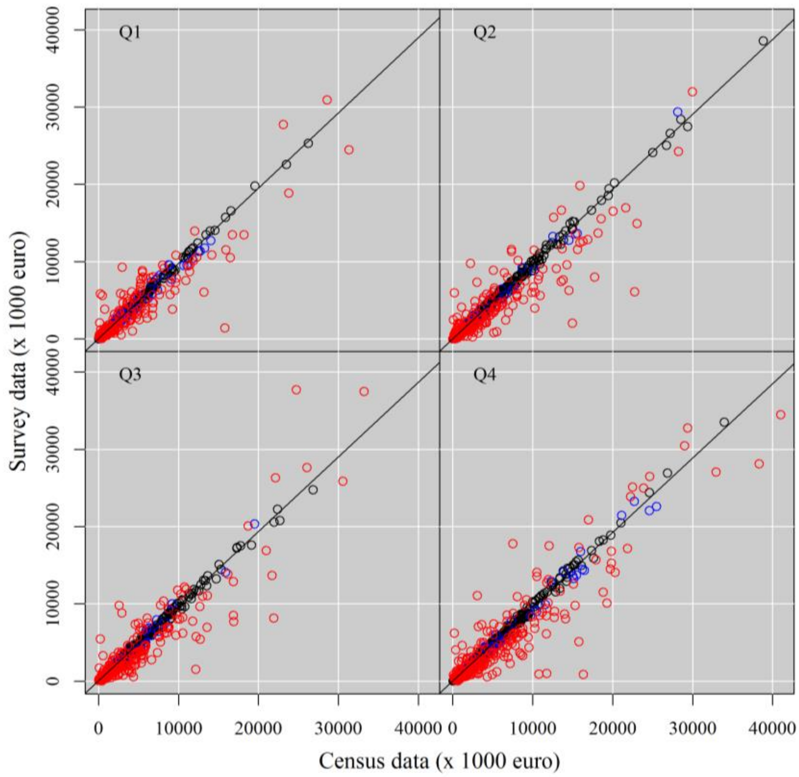

Figure 2. Mixture estimator the sector Construction in 2015 (see text)

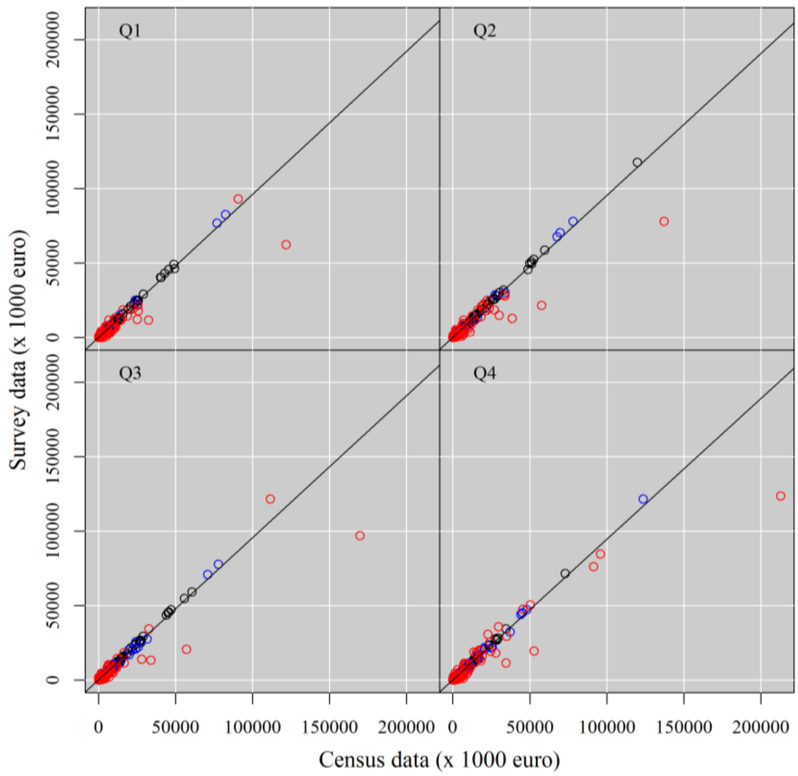

Figure 3. Mixture regressions the sector Retail trade in 2015 (see text).

3.4 Sensitivity on fraction of outlying units

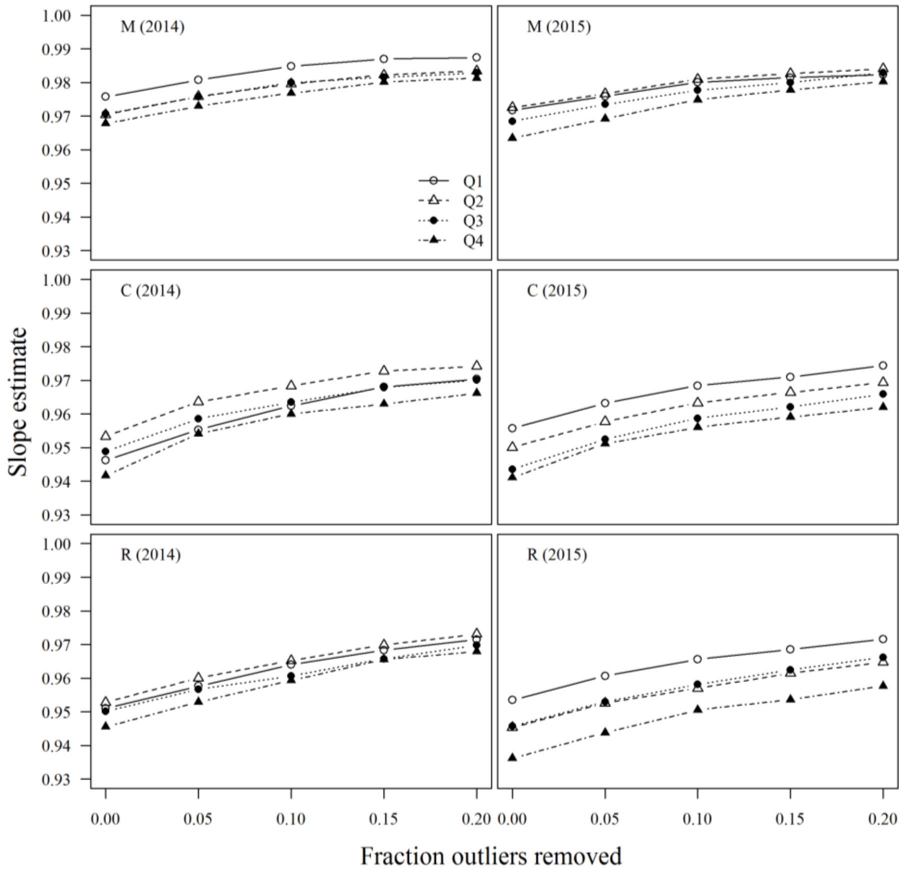

Figure 4. Sensitivity analyses on the fraction of outliers removed (Huber, Model B).

We tested the sensitivity of the quarterly slope (model B, Huber) on removing outliers from the data (Figure 4). We removed subsequently a larger fraction of the most outlying units. We removed outliers up to a fraction of 0.2 since results from the mixture estimator indicated that about 20 per cent of the data pairs were outlying ($\pi \approx 0.2$). With more outliers removed, the slope estimates for all four quarters of the year increased towards a value of 1. In most of the "sector by year" combinations the relative order of the slope estimates was not affected by removal of outliers, but the size of the effect became somewhat smaller with removal of outliers. In all three sectors, the slope of the fourth quarter of the year is the smallest one.

3.5 Adjusted growth rates

We are interested to compute the effect of adjusting the quarterly effects on the estimated qoq-growth rates. In statistical production two types of growth rates are computed:
1. The change between the turnover level estimate of the whole population in quarter q compared to that of the previous quarter $(q-1)$; and
2. The change between the ratio of turnover of a restricted population in quarter q compared to $q-1$. The restriction is that units that are prone to large administrative changes (mergers, splits etc.) are excluded from the analyses. This growth rate is used for the STS, to give an indication for the economic business cycle.
We will refer to the first estimator as the population level (PL) growth rate and to the second estimator as the STS growth rate.

The current study is mainly relevant for the STS growth rate, but its exact computation is done in a separate production system and is not trivial (van Delden et al., 2016). In the present study we therefore used the PL growth rate instead, to study the effect of adjusting the seasonal pattern in the reported VAT turnover on the estimated qoq-growth rates. The outcomes on the basis of the PL growth rates will not be identical to those of the STS growth rates, but we expect that they will be quite close.

To compute the VAT turnover for all units in the population (the PL estimate) that is adjusted for quarterly effects, we used two assumptions:

a. the seasonal distribution of turnover according to the survey data is correct.
b. the total yearly VAT turnover per enterprise is correct but the turnover is distributed incorrectly across the quarters.

We explain the procedure by an example illustrated in Figure 5. This figure displays the estimated linear relation between quarterly VAT and survey turnover. For simplicity, all lines in this figure pass through the origin. In Figure 5 quarterly VAT turnover is over reported in the fourth quarter and under reported in the first three quarters. According to the regression line in Q4 120 euro VAT turnover corresponds with 100 euro survey turnover, and according to the regression lines in Q1-Q3 100 euro VAT turnover corresponds with 100 euro survey turnover. This means that, on average, 105 euro quarterly VAT turnover corresponds with 100 euro quarterly survey turnover (line "average" in Figure 5).

We correct for this shift in quarterly turnover as follows. Suppose some unit i reports quarterly VAT values, x_i^q, of 300, 320, 250 and 360 euro for q= 1, … , 4, with a yearly total of 1230 euro. The expected quarterly survey turnover, denoted by $\hat{y}_i^q$, according to the linear regressions of Figure 5 for $q=4$ is 360 × (100/120) = 300; likewise for $q=1,..,3$ we obtain 300, 320 and 250. We cannot directly use these $\hat{y}_i^q$ figures , since the average quarterly VAT turnover does not have a one-to-one relationship with the survey turnover. Using the assumption a. the relative turnover for $q=4$ is $300/(300+320+250\ +300)=30/117$, according to $\hat{y}_i^q/\sum_q \hat{y}_i^q$. Likewise for q = 2, 3 and 4 we find $30/117$, $32/117$ and $25/117$. Let $\tilde{x}_i^q$ denote the adjusted quarterly turnover. Then, using assumption b. the adjusted quarterly turnover for $q = 4$ is $30/117 \times 1230 = 315{,}4$, according to $\tilde{x}_i^q = (\hat{y}_i^q / \sum_q \hat{y}_i^q) \times \sum_q x_i^q$. For $q = 1{,}2$ and 3 we obtain $315{,}4$, $336{,}4$ and $262{,}8$. Thus we adjusted the VAT turnover of the fourth quarter downwards and those of the other three quarters upwards.

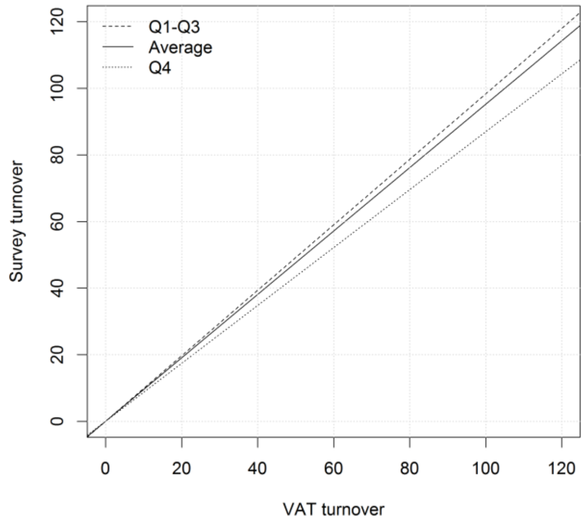

Figure 5. Illustration for seasonal correction of quarterly VAT figures (see text).

The adjusted quarterly turnover is then computed as:

$$\tilde{x}_i^q = \left(\frac{\hat{y}_i^q}{\sum_q \hat{y}_i^q}\right) \sum\nolimits_q x_i^q \tag{10}$$

where the expected quarterly survey turnover, $\hat{y}_i^q$, is computed using the selected models of Table 6 and the corresponding regression coefficients for the Huber estimator of Tables 7 – 9.

We also adjusted the 2014 figures, although in practice SN only plans to reconcile the monthly figures from 2015 onwards. Using the original (x_i^q) and the adjusted ($\tilde{x}_i^q$) turnover level estimates, we computed the original and adjusted quarter-on-quarter (qoq) and yoy growth rates. For the adjusted turnover levels of the census data, we only corrected the non-topX units of the industries that were judged to be plausible. Below the cut-off value of the sample survey, all census data are based on VAT values, so topX units are not observed by primary data collection. So, below this cut-off value all units are corrected when they are part of industries judged to be plausible. For the other units the original values were used. Further, recall that for each of the sectors a cut-off was used for the sample (see section 2.1). In the combined data set that was used to derive the regression coefficients, no data below the cut-off were available. We assumed that the regression coefficients are valid for all non-topX units.

The adjusted growth rates are shown in Table 10. Adjustments for Manufacturing were very small, because that sector has a limited contribution of non-topX turnover to the total turnover and quarterly effects were small. In Retail trade growth rates were corrected downwards in the fourth quarter of 2015, as expected, but also in the second quarter. In Construction, the qoq growth rate in the third quarter of 2015 was adjusted with as much as 7.4 per cent point, which is a very large correction. Also the adjustments in the second (-5.5) and fourth (-3.7) quarter are considerable. One explanation for this large effect is that the non-topX units form a large (65 %) proportion of the total turnover.

Table 10 Original and adjusted growth rates (%), selected models according to Table 6 (Huber).

Sr	Model	Year	Q	qoq			yoy		
				orig	adjust	diff	orig	adjust	diff
M	B	2014	2	2.8	2.7	-0.1			
		2014	3	-2.3	-2.3	0.0			
		2014	4	0.5	0.4	-0.1			
	B	2015	1	-5.0	-4.9	0.1	-4.2	-4.2	0.0
		2015	2	8.4	8.4	0.0	1.0	1.1	0.1
		2015	3	-2.4	-2.5	-0.1	0.9	0.9	0.0
		2015	4	1.4	1.3	-0.1	1.9	1.8	-0.1
C	A	2014	2	20.6	20.6	0.0			
		2014	3	-6.1	-6.1	0.0			
		2014	4	29.7	29.7	0.0			
	C	2015	1	-29.5	-29.4	0.2	3.5	3.8	0.3
		2015	2	25.5	20.0	-5.5	7.7	3.3	-4.4
		2015	3	-7.8	-0.4	7.4	5.7	9.5	3.8
		2015	4	25.6	21.9	-3.7	2.4	2.9	0.5
R	A	2014	2	11.1	11.1	0.0			
		2014	3	-3.0	-3.0	0.0			
		2014	4	8.5	8.5	0.0			
	B	2015	1	-10.7	-10.5	0.2	4.3	4.6	0.3
		2015	2	12.5	12.3	-0.3	5.7	5.7	0.0
		2015	3	-4.1	-4.1	0.0	4.5	4.5	0.0
		2015	4	8.3	7.9	-0.3	4.3	4.0	-0.3

To better understand the large adjustments for the sector Construction, we also computed the adjusted growth rates for model B in 2015 instead of model C , see Table 11. When model B is applied, the adjustments were smaller, but they were still as large as -2.1 per cent in the second quarter and 1.0 per cent in the first quarter. Further inspection of the data showed that Construction consists of many small enterprises. For these small enterprise, the estimated value of the *intercept* has a considerable effect on the actual adjusted turnover value. The problem however is that the intercept is estimated by extrapolation of the regression line, since only data pairs with 10 or more employees were available. Since we excluded all industries which were judged to give implausible level or change estimates, we have no reason to believe that the intercept differs (greatly) from zero. We therefore also computed model B for Construction, but now with the restriction that the intercept equals zero.

The results for model B with $\alpha = 0$ (for the sector Construction) are given in the lower part of Table 11. The adjustments of the growth rates were now much smaller compared to model B with intercept and compared to model C. The largest adjustment of the qoq growth rate was -0.6 %-points in the second quarter of 2015.

Table 11 Original and corrected growth rates (%) for Construction, model B (wit and without intercept) instead of model C for 2015.

Model	Year	Q	q-o-q			y-o-y		
			orig	adjust	diff	orig	adjust	diff
A	2014	2	20.6	20.6	0.0			
	2014	3	-6.1	-6.1	0.0			
	2014	4	29.7	29.7	0.0			
B	2015	1	-29.5	-28.6	1.0	3.5	4.9	1.4
$(\alpha \neq 0)$	2015	2	25.5	23.5	-2.1	7.7	7.4	-0.4
	2015	3	-7.8	-7.2	0.6	5.7	6.0	0.3
	2015	4	25.6	24.1	-1.4	2.4	1.5	-0.9
A	2014	2	20.6	20.6	0.0			
	2014	3	-6.1	-6.1	0.0			
	2014	4	29.7	29.7	0.0			
B	2015	1	-29.5	-29.1	0.4	3.5	4.1	0.6
$(\alpha = 0)^1$	2015	2	25.5	24.9	-0.6	7.7	7.9	0.2
	2015	3	-7.8	-8.1	-0.3	5.7	5.6	-0.2
	2015	4	25.6	25.3	-0.3	2.4	2.0	-0.4

[1] The estimated coefficients were $\hat{\beta}_1 = 0.967$, $\hat{\beta}_1 = 0.959$, $\hat{\beta}_3 = 0.953$ and $\hat{\beta}_4 = 0.949$.

Correcting the DRT figures for seasonal effects not necessarily reduces all reconciliation differences in the survey versus DRT index values. There are also other reasons for differences between those two values, such as sampling errors, the use of a size class cut-off value in the survey sampling data as opposed to observing the whole population in DRT and difference in the selection of units for the STS-growth rate estimates. We therefore compared the reconciliation differences between the indices before and after correcting for seasonal effects, to examine if and to what extend these difference have become smaller after corrections.

Table 12 shows the original survey indices (the release made after the year has ended) and the original DRT indices at economic sector level. In addition the corrected DRT-indices are given, which were computed as the original DRT indices x corrected DRT-turnover level/ original DRT turnover level (the overall correction at sector level). Note that this is an approximation of the actual correction, because in practice not all enterprises are included in the DRT-index computations.

The one-but-last column gives the original reconciliation difference (survey minus the census index) and the final column the reconciliation difference after the census data are corrected. Both for Manufacturing and for Construction, the reconciliation difference became larger in the first two quarters of 2015 and smaller in the third and fourth quarter of 2015. For Manufacturing these two effects nearly cancelled out. For Construction, the average absolute quarterly reconciliation difference was 0.55 before correction and 0.53 after correction. For Retail trade, the largest reconciliation difference was in Q4 of 2015 and it was -1.0 before correction and -0.7 after correction. The reconciliation difference of the first quarter was also smaller after the correction was applied. The average value of the absolute quarterly reconciliation difference was reduced from 0.55 to 0.43 index points.

Table 12 Effect of the correction on the reconciliation difference in Retail trade (2015).

Sector	Quarter	Survey Indices Original (SO)	DRT indices original (DO)	DRT indices Corrected (DC)	SO - DO	SO - DC
M	1	108.2	108.4	108.5	-0.20	-0.26
	2	116.8	117.4	117.5	-0.59	-0.69
	3	112.4	113.7	113.7	-1.30	-1.29
	4	112.6	114.1	114.0	-1.49	-1.34
C	1	81.6	81.6	82.1	0.05	-0.46
	2	102.3	102.7	102.8	-0.36	-0.51
	3	93.9	94.1	93.9	-0.20	-0.03
	4	117.3	118.9	118.4	-1.61	-1.12
R	1	91.1	90.7	90.9	0.43	0.19
	2	101.7	102.2	102.3	-0.53	-0.55
	3	97.9	98.1	98.1	-0.23	-0.27
	4	105.8	106.8	106.5	-1.00	-0.70

4. Possible explanations for the quarterly effects

In section 1 we expressed the possibility that businesses under report VAT in the first quarter of the year and over report in the final quarter such that the yearly value is correct. Although we did find some quarterly effects of this kind, we also found quarterly effects that we could not explain. In this section we consider three possible causes for the quarterly effects that we have found: (a) the presence of systematic patterns, (b) effects of new units in the observations and (c) effects from four-week reporters.

4.1 Systematic reporting patterns

We investigated whether the quarterly effects could have been caused by the systematic reporting patterns in the census data described by Ouwehand (2010), as mentioned in the introduction. In Ouwehand (2010) these patterns were counted at the level of the reported VAT data. In our study, for ease of computation, we identified systematic patterns directly in the data of VAT turnover per enterprise. One enterprise may consist of one or more VAT-reporting units. So if an enterprise consists of multiple units, and only a part of them has a systematic reporting pattern, or when they have a mixture of different reporting patterns then this might not be detected in the present analysis.

We identified enterprises with a systematic patterns that occurred in 2014 and /or 2015 within the sector Retail trade and eliminated those records from our data set. Next, we re-estimated the models with the Huber estimator.

In the whole population there were 1112 enterprises identified with a pattern, which corresponds to just 1 per cent of the total population. The turnover contribution was even smaller: only 0.02 per cent. Of the total available records for the analyses - 10762 (2014) and 10143 (2015) - only 19 (2014) and 12 (2015) had an identified systematic reporting pattern.

Table 13 Estimates of regression coefficients for Retail trade (Huber, model B) when systematic patterns are included (Pat In) versus excluded (Pat Out).

Year	Coef	Pat In			Pat Out		
		estim	se	p	estim	se	p
2014	α	0.169	0.055	0.002	0.160	0.055	0.004
	β_1	0.951	0.002	0.000	0.951	0.002	0.000
	$d\beta_2$	0.002	0.003	0.544	0.002	0.003	0.567
	$d\beta_3$	-0.001	0.003	0.754	-0.001	0.003	0.746
	$d\beta_4$	-0.006	0.003	0.070	-0.005	0.003	0.075
2015	α	0.340	0.067	0.000	0.325	0.066	0.000
	β_1	0.954	0.003	0.000	0.954	0.003	0.000
	$d\beta_2$	-0.008	0.003	0.015	-0.008	0.003	0.012
	$d\beta_3$	-0.008	0.003	0.023	-0.008	0.003	0.021
	$d\beta_4$	-0.017	0.004	0.000	-0.017	0.004	0.000

The results of the Wald test *with* versus *without* those patterns were nearly identical (not shown). The outcomes of the regression coefficients for model B showed a negligible effect on the estimated slopes and a very small effect on the intercept (Table 13). At least for Retail trade, the observed quarterly effect cannot be explained by the systematic reporting patterns in the census data, for patterns occurring at the level of the enterprise.

4.2 Units that are not present all months of the whole year

So far, we included all data pairs for which the survey data had response in all three months of a quarter. But during the same year not exactly the same set of units is available, this is due to births and deaths in the population and changes in the composition of the sample. We wanted to verify to what extent the quarterly effects that we found are due to these changes in the composition of the data. We compared the quarterly effects for the situation where all three months of response (survey, census) within a quarter should be available (scenario AM3) versus the situation that all twelve months within response should be available (scenario AM12).

For most of the sector x year combinations, the chi-squared values of the Wald tests were smaller for the AM12 than for the AM3 scenario. In other words, quarterly effect are smaller in the AM12 than in the AM3 scenario. According to the AM12 scenario only Manufacturing in 2015 and Construction in 2014 lead to model B at $p < 0.01$ for all other combinations model A (no quarterly effects) is the best model. Note that for Retail trade in 2015 and for Manufacturing in 2014 Model B was only just above at $p = 0.01$.

Table 14 Wald test (Huber weight) for the set of units that report 3 months within a quarter or 12 months of the year.

Sr	Year	Test	AM3				AM12			
			DF	Wald	p	Selected model	DF	Wald	p	Selected model
M	2014	A→B	10835	15.134	0.002	B	9787	10.633	0.014	A
	2014	B→C	10832	2.928	0.403		9784	3.845	0.279	
	2015	A→B	10738	22.373	0.000	B	9247	20.117	0.000	B
	2015	B→C	10735	2.333	0.506		9244	3.960	0.266	
C	2014	A→B	4296	11.042	0.012	A	3731	13.745	0.003	B
	2014	B→C	4293	14.015	0.003		3728	3.547	0.315	
	2015	A→B	3991	16.211	0.001	C	3208	6.578	0.087	A
	2015	B→C	3988	34.696	0.000		3205	5.610	0.132	
R	2014	A→B	10757	6.369	0.095	A	8489	6.911	0.075	A
	2014	B→C	10754	3.090	0.378		8486	2.875	0.411	
	2015	A→B	10138	24.138	0.000	B	6516	10.879	0.012	A
	2015		10135	0.153	0.985		6513	3.503	0.320	

We compared the estimated regression coefficients for model B scenario AM3 versus AM12. For Manufacturing the slope estimates for model B for both scenarios are nearly the same. For the sector Construction the positive value of $d\beta_2$ in 2014 is even more pronounced in the AM12 than in the AM3 scenario. In addition to that there is a decrease in the slope in the fourth quarter in both years, this effect is stronger in 2015 than in 2014, but in 2015 it is less pronounced in AM12 than in AM3. In Retail trade, the quarterly effects in 2014 were very close together for both scenarios, but in 2015 effects were smaller for AM12. In 2015, only a quarterly effect in the fourth quarter of the year was found.

Table 15 Estimates of regression coefficients (model B, Huber weight) when units should report 3 months within a quarter or 12 months of the year (2014).

Sector	coef	3 months			12 months		
		estim	se	p	estim	se	p
M	α	-1.066	0.597	0.074	-4.855	1.776	0.006
	β_1	0.976	0.002	0.000	0.978	0.002	0.000
	$d\beta_2$	-0.005	0.002	0.009	-0.005	0.002	0.024
	$d\beta_3$	-0.005	0.002	0.013	-0.004	0.002	0.032
	$d\beta_4$	-0.008	0.002	0.000	-0.007	0.002	0.002
C	α	10.848	2.069	0.000	11.697	1.745	0.000
	β_1	0.946	0.004	0.000	0.945	0.004	0.000
	$d\beta_2$	0.007	0.004	0.067	0.009	0.004	0.034
	$d\beta_3$	0.003	0.004	0.510	0.003	0.004	0.450
	$d\beta_4$	-0.005	0.004	0.235	-0.005	0.004	0.215
R	α	0.169	0.055	0.002	0.068	0.067	0.305
	β_1	0.951	0.002	0.000	0.955	0.002	0.000
	$d\beta_2$	0.002	0.003	0.544	0.003	0.003	0.339
	$d\beta_3$	-0.001	0.003	0.754	-0.003	0.003	0.342
	$d\beta_4$	-0.006	0.003	0.070	-0.005	0.003	0.152

Table 16 Estimates of regression coefficients (model B, Huber weight) when units should report 3 months within a quarter or 12 months of the year (2015).

Sector	coef	3 months			12 months		
		estim	se	p	estim	se	p
M	α	-0.209	0.706	0.768	-1.099	1.586	0.488
	β_1	0.972	0.002	0.000	0.972	0.002	0.000
	$d\beta_2$	0.001	0.002	0.677	0.001	0.002	0.526
	$d\beta_3$	-0.003	0.002	0.129	-0.003	0.002	0.187
	$d\beta_4$	-0.008	0.002	0.000	-0.008	0.002	0.001
C	α	13.084	2.519	0.000	17.497	1.815	0.000
	β_1	0.956	0.004	0.000	0.953	0.004	0.000
	$d\beta_2$	-0.006	0.004	0.147	-0.002	0.004	0.689
	$d\beta_3$	-0.012	0.004	0.003	-0.007	0.004	0.130
	$d\beta_4$	-0.015	0.004	0.000	-0.009	0.004	0.027
R	α	0.340	0.067	0.000	0.234	0.078	0.003
	β_1	0.954	0.003	0.000	0.956	0.003	0.000
	$d\beta_2$	-0.008	0.003	0.015	-0.001	0.004	0.811
	$d\beta_3$	-0.008	0.003	0.023	-0.002	0.004	0.540
	$d\beta_4$	-0.017	0.004	0.000	-0.011	0.004	0.004

4.3 Four-week reporters

As explained in the introduction, some of the businesses that report monthly turnover for their VAT declarations to the tax office in fact report eleven four-week values and one eight-week value. Scholtus (2012) developed a method to detect those four-week reporters. In practice in all sectors some of the enterprises are four-week reporters, but in some sectors it is difficult to detect them. The method by Scholtus (2012) was tested in a pilot for the sectors Manufacturing, Construction and Retail trade and in the majority of the industries within those sectors four-week reporters were detected (column three in Table 17). In practice, the method to detect four-week reporters is actually used in a much smaller set of industries (columns four and six in Table 17). The reason not to use this method is that the number of enterprises per industry was sometimes too small to produce reliable results, or the estimated effect of the four-week reporters on the outcomes was considered to be too low, or the variation in monthly and quarterly turnover values is so large that it is not possible to detect the four-week reporters reliably.
In principle the eight-week value can be reported in any period of the year and in theory this might have caused part of the difference between the true quarterly seasonal pattern and the reported one. In practice, we think that it is unlikely that they have had a major contribution to the shift in seasonal patterns, because in the current system we already correct for four-week reporters in those industries where their effect is traceable.

Table 17 Number of industries per sector with four-week reporters (see text).

Sector	Number of industries[2] (2012)			Number of industries (2016)	
	Total	four-week reporter in pilot	four-week detection in production	Total	four-week detection in production
E	-	-	-	13	0
M	103	76	12	107	12
C	16	16	0	19	0
R	55	54	13	55	13
I	-	-	-	1	0

5. Discussion

Preliminary results for Retail trade showed that reconciliation of monthly survey data to quarterly census data led to a downwards adjustment of the survey data in the

[2] The classification of a sector into industries in 2016 was slightly different from that in 2012.

first quarter of 2015 and to a large upwards adjustment in the fourth quarter of 2015. This latter adjustment was larger than the uncertainty margin around the survey data. We recommend to explore in the data what kind of reporting behaviour or conditions caused these differences. One possible explanation for this is that the seasonal pattern of the census data might differ from that of the survey data. More specifically, the VAT reporters might underreport in the first quarter of the year and over report in the last quarter. In addition, we should verify whether there are artificial reporting effects in the sample survey data, for instance units that report the same values throughout the year. In the present paper we aimed to investigate whether the relation between VAT and survey turnover (for the non-topX units) was affected by the quarter of the year. We did so for all economic sectors for which monthly values are published, not only Retail trade.

A major issue in this study was the presence of a considerable number of outlying units. In order to avoid that those outlying units would dominate our results we eliminated industries from our analyses for which the level or growth rate estimates were considered to be implausible. In addition, we used two types of robust regressions: a Huber and a mixture estimator. Finally, we verified the robustness of our outcomes by a sensitivity analyses in which we gradually removed more outliers. Because the results of the Huber and the mixture estimator were close together and because the sensitivity analyses showed that the relative quarterly effects remained stable when we removed increasingly more outliers, we are confident that the effects that we found are not due to outliers.

Before discussing the quarterly effects, we first wish to address that the estimated values for the slope(s) of the regressions were clearly below 1 for all quarters of the year; this was also the case when model A was applied (a common slope and a common intercept for all four quarters of the year). This raises the question whether we should correct the VAT values for this reduced value. If this reduced slope is the result of differences in definitions between the survey and the VAT variable, then indeed we might correct for this. In an earlier study, van Delden et al. (2016) linearly regressed yearly VAT turnover (dependent variable) on yearly survey turnover for 2009 and 2010 data. We looked into the slopes of those regressions for a control group without VAT regulations of 68 industries. We found median values of 0.985 (2009) and 0.968 (2010) and mean values of 1.005 (2009) and 0.968 (2010). This implies, that also without effects of differences in definitions, slopes below 1 were found.
It is also a well-known result, that in the presence of measurement errors in the variables x and y, the estimated slope is biased downwards. In that case, it can be shown that the slope can be written as $\beta = \beta_\tau \frac{\mathrm{var}(x_\tau)}{\mathrm{var}(x)}$, where β_τ stands for the slope for the situation without measurement errors, and $\mathrm{var}(x_\tau)$ stands for the variance of x without measurement errors and $\mathrm{var}(x)$ for the variance of x with measurement errors, see Biemer (2011). The sensitivity analyses showed that when 0.20 of the most extreme values were removed the slopes in the first quarter increased by about 0.02 points to 0.97 - 0.98. This "overall" reduction in the slope was larger than the quarterly effect that we found. Note that as long as the reduction in slope is equal from year-to-year, the estimated growth rates are not affected by it. But the reduced

slope might also imply that we have biased level estimates. It is therefore worthwhile, for future research, to investigate whether the reduced slope can be explained by the presence of measurement errors. If this is not the case, we may have to adjust the level estimates in future.

With the current data, we used the survey data as the gold standard for the seasonal effects. We did not check for artificial seasonal patterns in the survey data, such as respondents that fill in nearly the same value each reporting period. If fact we are interested in this study to know the true target turnover values. Scholtus et al. (2015) did an analysis with structural equation models where yearly turnover values from administrative and survey sources were compared and amended with results from a small audit sample. In this audit sample additional editing effort has been done to retrieve the true values. With this approach a correction factor from source to target variable could be derived, and they did not need the assumption that the survey data were the gold standard. It would be rather costly to repeat this analysis on the current quarterly data. An alternative approach might be to clean the survey data first from systematic patterns (if any) and use these cleaned data as the basis for the publications.

Returning to the aim of our paper: we indeed found some indication for over reporting of turnover in the fourth quarter of the year for Construction and Retail trade, but we found this in 2015 and not in 2014. Moreover, we also found effects of over reporting in Construction in the third quarter of 2015 and for Manufacturing in the second and fourth quarter of 2014. We tried to understand these quarterly effects from reporting behaviour. Excluding the systematic reporting patterns described by Ouwehand (2010) could not explain the results. We counted those patterns at the level of the enterprises, and it would be worthwhile to repeat this analysis but then at the level of the VAT-reporting units. Ouwehand (2010) did not examine to what extent the occurrence of systematic reporting patterns varies by economic sector. Beforehand however, we do not have any reason to believe that it varies greatly by economic sector, since we know for instance that four-week reporters occur in all economic sectors. Also, limiting the analyses to only those units that were in our data set for four quarters of a year, could not explain all the results. Overall, the effects were smaller for the AM12 scenario, but the effects in the fourth quarter were still present. It is also unlikely that the results can be explained from the presence of (uncorrected) four-week reporters.

When we correct the census data for the quarterly effects according to the regression analyses that were done in the present paper, the reconciliation differences for Retail trade indeed become smaller, as we demonstrated for the Retail trade in 2015. Effects of the corrections are limited, partly because they are only applied to non-topX units and their contribution varies among the sectors and years from 12 to 65 % in terms of the total turnover. However, we conclude that we should not yet apply an adjustment to the quarterly DRT figures, until we can explain the results that we found. For instance, why did we find significant quarterly effects in 2015 for the sectors Construction and Retail trade but not in 2014?

We believe that two topics should be addressed for research in the near future. Firstly, the study should be repeated for the 2016 figures, to examine whether the same seasonal effects as in 2015 are present. We should combine this with an analysis which set of units causes these effects. We expect that the effects are a combination of (a) a group of units with clear systematic reporting patterns, such as (0,0,0,x), (b) a group of units which do not satisfy one of the known systematic reporting patterns but for which the turnover declaration is postponed towards the third or fourth quarter of the year. The second topic is that we would like to develop an extended version of the mixture model that accounts for different groups in the population. We could allow the slope of the regression coefficients to vary with the groups of the population for instance. Furthermore, we could limit the regressions to model B with an intercept of zero. The latter is useful, since we have no reason to believe that the intercept should differ from zero whereas a non-zero intercept can lead to implausible results when it is based on an extrapolation of the regression line.

References

Biemer, P.P., 2011. Latent Class Analysis of Survey Error. Hoboken, New Yersey: John Wiley and Sons.

Bikker, R., J. Daalmans and N. Mushkudiani, 2013. Benchmarking Large Accounting Frameworks: a Generalised Multivariate Model. Economic Systems Research 25, pp. 390-408.

Boot, J. C. G., Feibes, W. and J. H. C. Lisman, 1967. Further Methods of Derivation of Quarterly Figures from Annual Data, Applied Statistics, 16, 65-75.

Delden, A. van, 2012. Marges voor de kortetermijnstatistieken op maandbasis bij steekproefreductie. CBS Nota (in Dutch).

Delden, A. van and P.P. de Wolf, 2013. A Production System for Quarterly Turnover Levels and Growth Rates Based on VAT Data. Paper presented at the New Techniques and Technologies for Statistics Conference, 5-7 March 2013, Brussels.

Delden, A. van, J. Pannekoek, R. Banning and A. de Boer, 2016. Analysing Correspondence between Administrative and Survey Data. Statistical Journal of the IAOS, accepted for publication.

Denton, F.T., 1971. Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization. Journal of the American Statistical Association 66, pp. 99-102.

Di Zio, M. and U. Guarnera, 2013. A Contamination Model for Selective Editing. Journal of Official Statistics 29, pp. 539–555.

Draper, N.R. and H. Smith, 1998. Applied Regression Analysis. Wiley, New York.

Eurostat, 2006. Methodology of short-term business statistics. Interpretation and guidelines. Publication of the European Community, 2006.

Fox, J. and S. Weisberg, 2013. Robust Regression. Course of the university of Minnesota, http://users.stat.umn.edu/~sandy/courses/8053/handouts/robust.pdf Accessed at 2016-10-28.

Kutner, M.H., C.J. Nachtsheim, J. Neter and W. Li, 2005. Applied Linear Statistical Models. McGraw-Hill International edition.

Little, R.J.A. and D.B. Rubin, 2002. Statistical Analysis with Missing Data (Second Edition). John Wiley & Sons, New York.

Ouwehand, P., 2010. Detectie en correctie van systematische fouten in BTW-aangiften. CBS Nota (in Dutch).

Scholtus, S., 2012. Schatten maandomzetten op basis van vierweeksomzetten. CBS Nota (in Dutch).

Scholtus, S. and P.P. de Wolf, 2011. Steekproefadvies Kortetermijnstatistieken. CBS Nota (in Dutch).

Scholtus, S., B.F.M. Bakker and A. van Delden, 2015. Modelling Measurement Error to Estimate Bias in Administrative and Survey Variables. CBS Discussion paper 2015-17.

Acknowledgements

We thank Arjen de Boer (SN) and Henk Koele for providing the data for the analyses and Quinten Meertens (SN) for counting the systematic reporting patterns in Retail trade. Furthermore, we thank Taco Prins (SN) who contributed to the computations for an early version of this paper. We thank Daniel Oberski (Utrecht University) for explaining that the mixture model is similar to the Huber estimator, since both can be expressed by a similar weight function. Finally, we thank Jeroen Pannekoek and Koert van Bemmel for their valuable comments to an earlier version of this paper.

Appendix A. Mixture model

The general assumption of the following approach is that the relation between quarterly VAT turnover x_i^q and quarterly survey turnover y_i^q can be described by a linear model of the form (1), (2) or (3), but that for some units this relation is distorted, for instance due to linking errors between statistical units and fiscal units or due to systematic reporting patterns in the VAT data. Under this assumption, the data can be modelled as a mixture of two distributions: one distribution which describes 'normal observations' and another distribution for 'unusual observations'. This leads to an alternative method for handling outliers instead of the robust regression approach based on Huber weights.

In this paper, we assume that the parameters of the two distributions are identical except for the variance of the disturbances ε_i^q, which may be different (larger) for 'unusual observations'. This leads to a relatively simple mixture model which has a strong resemblance to the 'contamination model' of Di Zio and Guarnera (2013). More advanced mixture models could also be developed, for instance by allowing the slope and intercept parameters to differ between the two distributions, or by introducing more than two distributions into the mixture. For now, we leave this as a topic for future research.

Models A, B and C given by (1), (2) and (3) are all instances of a general weighted linear regression model of the form $y_i^q = \boldsymbol{b}^T \boldsymbol{\xi}_i^q + \varepsilon_i^q$, with $\boldsymbol{b}$ a vector of regression coefficients, $\boldsymbol{\xi}_i^q$ a vector of covariates for unit i which includes at least a constant and the VAT turnover x_i^t, and ε_i^q a disturbance term which is assumed to be normally distributed with mean 0 and variance $\tilde{\sigma}^2/w_i^q$. For instance, model A is obtained by choosing $\boldsymbol{b} = (\alpha, \beta)^T$ and $\boldsymbol{\xi}_i^q = (1, x_i^q)^T$. The corresponding mixture model is defined as follows:

$$y_i^q = \boldsymbol{b}^T \boldsymbol{\xi}_i^q + \varepsilon_i^q + z_i^q e_i^q, \tag{11}$$

where $z_i^q \in \{0,1\}$ denotes an indicator with $P\left(z_i^q = 1\right) = \pi$ and e_i^q is an additional, normally distributed disturbance with mean 0 and variance $(\vartheta - 1)\tilde{\sigma}^2/w_i^q$ that only affects units with $z_i^q = 1$. It is assumed that ε_i^q, z_i^q and e_i^q are mutually independent.

Under model (11), the density of y_i^q is a mixture of two normal densities:

$$f(y_i^q) = (1 - \pi)\varphi(y_i^q; \boldsymbol{b}^T \boldsymbol{\xi}_i^q, \tilde{\sigma}^2/w_i^q) + \pi\varphi(y_i^q; \boldsymbol{b}^T \boldsymbol{\xi}_i^q, \vartheta\tilde{\sigma}^2/w_i^q), \tag{12}$$

with $\varphi(.; \mu, \sigma^2)$ the density of a normal distribution with mean μ and variance σ^2. Note that, under this model, the variance of the disturbance term for a given unit is inflated by a factor ϑ when $z_i^q = 1$. In comparison to the original model, there are now two additional parameters, namely the 'mixing probability' π and the 'inflation factor' ϑ. As before, we denote the vector of model parameters by $\boldsymbol{\theta} = (\pi, \boldsymbol{b}^T, \tilde{\sigma}^2, \vartheta)^T$.

To estimate $\boldsymbol{\theta}$, we can use maximum likelihood. If the indicator values z_i^q had been available in addition to the observations $(\boldsymbol{\xi}_i^q, y_i^q)$, estimation could have been based on the complete-data log likelihood, which is:

$$\log L(\boldsymbol{\theta}) = C + \log(1-\pi)\sum_{i,q}(1-z_i^q) - \frac{\log\tilde{\sigma}^2}{2}\sum_{i,q}(1-z_i^q) - \frac{1}{2\tilde{\sigma}^2}\sum_{i,q}(1-z_i^q)w_i^q\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2 + (\log\pi)\sum_{i,q} z_i^q - \frac{\log(\vartheta\tilde{\sigma}^2)}{2}\sum_{i,q} z_i^q - \frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{i,q} z_i^q w_i^q\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2.$$

Here, C contains all additional terms that do not depend on unknown parameters.

In practice, z_i^q is unobserved (i.e., we do not know in advance the assignment of units to the two mixture components). Following Di Zio and Guarnera (2013), the parameters can be estimated by means of an ECM algorithm (Expectation - Conditional Maximisation) for maximum likelihood estimation with incomplete data. Instead of $\log L$, we consider the following function in which each z_i^q is replaced by its conditional expectation $\tau_i^q = E(z_i^q|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta})$ under the model:

$$Q(\boldsymbol{\theta}) = C + \log(1-\pi)\left(n - \sum_{i,q}\tau_i^q\right) - \frac{\log\tilde{\sigma}^2}{2}\left(n - \sum_{i,q}\tau_i^q\right) - \frac{1}{2\tilde{\sigma}^2}\sum_{i,q}(1-\tau_i^q)w_i^q\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2 + (\log\pi)\sum_{i,q}\tau_i^q - \frac{\log(\vartheta\tilde{\sigma}^2)}{2}\sum_{i,q}\tau_i^q - \frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{i,q}\tau_i^q w_i^q\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2.$$

Here, n denotes the number of observations for the year and sector at hand. An expression for τ_i^q can be derived from (12) by means of Bayes' rule:

$$\tau_i^q = \frac{\pi\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \vartheta\tilde{\sigma}^2/w_i^q)}{(1-\pi)\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \tilde{\sigma}^2/w_i^q) + \pi\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \vartheta\tilde{\sigma}^2/w_i^q)}. \tag{13}$$

The algorithm iterates between an E step and an M step. The E step involves computing expression (13) for all observations, given the current parameter values in $\boldsymbol{\theta}$. The M step involves estimating new values for $\boldsymbol{\theta}$ by maximising the function $Q(\boldsymbol{\theta})$ (see the next paragraphs). The E and M steps are iterated until the parameter estimates have converged.

To find the maximum of $Q(\boldsymbol{\theta})$, we set the partial derivatives $\partial Q/\partial\pi$, $\partial Q/\partial\boldsymbol{b}^T$, $\partial Q/\partial\tilde{\sigma}^2$ en $\partial Q/\partial\vartheta$ equal to zero. For the above Q function, the resulting set of simultaneous equations cannot be solved analytically. Instead, we take on the more straightforward problem of solving each equation separately, while holding the other parameter values fixed. This leads to an ECM algorithm rather than an EM algorithm. See Little and Rubin (2002) for a discussion of E(C)M algorithms in general.

The M step consists of the following sub-steps:

6. Compute a new value for π:
$$\pi = \frac{1}{n}\sum_{i,q}\tau_i^q.$$
7. Compute new weights $v_i^q = \left(1-\tau_i^q\right)w_i^q + \tau_i^q w_i^q/\vartheta$. Next, compute new regression coefficients $\boldsymbol{b}$ by weighted least squares:

$$\boldsymbol{b} = \left(\sum_{i,q} v_i^q \boldsymbol{\xi}_i^q (\boldsymbol{\xi}_i^q)^T\right)^{-1} \left(\sum_{i,q} v_i^q \boldsymbol{\xi}_i^q y_i^q\right).$$

8. Compute a new value for $\tilde{\sigma}^2$:

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i,q} v_i^q \left(y_i^q - \boldsymbol{b}^T \boldsymbol{\xi}_i^q\right)^2.$$

9. Compute a new value for ϑ:

$$\vartheta = \frac{1}{n\pi\tilde{\sigma}^2} \sum_{i,q} \tau_i^q w_i^q \left(y_i^q - \boldsymbol{b}^T \boldsymbol{\xi}_i^q\right)^2.$$

In each computation, the most recent parameter values are used. (So, e.g., when $\tilde{\sigma}^2$ is computed in sub-step 3, we use the weights v_i^q and regression coefficients $\boldsymbol{b}$ that have just been computed in sub-step 2.)

To initialise the ECM algorithm, starting values for the parameters in $\boldsymbol{\theta}$ have to be chosen. In our application, we used the results of the robust regression approach to find appropriate starting values. We also repeated the algorithm several times with different starting values, to reduce the risk of converging to a local maximum rather than the actual maximum likelihood estimates. We found identical parameter estimates for the different starting values.

The estimates of 'mixing probability' π, the 'inflation factor' ϑ and the standard deviation $\tilde{\sigma}$ are given in Table 18.

Table 18 Final estimates of the parameters π, ϑ and $\tilde{\sigma}$ of the mixture model (model B).

Sector	Year	π	ϑ	$\tilde{\sigma}$
M	2014	0.191	298	12.4
	2015	0.213	353	11.0
C	2014	0.203	206	26.3
	2015	0.272	157	18.0
R	2014	0.186	94	17.1
	2015	0.166	91	25.8

Appendix B. List of implausible industries

Table 19 List with industries with implausible growth rates and with implausible level estimates. NACE(1) codes expressed 5-digit level.

Sector	NACE codes with implausible growth rate	NACE codes with implausible level
Mining and Quarrying	08000, 38300	09100, 09900, 35110X, 35120, 35130, 35140
Manufacturing	14100X, 21100,21200, 29100, 29200, 29300, 30100, 30200, 30300X, 30900, 32100X, 32400X, 32500, 32999X, 33110, 33120X, 33150X, 33200	10130, 11010X, 21200, 23300X, 29200, 32500, 32991
Construction	41100, 42910, 42990, 43110	42910, 42990
Retail trade	47290X, 47525, 47592, 47641, 47716X, 47730, 47742, 47761, 47770, 47789X, 47790, 47810, 47820, 47890, 47990	47110, 47190
Car Import		45111

(1) for the meaning of the NACE codes up to the 4-digit level, see the NACE classification at: http://ec.europa.eu/eurostat/ramon/nomenclatures.

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2015–2016	2015 to 2016 inclusive
2015/2016	Average for 2015 to 2016 inclusive
2015/’16	Crop year, financial year, school year, etc., beginning in 2015 and ending in 2016
2013/’14–2015/’16	Crop year, financial year, etc., 2013/’14 to 2015/’16 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Colofon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Studio BCO

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbsl.nl/information