Discussion paper

Analysing response differences between sample survey and VAT turnover

Arnout van Delden and
Sander Scholtus

May 2019

Content

Summary[1]

For a number of economic sectors, Statistics Netherlands (SN) produces two time series on turnover growth rates of businesses: a monthly series based on a sample survey and a quarterly series based on census data. The census data consist of Value Added Tax data (VAT) for most of the enterprises and of questionnaire data for a limited set of large or complex enterprises. To improve the quality of our output, SN aims to benchmark the monthly time series upon the quarterly one, using a Denton method. However, benchmarking has not been applied so far. One of the reasons is that a previous study on 2014 and 2015 data suggested that the two time series have different seasonal effects: the yearly distribution of quarterly turnover tends to be shifted more towards the fourth quarter of the year for the VAT data than for the sample survey data. In the present study, we aimed to analyse whether the earlier observed differences between the two times series are really due to seasonal differences or not. Further, we wanted to know whether the differences are due to units with specific reporting patterns and if it is caused by a limited number of units. To answer the first aim, we used different models to describe the quarterly relation between sample survey turnover and VAT turnover for 2014 – 2016. The analysis confirmed that the two time series show seasonal differences. We found only minor differences between the results of the different models, from which we conclude that the seasonal effects are not due to modelling assumptions. For the second aim, we classified the units to 81 different yearly reporting patterns. These patterns describe whether VAT is smaller, equal or larger than the sample survey for each of the four quarters of a year, plus an additional pattern. Unfortunately, we could not relate one or more of those reporting patterns to the seasonal effects. Finally, we found that a considerable amount of units contribute to the patterns each year. In the near future we aim to find the causes of those seasonal differences and we will seek background variables by which we can explain those differences. Finally, we aim to define measures such that the current differences in seasonal patterns no longer are an obstacle to benchmark the survey data to the VAT data.

Keywords
Measurement errors, reporting differences, tax data, seasonal patterns

[1] The authors thank the reviewers Koert van Bemmel, Danny van Elswijk and Jeroen Pannekoek for their useful comments to earlier versions of this paper.

1. Introduction

For a number of economic sectors, Statistics Netherlands (SN) produces two turnover time series: a monthly series based on sample survey turnover and a quarterly series based on census data. The census data consist of a combination of Value Added Tax data (VAT) for the smaller and simple enterprises and of survey data for the more complex enterprises. The smaller and simple enterprises are referred to as non-top X units and the more complex ones as top X enterprises. The census data are processed in the so-called DRT system. The sample survey data were processed in the IMPECT2 system until 2014 and in the KICR system from 2015 onwards.

The monthly series is used to publish output for the Short-term statistics (STS), whereas the sum of the quarterly level estimates based on the census data is used to calibrate the outcomes of the annual structural business statistics (SBS). The monthly STS data are used as input for the quarterly national accounts whereas the SBS is used as input for the annual national accounts. Differences between the two time series therefore contribute to differences between early and late releases of the national accounts figures. To improve the quality of our output, SN aims to benchmark the monthly time series upon the quarterly one, using a Denton method (Bikker et al., 2013; Denton, 1971).

SN aims to benchmark the two series from 2015 onwards. However, 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 (see Van Delden and Scholtus, 2017). For Retail trade adjustments of yoy growth rates of quarterly turnover for Q1 2015 up to 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. Since the 95 per cent sampling error margins for yoy growth rates of Retail trade are 0.7 per cent points (Van Delden, 2012; Scholtus and de Wolf, 2011), the adjustments in Q4 2015 and Q1 2016 are larger than this margin. Furthermore, Van Bemmel and Hoogland (2017) found that changes computed from the monthly time series differ systematically from changes computed directly as the ratios of two levels of the census data.
These findings lead to two complementary research questions:
1. What are the reasons for the systematic differences between the sample survey growth rates and the census growth rates?
2. Is the seasonal pattern based on the VAT data different from that based on the sample survey data for enterprises that report to both sources?

The first question is treated by Van Bemmel and Hoogland (2017). They quantified a number of causes for differences between the monthly and the quarterly time series. Some of those differences they identified lead to systematic differences:
– In the survey only enterprises above a certain size class are observed (cut-off sampling) whereas the census data concerns all size classes. More specifically,

new enterprises may enter the population frame at SN in a special size class "00", which refers to enterprises without any working persons. In practice enterprises are often put into this size class because the true number of employees is not yet known. Later, those enterprises move into a larger size class. The monthly survey with cut-off sampling is designed in such a way that this group of units is missed as births in the population, leading to an underestimation of the growth.
– The survey time series uses weights to aggregate from industry level towards economic sector level. Those weights are based on yearly turnover from the structural business statistics (SBS). In the past the final SBS estimates could differ from the estimates of the census data, leading to systematic differences between them. The new weights, that have been determined end of 2017, are based on SBS estimates that are calibrated upon the census data, thus now this problem has been solved.

Furthermore, Van Bemmel and Hoogland (2017) found that for a number of economic sectors, apart from systematic yearly differences, there were also systematic quarterly differences where the quarter-on-quarter growth rate based in the fourth quarter is larger for the census data than for the survey data. These quarterly differences, expressed in the second question, are addressed in the current report.

A first analysis on these quarterly differences has been given by Van Delden and Scholtus (2017). They linked the sample survey data to the VAT data for the smaller and simple units. Imputed values were left out of the analysis. Using a robust linear regression analysis and a mixture model, they found a slightly decreased slope for the relation between turnover in the sample survey data (dependent variable) and turnover in the VAT data (independent variable) in the fourth quarter of 2014 and 2015 and a slightly increased slope in the first and or second quarter. The intercept was not affected by the quarter. Van Delden and Scholtus (2017) first adjusted the seasonal VAT pattern using the results of the slopes, and then estimated what would be its implications on the benchmarking. They found that the downwards adjustment in the fourth quarter of the year was 0.7 instead of 1.0 per cent in the fourth quarter of 2015 for Retail trade. Also in other industries the sizes of the adjustments due to benchmarking were reduced. The seasonal effects found in Van Delden and Scholtus (2017) are relatively small and not entirely consistent over the different industries[2] that were tested: Manufacturing, Construction and Retail trade. We therefore concluded that we wanted to repeat the analysis for 2016 data to be more certain whether there are really seasonal effects.

A first objective of the present paper is to fine-tune the exact model that is used for the quarterly effects. We have used a robust linear regression and a mixture model in Van Delden and Scholtus (2017) and we compare those two approaches and their settings. A second objective is, given the selected model, to repeat the analysis including 2016 data, to determine whether the seasonal differences are

[2] Mining and quarrying and Import of new cars were also included in this study, but these economic sectors did not contain enough linked units to test the seasonal effects.

consistent over time and whether they are also found in another economic sector, namely Job Placement.

When the seasonal differences are due to a limited number of units, those units could be manually checked and errors can be corrected if needed. Instead, when it concerns a large number of units that contribute to those seasonal effects that it seems more realistic to develop a correction method at macro-level rather than manually checking the units. In its most simple form, this limited set of units concern units that have certain reporting patterns in common. A more complicated form is to directly account for the contribution of all units to the slope. The third objective is to investigate whether the quarterly slope-differences can be explained by groups of units that have a certain reporting pattern in common. The fourth objective of the present paper is therefore to investigate whether the quarterly slope-differences can be explained by a limited number of most influential enterprises.

The remainder of this paper is organised as follows. We start with section 2 that describes the data used in the present study. Section 3 addresses the effect of model settings on the estimated quarterly effects (objectives 1 and 2). Section 4 determines whether the quarterly effects can be explained by a limited number of units that either have a certain pattern in common or a limited set of most influential units (objectives 3 and 4). Section 5 discusses the results and gives main directions for future research. Additionally, Appendix 1 provides the formulas to estimate the mixture model and the model parameter lambda for the Huber model.

2. Data

2.1 Description

We compared survey turnover with VAT turnover on a quarterly basis, using 2014, 2015 and 2016 data of the economic sectors Manufacturing, Construction, Retail trade and Job placement. Manufacturing, Construction, Retail trade are sectors with a monthly survey, whereas Job placement is a quarterly survey. Compared to Van Delden and Scholtus (2017), we omitted the sectors Mining and quarrying and Import of new cars. The estimates of our model yielded unstable results for those two sectors due to the small non-top X population sizes in both sectors, relative to the other three economic sectors in that paper. Sizes of the non-top X population in 2014 were 2048 (Mining and quarrying), 56 572 (Manufacturing), 143 339 (Construction), 110 440 (Retail Trade) and 128 (Import of new cars), see Van Delden and Scholtus (2017; Table 3). Further, compared to Van Delden and Scholtus (2017), we added the economic sector Job placement because a study by Van Bemmel (2018) showed clear seasonal patterns in differences between survey and VAT turnover.

VAT data were linked to sample survey data at the level of the statistical units, the enterprises, using a unique enterprise identification number. The linked data have been processed each within their own production systems before linkage. More information on the data and the production systems can be found in Van Delden and Scholtus (2017). We did not use all of the linked data, but we made a few selections:

1. units that were likely to have a 'thousand error' were omitted (see section 2.3 in Van Delden and Scholtus, 2017);
2. units need to be in both data sets for all four quarters of a year;
3. units need to have reported turnover in all four quarters of the year;
4. industries for which the turnover level or change estimates based on VAT are considered unreliable, because of differences in definition between VAT and sample survey turnover, were omitted.

We have applied those selections, to ensure that the seasonal effects that we find are due to reporting differences and not due to other factors. In van Delden and Scholtus (2017) we verified the effect of those selections on the outcomes of the 2014 and 2015 data. Van Delden and Scholtus (2017) showed that a limited number of units were omitted due to the first three selection steps. Further, they also provide an analysis on the combined effect of the selections 2, 3 and 4 on the regression coefficients. Their analyses showed that the seasonal effects were not very sensitive to those selections.

For the current study, we maintained outlying units with our data set (with the exception of selection 1) and used robust regression analyses methods to deal with them; see the next section. When an analysis of seasonal effects is done as part of the production process for estimating quarterly turnover changes then one should first correct the large errors with a clear cause.

2.2 Basic figures

The quarterly turnover for the whole population (all size classes) and the quarterly population size, given as the average over the four quarters per year, varies considerably per economic sector (see **Table 1**). **Table 1** also shows that the turnover of the topX units forms a large part of the total turnover: the fraction is largest for Manufacturing (about 0.72) and smallest for Construction (0.35). All further tables in the present study only concern the non-topX population. Recall that the data for the topX units are obtained as a take-all part of the survey and the response is also used in the census data. Differences between the two sources only occur for the non-topX units.

In the remainder of the results, all our analysis concerns a selection from the population:

– non-topX units;
– units that exist all twelve months of the year;
– units that are within the sample survey size classes.

The set of units that remain after these selection steps are referred to as the selected units. The number of selected units is given in the final column of **Table 1**.

We computed an estimate of the quarterly turnover level for the population of non-topX units that exists all four quarters of the year based on the selected units. For each selected unit we computed the calibration weight d_i^q. The estimated non-topX turnover level is given by $\hat{Y}^q = \sum_i d_i^q y_i^q$ for the survey turnover and by $\hat{X}^q = \sum_i d_i^q x_i^q$ for the VAT turnover. The result is presented in **Figure 1**. For all quarters and economic sectors, the weighted VAT turnover was larger than the weighted survey turnover. Note that the weighted turnover levels are smaller than the average quarterly non-topX turnover levels given in **Table 1** because it concerns a smaller portion of the non-topX population.

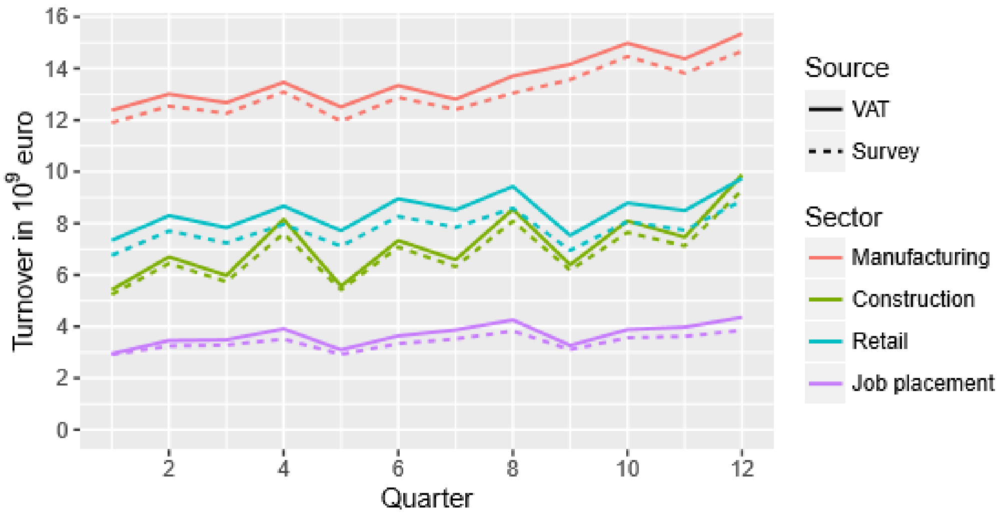

Figure 1. Estimated total non-topX turnover for all units that exist during a whole calendar year for VAT and Survey based on selected units that report both to the survey and the VAT data. Quarters are numbered from the first quarter of 2014 onwards.

Table 1. Some basic figures per economic sector and year based on the census data: average quarterly turnover and average quarterly number of enterprises for topX and non-topX units, and number of selected units for the analysis.

Sector	Turnover (10^9 euro)		Enterprises		Selected units
	topX	non-topX	topX	non-topX	(non-TopX)
Manufacturing					
2014	59.7	21.5	1554	56572	2296
2015	58.6	22.4	1495	58501	2187
2016	56.6	23.3	1417	60329	2271
Construction					
2014	6.9	12.3	534	143339	863
2015	7.1	13.1	515	149662	740
2016	7.2	14.3	430	156453	735
Retail trade					
2014	16.2	12.9	322	110439	2068
2015	17.0	13.5	316	115136	1590
2016	17.2	14.1	298	117836	1579
Job placement					
2014	2.0	3.6	168	12222	1290
2015	2.4	4.0	176	12469	936
2016	2.6	4.4	172	12775	1086

The number of non-topX units for a given quarter, above the survey threshold, for which we have reporting values during the whole year for both the survey and the VAT data, denoted by n^q, is shown in the final column of **Table 1**. It varies from 735 for Consumption in 2016 to 2296 for Manufacturing in 2014. In most of the comparisons (years × sectors), the difference between the quarterly VAT and survey turnover $\hat{X}^q - \hat{Y}^q$ was largest in the fourth quarter of the year. The quarterly VAT turnover growth rate, computed as $g_X^{q,q-1} = 100\left(\hat{X}^q/\hat{X}^{q-1} - 1\right)$ was often larger than the corresponding quarterly survey turnover growth rate $g_Y^{q,q-1}$ in the fourth quarter of the year and smaller in the first quarter of the year.

Table 2. The difference between in weighted non-topX totals (in 109 euros) for the survey ($\hat{Y}^q$) and the VAT turnover ($\hat{X}^q$) and their quarterly growth rates, for different quarters and years.

q	$\hat{X}^q - \hat{Y}^q$	$g_X^{q,q-1}$	$g_Y^{q,q-1}$	$\hat{X}^q - \hat{Y}^q$	$g_X^{q,q-1}$	$g_Y^{q,q-1}$
	Manu			Cons		
1	0.48			0.18		
2	0.47	5.1	5.4	0.26	23.3	22.7
3	0.41	-2.5	-2.2	0.24	-10.7	-10.9
4	0.37	6.2	6.7	0.55	36.4	32.5
5	0.55	-7.1	-8.6	0.14	-31.7	-28.6
6	0.47	6.7	7.6	0.25	31.7	30.5
7	0.39	-4.0	-3.6	0.27	-10.2	-10.7
8	0.66	7.1	5.1	0.46	29.7	28.0
9	0.59	3.3	4.0	0.21	-25.2	-23.5
10	0.52	5.8	6.6	0.45	26.6	23.6
11	0.55	-4.1	-4.4	0.34	-7.8	-6.9
12	0.69	6.8	6.1	0.60	32.4	30.3
	Retail				Job placement	
1	0.57			0.03		
2	0.6	13.1	13.9	0.21	17.3	11.3
3	0.6	-5.6	-6.0	0.20	0.8	1.2
4	0.71	10.6	9.9	0.40	12.3	6.9
5	0.6	-11.0	-10.6	0.18	-20.8	-16.7
6	0.68	16.0	16.3	0.30	17.3	14.1
7	0.67	-4.9	-5.1	0.35	6.4	5.4
8	0.84	10.7	9.4	0.43	10.0	8.7
9	0.58	-20.1	-19.1	0.16	-23.3	-18.7
10	0.71	16.7	16.1	0.32	18.8	14.6
11	0.76	-3.3	-4.2	0.36	2.5	1.5
12	0.86	14.7	14.9	0.50	9.6	6.7

3. Determine effects of model settings on quarterly slope effects

3.1 Methodology

3.1.1 General approach of the regression analysis

Underlying the economic sectors are a number of industries for which the outcomes are published separately. The regression analyses are done at the level of the economic sectors, so the data of the underlying industries are pooled. The reason is that the seasonal pattern effects are so subtle that we cannot estimate them accurately at industry level.

Throughout this paper we use VAT turnover as the independent variable and sample survey turnover as the dependent variable. The main reason for this is that the definition of survey turnover completely coincides with the definition of the target variable, while that of the VAT may differ (slightly). Furthermore, we are interested to understand the difference in seasonal pattern of VAT turnover versus sample survey turnover, where we have the hypothesis that enterprises may have shifted part of their declared sales for the VAT declarations towards the end of the year. We are well aware that in reality both sources may contain measurement errors and that neither one should act as the gold standard. We will come back to this issue at the end of this subsection and in section 5.

In the present paper we describe the relationship between survey and VAT turnover by a simple linear model with a number of assumptions, explained below. First of all, the motivation for the linear model is that in many industries the differences in definition between VAT turnover and survey turnover are either limited or they lead to a structural difference which can be corrected by a simple linear correction factor. This has been shown in Van Delden et al. (2016) who compared yearly survey and VAT turnover values and their definitions. Also, plots of VAT and survey turnover show a relationship which is close to linear (Van Delden et al., 2016; Van Delden and Scholtus, 2017). Other kinds of models that try to capture the measurement error in more detail are also possible, and will be used in future studies, see our explanation in section 5. In the current paper we will use a very simple version of such a measurement error model, as a start.

Van Delden and Scholtus (2017) found that the slope of a simple linear regression of survey turnover on VAT turnover depends on the quarter of the year while it is not necessary to let the intercept vary with the quarter of the year. We therefore applied the following linear regression model for the four quarters within a given year. Let x_i^q denote the VAT turnover for quarter q of enterprise i and let y_i^q be its sample survey turnover. Further, let α be the intercept, $\beta^{q=1}$ be the slope for quarter 1 and let $d\beta_1^{q=q*}$ stand for the difference in the slope between quarter

$q = q^*$ and quarter 1. Finally, let $\delta_{q*}^q \in \{0,1\}$ be a dummy variable that indicates whether $q = q^*$, with $q^* \in \{2,3,4\}$.

We used the following 'basic' model:

$$y_i^q = \alpha + \left(\beta^{q=1} + d\beta_1^{q=2}\delta_2^q + d\beta_1^{q=3}\delta_3^q + d\beta_1^{q=4}\delta_4^q\right)x_i^q + \varepsilon_i^q \tag{1}$$

where ε_i^q is a disturbance term. In what follows, we sometimes use $\boldsymbol{b} = (\alpha, \beta^{q=1}, d\beta_1^{q=2}, d\beta_1^{q=3}, d\beta_1^{q=4})^T$ as shorthand for the regression coefficients and $\boldsymbol{\xi}_i^q = \left(1, x_i^q, \delta_2^q x_i^q, \delta_3^q x_i^q, \delta_4^q x_i^q\right)^T$ so that (1) can be written as $y_i^q = \boldsymbol{b}^T \boldsymbol{\xi}_i^q + \varepsilon_i^q$. For simplicity, we assumed that ε_i^q is normally distributed with mean 0 and variance $\tilde{\sigma}^2 / w_i^q$. In reality, the error terms may be correlated over the quarters, implying that we may underestimate the significance of the regression parameters. In Ostlund (2018) the effect of these correlated errors on the estimates is shown for a similar model. The assumption that the variance of ε_i^q equals $\tilde{\sigma}^2 / w_i^q$ means that its size may vary with some properties of the units (heteroscedasticity).

The weights w_i^q used to correct for heteroscedasticity are given by:

$$w_i^q = 1/\left\{\max\left(x_i^q, 1\right)\right\}^\lambda. \tag{2}$$

Because some of the turnover values in the data were zero, we limited the maximum of this weight to 1. Here, $\lambda > 0$ is either a parameter of the model or a pre-defined constant (see Section 3.1.4).

This model corresponds to model B in Van Delden and Scholtus (2017), but with slightly different notation. 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.

For the estimation of the model (1), we have two further points that we account for. First of all, in all analyses, we estimate a model based on the 'selected set' which is a sample of the population. Since we want to estimate the outcomes at population level, we accounted for the calibration weights in the estimation procedure. The symbol d_i^q denotes a calibration weight, and stands for the ratio of the population size N_k^q in stratum k over the number of selected units n_k^q. A stratum k is given by the combination of an industry with a one-digit size class, see Scholtus and de Wolf (2011). Second, likewise to Van Delden and Scholtus (2017), we estimated the regression coefficients of the model (1) in a robust way, to avoid that results are affected by outliers. We compared two approaches: the use of a Huber model and the use of a mixture model.

We estimated the uncertainty in our parameter estimates in two ways. The *survey* package in R was used to compute uncertainty in the intercept and slope parameters of the linear regressions of both model types accounting for the actual calibration weights d_i^q; for details see Van Delden and Scholtus (2017). Furthermore, a bootstrap procedure was used to estimate uncertainty of *all* model parameters for the mixture model. This bootstrap approach used a fixed census of the large enterprises (50 and more employees) and simple random sampling with replacement from the medium and small enterprises (less than 50 employees), which is a simplified version of the actual sampling design.

In our estimation procedure we minimise the sum of the residuals ε_i^q rather than using an orthogonal regression. In our situation both the independent and the dependent variables may be prone to errors and the errors in the independent variable VAT may result in an underestimation of the quarterly slopes of the regression. If the size of the random error of the independent variable depends on the quarter then it might affect the outcomes. In additional studies we found this underestimation effect to be very small: in a mixture model where we estimated a group of units with no quarterly effect and a small variance the slopes were estimated to be 0.999 (not shown). Note that the alternative, orthogonal regression, is also sensitive to this effect because the bias in the regression parameters with orthogonal regression depends on the ratio of the random error in the dependent and independent variable. We use a regression model where we account for outliers in the regression. We expect quarterly differences in random error (if there are any) to be the largest for the units with the largest errors and that effect is corrected for in our model.

3.1.2 The Huber model

The Huber estimator is an outlier-robust estimator based on an iteratively weighted least squares procedure. In the remainder of this paper we will refer to this outlier-robust estimator as the "Huber model". This way, we can easier describe differences with results from the two-group mixture model.

For each regression we minimise:

$$\text{Min}_{\boldsymbol{b}} \sum_{q} \sum_{i \in s^q} \widehat{w}_i^q (\hat{u}_i^q)^2 \qquad (3)$$

where $\widehat{w}_i^q$ is an estimated weight for unit i which is corrected for outliers, accounts for the design weights and for the heteroscedasticity factor, and $\hat{u}_i^q$ is an estimated standardised residual, which is a residual divided by a robust estimate of its standard deviation, see Draper and Smith (1998). Furthermore s^q stands for the set of sample units that is available in quarter q. We will first explain the computation of the standard deviation, then give the computation of $\hat{u}_i^q$ and finally $\widehat{w}_i^q$.

A robust estimator for standard deviation $\tilde{\sigma} = \sqrt{\tilde{\sigma}^2}$ is the median absolute deviation (Rouseeuw and Croux, 1993):

$$\hat{\sigma}_{\text{MAD}} = a \text{ median} \left| \sqrt{\omega_i^q} \hat{\varepsilon}_i^q - \text{median}\left(\sqrt{\omega_i^q} \hat{\varepsilon}_i^q\right) \right| \qquad (4)$$

where the factor $a = \frac{1}{.6745}$ is needed to provide a consistent estimator of the standard deviation of the residuals of independent observations from a normal distribution and

$$\widehat{\omega}_i^q \equiv d_i^q / \{\max(x_i^q, 1)\}^{\widehat{\lambda}} \tag{5}$$

accounts for the effect of the calibration weight d_i^q and for the heteroscedasticity factor $1/\{\max(x_i^q, 1)\}^{\widehat{\lambda}}$. Note that in equation (4) the value of the term median$\left(\sqrt{\omega_i^q}\,\hat{\varepsilon}_i^q\right)$ is 0 in expectation, but the actual estimate for a given data set may differ from 0. A simplied, alternative, estimator for the standard deviation is therefore the median absolute residual:

$$\widehat{\sigma}_{\text{MAR}} = a \text{ median} \left| \sqrt{\omega_i^q}\,\hat{\varepsilon}_i \right| \tag{6}$$

The R function *rlm* that we used for the Huber model has implemented $\widehat{\sigma}_{\text{MAR}}$ (but the metadata of the package unjustly claimed it to be $\widehat{\sigma}_{\text{MAD}}$). In practice $\widehat{\sigma}_{\text{MAD}}$ and $\widehat{\sigma}_{\text{MAR}}$ are expected to be nearly identical.

For the purpose of robust estimation, the residuals are normalised as follows:

$$\hat{u}_{\text{M},i}^q = \frac{\sqrt{\omega_i^q}\,\hat{\varepsilon}_i^q}{\widehat{\sigma}_{\text{MAR}}}. \tag{7}$$

The estimated weights $\widehat{w}_i^q$ are computed according to:

$$\widehat{w}_i^q = \hat{g}_i^q\, d_i^q / \{\max(x_i^q, 1)\}^{\widehat{\lambda}} = \hat{g}_i^q \widehat{\omega}_i^q, \text{ with } \widehat{\omega}_i^q \text{ given in (5)}, \tag{8}$$

and $\hat{g}_i^q$ stands for the estimated Huber weight. The Huber weight is an outlier weight, such that non outlying units obtain a value of 1 while outlying values obtain a weight < 1. The Huber weight is estimated by:

$$\hat{g}_i^q = \begin{cases} 1 \text{ if } |\hat{u}_{\text{M},i}^q| \leq \gamma \\ \gamma / |\hat{u}_{\text{M},i}^q| \text{ else} \end{cases} \tag{9}$$

with γ = 1.345 (see p. 349 in Hastie et al., 2013). Thus, the estimated Huber weights are 1 for "normal" y_i^q values while outlying values obtain a weight < 1.

We estimated the regression coefficients of the Huber model with a maximum of 800 iterations. An iterative procedure is needed since the estimated Huber weights $\hat{g}_i^q$ depend on the estimated residuals, the estimated residuals depend on the estimated regression coefficients and the estimated regression coefficients result from minimisation of an estimated weighted sum of residuals. The procedure starts assuming that the Huber weights $\hat{g}_i^q = 1$ for all units. This leads to first parameter estimates and estimated residuals from which new Huber weights are estimated. This procedure is repeated until the estimates converge.

3.1.3 The two-group mixture model

Rather than using a robust model, one might also model the measurement errors in the population. Within the population, different units may have different types of systematic or random measurement errors. In the present paper, we use a very simple version of such a measurement error model. We plan to expand this approach in future work, see the discussion in section 5. In the current paper, we use a mixture model which assumes that the data are generated from a mixture of two populations: one set of units with a small error variance and another set of units with a larger error variance (the outlying units). We choose this simple mixture model because it closely resembles the Huber model: with both model types one estimates one regression line per quarter, while allowing for differences in variance among the units; see also Figure 1 in Van Delden and Scholtus (2017). In reality the population may consist of a mixture of units with more differences in measurement errors.

In fact equation (1) is replaced by:

$$y_i^q = \alpha + \left(\beta^{q=1} + d\beta_1^{q=2}\delta_2^q + d\beta_1^{q=3}\delta_3^q + d\beta_1^{q=4}\delta_4^q\right)x_i^q + \varepsilon_i^q + z_i^q e_i^q \qquad (10)$$

where $z_i^q \in \{0,1\}$ denotes an indicator with $P\left(z_i^q = 1\right) = \pi$, ε_i^q is normally distributed with mean 0 and variance $\tilde{\sigma}^2/w_i^q$ 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 $z_i^q = 1$. Note that the variance of ε_i^q, $\tilde{\sigma}^2/w_i^q$, is corrected for heteroscedasticity, according to expression (2). It is assumed that ε_i^q, z_i^q and e_i^q are mutually independent. Note that, under this model, the variance of the disturbance term for a given unit is inflated by a factor ϑ when $z_i^q = 1$. This means that the model separates outlying units (with a large variance) from inlying units without a large variance.

In Van Delden and Scholtus (2017) it is explained that from the maximum likelihood estimation of the model parameters it follows that the estimation of the mixture model in (10) is equivalent to using a weighted least squares procedure as in (3), but now with the weights:

$$\widehat{w}_i^q = \left(1 - \hat{\tau}_i^q\right)\widehat{\omega}_i^q + \hat{\tau}_i^q\widehat{\omega}_i^q/\hat{\vartheta} \qquad (11)$$

with $\widehat{\omega}_i^q \equiv d_i^q/\left\{\max\left(x_i^q, 1\right)\right\}^{\hat{\lambda}}$, according to equation (5). The parameter τ_i^q is the expectation of z_i^q given the covariates $\boldsymbol{\xi}_i^q$ for unit i, y_i^q and $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is the vector of model parameters with $\boldsymbol{\theta} = (\pi, \boldsymbol{b}^T, \tilde{\sigma}^2, \vartheta)^T$.

We can re-write equation (11) into

$$\widehat{w}_i^q = \left\{1 - \hat{\tau}_i^q\left(1 - 1/\hat{\vartheta}\right)\right\}\widehat{\omega}_i^q, \qquad (12)$$

which shows it has the same form of the Huber model in equations (8) and (9), but the outlier weights are different.

The parameters of the mixture model are estimated by (pseudo) maximum likelihood, using an Expectation Conditional Maximisation (ECM) algorithm; see Appendix 1. We estimated the regression coefficients of the mixture model with a maximum of 500 iterations of the ECM algorithm. This algorithm aims to maximise the expected log likelihood for $\boldsymbol{\theta}$ given the observed data: i.e. maximise $Q_d(\boldsymbol{\theta}) = E\{\log L_d(\boldsymbol{\theta}) \,|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}\}$.

Each iteration of this algorithm consists of two steps (see Appendix 1 for details):
1. Obtain $Q_d(\boldsymbol{\theta})$ by replacing each instance of z_i^q in the expression for $\log L_d(\boldsymbol{\theta})$ by its conditional expectation $\tau_i^q = E(z_i^q|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}) = P(z_i^q = 1|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta})$ given the current parameter estimates $\boldsymbol{\theta}$.
2. Obtain new parameter estimates $\boldsymbol{\theta}$ by maximising $Q_d(\boldsymbol{\theta})$.

3.1.4 Variations of the weights used in the regression

We performed a sensitivity analysis on the effect of the estimation of the weights $\widehat{w}_i^q$ on the estimated regression coefficients. We looked into two variations. The first variation concerned the heteroscedasticity correction factor λ. In Van Delden and Scholtus (2017), we used $\hat{\lambda} = 1$, thus assuming that the variance increases linearly with total VAT turnover. In the current study, we tried to verify this choice by inspecting plots of the residuals versus VAT turnover, but we found it was very hard to judge which value of λ is best (see for examples Figures 3–6 in Section 3.2). We therefore used another approach and estimated λ by maximum likelihood. In the mixture model, we estimate λ as part of the ECM algorithm; see Appendix 1. In a similar way, we included the estimation of λ in the iterative reweighting procedure of the Huber model. This is further explained in Appendix 1. We compare the results of estimating λ with $\hat{\lambda} = 1$ used in the previous report.

The second variation concerns the estimated weight for outlier detection: $\hat{g}_i^q$ in the Huber and in the mixture model $\{1 - \hat{\tau}_i^q(1 - 1/\hat{\vartheta})\}$. In Van Delden and Scholtus (2017), these outlier weight factors varied by the quarter of the year. We also considered the situation where a part of the units report differently in the sample survey compared to VAT throughout the whole year. We simulate this situation by giving the units the same 'outlier' weight during the whole year: $\hat{g}_i^+ = \hat{g}_i^q$, where the '+' denotes the yearly value.

In order to express an outlier weight $\hat{g}_i^+$ for the Huber model, we first consider the mean residual over the four quarters of the year: $\tilde{\varepsilon}_i^+ = \frac{1}{4}\sum_{q=1}^{4}\sqrt{\widehat{\omega}_i^q}\,\hat{\varepsilon}_i^q$. We assume that the mean quarterly residual $\tilde{\varepsilon}_i^+$ is normally distributed with mean 0 and $\mathrm{var}(\tilde{\varepsilon}_i^+) = \tilde{\sigma}^2/4$, with $\tilde{\sigma}$ estimated by $\hat{\sigma}_{\mathrm{MAR}}$. These values of $\tilde{\varepsilon}_i^+$ and $\mathrm{var}(\tilde{\varepsilon}_i^+)$ are subsequently used to compute the Huber weight g_i^+. That is done, by first computing a normalised residual:

$$\hat{u}_{\mathrm{M},i}^{+} = \frac{\tilde{\varepsilon}_i^+}{\hat{\tilde{\sigma}}/\sqrt{4}} \qquad (13)$$

followed by:

$$\hat{g}_i^+ = \begin{cases} 1 \text{ if } |\hat{u}_{\mathrm{M},i}^+| \leq \gamma \\ \gamma / |\hat{u}_{\mathrm{M},i}^+| \text{ else} \end{cases}. \tag{14}$$

In the mixture model with quarterly outlier weights, the parameter τ_i^q is based on a mixture of two normal distributions for quarter q. In order to estimate a yearly outlier weight, we replace the univariate distribution by a multivariate distribution to compute the simultaneous density for the four quarters of the year, where the four quarters are considered to be independent; see Appendix 1 for details. Using this approach we mainly identify units that have large residuals in all four quarters of a year or that have one very large residual in a single quarter as outlier.

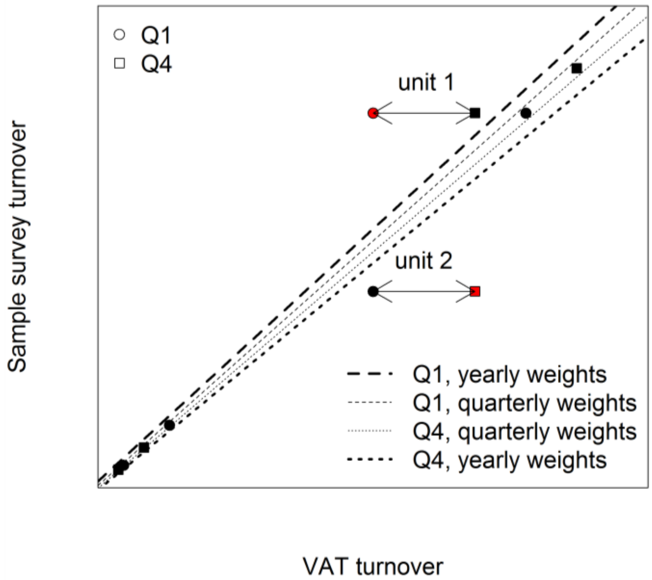

Figure 2. Illustration of effect of appointing outliers on a quarterly basis. Red points are outliers.

When the weights were estimated on a quarterly basis, one expects that slope differences are smaller than when the weights are estimated yearly. The reason for this effect is illustrated in Figure 2 that shows two units (marked 1 and 2) for which the reported VAT turnover in the fourth quarter of the year is relatively larger than in the first quarter (connected by an arrow). When units 1 and 2 are marked as yearly outliers and given a reduced weight of 0.75 in quarter 1 and 4, the thick regression lines are obtained. When units 1 and 2 are marked as quarterly outliers and given a reduced weight of 0.5 only in the quarter that they are outlying, the thin regression lines are obtained. It is seen that the difference in slope between quarters 1 and 4 is then attenuated. Therefore, use of yearly weights are expected to be more useful for a proper analysis of quarterly effects.

3.2 Results

3.2.1 Sensitivity analysis on outlier weights, on lambda and on model type

The outcomes of the sensitivity analysis on the computation of the outlier weights and on the heteroscedasticity correction are given in Tables 3–6. We found the following main observations in the results over the three years (2014 – 2016) and four economic sectors (Manufacturing, Construction, Retail trade and Job Placement):

- the estimated value of λ varied from year to year for a given economic sector, model type (Huber or mixture) and type of outlier weight (quarterly or yearly weights);
- the estimated values of λ for a given type of economic sector and outlier weight were larger for the Huber model than for the mixture model;
- the quarterly effects differed slightly between an estimated λ versus $\lambda = 1$;
- the values for $\widehat{d\beta}_1^{q=4}$ (the difference between the slope of the fourth and the first quarter) tended to be more extreme when yearly rather than quarterly outlier weights per unit were used (g_i^+ versus g_i^q) (with some exceptions);
- the values for $\widehat{d\beta}_1^{q=4}$ were often more extreme when the outliers were based on the Huber model compared to the mixture model (with some exceptions);

In order to better understand these five observations, we will first have a closer look into the effect of model type on the residuals. This was analysed by computing the estimated weighted residuals $\sqrt{\widehat{w}_i^q}\,\hat{\varepsilon}_i^q$ which should be homogeneous if the weighting correction works well and if λ is estimated correctly. For the Huber models we found that the computed reduction in weights for outliers was not large enough to achieve homogeneity of variances. This is illustrated in **Figure 3** for Construction in 2015. We found that the 5 per cent most outlying units (red points) despite having reduced weights still had a larger variance than the 5-15% most outlying units (orange points). The latter group had a larger variance than the remaining units (blue points). In contrast, for the Mixture model with two groups we found that after correction with weights the variance of outlying units was comparable to that of less outlying units (compare the red, orange and blue points in **Figure 4**). The same difference between the two model types was found in other years and in the other economic sectors. Therefore, results in the remainder of this paper will be based on the Mixture model. The results of the two-group mixture model are not completely homoscedastic: in a future study we will show that this is due to the fact that the population consists of more than two groups with different reporting behaviour.

Note that it is likely that the differences in the estimated quarterly effects between the Huber and the Mixture model (our fifth point) are caused by the larger influence of the outlying units in the Huber model on those quarterly slopes than is the case with the Mixture model.

A second important aspect is the difference between quarterly versus yearly weights. When the weights were estimated on a quarterly basis, the slope differences tended to be smaller than when estimated on a yearly basis. Recall that this result is in line with the illustration in **Figure 2** (section 3.1.4) that shows that quarterly appointment of the weights attenuates quarterly slope differences. Therefore, in the remainder of this document we use yearly outlier weights rather than quarterly ones since we believe that it approaches the source of the measurement errors more closely.

The third important aspect is the desired value of λ. Figures 3-6 show that, as expected, very small values of λ (close to 0) resulted in an increase of the variance of the residuals with size of the independent variable while vary large values of λ (close to 2) generally resulted in the opposite effect. The selected estimated λ value deviated somewhat from 1 (depending on the year and economic sector), but the estimated value was also affected by some outlying points with τ=1 that had still a larger variance than the remainder of the units. Note that the units with τ=1 (the crosses) mostly occurred for the red points, but sometimes also for orange and blue points. A future study will show that a mixture model with more groups better fits the data: there are units with an even larger variance than is captured in the two-mixture model. Concerning the value for λ we want to avoid that its value depends too much on the actual selected data, we therefore prefer to use a fixed value. When those outlying units are ignored $\lambda = 1$ is an adequate value to achieve homoscedasticity.

Table 3. Regression coefficients and lambda for Huber and Mixture models, for Manufacturing[1].

Year	Coef	Huber				Mixture			
		EQ	FQ	EY	FY	EQ	FQ	EY	FY
2014	$\hat{\lambda}$	1.244	1.000	1.352	1.000	1.013	1.000	0.973	1.000
	$\hat{\alpha}$	-0.219	-5.830	4.797	3.843	-17.854	-18.151	-11.800	-14.260
	$\hat{\beta}_1^{\square}$	0.975	0.979	0.975	0.972	0.990	0.990	0.990	0.990
	$\widehat{d\beta}_1^{q=2}$	-0.004	-0.004	-0.003	0.000	-0.003	-0.003	-0.004	-0.004
	$\widehat{d\beta}_1^{q=3}$	-0.004	-0.004	-0.002	0.003	-0.005	-0.005	-0.004	-0.004
	$\widehat{d\beta}_1^{q=4}$	-0.007	-0.006	-0.003	0.005	-0.006	-0.006	-0.007	-0.007
2015	$\hat{\lambda}$	1.141	1.000	1.212	1.000	0.842	1.000	0.779	1.000
	$\hat{\alpha}$	2.930	-1.813	18.836	10.607	-22.000	-17.459	-24.559	-14.584
	$\hat{\beta}_1^{\square}$	0.970	0.972	0.968	0.975	0.986	0.983	0.992	0.984
	$\widehat{d\beta}_1^{q=2}$	0.001	0.001	0.001	-0.001	0.002	0.002	-0.003	-0.002
	$\widehat{d\beta}_1^{q=3}$	-0.001	-0.001	-0.002	-0.004	0.001	-0.002	-0.004	-0.006
	$\widehat{d\beta}_1^{q=4}$	-0.008	-0.008	-0.011	-0.018	-0.002	-0.004	-0.008	-0.011
2016	$\hat{\lambda}$	1.148	1.000	1.316	1.000	0.915	1.000	0.904	1.000
	$\hat{\alpha}$	0.004	-0.970	0.298	1.428	-7.878	-4.286	-6.943	-3.584
	$\hat{\beta}_1^{\square}$	0.976	0.977	0.981	0.980	0.987	0.985	0.988	0.986
	$\widehat{d\beta}_1^{q=2}$	-0.003	-0.003	-0.006	-0.005	-0.003	-0.003	-0.004	-0.005
	$\widehat{d\beta}_1^{q=3}$	-0.005	-0.004	-0.007	-0.006	-0.002	-0.003	-0.004	-0.004
	$\widehat{d\beta}_1^{q=4}$	-0.009	-0.009	-0.012	-0.011	-0.005	-0.005	-0.008	-0.009

[1] EQ: λ estimated and quarterly outlier weights; FQ: λ = 1 and quarterly outlier weights; EY: λ estimated and yearly outlier weights; FQ: λ = 1 and yearly outlier weights.

Table 4. Regression coefficients and lambda for Huber and Mixture models, for Construction.

Year	Coef	Huber				Mixture			
		EQ	FQ	EY	FY	EQ	FQ	EY	FY
2014	$\hat{\lambda}$	1.311	1.000	1.352	1.000	0.639	1.000	0.580	1.000
	$\hat{\alpha}$	0.599	3.185	4.706	10.654	-2.239	-0.397	-0.497	1.104
	$\hat{\beta}_1$	0.968	0.965	0.970	0.963	0.985	0.980	0.983	0.980
	$\widehat{d\beta}_1^{q=2}$	0.007	0.008	0.006	0.006	0.005	0.005	0.002	0.002
	$\widehat{d\beta}_1^{q=3}$	0.003	0.003	0.005	0.006	0.002	0.000	0.001	-0.001
	$\widehat{d\beta}_1^{q=4}$	-0.007	-0.007	-0.011	-0.010	-0.001	-0.001	-0.002	-0.007
2015	$\hat{\lambda}$	1.457	1.000	1.633	1.000	0.772	1.000	0.799	1.000
	$\hat{\alpha}$	0.200	5.599	0.049	10.139	0.113	0.695	1.948	1.330
	$\hat{\beta}_1$	0.981	0.973	0.990	0.977	0.990	0.988	0.983	0.984
	$\widehat{d\beta}_1^{q=2}$	-0.006	-0.003	-0.010	-0.004	-0.006	-0.007	-0.003	-0.004
	$\widehat{d\beta}_1^{q=3}$	-0.006	-0.005	-0.009	-0.009	-0.003	-0.003	-0.002	-0.002
	$\widehat{d\beta}_1^{q=4}$	-0.013	-0.010	-0.017	-0.025	-0.006	-0.008	-0.008	-0.010
2016	$\hat{\lambda}$	1.353	1.000	1.510	1.000	0.757	1.000	0.718	1.000
	$\hat{\alpha}$	0.298	2.535	0.086	4.819	-3.116	-0.868	-3.338	-0.475
	$\hat{\beta}_1$	0.974	0.971	0.982	0.976	0.991	0.987	0.989	0.984
	$\widehat{d\beta}_1^{q=2}$	0.000	0.000	-0.005	-0.004	0.001	0.002	-0.001	0.001
	$\widehat{d\beta}_1^{q=3}$	-0.005	-0.006	-0.010	-0.010	0.000	0.000	-0.002	-0.003
	$\widehat{d\beta}_1^{q=4}$	-0.010	-0.010	-0.016	-0.019	-0.005	-0.004	-0.008	-0.007

Table 5. Regression coefficients and lambda for Huber and Mixture models, for Retail trade.

Year	Coef	Huber				Mixture			
		EQ	FQ	EY	FY	EQ	FQ	EY	FY
	$\hat{\lambda}$	1.446	1.000	1.379	1.000	1.081	1.000	1.023	1.000
2014	$\hat{\alpha}$	0.005	0.068	0.047	0.172	-0.244	-0.310	-0.311	-0.325
	$\hat{\beta}_1$	0.952	0.955	0.963	0.963	0.970	0.974	0.976	0.976
	$\widehat{d\beta}_1^{q=2}$	0.002	0.004	0.003	0.004	-0.001	-0.001	0.002	0.002
	$\widehat{d\beta}_1^{q=3}$	-0.003	-0.002	0.000	0.002	-0.006	-0.006	-0.002	-0.002
	$\widehat{d\beta}_1^{q=4}$	-0.004	-0.004	-0.004	-0.008	-0.002	-0.003	-0.003	-0.003
2015	$\hat{\lambda}$	1.678	1.000	1.506	1.000	1.350	1.000	1.312	1.000
	$\hat{\alpha}$	0.088	0.240	0.306	0.419	0.044	-0.056	0.014	-0.053
	$\hat{\beta}_1$	0.959	0.957	0.967	0.965	0.963	0.971	0.972	0.975
	$\widehat{d\beta}_1^{q=2}$	-0.006	-0.001	-0.004	-0.001	-0.004	-0.004	-0.002	-0.001
	$\widehat{d\beta}_1^{q=3}$	-0.007	-0.002	-0.005	-0.002	-0.004	-0.003	-0.002	-0.002
	$\widehat{d\beta}_1^{q=4}$	-0.013	-0.011	-0.012	-0.017	-0.012	-0.011	-0.009	-0.010
2016	$\hat{\lambda}$	1.474	1.000	1.347	1.000	1.137	1.000	1.054	1.000
	$\hat{\alpha}$	0.005	-0.027	0.114	0.140	-0.197	-0.350	-0.250	-0.328
	$\hat{\beta}_1$	0.948	0.952	0.963	0.961	0.957	0.964	0.965	0.968
	$\widehat{d\beta}_1^{q=2}$	-0.001	0.000	-0.003	-0.001	0.002	0.001	0.004	0.004
	$\widehat{d\beta}_1^{q=3}$	-0.005	-0.004	-0.006	-0.003	-0.001	0.000	-0.001	-0.001
	$\widehat{d\beta}_1^{q=4}$	-0.007	-0.008	-0.008	-0.007	-0.006	-0.004	-0.006	-0.006

Table 6. Regression coefficients and lambda for Huber and Mixture models, for Job placement.

Year	Coef	Huber				Mixture			
		EQ	FQ	EY	FY	EQ	FQ	EY	FY
2014	$\hat{\lambda}$	1.541	1.000	1.460	1.000	1.088	1.000	1.659	1.000
	$\hat{\alpha}$	0.062	0.195	0.140	0.308	-0.009	-0.012	0.069	0.030
	$\hat{\beta}_1$	0.986	0.989	0.992	0.989	0.999	0.999	0.964	1.000
	$\widehat{d\beta}_1^{q=2}$	-0.021	-0.025	-0.026	-0.025	-0.001	0.000	-0.013	-0.017
	$\widehat{d\beta}_1^{q=3}$	-0.021	-0.025	-0.026	-0.025	0.000	0.000	-0.031	-0.022
	$\widehat{d\beta}_1^{q=4}$	-0.047	-0.059	-0.062	-0.067	-0.004	-0.002	-0.060	-0.035
2015	$\hat{\lambda}$	1.694	1.000	1.643	1.000	1.232	1.000	1.188	1.000
	$\hat{\alpha}$	0.067	0.364	0.231	0.430	0.177	0.155	0.141	0.125
	$\hat{\beta}_1$	0.975	0.973	0.980	0.979	0.983	0.994	0.987	0.990
	$\widehat{d\beta}_1^{q=2}$	-0.011	-0.018	-0.015	-0.019	-0.004	-0.004	-0.001	-0.003
	$\widehat{d\beta}_1^{q=3}$	-0.019	-0.029	-0.020	-0.027	-0.008	-0.004	-0.010	-0.010
	$\widehat{d\beta}_1^{q=4}$	-0.028	-0.044	-0.030	-0.056	-0.011	-0.008	-0.017	-0.012
2016	$\hat{\lambda}$	1.892	1.000	1.886	1.000	1.761	1.000	1.828	1.000
	$\hat{\alpha}$	0.009	0.172	0.011	0.210	0.007	-0.004	0.018	0.026
	$\hat{\beta}_1$	0.987	0.972	0.998	0.976	0.990	0.999	0.973	0.993
	$\widehat{d\beta}_1^{q=2}$	-0.016	-0.021	-0.031	-0.022	-0.022	-0.001	-0.041	-0.009
	$\widehat{d\beta}_1^{q=3}$	-0.027	-0.026	-0.038	-0.022	-0.029	-0.001	-0.067	-0.018
	$\widehat{d\beta}_1^{q=4}$	-0.037	-0.056	-0.046	-0.063	-0.043	-0.004	-0.055	-0.025

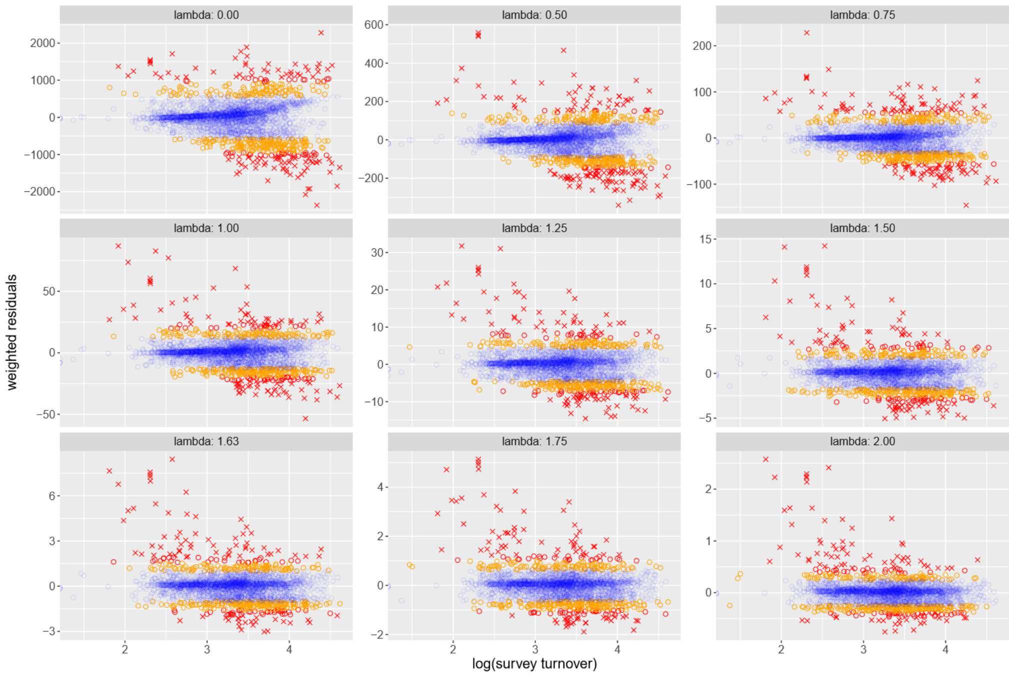

Figure 3. Weighted residuals for Construction 2015 estimated with the Huber model for different values of λ, with λ=1.63 as the estimated value in the EY model. Red symbols: top 5% smallest Huber weight values, orange 5-15% smallest Huber weight values, blue: all other Huber weights. Cross means Huber weight < 0.1, circle Huber weight ≥ 0.1.

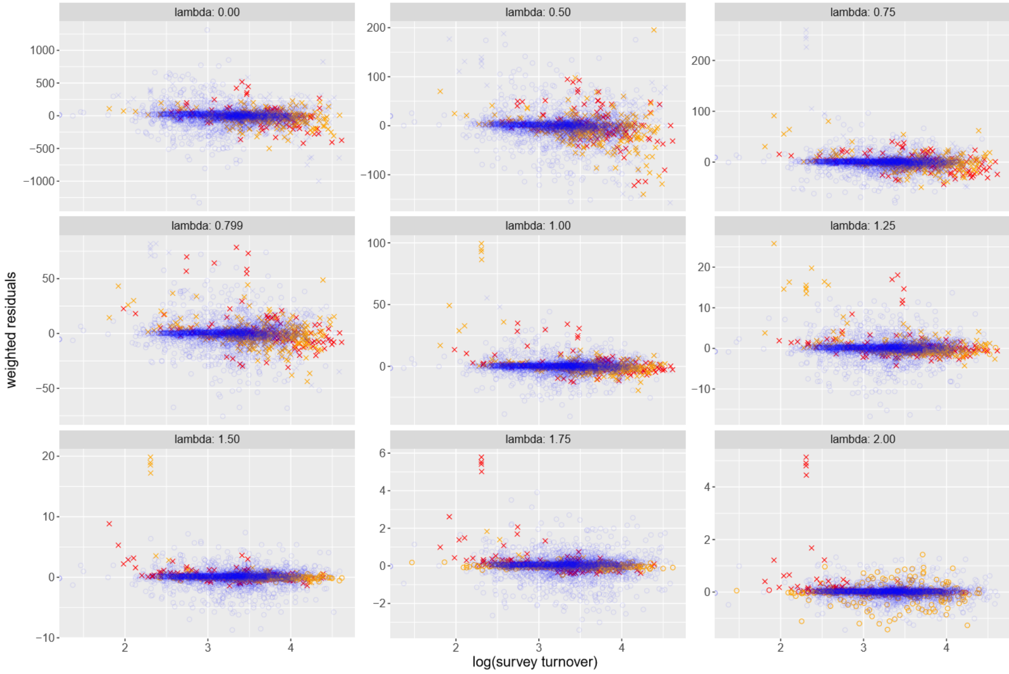

Figure 4. Weighted residuals for Construction 2015 estimated with the Mixture model for different values of λ, with λ=0.799 as the estimated value in the EY model. Red symbols: top 5% smallest Huber weight values, orange 5-15% smallest Huber weight values, blue: all other Huber weights. Cross means τ = 1, circles stand for τ < 1.

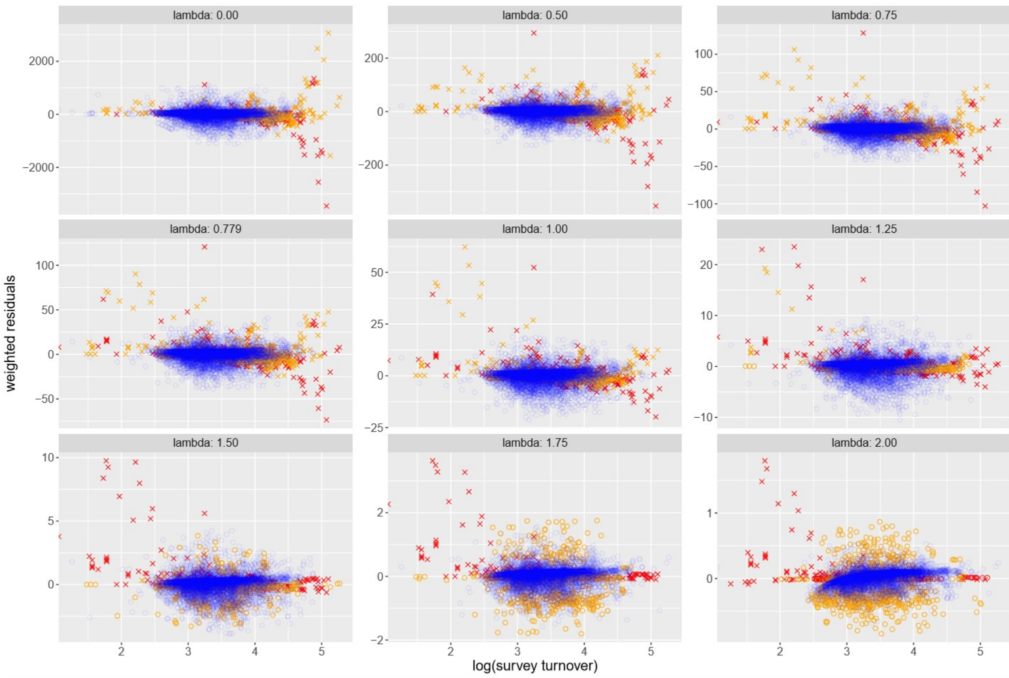

Figure 5. Weighted residuals for Manufacturing 2015 estimated with the Mixture model for different values of λ, with λ=0.799 as the estimated value in the EY model. Red symbols: top 5% smallest Huber weight values, orange 5-15% smallest Huber weight values, blue: all other Huber weights. Cross means τ = 1, circles stand for τ < 1.

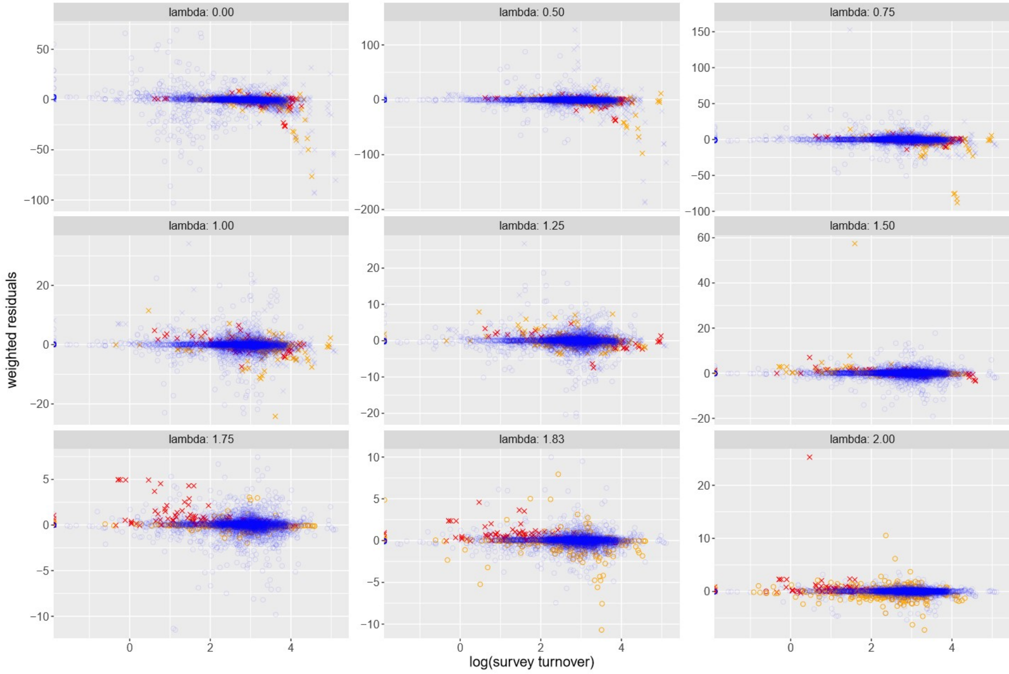

Figure 6. Weighted residuals for Job Placement 2016 estimated with the Mixture model for different values of λ, with λ=1.83 as the estimated value in the EY model. Red symbols: top 5% smallest Huber weight values, orange 5-15% smallest Huber weight values, blue: all other Huber weights. Cross means τ = 1, circles stand for τ < 1.

We finally remark that Van Delden and Scholtus (2017) also estimated the linear regressions for 2014 and 2015 for the Huber model where units should report 12 months of the year (their Tables 15 and 16). The 2015 data used in the current paper are slightly different, because in the meantime some additional editing has been done on the data during regular production. Further, the model in Van Delden and Scholtus (2017) was slightly different since in their paper they did not completely correctly select the units that were non-top X units all four quarter of the year, and units without imputed values for the VAT data for all four quarter of the year (for some quarters the data could be imputed). Comparing the columns labelled "Huber FQ" in Tables 3 – 5 (current paper) with the column "12 months estim" in Tables 15 and 16 (van Delden and Scholtus, 2017) we find that results for Manufacturing and Retail trade were nearly the same. Only for Construction we found that the estimated intercept is larger and the slope of the first quarter smaller in Van Delden and Scholtus (2017) than in the current paper.

3.2.2 Parameter estimates for the chosen model

The estimated parameters of the chosen mixture model according to equation (10) are given in Table 7 (for Manufacturing), Table 8 (for Construction) and Table 9 (for Retail trade) and Table 10 (for Job Placement). Tables 7-10 show that the standard errors for the intercept and slope parameters based on the bootstrap procedure and those estimated by the *survey* package based on the design weights were close together. This strengthens the confidence in both estimated standard errors.

We will first describe the regression coefficients, and then we will discuss the model parameters of the mixture model. In the fifth column of these tables we give the probability (p-value) based on a Student t distribution, that a regression coefficient equals its reference value (symbol r). Small p-values indicate that a regression coefficient differs from its reference value. As a reference value we used $r = 0$ for all regression coefficients except for $\hat{\beta}^1$ where we used $r = 1$.

In most of the twelve cases (four economic sectors times three years), the p-value of the slope effect coefficient ($d\beta_1^{q=4}$) of the fourth quarter of the year was the smallest. In two cases (Job Placement 2014 and 2016), the slope effect of a second and/or third quarter had an equally small p-value. The size of $d\beta_1^{q=4}$ was largest for Job placement (range: -0.012 to -0.035), followed by Manufacturing (range: -0.007 to -0.011), Construction (range: -0.007 to -0.010) and Retail trade (range: -0.003 to -0.010). The effects for the fourth quarter were strong for Manufacturing and Job Placement (all p-values < 0.01), weaker for Construction (p-values varying from 0.03 to 0.06), and for Retail trade only in 2015 the p-value for $d\beta_1^{q=4}$ was small. Note that the slope effect coefficients had a relatively large uncertainty, leading to a large 95%-uncertainty interval.

Table 7. Parameters of the Mixture model ($\lambda = 1$), for Manufacturing.[1]

Year	Parameter	Value	Survey package		Bootstrap		
			SE	p-value	SE	L95	U95
2014	$\hat{\pi}$	0.262			0.007	0.248	0.276
	$\hat{\hat{\sigma}}^2$	9.322			0.167	8.983	9.631
	$\hat{\vartheta}$	199.004			6.909	186.175	213.370
	$\hat{\alpha}$	-14.260	1.658	0.000	1.797	-17.677	-10.953
	$\hat{\beta}^1$	0.990	0.001	0.000	0.002	0.987	0.993
	$\widehat{d\beta}_1^{q=2}$	-0.003	0.002	0.062	0.002	-0.007	0.000
	$\widehat{d\beta}_1^{q=3}$	-0.004	0.002	0.065	0.002	-0.007	0.001
	$\widehat{d\beta}_1^{q=4}$	-0.007	0.002	0.001	0.002	-0.010	-0.003
	$\hat{\beta}^+$	0.987					
2015	$\hat{\pi}$	0.222			0.007	0.209	0.236
	$\hat{\hat{\sigma}}^2$	13.176			0.234	12.713	13.607
	$\hat{\vartheta}$	281.412			10.033	261.976	302.038
	$\hat{\alpha}$	-14.584	2.165	0.000	2.328	-19.197	-10.197
	$\hat{\beta}^1$	0.984	0.002	0.000	0.002	0.981	0.987
	$\widehat{d\beta}_1^{q=2}$	-0.002	0.002	0.303	0.002	-0.007	0.002
	$\widehat{d\beta}_1^{q=3}$	-0.006	0.002	0.017	0.002	-0.010	-0.002
	$\widehat{d\beta}_1^{q=4}$	-0.011	0.003	0.000	0.002	-0.015	-0.006
	$\hat{\beta}^+$	0.979					
2016	$\hat{\pi}$	0.241			0.007	0.227	0.254
	$\hat{\hat{\sigma}}^2$	12.175			0.212	11.762	12.611
	$\hat{\vartheta}$	161.771			5.681	151.072	173.243
	$\hat{\alpha}$	-3.584	0.985	0.000	1.041	-5.687	-1.511
	$\hat{\beta}^1$	0.986	0.002	0.000	0.002	0.983	0.989
	$\widehat{d\beta}_1^{q=2}$	-0.004	0.002	0.029	0.002	-0.008	-0.001
	$\widehat{d\beta}_1^{q=3}$	-0.004	0.002	0.036	0.002	-0.009	0.000
	$\widehat{d\beta}_1^{q=4}$	-0.009	0.002	0.000	0.002	-0.013	-0.005
	$\hat{\beta}^+$	0.982					

[1] The parameters $\hat{\pi}$, $\hat{\hat{\sigma}}^2$, $\hat{\vartheta}$, $\hat{\alpha}$, $\hat{\beta}^1$, $\widehat{d\beta}_1^{q=2}$, $\widehat{d\beta}_1^{q=3}$ and $\widehat{d\beta}_1^{q=4}$ are according to equation (10), $\hat{\beta}^+$ is according to equation (23) defined below.

Table 8. Parameters of the Mixture model ($\lambda = 0.5$), for Construction.

Year	Parameter	Value	Survey package		Bootstrap		
			SE	*p*-value	SE	L95	U95
2014	$\hat{\pi}$	0.286			0.020	0.246	0.325
	$\hat{\bar{\sigma}}^2$	3.933			0.152	3.637	4.217
	$\hat{\vartheta}$	213.246			14.859	186.591	244.969
	$\hat{\alpha}$	1.104	0.910	0.225	1.188	-1.696	3.426
	$\hat{\beta}^1$	0.980	0.003	0.000	0.003	0.974	0.986
	$\widehat{d\beta}_1^{q=2}$	0.002	0.003	0.649	0.004	-0.006	0.010
	$\widehat{d\beta}_1^{q=3}$	-0.001	0.004	0.799	0.004	-0.009	0.007
	$\widehat{d\beta}_1^{q=4}$	-0.007	0.004	0.063	0.004	-0.015	0.000
	$\hat{\beta}^+$	0.978					
2015	$\hat{\pi}$	0.219			0.020	0.183	0.261
	$\hat{\bar{\sigma}}^2$	5.076			0.219	4.659	5.505
	$\hat{\vartheta}$	164.765			13.213	139.658	191.620
	$\hat{\alpha}$	1.330	0.813	0.102	1.602	-2.384	4.557
	$\hat{\beta}^1$	0.984	0.003	0.000	0.004	0.978	0.991
	$\widehat{d\beta}_1^{q=2}$	-0.004	0.004	0.324	0.005	-0.013	0.004
	$\widehat{d\beta}_1^{q=3}$	-0.002	0.004	0.575	0.005	-0.012	0.007
	$\widehat{d\beta}_1^{q=4}$	-0.010	0.005	0.032	0.004	-0.019	-0.002
	$\hat{\beta}^+$	0.980					
2016	$\hat{\pi}$	0.282			0.022	0.241	0.325
	$\hat{\bar{\sigma}}^2$	2.855			0.132	2.591	3.119
	$\hat{\vartheta}$	286.006			21.991	246.665	330.219
	$\hat{\alpha}$	-0.475	0.517	0.358	1.276	-3.139	2.570
	$\hat{\beta}^1$	0.984	0.002	0.000	0.003	0.979	0.989
	$\widehat{d\beta}_1^{q=2}$	0.001	0.003	0.832	0.004	-0.006	0.008
	$\widehat{d\beta}_1^{q=3}$	-0.003	0.003	0.313	0.004	-0.011	0.004
	$\widehat{d\beta}_1^{q=4}$	-0.007	0.003	0.016	0.003	-0.014	-0.001
	$\hat{\beta}^+$	0.981					

Table 9. Parameters of the Mixture model ($\lambda = 1$), for Retail trade.

Year	Parameter	Value	Survey package		Bootstrap		
			SE	*p*-value	SE	L95	U95
2014	$\hat{\pi}$	0.243			0.013	0.220	0.269
	$\hat{\sigma}^2$	1.034			0.027	0.982	1.083
	$\hat{\vartheta}$	112.667			5.333	102.760	122.989
	$\hat{\alpha}$	-0.325	0.060	0.000	0.081	-0.496	-0.180
	$\hat{\beta}^1$	0.976	0.002	0.000	0.002	0.972	0.981
	$\widehat{d\beta}_1^{q=2}$	0.002	0.003	0.391	0.003	-0.004	0.008
	$\widehat{d\beta}_1^{q=3}$	-0.002	0.003	0.442	0.003	-0.008	0.003
	$\widehat{d\beta}_1^{q=4}$	-0.003	0.003	0.346	0.003	-0.008	0.003
	$\hat{\beta}^+$	0.976					
2015	$\hat{\pi}$	0.245			0.014	0.220	0.275
	$\hat{\sigma}^2$	1.119			0.033	1.051	1.179
	$\hat{\vartheta}$	77.157			4.148	69.778	86.228
	$\hat{\alpha}$	-0.053	0.081	0.513	0.105	-0.262	0.148
	$\hat{\beta}^1$	0.975	0.002	0.000	0.003	0.970	0.980
	$\widehat{d\beta}_1^{q=2}$	-0.001	0.003	0.752	0.004	-0.008	0.006
	$\widehat{d\beta}_1^{q=3}$	-0.002	0.004	0.521	0.004	-0.010	0.006
	$\widehat{d\beta}_1^{q=4}$	-0.010	0.004	0.004	0.004	-0.017	-0.003
	$\hat{\beta}^+$	0.971					
2016	$\hat{\pi}$	0.218			0.013	0.193	0.244
	$\hat{\sigma}^2$	1.419			0.041	1.335	1.495
	$\hat{\vartheta}$	123.400			6.837	110.953	136.939
	$\hat{\alpha}$	-0.328	0.090	0.000	0.118	-0.545	-0.096
	$\hat{\beta}^1$	0.968	0.003	0.000	0.003	0.963	0.973
	$\widehat{d\beta}_1^{q=2}$	0.004	0.004	0.307	0.004	-0.003	0.011
	$\widehat{d\beta}_1^{q=3}$	-0.001	0.004	0.745	0.004	-0.009	0.006
	$\widehat{d\beta}_1^{q=4}$	-0.006	0.004	0.155	0.004	-0.013	0.001
	$\hat{\beta}^+$	0.967					

Table 10. Parameters of the Mixture model ($\lambda = 1$), for Job placement.

Year	Parameter	Value	Survey package		Bootstrap		
			SE	*p*-value	SE	L95	U95
2014	$\hat{\pi}$	0.337			0.022	0.294	0.379
	$\hat{\sigma}^2$	0.921			0.042	0.847	1.003
	$\hat{\vartheta}$	188.487			12.705	164.687	214.511
	$\hat{\alpha}$	0.030	0.036	0.411	0.071	-0.106	0.175
	$\hat{\beta}^1$	1.000	0.004	0.993	0.003	0.994	1.005
	$\widehat{d\beta}_1^{q=2}$	-0.017	0.005	0.000	0.004	-0.025	-0.009
	$\widehat{d\beta}_1^{q=3}$	-0.022	0.005	0.000	0.004	-0.030	-0.015
	$\widehat{d\beta}_1^{q=4}$	-0.035	0.005	0.000	0.004	-0.043	-0.028
	$\hat{\beta}^+$	0.980					
2015	$\hat{\pi}$	0.375			0.031	0.316	0.441
	$\hat{\sigma}^2$	0.515			0.035	0.446	0.583
	$\hat{\vartheta}$	299.031			28.464	249.894	362.582
	$\hat{\alpha}$	0.125	0.049	0.010	0.072	-0.015	0.272
	$\hat{\beta}^1$	0.990	0.003	0.001	0.003	0.984	0.996
	$\widehat{d\beta}_1^{q=2}$	-0.003	0.005	0.565	0.004	-0.011	0.006
	$\widehat{d\beta}_1^{q=3}$	-0.010	0.005	0.029	0.004	-0.018	-0.001
	$\widehat{d\beta}_1^{q=4}$	-0.012	0.005	0.008	0.004	-0.021	-0.004
	$\hat{\beta}^+$	0.983					
2016	$\hat{\pi}$	0.292			0.028	0.241	0.347
	$\hat{\sigma}^2$	0.815			0.055	0.708	0.914
	$\hat{\vartheta}$	462.435			43.772	390.350	564.006
	$\hat{\alpha}$	0.026	0.060	0.657	0.091	-0.136	0.202
	$\hat{\beta}^1$	0.993	0.003	0.041	0.004	0.986	1.000
	$\widehat{d\beta}_1^{q=2}$	-0.009	0.005	0.084	0.005	-0.019	0.001
	$\widehat{d\beta}_1^{q=3}$	-0.018	0.005	0.000	0.005	-0.027	-0.008
	$\widehat{d\beta}_1^{q=4}$	-0.025	0.005	0.000	0.005	-0.033	-0.015
	$\hat{\beta}^+$	0.980					

Further, for Job placement, reduced slopes at p-value < 0.05 were also found for the third quarter in all three years, and for the second quarter in 2014. Also, for Manufacturing, a reduced slope at a p-value < 0.05 was found in the third quarter of 2015 and in the second quarter for Manufacturing 2016.

We found that the slope of the first quarter of the year was significantly different (in fact: smaller) from one at a p-value < 0.05 in all years for all four economic sectors, except for Job placement in 2014. That result is consistent with **Table 1** showing that the quarterly VAT turnover was larger than the quarterly survey turnover, also in the first quarter of the year. We also found that the intercepts were differing significantly from zero in all years for Manufacturing, in 2014 and 2016 for Retail Trade and in 2015 for Job placement at a p-value < 0.01 although the intercepts were generally very small. We prefer to include an intercept into the model because in this way we can verify the value of the intercept of the estimated relationship between survey and VAT and avoid that an estimated effect on the slope is in fact due to a missing intercept term in the model.

Tables 7 – 10 also mention the slope $\hat{\beta}^{+}$. That slope will be explained in section 4.1.2.

In the chosen mixture model according to equation (10), the proportion of units with an enlarged disturbance term, π, was estimated to be similar for Retail trade (range: 0.218 – 0.245), Construction (range: 0.219 – 0.286) and Manufacturing (range: 0.222 – 0.262), whereas it was somewhat larger for Job placement (range: 0.292 – 0.375). The inflation factor ϑ that describes the ratio of the variance of outliers to that of the non-outlying units, was smallest for Retail trade (range: 77.1 to 123.4), followed by Manufacturing (range: 161.8 to 281.4) and Construction (range: 164.8 to 286.0). For Job placement the inflation factor varied widely between years (range: 188.5 to 462.4). The estimated variance $\tilde{\sigma}^2$ was around a value of '10' for Manufacturing and around of value of '1' for the other three economic sectors. The relative uncertainty (coefficient of variation: SE-boot/parameter value) was much smaller for π, ϑ and $\tilde{\sigma}^2$ than for α and for the slope effect parameters $d\beta_1^q$.

The above estimated π-, and ϑ-values were similar to those obtained for the mixture model (FQ) in Van Delden and Scholtus (2017; Table 18). The values for the estimated $\tilde{\sigma}^2$ of the current report are smaller than found in Van Delden and Scholtus (2017); note that the header of the last column of their Table 18 is $\tilde{\sigma}$ but should have been $\tilde{\sigma}^2$. The estimation of the parameters in the current report has changed slightly compared to Van Delden and Scholtus (2017) in the sense that the design weights are now included in the estimation procedure according to pseudo maximum likelihood estimation; see Appendix 1.

4. Determine if quarterly slope effects can be explained by a limited number of enterprises

4.1 Methodology

To explain which units contribute most to the differences in quarterly slopes our approach is the following:
- We start by finding an expression for contribution of each unit to the quarterly slopes. This contribution can be interpreted as the "unit level quarterly slope" (eq. (21) in section 4.1.1);
- Define an overall slope common to all quarters and units (eq. (23))
- For each unit and quarter, we define the "quarterly unit slope effect" as the differences between the unit level quarterly slope and the overall slope (eq. (25));

– Now define a "normalised quarterly unit slope effect" by subtracting from the unit slope effects their weighted means (eq. (27));
– The sum of these (normalised) quarterly unit level effects over k units is their combined effect on slope differences (eq. (28));
– We can also accumulate the contributions of all units in a response pattern to the response pattern effect (section 4.1.3). Likewise to the expressions at unit level, we distinguish among a "quarterly pattern slope effect" and a "normalised quarterly pattern slope effect", and we have the sum of these (normalised) quarterly pattern level effects over P patterns as their combined effect on slope differences.

4.1.1 Contribution of units to the slope of the regression

We are interested to explain which units contribute most to the differences in quarterly slopes. We will first explain how this can be done in the case of a simple linear regression without an intercept. Then, we will derive how this can be done in our situation.

Consider the simple linear regression without an intercept:

$$y_i^q = \hat{\beta}^q x_i^q + \hat{\varepsilon}_i^q. \tag{15}$$

The ordinary least squares (OLS) estimate for the slope is then equivalent to:

$$\hat{\beta}^q = \frac{\sum_i y_i^q x_i^q}{\sum_i x_i^q x_i^q} \tag{16}$$

The linear regression model of equation (1) can be re-written in a similar form. Weighted least squares of equation (1) with the weights w_i^q is equivalent to estimating:

$$\begin{aligned}\sqrt{w_i^q} y_i^q = \sqrt{w_i^q}\alpha \\ + \left(\beta^{q=1} + d\beta_1^{q=2}\delta_2^q + d\beta_1^{q=3}\delta_3^q + d\beta_1^{q=4}\delta_4^q\right)\left(\sqrt{w_i^q} x_i^q\right) \\ + \sqrt{w_i^q}\varepsilon_i^q\end{aligned} \tag{17}$$

with OLS. In practice, we can first estimate equation (1) with weighted least squares and save the retrieved weights $\widehat{w}_i^q$ and the estimated intercept $\hat{\alpha}$. We then define $\check{y}_i^q = \sqrt{\widehat{w}_i^q}\left(y_i^q - \hat{\alpha}\right)$, $\check{x}_i^q = \sqrt{\widehat{w}_i^q} x_i^q$ and $\check{\varepsilon}_i^q = \sqrt{\widehat{w}_i^q}\varepsilon_i^q$. Using these new variables, we obtain the linear equation:

$$\check{y}_i^q = \left(\beta^{q=1} + d\beta_1^{q=2}\delta_2^q + d\beta_1^{q=3}\delta_3^q + d\beta_1^{q=4}\delta_4^q\right)\check{x}_i^q + \check{\varepsilon}_i^q \tag{18}$$

When we estimate the regression coefficients in equation (18) with OLS, we obtain exactly the same parameter estimates for the slopes as are obtained by estimating equation (1) with weighted least squares. Furthermore, the slope parameters $\beta^{q=1}$ and $\beta^{q^*} = \beta^{q=1} + d\beta_1^{q=q^*}$ in (18) can also be estimated from simple linear regressions of the form (15) for each quarter separately (using the already estimated intercept $\hat{\alpha}$ from (17) as "input"). This is true because all sums of cross-products of covariates in (18) for different quarters are zero; e.g., $\sum_i \left(\breve{x}_i^q \delta_2^q\right)\left(\breve{x}_i^q \delta_3^q\right) = 0$ for all q. The quarterly slope β^q can thus be estimated, analogously to (16), by:

$$\hat{\beta}^q = \frac{\sum_i \breve{y}_i^q \breve{x}_i^q}{\sum_i \breve{x}_i^q \breve{x}_i^q} = \frac{\sum_i \sqrt{\widehat{w}_i^q}\left(y_i^q - \hat{\alpha}\right)\sqrt{\widehat{w}_i^q}\, x_i^q}{\sum_i \sqrt{\widehat{w}_i^q}\, x_i^q \sqrt{\widehat{w}_i^q}\, x_i^q} = \frac{\sum_i \widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q}{\sum_i \widehat{w}_i^q \left(x_i^q\right)^2} \tag{19}$$

Recall that $\widehat{w}_i^q = \hat{g}_i^q\, d_i^q / \left\{\max\left(x_i^q, 1\right)\right\}^{\hat{\lambda}}$. Ignoring the exceptional cases with $x_i^q < 1$, in the special situation that $\hat{\lambda} = 1$, we can simplify (19) to:

$$\hat{\beta}^q = \frac{\sum_i \hat{g}_i^q\, d_i^q / x_i^q\, \left(y_i^q - \hat{\alpha}\right) x_i^q}{\sum_i \hat{g}_i^q\, d_i^q / x_i^q\, \left(x_i^q\right)^2} = \frac{\sum_i \hat{g}_i^q d_i^q \left(y_i^q - \hat{\alpha}\right)}{\sum_i \hat{g}_i^q d_i^q x_i^q} \tag{20}$$

Notice that in the denominator in (20) one recognises the term $\hat{X} = \sum_i d_i^q x_i^q$ which stands for an estimator of the total VAT turnover. The actual denominator also has a correction for outliers. Similarly, the term $\hat{Y} = \sum_i d_i^q y_i^q$ is an estimator of the total survey turnover. The numerator also has corrections for outliers and for the common intercept.

We re-express equation (19) to see the contribution of the units to the slope, we refer to this as the "quarterly unit slope effect". Let $\hat{c}_i^q = \widehat{w}_i^q \left(x_i^q\right)^2$ be the term of each unit in the denominator and let $\hat{r}_i^q = \hat{c}_i^q / \sum_i \hat{c}_i^q$ be the relative contribution of each unit to the denominator. We further define:

$$\hat{\beta}_i^q = \frac{\widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q}{\widehat{w}_i^q \left(x_i^q\right)^2} = \frac{y_i^q - \hat{\alpha}}{x_i^q} \tag{21}$$

as being the equivalent of $\hat{\beta}^q$ in (19) but now at unit level. It corresponds to the ratio of survey to VAT turnover, but now corrected for the intercept. We obtain:

$$
\begin{aligned}
\hat{\beta}^q &= \frac{\sum_i \widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q}{\sum_i \hat{c}_i^q} \\
&= \sum\nolimits_i \frac{1}{\sum_i \hat{c}_i^q} \widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q \\
&= \sum\nolimits_i \frac{\hat{c}_i^q}{\sum_i \hat{c}_i^q} \{\widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q\} / \hat{c}_i^q \\
&= \sum\nolimits_i \hat{r}_i^q \hat{\beta}_i^q
\end{aligned}
\tag{22}
$$

Notice that when $x_i^q = 0$, $\hat{\beta}_i^q$ is not defined. But in that situation $\hat{r}_i^q = 0$ and $\frac{1}{\sum_i c_i} \widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q = 0$ (second line in equation (22) so those units do not contribute to $\hat{\beta}^q$. For all other units i, the contribution to $\hat{\beta}^q$ is defined to be $\hat{r}_i^q \hat{\beta}_i^q$.

4.1.2 Contribution of units to the quarterly effects of slopes

We are interested to identify the units that contribute to the differences in the slopes between the four quarters of a year. Additionally, we would like to find out which of the patterns, to be defined below in section 4.1.3, contribute most to the quarterly differences in slopes.

Let $\hat{\beta}^+$ be a weighted version of the four quarterly slopes:

$$
\begin{aligned}
\hat{\beta}^+ &= \frac{\sum_q \sum_i \widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q}{\sum_q \sum_i \hat{c}_i^q} \\
&= \frac{\sum_q \left(\sum_i \hat{c}_i^q\right) \left\{\sum_i \widehat{w}_i^q (y_i^q - \hat{\alpha}) x_i^q / \sum_i \hat{c}_i^q\right\}}{\sum_q \sum_i \hat{c}_i^q} \\
&= \frac{\sum_q \sum_i \hat{c}_i^q \hat{\beta}^q}{\sum_q \sum_i \hat{c}_i^q} = \sum\nolimits_q \hat{r}^{q|+} \hat{\beta}^q
\end{aligned}
\tag{23}
$$

with $\hat{r}^{q|+} \equiv \frac{\sum_i \hat{c}_i^q}{\sum_q \sum_i \hat{c}_i^q}$.

In the empirical data at hand, the outcome of $\hat{\beta}^+$ differed at most 0.001 from the estimated slope $\hat{\beta}^0$ that results from fitting a regression with an intercept and one common slope to the data for all four quarters (see Tables 5–7):

$$
y_i^q = \alpha^0 + \beta^0 x_i^q + \varepsilon_i^q \text{ with } q = 1, \dots, 4 \tag{24}
$$

Note that estimation of α^0 in equation (24) will probably lead to another value than in equation (1).

We analyse the differences $\hat{\beta}^q - \hat{\beta}^+$:

$$\hat{\beta}^q - \hat{\beta}^+ = \sum\nolimits_i \hat{r}_i^q \hat{\beta}_i^q - \hat{\beta}^+ = \sum\nolimits_i \hat{r}_i^q \left(\hat{\beta}_i^q - \hat{\beta}^+\right) = \sum\nolimits_i \hat{r}_i^q \tilde{\beta}_i^q \tag{25}$$

where $\tilde{\beta}_i^q$ is defined as: $\tilde{\beta}_i^q = \hat{\beta}_i^q - \hat{\beta}^+$. The right-hand side of equation (25) holds because $\sum_i \hat{r}_i^q = 1$. Expression (25) gives the absolute difference of the 'slope' $\hat{\beta}_i^q$ compared to $\hat{\beta}^+$ for each unit i. However, to go one step further we are interested to analyse how this difference varies with the four quarters. For instance, assume that we found $\hat{\alpha} = 0$, $\hat{\beta}^+ = 1$ and for unit $i = i_0$ we found $\hat{\beta}_{i=i_0}^{q=1} = 1.1$, $\hat{\beta}_{i=i_0}^{q=2} = 1.11$, $\hat{\beta}_{i=i_0}^{q=3} = 1.09$ and $\hat{\beta}_{i=i_0}^{q=4} = 1.01$. Then for all of the four quarters the sample survey turnover is larger than the VAT turnover, but in the fourth quarter of the year relatively more VAT turnover is reported than in the other three quarters. The seasonal VAT turnover pattern for this unit is shifted towards the fourth quarter of the year, relative to the survey turnover.

We can quantify this effect as follows. Define the weighted average value of the quarterly values for $\tilde{\beta}_i^q$ as:

$$\bar{\tilde{\beta}}_i = \frac{1}{4}\sum\nolimits_{q=1}^{4} \hat{r}_i^q \tilde{\beta}_i^q \tag{26}$$

Now, we consider the differences:

$$d\tilde{\beta}_i^q = \hat{r}_i^q \tilde{\beta}_i^q - \bar{\tilde{\beta}}_i \tag{27}$$

A negative value of $d\tilde{\beta}_i^{q=4}$ implies the slope of the regression between turnover and VAT is relatively smaller in the fourth quarter that the weighted average over the year, a positive value implies the opposite. We refer to expression (27) as the "normalised quarterly unit slope effect".

The sum of $d\tilde{\beta}_i^q$ over all units is not equal to the quarterly slope effect, $\sum_i d\tilde{\beta}_i^q \neq \hat{\beta}^q$, because we have 'normalised' the quarterly unit slope effect with the yearly mean at unit level ($\bar{\tilde{\beta}}_i$). In fact,

$$\begin{aligned}
\sum\nolimits_i d\tilde{\beta}_i^q &= \sum\nolimits_i \hat{r}_i^q \tilde{\beta}_i^q - \sum\nolimits_i \bar{\tilde{\beta}}_i \\
&= \sum\nolimits_i \hat{r}_i^q \tilde{\beta}_i^q - \frac{1}{4}\sum\nolimits_{q^*=1}^{4} \sum\nolimits_i \hat{r}_i^{q^*} \tilde{\beta}_i^{q^*} \\
&= \left(\hat{\beta}^q - \hat{\beta}^+\right) - \frac{1}{4}\sum\nolimits_{q^*=1}^{4} \sum\nolimits_i \left(\hat{\beta}^{q^*} - \hat{\beta}^+\right) \\
&= \hat{\beta}^q - \frac{1}{4}\sum\nolimits_{q^*=1}^{4} \hat{\beta}^{q^*}.
\end{aligned}$$

But similarly it is easy to show that difference of the sum of $d\tilde{\beta}_i^q$ over all units between two quarters, say quarter $q = a$ versus $q = b$, equals the difference in slopes between those two quarters:

$$\begin{aligned}
&\sum\nolimits_i d\tilde{\beta}_i^{q=a} - \sum\nolimits_i d\tilde{\beta}_i^{q=b} \\
&\qquad = \sum\nolimits_i \left(\hat{r}_i^{q=a} \tilde{\beta}_i^{q=a} - \bar{\tilde{\beta}}_i\right) - \left\{\sum\nolimits_i \left(\hat{r}_i^{q=b} \tilde{\beta}_i^{q=b} - \bar{\tilde{\beta}}_i\right)\right\} \qquad (28) \\
&\quad = \hat{\beta}^{q=a} - \hat{\beta}^{q=b} \text{ with } (a, b) \in \{1, 2, 3, 4\}
\end{aligned}$$

The value $d\tilde{\beta}_i^q$ thus describes the contribution of each unit to the quarterly slope-effect. We sort all units in the population by the absolute value of $d\tilde{\beta}_i^q$ and then plot the cumulative value $\sum_{i=1}^{k} d\tilde{\beta}_i^{q=a}$ up to the first k units. We can then see what the combined effect of the k units on the quarterly slope effect is.

The slope effect does not have to be zero, we do accept a certain difference between the two sources. If this difference occurs year after year this concerns a bias rather than a variance. The bias of the survey design is expected to be close to zero but requiring a bias of zero would be a very strict norm. The business statistics division who compiles the output uses the norm that the month-on-month growth rates of Manufacturing, Retail trade and Construction should be changed by at most 0.4 per cent points due to the benchmarking process. The data that we have available for our analysis cannot directly be used to compare with the 0.4 per cent norm, for three reasons. First, the data that we analyse in the present study concern only the non-topX part of the population. Because the top-X part does not lead to an adjustment in benchmarking, this implies we should use a wider norm for our data. Second, we use quarterly data whereas the norm refers to monthly data. Effects of benchmarking on monthly growth rates will be at least as large as effects on the quarterly growth rates, since a quarterly index is the average over three monthly indices. This implies that a norm for quarterly growth rates should be stricter. Third, in our data we analyse the effect of units for specific quarters. That effect will be a mixture of random and structural measurement effects, instead of only a structural measurement effect.

In order to get an *indication* of the effects of patterns and of units on the estimated slopes, we used the following approach to find a margin for the cumulative value $\sum_{i=1}^{k} d\tilde{\beta}_i^{q=a}$. If the turnover in the first quarter is underestimated by 0.0015 and the fourth quarter is overestimated by 0.0015, this leads to a shifted quarterly growth rate of 0.3 per cent points. Since the intercept is close to zero, this effect of under- and overestimation can be achieved by a shift in the slope of ± 0.0015. If the cumulative effect remains within the margin then the quarterly benchmarking revision remains within 0.3 per cent points. (Note, the Figures 8–11 and 16–20 show that when the cumulative effect stays within the margin, they stay within the upper or lower margin.) This is slightly stricter than the above-mentioned 0.4 per cent norm. Because this norm is not directly applicable to our situation as noted above, we have chosen a conservative margin.

4.1.3 Contribution of the patterns to the quarterly effects of slopes

As explained before, we are interested to understand the causes of the differences in quarterly slopes within a year. The simplest situation would be when quarterly slope-differences are caused by units with certain reporting patterns. In that case we might try to find the cause of those patterns. Are they due to measurement errors in the survey or in the VAT data? We might use this to correct the data with a specific pattern. In this section we will first define a set of reporting patterns. Then we will show how the compute the contribution of units with a reporting patterns to the quarterly slope effects.

Defining the reporting patterns. We used a simple approach to appoint enterprises to a reporting pattern for each year. In each quarter q we compare the difference 'quarterly sample survey minus quarterly VAT turnover' (both expressed in 1000 euros, $y_i^q - x_i^q$) with $\hat{\sigma}_i^q$, the estimated standard deviation obtained from the regression. The estimated standard deviation for unit i in quarter q corrected for heteroscedasticity, for the design weight and for the outlier weight is $\hat{\sigma}_i^q = \sqrt{\hat{\bar{\sigma}}^2/\widehat{w}_i^q}$. The design weight is included because we aim to estimate the standard deviation of the disturbances at population level rather than at sample level. For a quarter q we considered VAT to be larger ("L") than the sample survey turnover if $y_i^q - x_i^q > \hat{\sigma}_i^q$, smaller ("S") if $y_i^q - x_i^q < -\hat{\sigma}_i^q$ and equal ("E") otherwise. For instance, we might obtain the pattern "LLLS" which implies that VAT is larger in quarter 1–3 and smaller in quarter 4 than the survey turnover by at least one standard deviation. This way a total of 3^4 = 81 different patterns is obtained for each year. Note that $\hat{\sigma}_i^q$ is the residual obtained in the regression for comparing $y_i^q - \hat{y}_i^q$. In our situation, the obtained quarterly regressions yielded an intercept, slope combination close to (0, 1), so the expected value $\hat{y}_i^q$ for y_i^q is close to x_i^q. We therefore used $\hat{\sigma}_i^q$ as an approximation for the (weighted) standard deviation of $y_i^q - x_i^q$. In a preliminary study we first computed the relative proportion of quarterly turnover within a year for both sources and then determined the categories larger, equal and smaller. We found that a part of the units were now appointed to another pattern, but that the two main conclusions from this analysis were not changed. Those two main conclusions are that a large number of units contribute to quarterly effects and the units have different patterns over time. We therefore decided to only present the results based on the actual turnover levels (without normalising to the yearly turnover values).

There is one special reporting situation to be aware of. Sometimes persons report turnover on the survey questionnaire that excludes the VAT payment itself, whereas the survey requested to report the turnover that includes the VAT payment (for sector Retail Trade). Also the opposite might occur, people reporting turnover that includes the VAT payment whereas the survey asks to exclude the amount of VAT paid (the other economic sectors). In a preliminary computation we found that this reporting error occurred rarely, with a maximum of 1% at industry level. We have therefore not classified this as a separate pattern.

Contribution of units within a pattern to slope effects. Let $\mathcal{G}_p$ denote a set of units that are appointed to the same pattern p. We are interested to quantify the contribution of each of the patterns to the quarterly slope effects.

Analogously to equation (22) we express the overall quarterly slope $\hat{\beta}^q$ as a linear sum of a quarterly 'slopes per pattern' $\hat{\beta}_p^q$:

$$
\begin{aligned}
\hat{\beta}^q &= \frac{\sum_i \widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q}{\sum_i \hat{c}_i^q} \\
&= \frac{\sum_p \sum_{i \epsilon \mathcal{G}_p} \widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q}{\sum_i \hat{c}_i^q} \\
&= \sum_p \frac{\sum_{i \epsilon \mathcal{G}_p} \hat{c}_i^q}{\sum_i \hat{c}_i^q} \left(\sum_{i \epsilon \mathcal{G}_p} \widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q \Big/ \sum_{i \epsilon \mathcal{G}_p} \hat{c}_i^q \right) \\
&= \sum_p \hat{r}_p^q \hat{\beta}_p^q
\end{aligned}
\tag{29}
$$

where $\hat{\beta}_p^q$ is defined as:

$$
\hat{\beta}_p^q = \frac{\sum_{i \epsilon \mathcal{G}_p} \widehat{w}_i^q \left(y_i^q - \hat{\alpha}\right) x_i^q}{\sum_{i \epsilon \mathcal{G}_p} \hat{c}_i^q} \tag{30}
$$

and $\hat{r}_p^q$ as:

$$
\hat{r}_p^q = \frac{\sum_{i \epsilon \mathcal{G}_p} \hat{c}_i^q}{\sum_i \hat{c}_i^q} . \tag{31}
$$

Likewise to the analysis of the quarterly effects of the units, we are interested in the quarterly effects of the patterns compared to the overall common yearly slope. First of all, we define the differences

$$
\tilde{\beta}_p^q = \hat{\beta}_p^q - \hat{\beta}^+ . \tag{32}
$$

These differences are referred to as the "quarterly pattern slope effect".

Analogously to (26) and (27), for each pattern, we compute the weighted mean of $\tilde{\beta}_p^q$ over the four quarters of the year:

$$
\bar{\tilde{\beta}}_p = \frac{1}{4} \sum_{q=1}^{4} \hat{r}_p^q \tilde{\beta}_p^q \tag{33}
$$

and we consider the difference:

$$
d\tilde{\beta}_p^q = \hat{r}_p^q \tilde{\beta}_p^q - \bar{\tilde{\beta}}_p \tag{34}
$$

This latter difference is referred to as the "normalised quarterly pattern slope effect".

A negative value of $d\tilde{\beta}_p^{q=4}$ implies that the component $\hat{r}_p^q \hat{\beta}_p^q$ is relatively smaller in the fourth quarter of the year than the weighted average $\bar{\tilde{\beta}}_p$, thus that for units in $\mathcal{G}_p$ the ratio of survey to VAT turnover (ignoring the intercept) is smaller in the fourth quarter.

Likewise to $d\tilde{\beta}_i^q$, the sum of $d\tilde{\beta}_p^q$ over all patterns is not equal to the quarterly slope, $\sum_p d\tilde{\beta}_p^q \neq \hat{\beta}^q$. But the difference of the sum of $d\tilde{\beta}_p^q$ over all patterns between two quarters, say quarter $q = a$ versus $q = b$, equals the difference in slopes between those two quarters:

$$\sum\nolimits_p d\tilde{\beta}_p^{q=a} - \sum\nolimits_p d\tilde{\beta}_p^{q=b}$$
$$= \sum\nolimits_p \left(\hat{r}_p^{q=a}\tilde{\beta}_p^{q=a} - \bar{\tilde{\beta}}_p\right) - \left\{\sum\nolimits_p \left(\hat{r}_p^{q=b}\tilde{\beta}_p^{q=b} - \bar{\tilde{\beta}}_p\right)\right\} \tag{35}$$
$$= \hat{\beta}^{q=a} - \hat{\beta}^{q=b} \text{ with } (a, b) \in \{1, 2,3, 4\}$$

The value $d\tilde{\beta}_p^q$ thus describes the normalised contribution of each pattern to the quarterly slope-effect. For each quarter, we sort all patterns in the population by the absolute value of $d\tilde{\beta}_p^q$ and then plot the cumulative value $\sum_{p=1}^{P} d\tilde{\beta}_p^{q=a}$ up to the first P patterns. We can then analyse what the accumulated normalised effect of the patterns on the quarterly slope effects is.

In the current paper, we used the differences $\tilde{\beta}_p^q = \hat{\beta}_p^q - \hat{\beta}^+$, $\hat{r}_p^q\tilde{\beta}_p^q - \bar{\tilde{\beta}}_p$, $\tilde{\beta}_i^q = \hat{\beta}_i^q - \hat{\beta}^+$, and $\hat{r}_i^q\tilde{\beta}_i^q - \bar{\tilde{\beta}}_i$. Note that we have considered to use a number of other 'differences', but they did not work out well. For instance, $\hat{\beta}_i^q - \hat{\beta}^q$ and $\hat{\beta}_i^q - \hat{\beta}_p^q$ have the disadvantage that they do not show the effect on the quarterly slope differences. Furthermore, we tried to use $\hat{\beta}_p^q - \hat{\beta}_p$, to analyse the quarterly effects per pattern: The disadvantage of the latter expression is that the difference between two quarters ($q = a$, $q = b$) for the sum over all patterns – the equivalent of equation (35) – has no clear interpretation because $\sum_p \hat{r}_p^{q=a}\hat{\beta}_p \neq \sum_p \hat{r}_p^{q=b}\hat{\beta}_p$.

4.2 Results

We will now analyse whether a limited number of units contribute to the quarterly slope effects or not. The simplest form would be when this limited set of units have certain reporting patterns in common. That analysis is described in section 4.2.1. A more complicated form is to directly account for the contribution of all units to the slope. That is discussed in sections 4.2.2 and 4.2.3.

4.2.1 Contribution of patterns to slope differences

The distribution of the units over the different patterns, as defined in section 4.1 is given in **Figure 7**. For all economic sectors and years, the largest group of units is appointed to pattern "EEEE" (range: 16.2 – 53.7 per cent). Recall that a unit is appointed to pattern "E" in a single quarter when the difference between the survey and the VAT value is within one (weighted) standard deviation. For a normal distribution, the standard deviation covers 68 per cent of the data. The second most frequent pattern was "LLLL" (range 6.1 – 22.1 per cent), that occurred far less frequently than "EEEE". We computed the average frequencies over the economic sectors and years. In reversed order we thus obtained, with percentage between brackets: “EEEE” (33.6), “LLLL” (11.2), “EEEL” (6.2), “EELE” (2.9), “LEEE” (2.8), “ELEE” (2.6),”ELLL”(2.5), “EEES” (2.3), “SEEE”(2.3) and "EELL"(2.1). Other patterns had an average contribution smaller than 2 per cent. It is consistent with our earlier finding that the VAT turnover totals are larger than the survey turnover totals (**Table 1**).

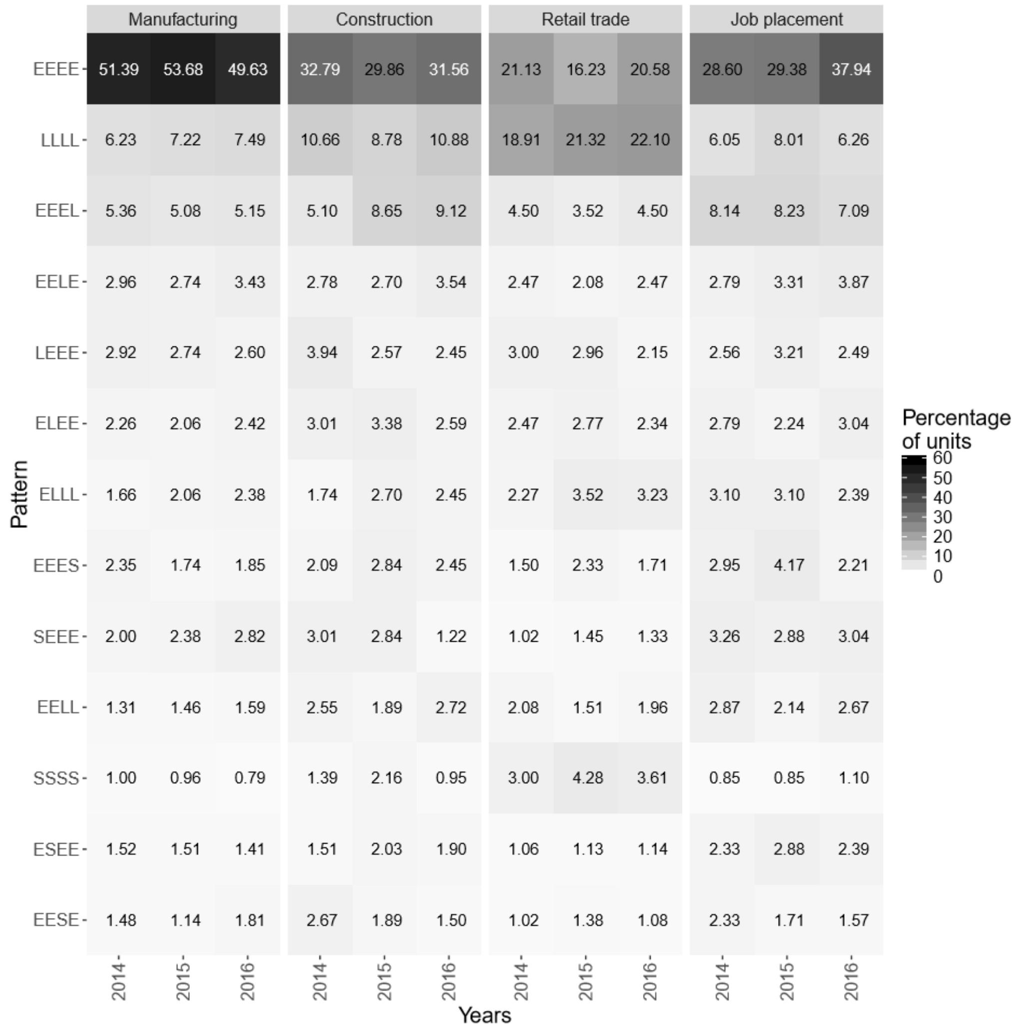

Figure 7. Distribution of the units over the patterns (in percentages), for the four industries and three years. Patterns are sorted by their average fraction (over the three years and three economic sectors), and only patterns with an average >=1.5 per cent are shown.

We also verified whether the patterns per unit were stable over time or not. We selected units that reported turnover for eight subsequent quarters. This concerned about half the number of units that report turnover for four subsequent quarters. **Table 11** gives for each pattern, the total number of units and the number with the same pattern in the subsequent year, both for the 2014 – 2015 comparison and for the 2015 – 2016 comparison. **Table 11** shows that for most of the units the pattern changes from one year to the next.

The table also reveals that for the patterns "EEEE", "LLLL" and "SSSS" there are relatively many units that do remain in the same category in two subsequent years. A further inspection of the microdata revealed within the units that kept the "EEEE" pattern for three years, there is a group of enterprises which report nearly the same values for the quarterly VAT turnover, but there are also enterprises within this group with larger differences. Further, we found that some of the enterprises that report nearly the same turnover values for VAT as for the survey do so for 12 subsequent quarters, others do so in 11 of the 12 quarters and others do so for one or two years but not in the other ones. This means that 'reporting nearly the same quarterly value' is not necessarily done for each year again.

Table 11. Total number of units per pattern for units in the selection for eight consecutive quarters and the number with same pattern, for two periods. (Top 25 most frequent patterns).

Pattern	2014 – 2015		2015 – 2016	
	Total	Same	Total	Same
EEEE	1660	1008	1207	785
LLLL	438	265	330	189
EEEL	224	27	178	25
EELE	120	8	93	4
LEEE	124	4	76	2
ELEE	110	10	89	7
EEES	102	8	69	5
ELLL	89	11	74	9
SEEE	88	5	57	4
EELL	87	3	48	2
SSSS	66	40	57	31
EESE	68	2	43	4
ESEE	63	4	42	3
LELL	56	1	42	2
ELEL	52	3	38	1
LLLE	57	4	32	0
LEEL	49	3	34	0
LLEL	54	4	26	3
LLEE	42	0	32	0
ELLE	49	1	22	0
LELE	43	2	21	1
SEEL	32	2	20	2
EESL	31	2	14	1
SLLL	30	3	14	0
SLEE	18	1	15	0

Analysis of the microdata further showed that within the "LLLL" pattern there is a group of units that report structurally larger VAT values in three subsequent years. But this behaviour is also not consistent over time, since there are also units with large deviations in one year but not in the other two years. Nearly all of the enterprises in the pattern "SSSS" had structurally smaller VAT values in the three subsequent years.

These results suggest that the reporting behaviour in terms of the patterns as defined in the current paper are not consistently applied by the enterprises over time.

Further, we analysed the contribution of each of the patterns to the differences between the quarterly slope and the slope that would have been obtained when we do not allow for a quarterly effect, according to equation (29) and (32): $\hat{\beta}^q - \hat{\beta}^+ = \sum_p \hat{r}_p^q (\hat{\beta}_p^q - \hat{\beta}^+) = \sum_p \hat{r}_p^q \tilde{\beta}_p^q$. The quantity $\hat{r}_p^q \tilde{\beta}_p^q$ describes the contribution of pattern p to the quarterly slope effect, also referred to as the "quarterly pattern slope effect". For each quarter, we sort all patterns in the population by the absolute value of $\hat{r}_p^q \tilde{\beta}_p^q$ and then plot the cumulative value $\sum_{p=1}^{P} \hat{r}_p^q \tilde{\beta}_p^q$ up to the first P patterns. We can then analyse the accumulated effect of the patterns on the quarterly slope effects. We refer to this in short as the cumulative effects of the patterns.

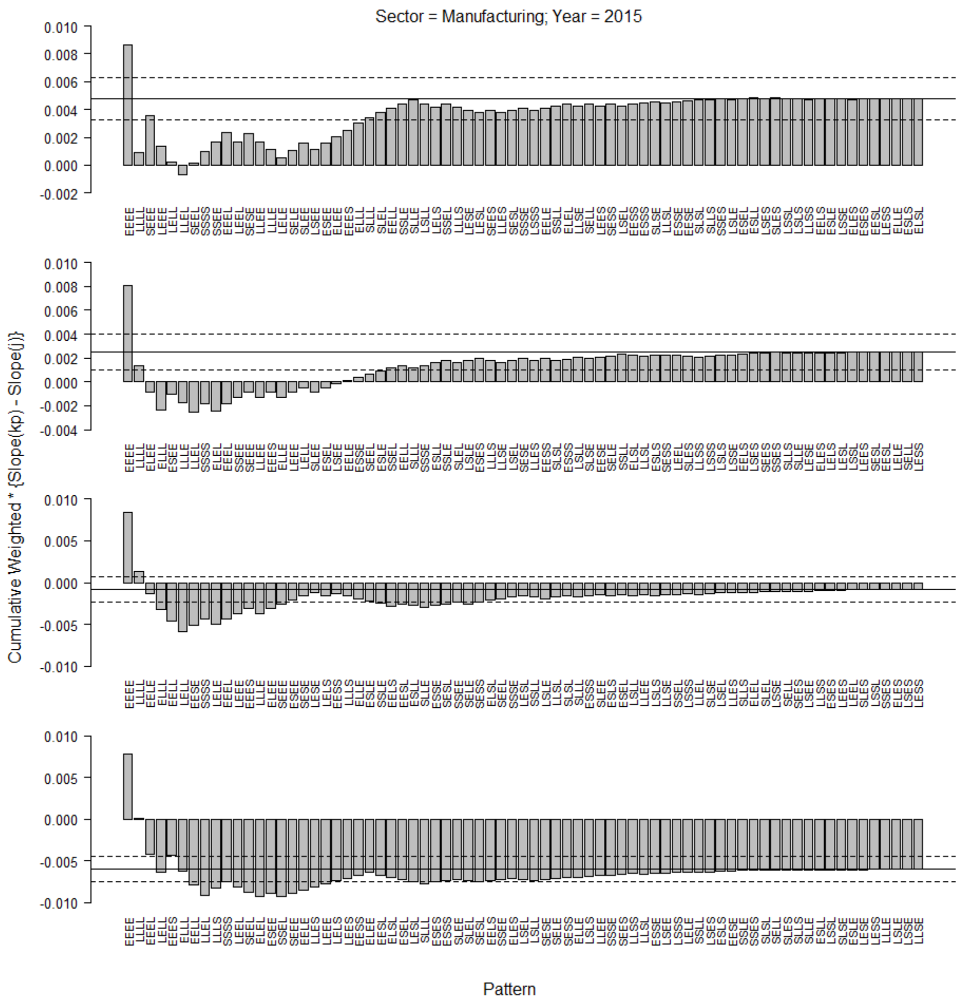

Figure 8. Cumulative effect of the patterns to the difference between the quarterly slope and the yearly slope, for Manufacturing 2015. From top to bottom: quarter 1 to 4. Straight line represents the final quarterly slope difference and the dotted lines represent 0.0015 margin line.

Here we give an example for each Industry: manufacturing (2015), Construction (2015), Retail Trade (2015) and Job placement (2016). The selected economic sector-year combinations each have a significantly smaller slope in the fourth quarter of the year.

Figure 8 –11 show that, overall, patterns “EEEE”, “LLLL” and “EEEL” are the three patterns that have the largest effect on the slope of the fourth quarter, where “EEEE” has a positive slope effect while “LLLL” and “EEEL” have a negative slope effect. All of the patterns have a small effect on the slope. If we accept a slope effect that lies within the 0.0015 margin lines, then still we need to consider 34 different patterns in case of the fourth quarter for Manufacturing 2015. For Construction 2015 37 patterns, for Retail trade 2015 57 patterns and for Job placement 2016 35 patterns were needed for the cumulative effect of the fourth quarter to stay within the margins. For the other quarter, also a large number of patterns was needed before the cumulative slope effect remained within its margins. This implies that the quarterly effect cannot be explained by considering all units within a limited set of patterns.

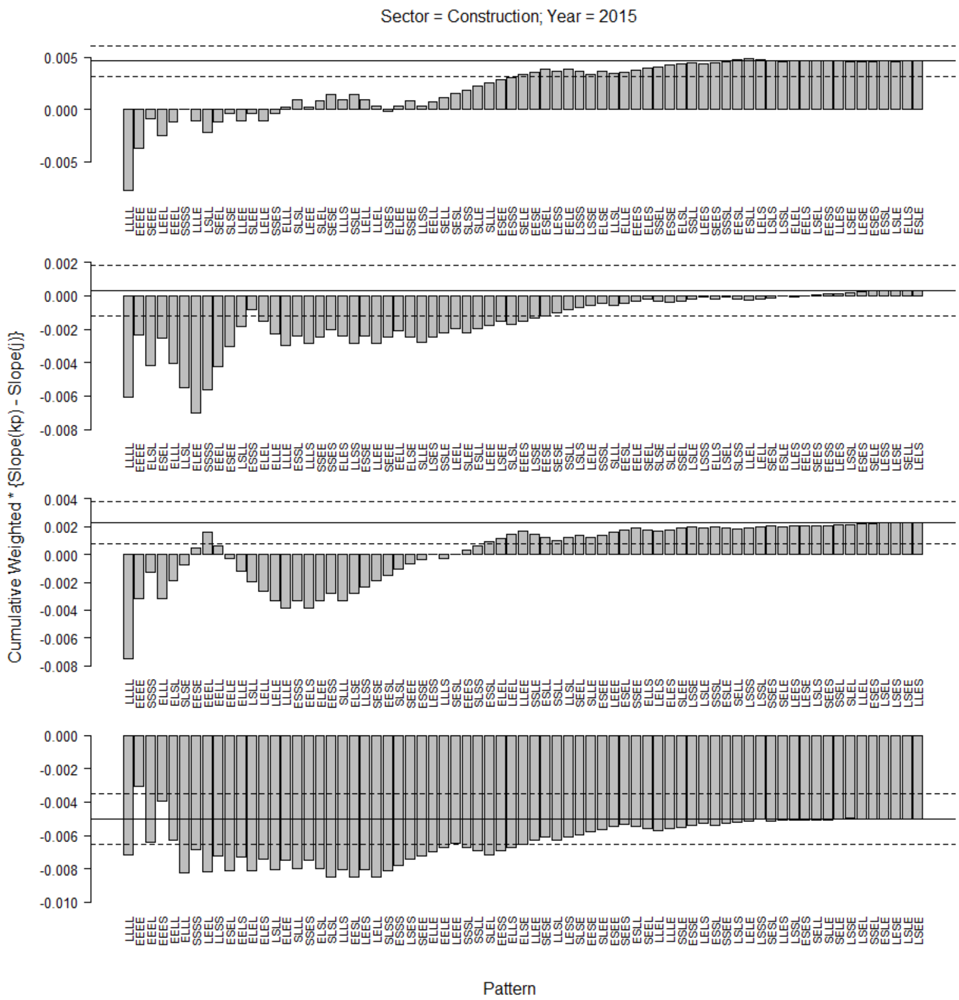

Figure 9. Cumulative effect of the patterns to the difference between the quarterly slope and the yearly slope for Construction, 2015. Straight line represents the final quarterly slope difference and the dotted lines represent 0.0015 margin line.

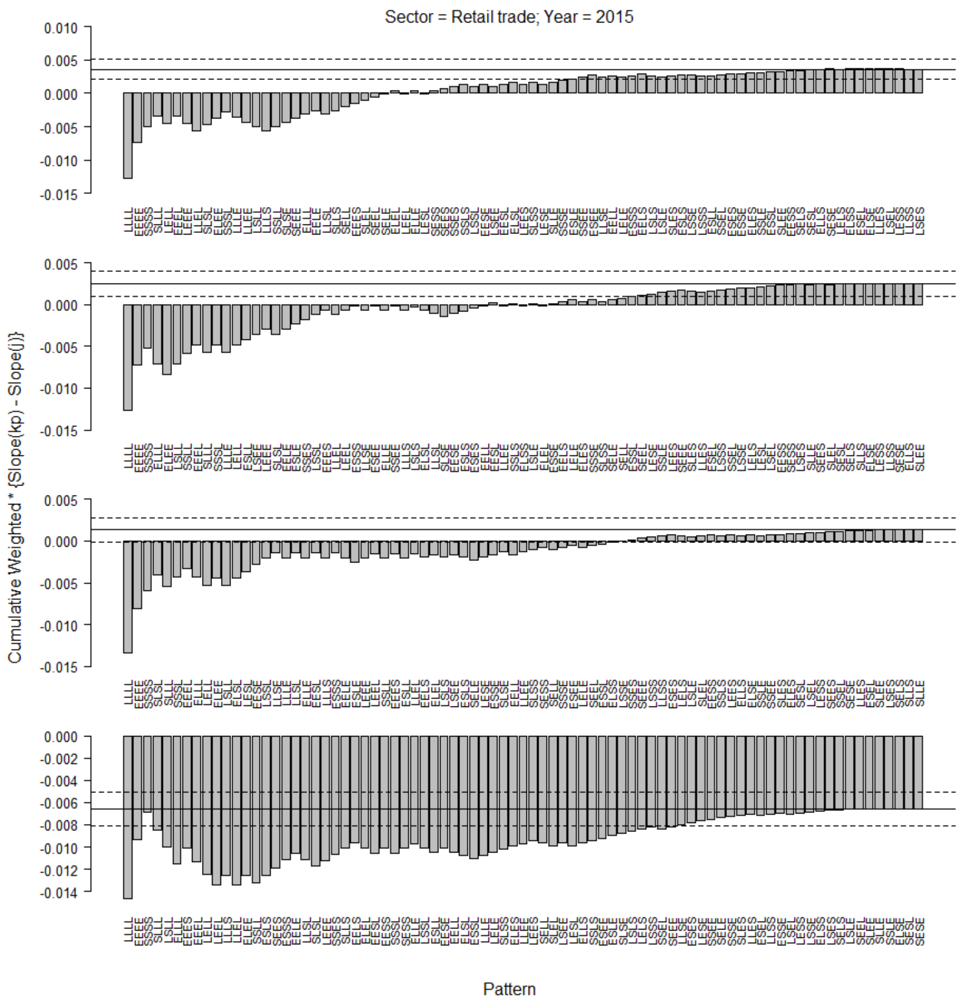

Figure 10. Cumulative effect of the patterns to the difference between the quarterly slope and the yearly slope for Retail trade, 2015. Straight line represents the final quarterly slope difference and the dotted lines represent 0.0015 margin line.

Next, we verified the normalised contribution of each of the patterns to the quarterly effect, in terms of $d\tilde{\beta}_p^q$ from (34). These effects are difficult to see in the Figures 8–11, since those figures show the cumulative effect. The effect in terms of $d\tilde{\beta}_p^q$ is illustrated in Figure 2 where units 1 and 2 both support a regression line with a smaller slope in the fourth quarter relative to the first quarter. We normalised the quarterly effects per patterns to the average quarterly effect per pattern. The result is shown in Figure 12 for Manufacturing 2015. The patterns are sorted by the absolute value of $d\tilde{\beta}_p^q$ in the fourth quarter ($q = 4$). The largest downwards effect on the slope of the fourth quarter is due to units in pattern "EEEL". When we look back to Figure 8 with the cumulative pattern effects we see for "EEEL" that it increased the cumulative effect in Q1, Q2 and Q3 whereas it strongly decreased the cumulative effect in Q4, a result that is consistent with Figure 12. From Figure 8 you can see that units in pattern "EEEE" are above the regression line in all four quarters, combining this with Figure 12 one can see that Q1 lies above its average whereas Q4 lies below its average.

The quarterly effects of the other three economic sectors are found in Figures 13–15. In all four figures it is seen that pattern "EEEL" has the largest normalised quarterly effect $d\tilde{\beta}_p^q$ in the fourth quarter ($q = 4$). Additionally, we find that there are a number of other quarterly patterns that contribute to a quarterly effect in the fourth quarter.

Furthermore, Figure 12 confirms that there are many patterns, each with small additional contributions to the quarterly slope. Figures 12 – 15 show that “EELE” leads to a quarterly effect in the third quarter, “ELEE” to an effect in the second quarter and “LEEE” to an effect in the first quarter. The figures show that pattern “EEEE” also has some (small) quarterly effects, but the direction and extent of its effect differs between the economic sectors and for a given economic sector it varies between years (not shown).

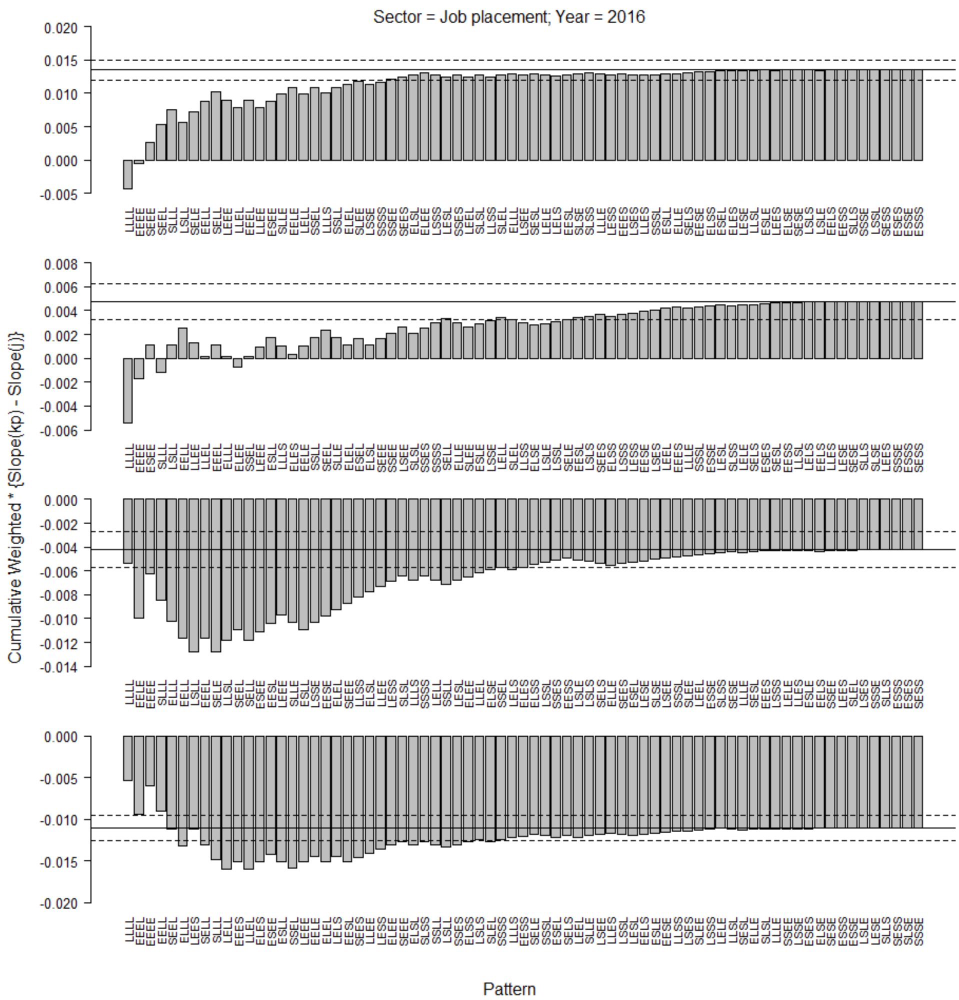

Figure 11. Cumulative effect of the patterns to the difference between the quarterly slope and the yearly slope for Job placement, 2016. Straight line represents the final quarterly slope difference and the dotted lines represent 0.0015 margin line.

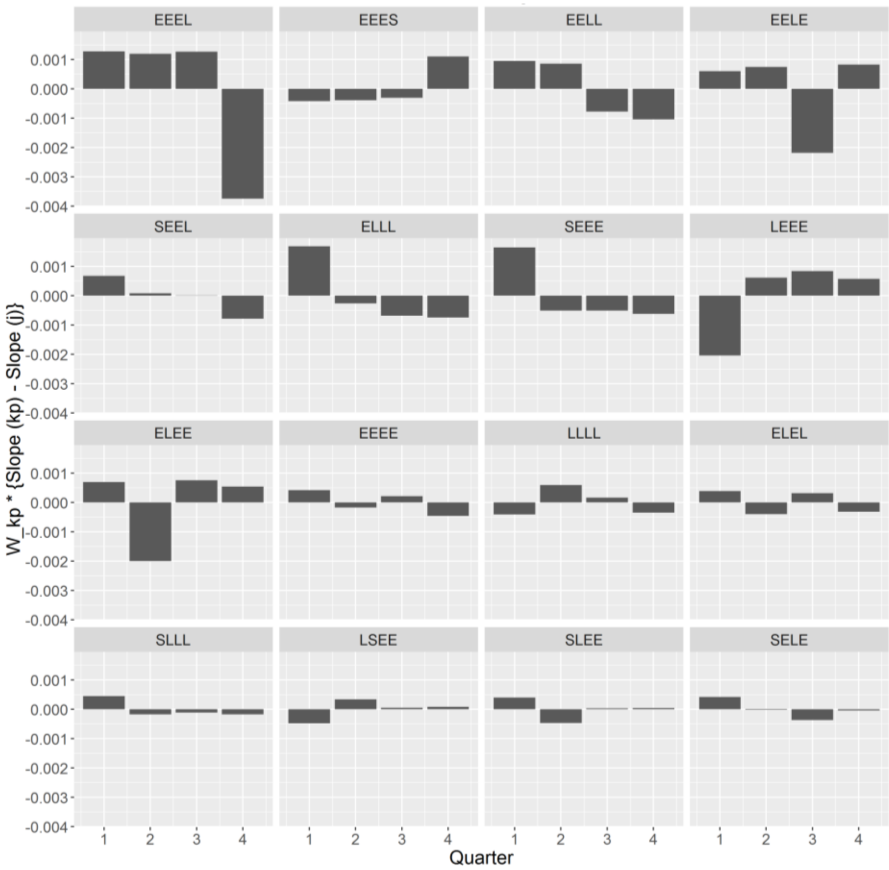

Figure 12. Normalised quarterly effect for first 16 patterns, Manufacturing 2015.

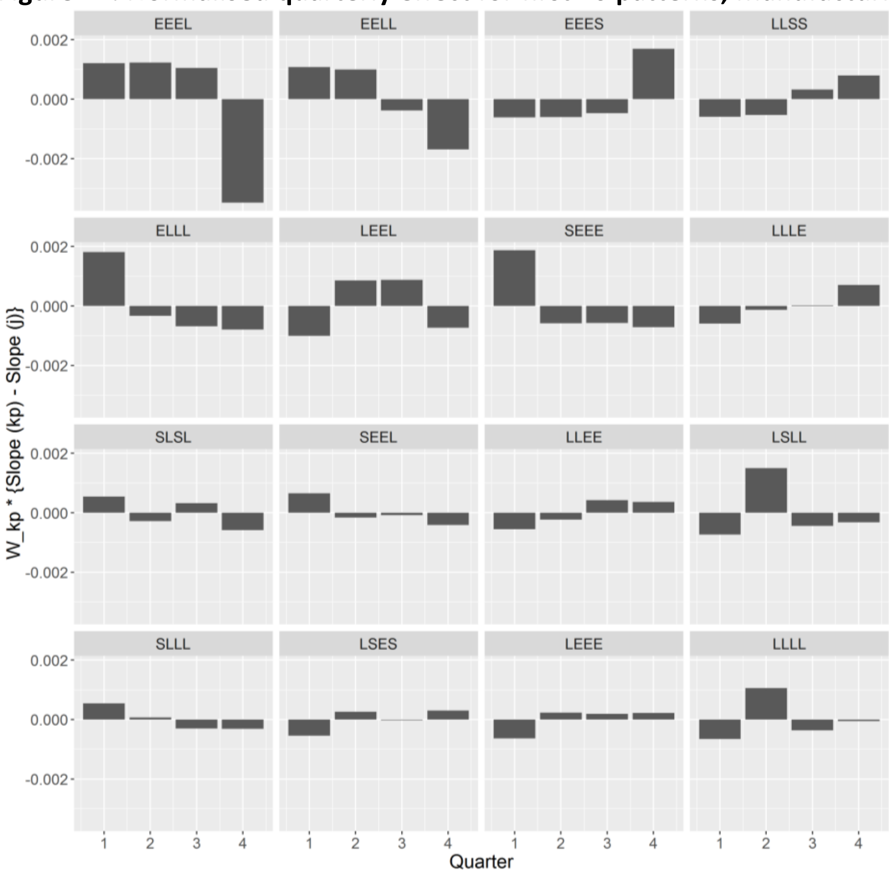

Figure 13. Normalised quarterly effect for first 16 patterns, Construction 2015.

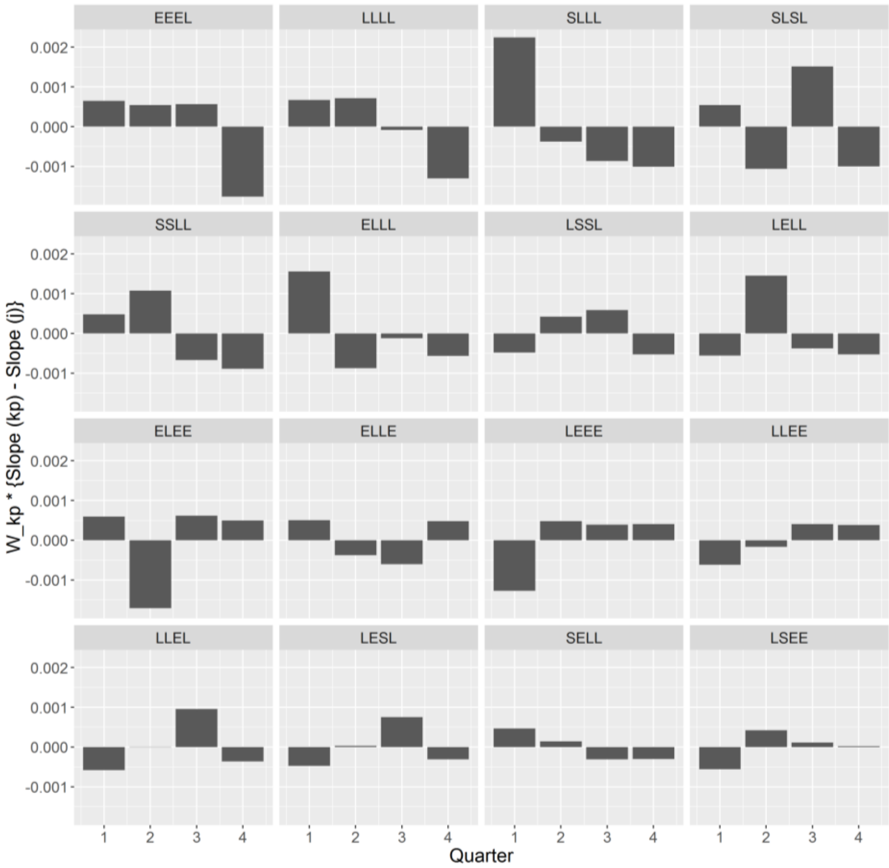

Figure 14. Normalised quarterly effect for first 16 patterns, Retail trade 2015.

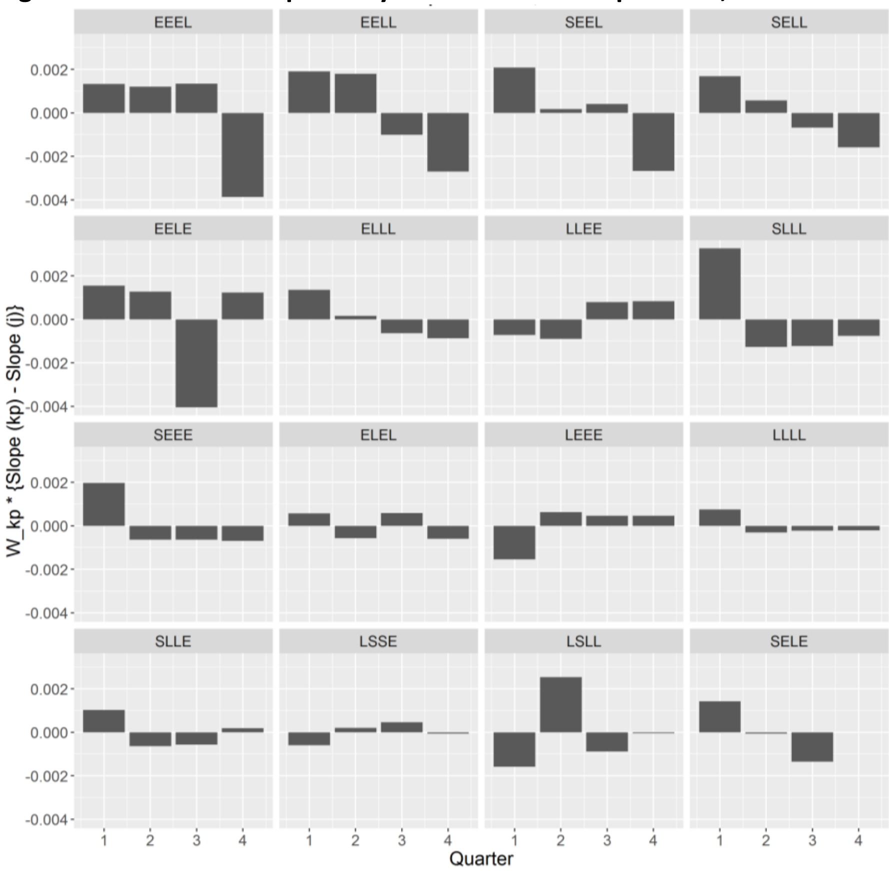

Figure 15. Normalised quarterly effect for first 16 patterns, Job Placement 2016.

An alternative to **Figure 12** – 15 is to plot their absolute contribution to the slopes, thus $\hat{r}_p^q \tilde{\beta}_p^q$ from (30) rather than their normalised versions $d\tilde{\beta}_p^q$. These were not shown, because then the relative effects on the quarterly slopes are more difficult to see. But the absolute contributions make it easier to analyse whether units within a certain pattern are above or below the regression line. To avoid an abundance of figures, these were omitted here.

4.2.2 Results: cumulative contribution of units to slope differences

Analogously to the cumulative effect of the patterns to the difference between the quarterly and the yearly slopes, we analysed the cumulative effect of the units themselves to that difference. The reason behind this is that maybe, within each pattern, only a limited number of units can explain the effects already. We analysed the cumulative value of the first k units of the "quarterly unit slope effect", i.e. $\sum_{i=1}^{k} \hat{r}_i^q \tilde{\beta}_i^q$ according to (25) and the cumulative value of the "normalised quarterly unit slope effect" $\sum_{i=1}^{k} d\tilde{\beta}_i^q$ according to (27). This first will be referred to in short as the cumulative effect and the second as the cumulative normalised effect. The cumulative effect is plotted with black lines (label "betatilde") and the cumulative normalised effect in blue lines (label "dbetatilde") in Figures 16–20.

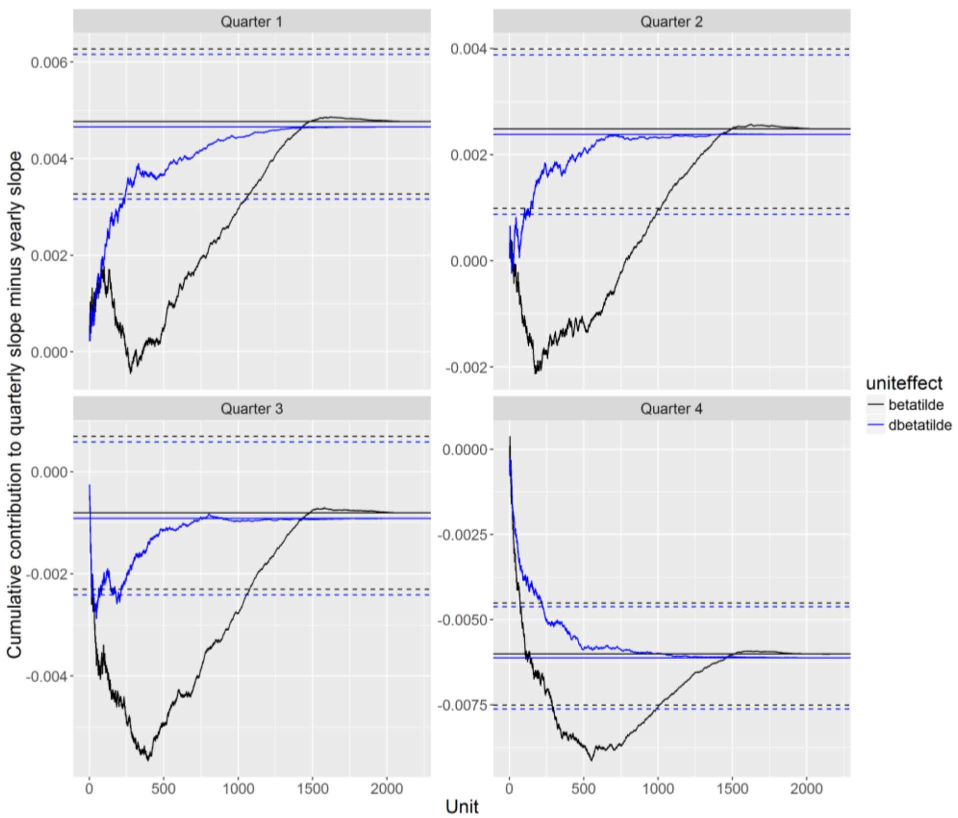

Figure 16. Cumulative effect of the units to the difference between the quarterly slope and the yearly slope, for Manufacturing 2015. Straight line represents the final quarterly slope difference and the dotted lines represent 0.0015 margin lines.

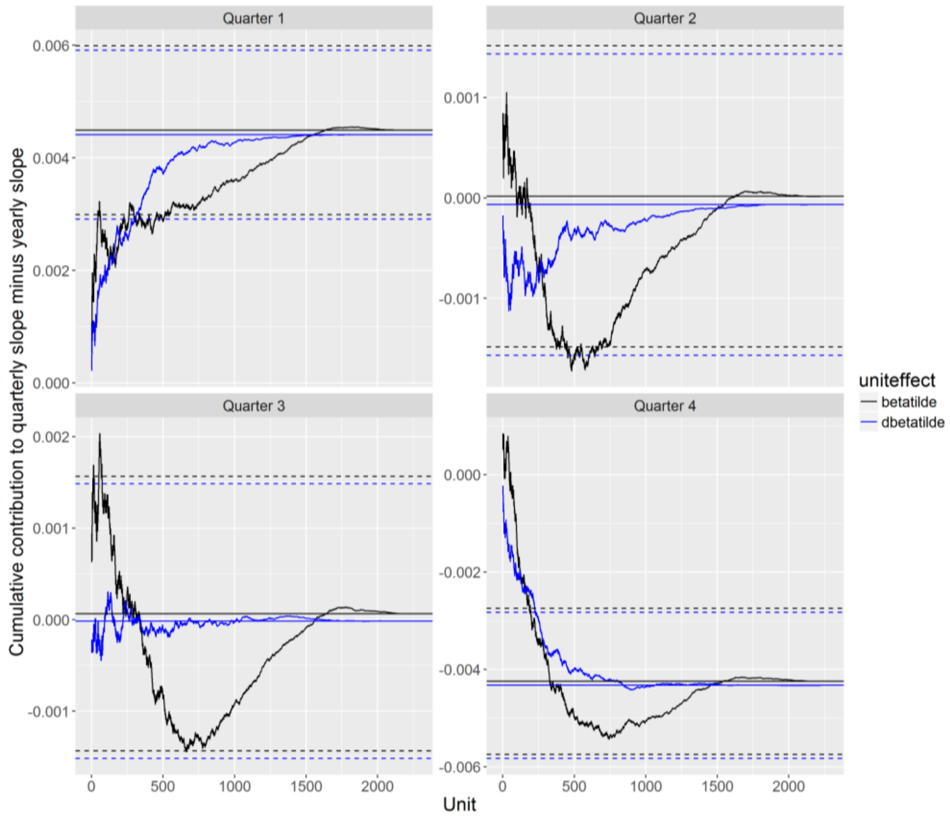

Figure 17. As figure 16, for Manufacturing 2016.

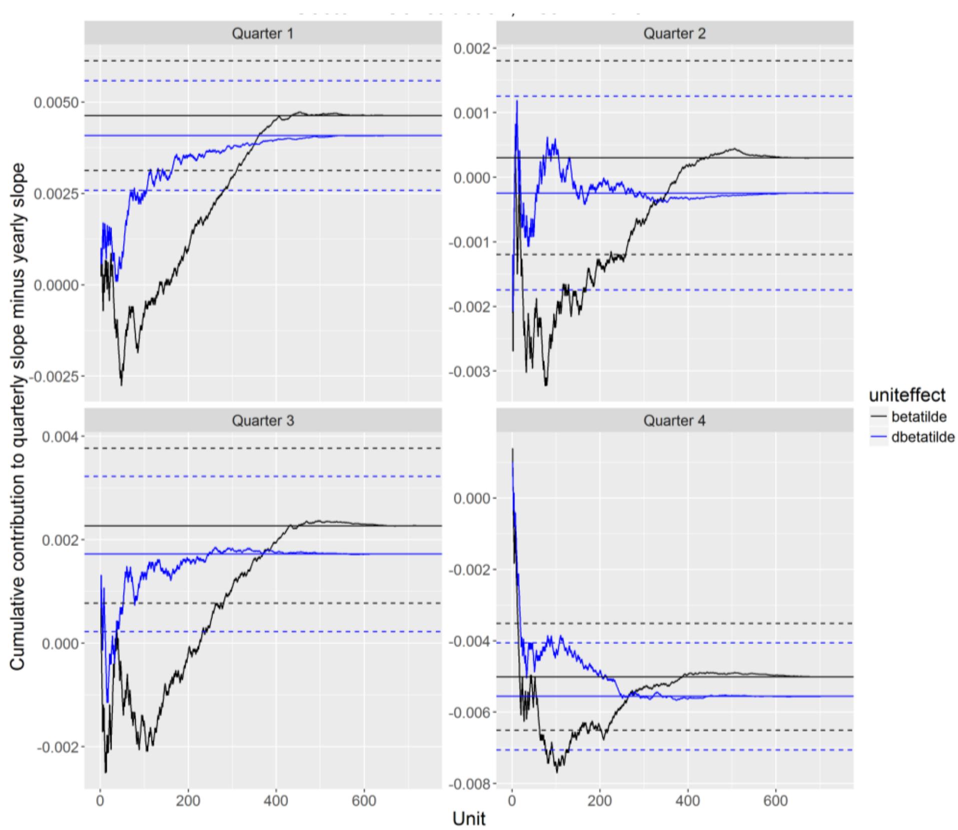

Figure 18. As Figure 16, for Construction 2015.

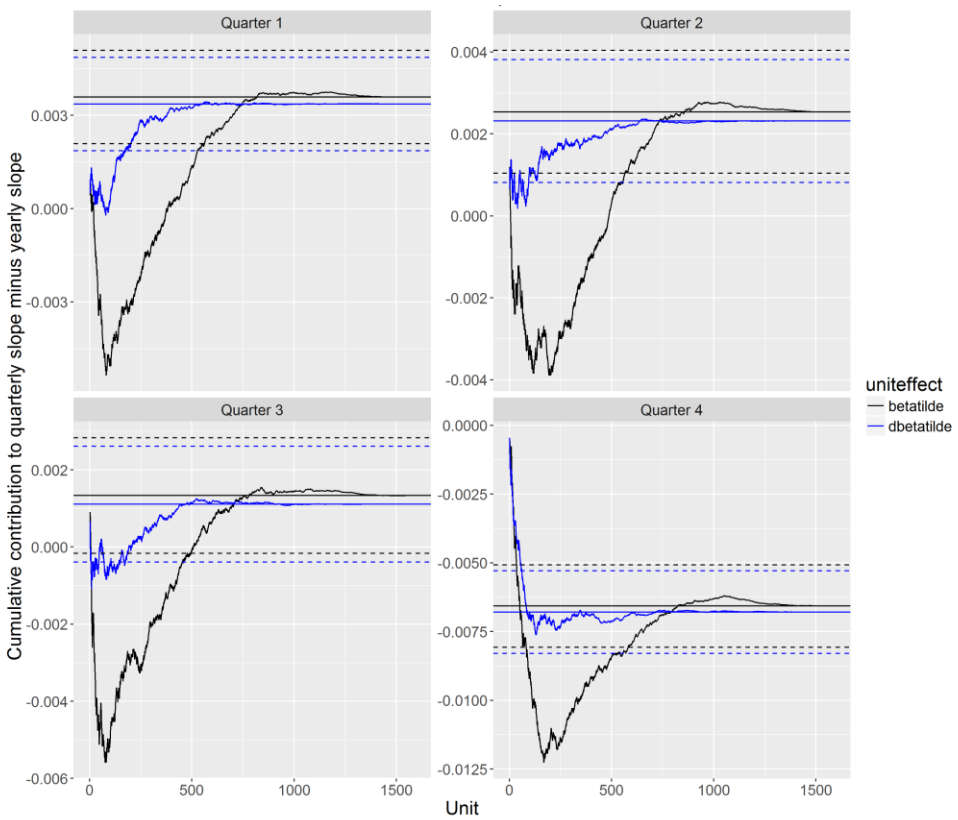

Figure 19. As Figure 16 but now for Retail trade 2015.

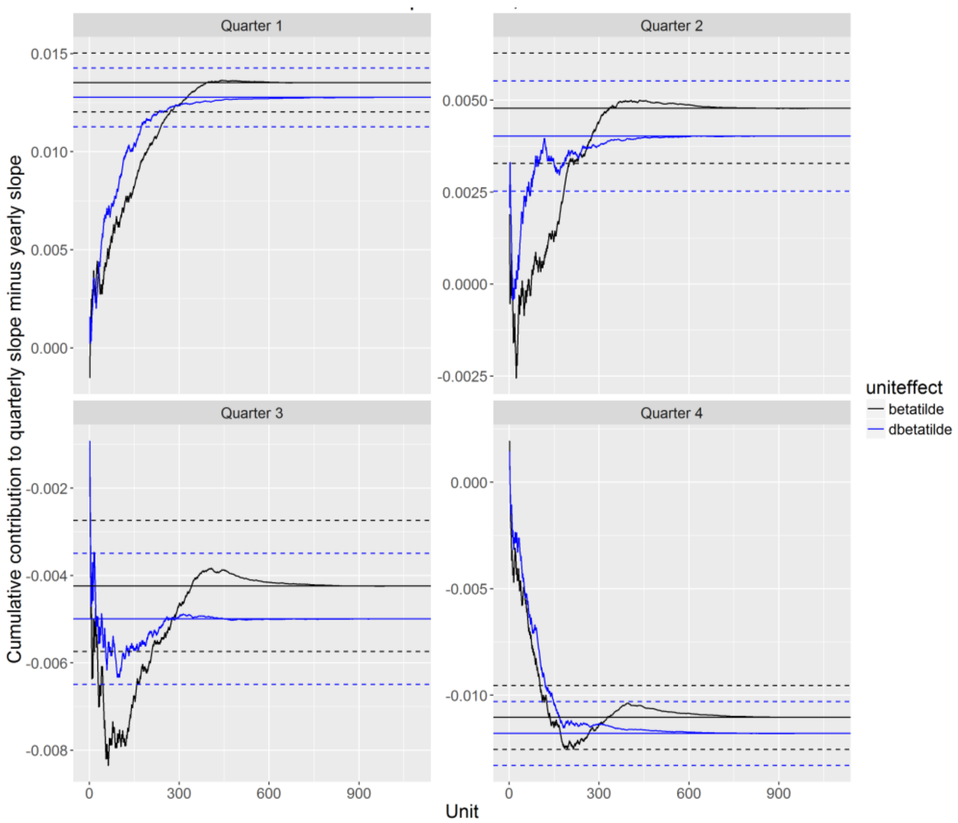

Figure 20. As Figure 16 but now for Job placement 2016.

For Manufacturing in 2014 (not shown) and 2015 (**Figure 16**), about 900 – 1000 units were passed before the cumulative effect for the different quarters stays within the margins. For 2016 (**Figure 17**), far less units were needed to stay within the margins. For instance, after about 200 units the cumulative effect of the fourth quarter stayed within the margins. For Construction the curves for all three years are similar, only 2015 is shown (**Figure 18**). After about 300 (2014) and 200 (2015 and 2016) units the effect in the fourth quarter stays within the margins. In the other quarters it also takes 200 - 300 units till the cumulative effect stays within the margins. For Retail trade, about 500 – 700 units are passed until the quarterly effects stay within the margins (see **Figure 19** for an example). This holds for all four quarters. For Job placement it varied between quarters and years from 50 – 300 units until the effect stayed within the margins. The smallest number of units was needed for Q4, 2015, see **Figure 20**. The exact number of units until the cumulative effect stayed within the margins is given in **Table 12**.

The number of units that was needed before the cumulative normalised effects ($d\beta_i^q$) stayed within the margins was often considerably smaller than for the cumulative effects ($\hat{r}_i^q \beta_i^q$). For instance, for Manufacturing the fourth quarter in 2015, 223 units were needed for $d\beta_i^q$ whereas 1006 for $\hat{r}_i^q \beta_i^q$ until the cumulative effect stayed within the margins, see **Table 12**. The actual number of units before the cumulative normalised effects ($d\beta_i^q$) stayed within the margins varied greatly between economic sectors and years. For instance, for the fourth quarter this ranged from 36 – 671.

We computed the percentage of units until $d\beta_i^q$ stayed within the margins, as a function of size class and economic sector, averaged over 2014–2016, see **Table 13**. For Manufacturing, the relative fractions increased somewhat with size class, for Construction, Retail Trade and Job Placement the relative fractions decreased with size class. The main conclusion however is that for each economic sector, units with the largest effects are found in any of the size classes (that are included in the selected units).

Table 12. Number of units until the cumulative quarter effects (Cum) and cumulative normalised quarter effects (Norm) within the margins.

		Manufacturing		Construction		Retail trade		Job placement	
Year	Q	Cum	Norm	Cum	Norm	Cum	Norm	Cum	Norm
2014	1	931	187	333	0	711	0	299	281
2014	2	924	0	292	69	717	99	290	52
2014	3	963	0	317	27	672	0	302	96
2014	4	871	93	290	90	671	36	122	260
2015	1	1072	235	308	104	560	177	157	79
2015	2	1017	139	257	1	565	93	114	45
2015	3	1075	208	279	34	493	173	94	1
2015	4	1006	223	216	115	579	59	35	69
2016	1	542	303	268	17	579	29	268	171
2016	2	695	0	255	99	537	165	214	69
2016	3	663	0	224	54	584	15	208	16
2016	4	191	228	193	62	543	46	101	146

Table 13. Percentage of units per size class, averaged over 2014–2016, until the cumulative normalised quarter effects in Q4 stay within the margins.

Size class	Manufacturing	Construction	Retail trade	Job placement
1				24.3
2			4.8	18.3
3			2.7	21.4
4		20.3	0.9	20.6
5	5.5	15.6	1.0	15.2
6	9.7	4.1	0.0	8.5
7	16.1	5.3	0.0	9.7
8	23.7	1.6	1.3	8.8
9			0.0	8.5

4.2.3 Cumulative contribution of units to slope differences: simulation

Introduction

For many years and economic sectors, the cumulative effect of $\hat{r}_i^q \beta_i^q$ has a typical shape whereby it first reaches a minimum and/or maximum and then it reaches its final value. If the residuals are normally distributed around the line, then one would expect that the cumulative effect moves toward its final value, with only one minimum or maximum. Van Delden et al. (2019) have enlarged the mixture model with additional groups and found that a mixture model with more groups had a larger likelihood than the two-group model used in the current paper. The key point that they found is that the population is a mixture of three groups that differ in their quarterly reporting behaviour:

– a group of units which reports the same value for VAT as for survey turnover;
– a group of units with relatively large quarterly effects and for which VAT is clearly larger than survey turnover (also in the first quarter);
– a group of units with smaller quarterly effects and for which VAT is closer to the survey turnover than the second group.

Methodology

In the current section we investigate the effect of having one, two or three groups with different reporting behaviour on the shape of the cumulative effect of $\hat{r}_i^q \beta_i^q$. We investigated this effect on data where the VAT turnover values were obtained from the population of Job placement 2016. The survey turnover values were simulated, using parameters underlying Figure 4.2-1 in Van Delden et al. (2019) for Job placement 2016.

We first describe the three group scenario. For this scenario the units of Job Placement 2016 were stratified into large enterprises of 1-digit size class 6 and higher and small and medium sized enterprises (1-digit size class 5 and smaller). Within each of those two size class strata, the enterprises were divided randomly over three groups. Let p_k be the proportion of units of group k, then $n_k = np_k$ is the stratum group size. The proportions are given in **Table 14**. For each unit in the population, we first drew a random number $u_i \sim unif[0,1]$, and then sorted the units on u_i. Next the first $n_{k=1}$ units were appointed to group 1, the next $n_{k=2}$ units were appointed to group 2, and the remaining units to group 3. Next, the quarterly survey turnover values, y_i, were generated from a linear regression with a yearly slope for group 1 and quarterly slopes for groups 2 and 3 with an intercept of 0, slopes given in **Table 14** and random error ε_i. For simplicity, we assumed that ε_i^q is normally distributed with mean 0 and variance $\tilde{\sigma}^2/w_i^q$, with w_i^q given by equation (2). $\tilde{\sigma}^2$ was taken to be 0.95. The estimated value for ε_i, the disturbance terms $\hat{\varepsilon}_i$, were saved, and re-used for the two- and one-group scenario. For the estimation of the slopes, the design weights were taken into account.

For the two-group scenario, units appointed to group 1 were left unchanged and units of group 2&3 were combined into one group. Using the y_i-values and model structure of the three group model, we estimated their common quarterly slopes, the parameters are given in **Table 14**. Next, new y_i values were generated for the combined group 2&3 units using these common quarterly slopes and using exactly the same disturbance terms as for the three group model.

Finally for the one-group scenario the same approach was taken as for the two-group scenario. First, all units were combined into one common group (group 1&2&3). Using the y_i-values and model structure of the three group model, we estimated the common quarterly slopes (for all units together), the parameters are given in **Table 14**. Next, new y_i values were generated using these common quarterly slopes and using exactly the same disturbance terms as for the three group model.

Table 14. Set up of three artificial populations: with three, with two and with one group of units.

Mixture	Group	Proportion	Yearly slope	Quarterly slope			
				Q1	Q2	Q3	Q3
Three	1	23.1	0.999				
	2	36.7		0.908	0.861	0.855	0.807
	3	40.2		0.988	0.968	0.968	0.935
Two	1	23.1	0.999				
	2&3	76.9		0.949	0.915	0.917	0.874
One	1&2&3	100		0.963	0.938	0.938	0.905

Results

As expected, for the population where all units have the same reporting behaviour (one group scenario), the cumulative value of $\hat{r}_i^q \beta_i^q$ continuously increased toward its maximum, or continuously decreased towards its minimum for each of the four quarters of the year. When the population was a mixture of two groups with a different reporting behaviour, this shape was also found in the first and fourth quarter whereas in the second and third quarter the cumulative effect first increased towards a maximum and then decreased again. For a population with a mixture of three reporting groups, the shape of the cumulative value of $\hat{r}_i^q \beta_i^q$ in the second and third quarter became much more irregular than for the population with one reporting group. In the fourth quarter the cumulative value of $\hat{r}_i^q \beta_i^q$ for the population with three groups first reached a local minimum and the increased again till its final value. Furthermore, for the population with three groups, in the fourth quarter, the first 30 largest absolute residuals belong to units of group 1 and 2 (of which six of group 1) thereafter residuals belonging to units of group 3 are found.

From this simulation analysis we can conclude that the shape of the cumulative distribution of $\hat{r}_i^q \beta_i^q$ becomes more irregular when the population in reality consists of a mixture of groups with different reporting patterns.

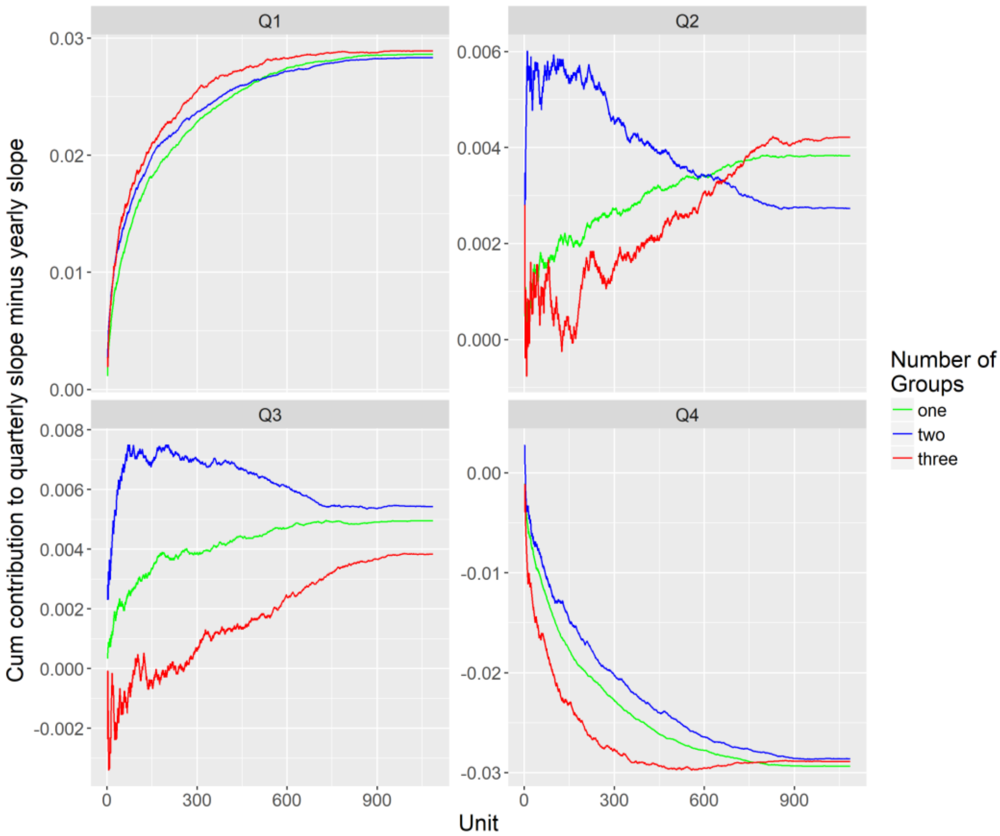

Figure 21. Cumulative effect of the units to the difference between the quarterly slope and the yearly slope, for a simulated population. Parameters are based on Job Placement 2016.

5. Conclusions and discussion

A first objective of the present paper was to fine-tune the exact model that is used for the quarterly effects. We looked into the residuals of outlying versus non-outlying units, to the appointment of weights on quarterly versus yearly basis and to the factor lambda. Firstly, we found that the residuals over the different units were clearly more homoscedastic with the two-group mixture model than with the Huber model. Secondly, we showed that the weights (that distinguish outliers from 'non-outliers') should be appointed on a yearly basis rather than quarterly, because otherwise the quarterly slope effects can be unjustly attenuated due to the method to compute the weights. Thirdly, we found that setting the heteroscedasticity correction factor lambda to 1 resulted in approximately homoscedastic disturbance terms. We decided to use fixed, rather than estimated, values for the parameter lambda to make the results over time better comparable. The remainder of the results were therefore based on a two-group mixture model, with group membership weights appointed yearly and a factor lambda of 1. Examples in the literature of the use of mixture models to model measurement errors can be found in Di Zio and Guarnera (2013).

For the three years and the four economic sectors, we consistently found that the weighted turnover was larger for the VAT data than for the survey data. Analysis of linked VAT-survey microdata by van Bemmel (2018) on Job Placement showed that in some cases the turnover that was reported by enterprises matched with the turnover of only a part of the total number of VAT units that belonged to the enterprises. In those cases, it may not be clear to the enterprise for which (VAT / legal) units we would like them to report survey data. Possibly, explicitly summing up all legal units on the survey form reduces the extent of this problem. We do not know whether missing a part of the legal units in the survey turnover is the *only* explanation for this turnover difference. We aim to contact a number of enterprises in the near future, to gain a better understanding of the causes of this difference.

Van Delden and Scholtus (2017) analysed the data of 2014 and 2015 and found quarterly effects for Manufacturing, Construction and Retail Trade. Because the quarterly effect that they found was small compared to noise in the data, the second objective of this paper was to repeat the analysis including 2016 data, to determine whether the seasonal differences are consistent over time and whether it is also found in the economic sector Job Placement. Indeed, the results in the current paper over the three years and all four economic sectors consistently showed a quarterly effect: the slope of the fourth quarter of the year was smaller than that of the first quarter. The effects were strongest for Manufacturing and Job placement, and weaker for Construction and Retail Trade. Our findings thus strongly suggest that the relation between VAT and survey turnover varies with the quarter of the year, although the effect is not so large.

We attempted to quantify whether quarterly effects of this size are relevant for the published results. The business statistics division who compiles the output uses the norm that month-on-month growth rates of the economic sectors should be changed by at most 0.4 per cent points due to the benchmarking process. That value is to be understood as a norm for the bias, for a systematic effect. We did not use this norm directy but used the conservative norm of a maximum difference in the quarterly slope compared to the yearly average slope of ± 0.0015. If the slope in the first quarter is shifted upwards (by at most 0.0015) and in the fourth quarter downwards (by at most -0.0015) then the benchmarking revision of the quarter-on-quarter growth rate for the first quarter remains within 0.3 per cent points, and therefore should be considered irrelevant for statistical production. Tables 7–10 showed that the difference between the slope of the fourth and the first quarter in all years and economic sectors was at least 0.3 per cent points.

We have checked whether definitional differences between survey and VAT turnover can explain the quarterly effects. The survey turnover concerns totals invoiced during the reference period (Commission Regulation, 2006). The VAT also concerns totals invoiced, but the interesting case is the situation of goods with a long production time (roads, ships and so on). For the yearly taxes on profit and loss, each tax declaration should also account for the returns and costs of the part of large projects that has been completed at the end of the year (see Belastingdienst, 2018). For the VAT tax, however, the turnover has to be declared as soon as an invoice is send, but a company can decide to send the invoice after the project has been completed (see HigherLevel, 2014). Since the survey turnover also concerns totals invoiced, this should not yield definitional differences. In practical reporting behaviour however, differences might occur.

We would like to know more about practical reporting behaviour of enterprises in order to understand the quarterly effects. To this end, we plan to contact a number of enterprises with a pronounced quarterly effect, preferably, for a subsequent number of subsequent years. We will ask them to explain how they fill in both the survey and the VAT form and how the differences in the values occur. We also like to know which of the two sources has the most reliable quarterly effect. Hopefully, this can be deduced when we have a better understanding of the reporting behaviour of the enterprises.

The third objective of the paper was to investigate whether the quarterly slope-differences can be explained by groups of units that have a certain reporting pattern in common. To this end, we grouped the units into 81 reporting patterns, defined in section 4.1. By design, the pattern with the largest fraction was "EEEE". The next five patterns with the largest fractions were “LLLL”, “EEEL”, “EELE”, “LEEE”, ELEE””, which is consistent with the observation that the weighted total turnover derived from VAT declarations is larger than for the survey. For each of the twelve combinations of economic sectors and years we found that a large number of different patterns each contributed to the total estimated quarterly effects. This implies that quarterly effects cannot be attributed to units with a limited set of specific patterns. Moreover, we found that a large fraction of the enterprises were not appointed to the same pattern for two subsequent years. This implies that the distinguished patterns apparently do not capture inherent reporting behaviour of enterprises. This results underlines the need to contact the enterprises, to gain insight into the underlying causes of the quarterly differences between VAT and survey turnover.

The fourth and last objective of the present paper was to investigate whether the quarterly slope-differences can be explained by a limited number of most influential enterprises. To that end, we analysed the contribution of individual units to the differences in quarterly slopes. When sorted on the basis of their absolute effect on the quarterly slope in the fourth quarter, there were more units needed until the cumulative value remained within the selected threshold, than when sorted on the basis of the normalised quarterly effect per unit. The normalised quarterly effect is the most relevant one for explaining the quarterly effects. Using those normalised effects, depending on the economic sector and year 36 – 671 units were needed to explain the quarterly effects up to a threshold. This concerns too many units to analyse manually (each production cycle) and correct when needed, to get rid of the quarterly effect. Recall that the sorting of the units varied for each of the four quarters, so in practice even more units need to be checked to cancel out all quarterly effects.

Recall from the introduction that we would like to correct the seasonal effects in either the survey or in the VAT data, or in both, in order to facilitate future benchmarking. First of all, outcomes of the causes of the differences in seasonal effects may be used to reduce the measurement errors in the survey data, for instance by adjusting survey questions. Secondly, maybe the editing process of the survey and/or the VAT data may be adjusted such that effects of seasonal differences are attenuated. Finally one might try to derive a model-based correction for the seasonal patterns or one might base quarterly growth rates solely on the data set for which the seasonal pattern is most reliable.

In order to derive such an automatic correction method we have three future points of investigation. The first point is that we want to know to what extent the two turnover time series have a seasonal bias. By contacting a number of enterprises, as mentioned earlier, we hope to gain insight into this seasonal bias and its causes. The second point is that we would like to know whether the quarterly effects can be explained by background variables or not. An example of such a background variable is the difference, in days, between the day of recipience of the VAT report versus that of the survey. A larger difference in the fourth quarter suggests that more work is needed to complete the VAT declaration for instance because turnover corrections need to be made. Relevant background variables might be included in a future bias-correction method. The third point of investigation concerns the number and kind of groups that is accounted for in the mixture model. The analysis of the residuals (section 4.2.3) suggested that is it useful to check whether a mixture model with more than two groups fits the data better. A first analysis on estimating such an extended mixture model can be found in Ostlund (2018) and in Van Delden et al. (2019).

6. References

Belastingdienst, 2018. Website of the Dutch tax office. Explanation on tax declarations for Dutch profit and loss taxes of ongoing projects: https://www.belastingdienst.nl/wps/wcm/connect/bldcontentnl/belastingdienst/zakelijk/winst/vennootschapsbelasting/waardering_onderhanden_werk/. Accessed at 2019-01-29.

Bemmel, K.J.H. van and J.J. Hoogland (2017). Verschillen DRT-KICR. Analyseverslag. Versie 1.0p1. CBS rapport, 20-10-2017. (in Dutch)

Bemmel, K.J.H. van (2018). Verschillen DRT-KICR Uitzendbranche. Analyseverslag. Versie 2.0. CBS rapport, 04-04-2018. (in Dutch)

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

Chambers, R.L. and C.J. Skinner (eds.) (2003). Analysis of Survey Data. John Wiley & Sons, Chichester.

Commission Regulation AC 1053/2006, 2006. Implementing and amending Council Regulation (EC) 1165/98 concerning short-term statistics as regards definitions of variables, lists of variables and frequency of data compilation. Official Journal of the European Union L 281/15, 12-10-2016.

Daalmans, J. (2017). Inpassen van KICR en DRT; de kwartaalinpassing. Internal CBS report, 31-7-2017. (in Dutch)

Delden, A. van (2012). Marges van kortetermijnstatistieken op maandbasis bij steekproefreductie. Internal CBS report, 12 April 2012. (in Dutch)

Delden, A. van and S. Scholtus (2017). Correspondence between survey and administrative data on quarterly turnover. CBS Discussion paper 2017-3.

Delden, A. van, Pannekoek, J. Banning, R, and Boer, A. de (2016). Analysing correspondence between administrative and survey data. Statistical Journal of the IAOS 32, 569–584.

Delden, A. van, Scholtus S. and Ostlund, N. (2019). Modelling measurement errors to enable consistency between monthly and quarterly turnover growth rates. Proceedings of Statistics Canada Symposium 2018. Combine to Conquer: Innovations in the Use of Multiple Sources of Data.

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.

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

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

Hastie, T., Tibshirani, R. and J. Friedman (2013). The elements of statistical learning. Second edition. Springer Series in Statistics, Springer New York.

HigherLevel (2014). Website of entrepreneurship and innovation (In Dutch). Explanation on Dutch VAT tax declarations of ongoing projects: https://www.higherlevel.nl/forums/topic/44194-onderhanden-werk-en-btw/ Accessed at 2019-01-29.

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

Ostlund, N. (2018). Capturing Measurement Error in Linked Survey and Administrative Data. Verslag Master thesis.

Rouseeuw, P.J. and C. Croux (1993). Alternative to the median absolute deviation. Journal of the American Statistical Association 88, 1273 –1283.

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

Skinner, C.J., D. Holt and T.M.F. Smith (eds.) (1989). Analysis of Complex Surveys. John Wiley & Sons, Chichester.

7. Appendix 1: Details on model estimation

7.1 Mixture model, outlier detection by quarter

The parameters of the mixture model given by equation (10) are estimated by maximum likelihood, using an Expectation Conditional Maximisation (ECM) algorithm. For a fixed value of λ (the parameter that describes the degree of heteroscedasticity), this algorithm was given in Appendix A of Van Delden and Scholtus (2017). An updated version of this algorithm will be described here, but we now include an optional extension to estimate λ by maximum likelihood as well.

Let $\varphi(.;\mu,\sigma^2)$ denote the density of a univariate normal distribution with mean μ and variance σ^2. Under model (10), the observed density of y_i^q is a mixture of two normal densities:

$$f(y_i^q;\boldsymbol{\theta}) = (1-\pi)\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\tilde{\sigma}^2/w_i^q) + \pi\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\vartheta\tilde{\sigma}^2/w_i^q). \qquad (36)$$

Recall that $w_i^q = (\tilde{x}_i^q)^{-\lambda}$, with $\tilde{x}_i^q = \max(x_i^q,1)$. Furthermore, $\boldsymbol{\theta}$ denotes the vector of model parameters. Depending on whether λ is fixed or considered as an estimable parameter, $\boldsymbol{\theta}$ is either given by $\boldsymbol{\theta} = (\pi,\boldsymbol{b}^T,\tilde{\sigma}^2,\vartheta)^T$ or $\boldsymbol{\theta} = (\pi,\boldsymbol{b}^T,\tilde{\sigma}^2,\vartheta,\lambda)^T$.

Density (36) arises because the assignment of each observation $(\boldsymbol{\xi}_i^q, y_i^q)$ to the first or second group of the mixture model, represented by the indicator z_i^q, is not known. If the values z_i^q had been observed as well, maximum likelihood estimation could have been based instead on the complete-data log likelihood function, which is:

$$\begin{aligned}
\log L(\boldsymbol{\theta}) &= \sum_{q=1}^{4}\sum_{i\in s^q}\left\{\left(1-z_i^q\right)\log[(1-\pi)\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\tilde{\sigma}^2/w_i^q)]\right.\\
&\qquad \left. + z_i^q\log[\pi\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\vartheta\tilde{\sigma}^2/w_i^q)]\right\}\\
&= -\frac{n\log(2\mathbf{pi})}{2} - \frac{\lambda}{2}\sum_{q,i}\log\tilde{x}_i^q + \log(1-\pi)\sum_{q,i}\left(1-z_i^q\right) + (\log\pi)\sum_{q,i} z_i^q\\
&\qquad -\frac{n\log\tilde{\sigma}^2}{2} - \frac{\log\vartheta}{2}\sum_{q,i} z_i^q - \frac{1}{2\tilde{\sigma}^2}\sum_{q,i}\left(1-z_i^q\right)\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q-\boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2\\
&\qquad -\frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{q,i} z_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q-\boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2.
\end{aligned}$$

Here $n = \sum_{q=1}^{4}|s^q|$ is the total number of observations and $\mathbf{pi}$ denotes the numerical constant 3.1415… (not to be confused with the probability π).

The above log likelihood function is based on the assumption that the data consist of independent, identically distributed observations. To account for the sampling design of the survey data used here, we can apply pseudo maximum likelihood (PML); see Skinner et al. (1989, Section 3.4.4) or Chambers and Skinner (2003, Section 2.4). For this particular application, PML is equivalent to maximising the *weighted* complete-data log likelihood function

$$\begin{aligned}
\log L_d(\boldsymbol{\theta}) &= \sum_{q=1}^{4}\sum_{i\in s^q} d_i^q\left\{\left(1-z_i^q\right)\log[(1-\pi)\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\tilde{\sigma}^2/w_i^q)]\right.\\
&\qquad \left. + z_i^q\log[\pi\varphi(y_i^q;\boldsymbol{b}^T\boldsymbol{\xi}_i^q,\vartheta\tilde{\sigma}^2/w_i^q)]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{N\log(2\mathbf{pi})}{2} - \frac{\lambda}{2}\sum_{q,i} d_i^q \log\tilde{x}_i^q + \log(1-\pi)\sum_{q,i} d_i^q\left(1-z_i^q\right) \\
&\quad + (\log\pi)\sum_{q,i} d_i^q z_i^q - \frac{N\log\tilde{\sigma}^2}{2} - \frac{\log\vartheta}{2}\sum_{q,i} d_i^q z_i^q \\
&\quad - \frac{1}{2\tilde{\sigma}^2}\sum_{q,i}\left(1-z_i^q\right)d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2 \\
&\quad - \frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{q,i} z_i^q d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2,
\end{aligned}
$$

with d_i^q the calibration weight of unit i in quarter q, and $N = \sum_{q=1}^4 \sum_{i\in s^q} d_i^q$. In general, PML yields consistent estimates of the parameter values that would be obtained by regular maximum likelihood estimation if the entire target population had been observed.

As z_i^q is unobserved, we cannot maximise the complete-data log likelihood directly. Instead, we can apply PML estimation for incomplete data using an ECM algorithm. This means that we maximise $Q_d(\boldsymbol{\theta}) = E\left\{\log L_d(\boldsymbol{\theta})\,\middle|\,\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}\right\}$, i.e., the conditional expectation of the complete-data log likelihood, given the observed data. For model (10), the function $Q_d(\boldsymbol{\theta})$ is obtained by replacing each instance of z_i^q in the above expression for $\log L_d(\boldsymbol{\theta})$ by its conditional expectation $\tau_i^q = E\left(z_i^q\middle|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}\right) = P\left(z_i^q = 1\middle|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}\right)$. Hence, $Q_d(\boldsymbol{\theta})$ is given by:

$$
\begin{aligned}
Q_d(\boldsymbol{\theta}) &= -\frac{N\log(2\mathbf{pi})}{2} - \frac{\lambda}{2}\sum_{q,i} d_i^q \log\tilde{x}_i^q + \log(1-\pi)\sum_{q,i} d_i^q\left(1-\tau_i^q\right) \\
&\quad + (\log\pi)\sum_{q,i} d_i^q \tau_i^q - \frac{N\log\tilde{\sigma}^2}{2} - \frac{\log\vartheta}{2}\sum_{q,i} d_i^q \tau_i^q \\
&\quad - \frac{1}{2\tilde{\sigma}^2}\sum_{q,i}\left(1-\tau_i^q\right)d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2 \\
&\quad - \frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{q,i} \tau_i^q d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2.
\end{aligned}
\tag{37}
$$

By means of Bayes' rule, it follows from (36) that τ_i^q is given by:

$$
\begin{aligned}
\tau_i^q &= P\left(z_i^q = 1\middle|\boldsymbol{\xi}_i^q, y_i^q, \boldsymbol{\theta}\right) \\
&= \frac{P\left(z_i^q = 1\middle|\boldsymbol{\theta}\right)f\left(y_i^q\middle|z_i^q = 1; \boldsymbol{\xi}_i^q, \boldsymbol{\theta}\right)}{P\left(z_i^q = 0\middle|\boldsymbol{\theta}\right)f\left(y_i^q\middle|z_i^q = 0; \boldsymbol{\xi}_i^q, \boldsymbol{\theta}\right) + P\left(z_i^q = 1\middle|\boldsymbol{\theta}\right)f\left(y_i^q\middle|z_i^q = 1; \boldsymbol{\xi}_i^q, \boldsymbol{\theta}\right)} \\
&= \frac{\pi\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \vartheta\tilde{\sigma}^2/w_i^q)}{(1-\pi)\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \tilde{\sigma}^2/w_i^q) + \pi\varphi(y_i^q; \boldsymbol{b}^T\boldsymbol{\xi}_i^q, \vartheta\tilde{\sigma}^2/w_i^q)}.
\end{aligned}
\tag{38}
$$

The ECM algorithm alternates between an E step and an M step. The E step involves computing expression (38) 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 $Q_d(\boldsymbol{\theta})$. The E and M steps are iterated until the parameter estimates have converged.

To find the maximum of $Q_d(\boldsymbol{\theta})$, we set the partial derivatives $\partial Q_d/\partial \pi$, $\partial Q_d/\partial \boldsymbol{b}^T$, $\partial Q_d/\partial \tilde{\sigma}^2$, $\partial Q_d/\partial \vartheta$, and (optionally) $\partial Q_d/\partial \lambda$ equal to zero. For the above Q_d 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 (Little and Rubin, 2002).

The M step consists of the following sub-steps:
M1. Compute a new value for π by solving $\partial Q_d/\partial \pi = 0$. The solution is given by:

$$\pi = \frac{\sum_{q,i} d_i^q \tau_i^q}{\sum_{q,i} d_i^q}.$$

M2. Compute new regression coefficients $\boldsymbol{b}$ by solving $\partial Q_d/\partial \boldsymbol{b}^T = \mathbf{0}$. It turns out that $\boldsymbol{b}$ can be obtained by weighted least squares, using the weights $v_i^q = d_i^q w_i^q \left(1 - \tau_i^q + \tau_i^q/\vartheta\right)$:

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

M3. Compute a new value for $\tilde{\sigma}^2$ by solving $\partial Q_d/\partial \tilde{\sigma}^2 = 0$. The solution is:

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

M4. Compute a new value for ϑ by solving $\partial Q_d/\partial \vartheta = 0$. The solution is given by:

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

M5 (optional). Compute a new value for λ by solving $\partial Q_d/\partial \lambda = 0$, i.e.:

$$\frac{1}{\tilde{\sigma}^2} \sum_{q,i} \left(1 - \tau_i^q + \tau_i^q/\vartheta\right) d_i^q \left(\tilde{x}_i^q\right)^{-\lambda} \left(\log \tilde{x}_i^q\right) \left(y_i^q - \boldsymbol{b}^T \boldsymbol{\xi}_i^q\right)^2 - \sum_{q,i} d_i^q \log \tilde{x}_i^q = 0.$$

This last equation does not have an analytical solution for λ, but it can be solved numerically. (In our application, we used the R function *uniroot* for this.) If this sub-step is run, then we must also update the weights $w_i^q = (\tilde{x}_i^q)^{-\lambda}$.

In each computation, the most recent parameter values are used. (So, e.g., when $\tilde{\sigma}^2$ is computed in sub-step M3, we use the weights v_i^q and regression coefficients $\boldsymbol{b}$ that have just been computed in sub-step M2.) If λ is considered fixed, sub-step M5 can simply be left out of the algorithm. In that case, the weights $w_i^q = (\tilde{x}_i^q)^{-\lambda}$ do not change.

For fixed λ, the above algorithm corresponds to that of Van Delden and Scholtus (2017), with minor corrections so that PML estimation is applied consistently. For regular ML estimation (without survey weights), this algorithm can be seen as a special case of the ECM algorithm for a normal contamination model described by Di Zio and Guarnera (2013).

To initialise the ECM algorithm, starting values have to be chosen for the parameters in $\boldsymbol{\theta}$. In our application, we used the results of the robust regression approach to find appropriate starting values for π, $\boldsymbol{b}$, $\tilde{\sigma}^2$, and ϑ. In the scenarios that involved estimating λ, we always used $\lambda = 1$ as a starting value. Van Delden and Scholtus (2017) investigated whether the ECM algorithm might converge to a local rather than a global maximum, by repeating the algorithm with different starting values. They did not find any evidence for convergence to local maxima.

7.2 Mixture model, outlier detection by year

As was described in Section 3.1.4, we also used a version of the mixture model in which units i were assigned to the same group (outlier or non-outlier) for all quarters q of the same year. For a given year, this amounts to introducing the restriction that $z_i^1 = z_i^2 = z_i^3 = z_i^4 = z_i^+$, say, for all units i. As we will now show, this restriction affects the E step of the above ECM algorithm but not the M step.

Let $\varphi_p(.\,; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the density of a p-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Under model (10) with the restriction $z_i^1 = z_i^2 = z_i^3 = z_i^4 = z_i^+$, the observed density of $\boldsymbol{y}_i = \left(y_i^1, y_i^2, y_i^3, y_i^4\right)^T$ is a mixture of two four-dimensional normal densities:

$$\begin{aligned} f(\boldsymbol{y}_i; \boldsymbol{\theta}) &= (1-\pi)\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + \pi\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \vartheta\boldsymbol{\Sigma}_i), \\ \boldsymbol{\mu}_i &= \left(\boldsymbol{b}^T\boldsymbol{\xi}_i^1, \boldsymbol{b}^T\boldsymbol{\xi}_i^2, \boldsymbol{b}^T\boldsymbol{\xi}_i^3, \boldsymbol{b}^T\boldsymbol{\xi}_i^4\right)^T = \boldsymbol{X}_i\boldsymbol{b}, \\ \boldsymbol{\Sigma}_i &= \mathrm{diag}\{\tilde{\sigma}^2/w_i^1, \tilde{\sigma}^2/w_i^2, \tilde{\sigma}^2/w_i^3, \tilde{\sigma}^2/w_i^4\}. \end{aligned} \tag{39}$$

with $\boldsymbol{X}_i^T = \left(\boldsymbol{\xi}_i^1, \boldsymbol{\xi}_i^2, \boldsymbol{\xi}_i^3, \boldsymbol{\xi}_i^4\right)$. Note that within each group (i.e., conditional on $z_i^+ = 0$ or $z_i^+ = 1$), the quarterly observations of unit i are still considered independent, because the matrix $\boldsymbol{\Sigma}_i$ is diagonal.

In practice, the calibration weights are fixed throughout a year, so we can write $d_i^1 = d_i^2 = d_i^3 = d_i^4 = d_i^+$ for each unit i. The weighted complete-data log likelihood and the function $Q_d(\boldsymbol{\theta})$ are now given by:

$$\begin{aligned} \log L_d(\boldsymbol{\theta}) = \sum_i d_i^+\{&(1-z_i^+)\log[(1-\pi)\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)] \\ &+ z_i^+\log[\pi\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \vartheta\boldsymbol{\Sigma}_i)]\}, \end{aligned}$$

$$
\begin{aligned}
Q_d(\boldsymbol{\theta}) = &-\frac{N\log(2\mathbf{pi})}{2} - \frac{\lambda}{2}\sum_{q,i} d_i^+ \log \tilde{x}_i^q + \log(1-\pi)\sum_i d_i^+(1-\tau_i^+) \\
&+ (\log \pi)\sum_i d_i^+\tau_i^+ - \frac{N\log\tilde{\sigma}^2}{2} - \frac{\log\vartheta}{2}\sum_i 4d_i^+\tau_i^+ \\
&- \frac{1}{2\tilde{\sigma}^2}\sum_{q,i}(1-\tau_i^+)d_i^+\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2 \\
&- \frac{1}{2\vartheta\tilde{\sigma}^2}\sum_{q,i}\tau_i^+ d_i^+\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2,
\end{aligned}
$$

with $\tau_i^+ = E(z_i^+|\boldsymbol{X}_i, \boldsymbol{y}_i, \boldsymbol{\theta})$. Analogously to (38), we obtain from (39):

$$
\tau_i^+ = \frac{\pi\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \vartheta\boldsymbol{\Sigma}_i)}{(1-\pi)\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + \pi\varphi_4(\boldsymbol{y}_i; \boldsymbol{\mu}_i, \vartheta\boldsymbol{\Sigma}_i)}. \tag{40}
$$

In the E step of the ECM algorithm, the probabilities τ_i^+ are now updated using (40) instead of (38). For the M step, we note that maximising the above function $Q_d(\boldsymbol{\theta})$ is equivalent to maximising the function (37) given in the previous subsection, provided that we set $\tau_i^1 = \tau_i^2 = \tau_i^3 = \tau_i^4 = \tau_i^+$ at the end of each E step. Hence, with this modification, the M step can remain the same as before.

7.3 Huber model

As was described in Section 3, the Huber models are estimated by an iteratively weighted least squares procedure. For Huber models with fixed λ, we used the standard R function *rlm*. To estimate λ for Huber models that contain this additional parameter, we used the following heuristic approach, which mimics the way λ is estimated for the mixture model.

First, consider the Huber model with outliers assigned on a quarterly basis. In each iteration, the regression coefficients are estimated by weighted least squares, by minimising the loss function given by (3). This loss function is approximately equivalent to the log likelihood (under PML estimation) of a normal weighted linear regression model with weights $w_i^q = g_i^q\omega_i^q = g_i^q d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}$:

$$
\begin{aligned}
\Lambda_d = &-\frac{N\log(2\mathbf{pi})}{2} + \frac{1}{2}\sum_{q,i} d_i^q \log g_i^q - \frac{\lambda}{2}\sum_{q,i} d_i^q \log \tilde{x}_i^q - \frac{N\log\tilde{\sigma}^2}{2} \\
&- \frac{1}{2\tilde{\sigma}^2}\sum_{q,i} g_i^q d_i^q\left(\tilde{x}_i^q\right)^{-\lambda}\left(y_i^q - \boldsymbol{b}^T\boldsymbol{\xi}_i^q\right)^2.
\end{aligned}
$$

Note the similarity of this expression to (37). Having obtained estimates for the regression parameters $\boldsymbol{b}$ and the Huber weights g_i^q in the current iteration by the standard *rlm* procedure, we then compute an estimate for $\tilde{\sigma}^2$ based on Λ_d (cf. sub-step M3 of the ECM algorithm):

$$\tilde{\sigma}^2 = \frac{\sum_{q,i} d_i^q g_i^q \left(\tilde{x}_i^q\right)^{-\lambda} \left(y_i^q - \boldsymbol{b}^T \boldsymbol{\xi}_i^q\right)^2}{\sum_{q,i} d_i^q}.$$

Next, λ is estimated analogously to sub-step M5 of the ECM algorithm, by numerically solving:

$$\frac{1}{\tilde{\sigma}^2} \sum_{q,i} d_i^q g_i^q \left(\tilde{x}_i^q\right)^{-\lambda} \left(\log \tilde{x}_i^q\right) \left(y_i^q - \boldsymbol{b}^T \boldsymbol{\xi}_i^q\right)^2 - \sum_{q,i} d_i^q \log \tilde{x}_i^q = 0.$$

Finally, the weights ω_i^q are updated by $\omega_i^q = d_i^q (\tilde{x}_i^q)^{-\lambda}$. We then proceed to the next iteration, until convergence. Again, we used $\lambda = 1$ as a starting value.

Note that we used the above PML estimate for $\tilde{\sigma}^2$ only to derive the above equation for λ; for the final estimate we used $\widehat{\sigma}_{\mathrm{MAR}}$ from (6). In fact, the iterative algorithm did not converge if we used $\widehat{\sigma}_{\mathrm{MAR}}$ to estimate λ. Intuitively, using a non-robust estimator for $\tilde{\sigma}^2$ to estimate λ makes sense, because λ is used to define the heteroscedasticity weight $\omega_i^q = d_i^q (\tilde{x}_i^q)^{-\lambda}$ prior to any correction for outlying observations.

For the Huber model with outliers assigned on a yearly basis, the estimation procedure is virtually identical. The only difference is that the Huber weights g_i^q are now restricted such that $g_i^1 = g_i^2 = g_i^3 = g_i^4 = g_i^+$, with g_i^+ given by (14).

Explanation of symbols

Empty cell	Figure not applicable
.	Figure is unknown, insufficiently reliable or confidential
*	Provisional figure
**	Revised provisional figure
2017–2018	2017 to 2018 inclusive
2017/2018	Average for 2017 to 2018 inclusive
2017/’18	Crop year, financial year, school year, etc., beginning in 2017 and ending in 2018
2013/’14–2017/’18	Crop year, financial year, etc., 2015/’16 to 2017/’18 inclusive

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

Colophon

Publisher
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

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