Exploring the Effect of Time-Related Classification Errors on the Accuracy of Growth Rates in Business Statistics

Arnout van Delden, Sander Scholtus and Joep Burger[1]

Abstract

Producing reliable, undisputed statistical figures is the backbone of national statistical institutes (NSIs). One of the core publications in official statistics are short-term indicators for the economic business cycle, such as the quarterly turnover growth of economic sectors. In a number of countries turnover growth is estimated from administrative data, such as value added tax data. The data based on administrative units need to be linked to statistical units and classified by economic activity code. Determining the correct economic activity is often difficult. One of the reasons is that statistical units often consist of multiple legal units and likewise consist of multiple administrative units. The current paper addresses the problem of estimating the accuracy of turnover *growth rates*, as affected by classification errors in industry code. It consists of two main parts. First, we describe an approach for estimating the effect of the time-related classification errors on growth rates given that we know the size of the classification errors. Second, we describe how we collected data to estimate the size of the time-related classification errors. We will show some explorations on the impact of those errors on the accuracy of the growth rates by means of bootstrap simulation.

Keywords: accuracy, bootstrap, mixed-source statistics, NACE code

1. Introduction

Short-term indicators for the economic business cycle are an important part of the output of national statistical institutes (NSIs). One such indicator is the yearly growth rate of quarterly turnover by economic sector, that is published as part of the European STS regulation. It is crucial that NSIs know how accurate their estimated quarterly growth rates are, to be able to validate their estimates.

Many countries base their turnover growth rates on administrative data or on a combination of administrative and survey data (Costanzo, 2011). The data are often linked to a general business register (GBR) that contains background information such as the economic activity code. The latter is used to produce output by economic sector. The economic activity code in the GBR is not always of a high precision (Christensen, 2008). This is due to a combination of reasons. One reason is that the statistical units may consist of multiple legal units (e.g. Struijs, 2015), and each legal unit has its own economic activity. A common procedure at statistical offices is then to estimate a main activity code for the statistical unit. Furthermore, NSIs often use administrative sources, such as chamber of commerce data, where companies register their economic activity codes. However that information may be erroneous or outdated when a company changes its activity but does not report this. Errors in the economic activity are especially expected for the smaller units within a GBR, as NSIs often do not have the means to check the data of all those smaller companies.

Errors in the economic activity may affect the accuracy of the turnover growth rates per industry. It is, however, not straightforward to quantify this effect. Van Delden et al. (2015, 2016) developed an approach to quantify the effect of classification errors on level estimates. In the current paper we extend that approach to growth rate estimates. We address two main issues. First we describe an approach how we can estimate the effect of time-related classification errors on growth rates given that we know the size of the classification errors (Sections 2–4). Second, we describe how we collected data to estimate the size of the time-related classification errors (Section 5). We will show some

[1] Sander Scholtus and Arnout van Delden, Statistics Netherlands, Henri Faasdreef 312, 2492 JP The Hague, The Netherlands. Joep Burger, Statistics Netherlands, CBS-weg 11, P.O. Box 4481, 6401 CZ Heerlen, The Netherlands. Emails: s.scholtus@cbs.nl, a.vandelden@cbs.nl, j.burger@cbs.nl.

explorations on the impact of those errors on the accuracy of the growth rates by means of bootstrap simulation.

2. Estimating the accuracy

Consider a population of units ($i = 1, \dots, N$) that is divided into industries based on economic activity as derived in a GBR. Denote the total set of industries by $\mathcal{H}_{\text{full}}$. Denote the variable year as t and denote the starting year of the computations as $t = T_0$ and the following years as $T_0 + 1$, $T_0 + 2$, etc.

For each year t, each active unit (enterprise) i has an unknown true industry code $s_i^t = g$ and an observed industry code $\hat{s}_i^t = h$, where $g, h \in \mathcal{H}_{\text{full}}$. The true and observed industry codes are kept constant during a year; this is called the coordinated industry code at Statistics Netherlands. Between 31 December of year T_0 and 1 January of year $T_0 + 1$, or generally between 31 December of year $t - 1$ and 1 January of year t, the true and observed industry codes are updated for the units that are present in both $t - 1$ and t (continuing units).

Due to classification errors some of the observed industry codes may differ from the true ones. Those classification errors may affect the publication figures. In this paper, we consider the relatively simple case where classification errors are the only errors that occur. In particular, we assume that the target variable is observed for all units in the population. This can for instance be the case when administrative data are available, an example will be discussed in Section 5.

We are interested in changes in quarterly turnover per industry. First denote the true turnover for industry $h \in \mathcal{H}_{\text{full}}$ in quarter q of year t by $Y_h^{t,q} = \sum_{i=1}^{N^{t,q}} a_{hi}^t y_i^{t,q}$, where $N^{t,q}$ stands for the size of the population in quarter q of year t, $y_i^{t,q}$ denotes the turnover of unit i in this quarter and a_{hi}^t is a dummy variable with

$$a_{hi}^t = I(s_i^t = h) = \begin{cases} 1 & \text{if } s_i^t = h, \\ 0 & \text{if } s_i^t \neq h. \end{cases}$$

Recall that a_{hi}^t does not depend on q because the industry code is kept constant during the year. In practice, $Y_h^{t,q}$ is estimated by $\hat{Y}_h^{t,q} = \sum_{i=1}^{N^{t,q}} \hat{a}_{hi}^t y_i^{t,q}$, with $\hat{a}_{hi}^t = I(\hat{s}_i^t = h)$. In the remainder of the paper we use a single time index in the notation (thus either t or q) unless both indices are needed to avoid confusion.

We denote the change in turnover of quarter q to a previous quarter $q - u$ as $G_h^{q,q-u} = \frac{Y_h^q}{Y_h^{q-u}}$, where $u = 1$ gives the change with respect to the previous quarter and $u = 4$ the change with respect to the same quarter in the previous year. $G_h^{q,q-u}$ is estimated as $\hat{G}_h^{q,q-u} = \frac{\hat{Y}_h^q}{\hat{Y}_h^{q-u}}$. The corresponding relative changes are expressed as $g_h^{q,q-u} = 100(G_h^{q,q-u} - 1)$ and $\hat{g}_h^{q,q-u} = 100(\hat{G}_h^{q,q-u} - 1)$.

We would like to assess the bias and variance of $\hat{g}_h^{q,q-u}$ as an estimator for $g_h^{q,q-u}$, i.e.,

$$B\left(\hat{g}_h^{q,q-u}\right) = E\left(\hat{g}_h^{q,q-u}\right) - g_h^{q,q-u}, \tag{1}$$

$$V\left(\hat{g}_h^{q,q-u}\right) = E\left[\left(\hat{g}_h^{q,q-u} - E\left(\hat{g}_h^{q,q-u}\right)\right)^2\right]. \tag{2}$$

In order to estimate this bias and variance we need to model the classification errors that occur in the GBR over time. In the remainder of this section we introduce such a model and discuss a bootstrap method to estimate (1) and (2).

For the observed industry codes at the start of year $t = T_0$, we suppose that random classification errors occur, independently across units, according to a known (or previously estimated) transition

matrix $\mathbf{P}_i^{OL} = (p_{ghi}^{OL})$, with $p_{ghi}^{OL} = P(\hat{s}_i^t = h | s_i^t = g)$. In this notation superscript O stands for observed industry—conditional on the true value—and L stands for level (the cross-sectional situation). Note that we consider the true industry codes as fixed and the observed industry codes as stochastic, in line with, e.g., Kuha and Skinner (1997). We further assume that all newborn units in year $t \in \{T_0, T_0+1, T_0+2, \cdots\}$ also have an observed industry code according to $\mathbf{P}_i^{OL}$ in their first year, where the values of probabilities p_{ghi}^{OL} are for the moment assumed to be independent of t.

To model the time dependency of classification errors in the observed industry codes for continuing units, we introduce an additional transition matrix $\mathbf{P}_i^{OC} = (p_{jklhi}^{OC})$, with $p_{jklhi}^{OC} = P(\hat{s}_i^t = h | s_i^{t-1} = j, s_i^t = k, \hat{s}_i^{t-1} = l)$. In this notation superscript C stands for change. Again, we assume that the probabilities p_{jklhi}^{OC} are constant across time. Some additional simplifying assumptions on the structure of $\mathbf{P}_i^{OL}$ and $\mathbf{P}_i^{OC}$ will be introduced below. The true and observed industry codes with their dependencies are shown in the left part of Figure 1.

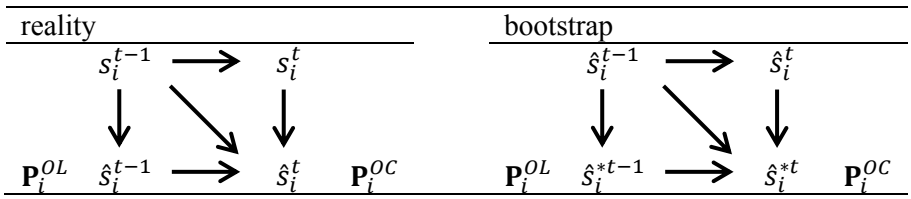

Figure 1: Classification errors over time: reality versus bootstrap.

Given the matrices $\mathbf{P}_i^{OL}$ and $\mathbf{P}_i^{OC}$ we can estimate the accuracy (bias and variance) of the growth rate $\hat{g}_h^{q,q-u}$. In the present paper, likewise to Van Delden et al. (2016) and Burger et al. (2015), we use a bootstrap approach for this. We use a bootstrap, because we can then, in future, also include other non-sampling errors besides classification errors, such as measurement, linkage, and coverage errors, as well as combinations thereof. Furthermore, we aim to investigate whether analytical expressions can be derived, using simplifying assumptions that approximate the bias and the variance. The results of those analytic expressions could then be compared with the bootstrap estimates to judge how well they perform.

In the bootstrap approach, we start by simulating the situation of the first quarter $q = 1$ in year $t = T_0$. For $q = 1$ we apply the transition matrix $\mathbf{P}_i^{OL}$, as in Van Delden et al. (2016), to the observed $\hat{s}_i^t$, which results in a new industry assignment variable denoted by $\hat{s}_i^{*t}$ (see the right part of Figure 1). That is to say, we consider realisations of the alternative classification error model given by:

$$P(\hat{s}_i^{*t} = h | \hat{s}_i^t = g) \equiv P(\hat{s}_i^t = h | s_i^t = g) = p_{ghi}^{OL} \ (t = T_0). \quad (3)$$

For the other quarters in year T_0, each unit keeps the same industry code, thus $\hat{s}_i^{*t} = h$. For any newborn units within year T_0 also transition matrix $\mathbf{P}_i^{OL}$ is applied to derive $\hat{s}_i^{*t}$. Next, for the first quarter of the next year ($t = T_0 + 1$, $q = 1$) we apply transition matrix $\mathbf{P}_i^{OC}$ to obtain $\hat{s}_i^{*t}$ (with $t = T_0 + 1$) given the values of $\hat{s}_i^{t-1}$, $\hat{s}_i^t$ and $\hat{s}_i^{*t-1}$ (see the right part of Figure 1). Thus, likewise to $\mathbf{P}_i^{OL}$, we consider realisations of the alternative classification error model

$$P(\hat{s}_i^{*t} = h | \hat{s}_i^{t-1} = j, \hat{s}_i^t = k, \hat{s}_i^{*t-1} = l) \equiv P(\hat{s}_i^t = h | s_i^{t-1} = j, s_i^t = k, \hat{s}_i^{t-1} = l) = p_{jklhi}^{OC}. \quad (4)$$

These new codes $\hat{s}_i^{*t}$ are again kept fixed for the remaining quarters in year $t = T_0 + 1$.

We continue this whole procedure for $t = T_0, T_0 + 1, \ldots$ as a Markov chain, in the sense that estimates for the quarters within the current year t depend on values of the previous year $t - 1$, but not of earlier

years. Next, we define: $\hat{a}_{hi}^{*t} = I(\hat{s}_i^{*t} = h)$. For one bootstrap replicate r, we obtain the sequence of estimated turnover levels in industry h: $\hat{Y}_{hr}^{*t,q} = \sum_{i=1}^{N^{t,q}} \hat{a}_{hir}^{*t} y_i^{t,q}$ $(t = T_0, T_0+1, \ldots; q = 1, \ldots, 4)$. Next we derive the sequence of growth rates $\hat{g}_{hr}^{*q,q-u} = 100(\hat{Y}_{hr}^{*q}/\hat{Y}_{hr}^{*q-u} - 1)$ $(u = 1,4)$ (here we omitted superscript t). We then repeat the whole procedure R times (for some large R).

The bootstrap bias and variance of the estimated growth rates are then estimated as follows (Efron and Tibshirani, 1993):

$$\hat{B}_R^*(\hat{g}_h^{q,q-u}) = m_R(\hat{g}_h^{*q,q-u}) - \hat{g}_h^{q,q-u}, \tag{5}$$

$$\hat{V}_R^*(\hat{g}_h^{q,q-u}) = \frac{1}{R-1}\sum_{r=1}^{R}\{\hat{g}_{hr}^{*q,q-u} - m_R(\hat{g}_h^{*q,q-u})\}^2. \tag{6}$$

with $m_R(\hat{g}_{hr}^{*q,q-u}) = \frac{1}{R}\sum_{r=1}^{R} \hat{g}_{hr}^{*q,q-u}$.

3. Modelling classification errors for the level transition matrix

The total number of industries in $\mathcal{H}_{\text{full}}$ is large – about 300 in The Netherlands. We therefore limit ourselves to estimating the accuracy of growth rates for a subset of nine target industries. Note that we do take the effect of misclassifications between target and non-target industries into account. We use $\mathcal{H}$ to denote the set of target industries, for which we want to compute (5) and (6), and $\mathcal{H}_{\text{full}} \backslash \mathcal{H}$ to denote the other industries.

To estimate the transition matrices $\mathbf{P}_i^{OL}$ and $\mathbf{P}_i^{OC}$, we used audit samples of units for which both $\hat{s}_i^t$ and s_i^t are observed (see Section 5). As these audit samples are small, we introduced parsimonious models for the probabilities in these matrices. The model for $\mathbf{P}_i^{OL}$ is described in this section and the model for $\mathbf{P}_i^{OC}$ will be described in Section 4.

The transition matrix $\mathbf{P}_i^{OL}$ (Table 1) was modelled and estimated in the same way as described previously in Van Delden et al. (2016), so we only discuss this briefly. We divided the transition matrix into three parts: (1) the diagonal elements within $\mathcal{H}$ $(h = 1, \cdots, H)$; (2) the off-diagonal elements within $\mathcal{H}$; and (3) the elements that belong to $\mathcal{H}_{\text{full}} \backslash \mathcal{H}$. The latter stratum is also denoted by $h = H + 1$.

Table 1: Transition probabilities (subscript i omitted) for $\mathbf{P}_i^{OL}$.

True industry	Observed industry				
	1	2		H	$H+1$
1	p_{11}	p_{12}	$\cdots$	p_{1H}	$p_{1,H+1}$
2	p_{21}	p_{22}	$\cdots$	p_{2H}	$p_{2,H+1}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
H	p_{H1}	p_{H2}	$\cdots$	p_{HH}	$p_{H,H+1}$
$H+1$	$p_{H+1,1}$	$p_{H+1,2}$	$\cdots$	$p_{H+1,H}$	$p_{H+1,H+1}$

3.1 The diagonal elements

For the diagonal elements we estimated the probability π_i of unit i to be classified correctly, $\pi_i = P(\hat{s}_i^t = g | s_i^t = g)$ by means of a logistic regression on a number of independent variables (McCullagh and Nelder, 1989), namely size class, number of chamber of commerce units and the observed industry code. We estimated the probabilities π_i from an audit sample that we drew on 1 July

2014, further referred to as the '2014 audit sample'. For the '2014 audit sample', those three variables described the π_i sufficiently well (see Van Delden et al., 2016).

3.2 The off-diagonal elements

For the off-diagonal elements, the starting point is $1-\pi_i$, which stands for the probability that the observed industry code is misclassified. Next we estimate, given that a unit is misclassified, the probability distribution over the other observed industry codes, according to

$$P\left(\hat{s}_i^t = h \middle| s_i^t = g, \hat{s}_i^t \neq g\right) = \frac{P\left(\hat{s}_i^t = h \middle| s_i^t = g\right)}{1-\pi_i} \equiv \psi(g,h), \quad (g \neq h). \tag{7}$$

We assumed that the conditional probabilities $\psi(g,h)$ are the same for all units. We also assumed that the numbers of misclassified units in the off-diagonal cells follow a log-linear model. We estimated the parameters from the '2014 audit sample'. To further reduce the number of parameters, we grouped the off-diagonal cells into five clusters, where cells within the same cluster are supposed to have a comparable probability of misclassification. The estimation procedure is explained in Van Delden et al. (2016).

3.3 The $H+1$ colums and rows

Finally, we estimated the overall probabilities $p_{g,H+1}$ and $p_{H+1,h}$, the last column and last row in Table 1. The units in the last row were observed in the audit sample, so the corresponding probabilities could be estimated directly from the log-linear model of Section 3.2. Note that direct estimation of the probabilities in the last column would require an additional, very large, audit sample from units observed within the non-target industries. Instead we used an indirect approach to approximate the probabilities in the last column.

Let $B = \sum_{g=1}^{H} N_{g,H+1} / \sum_{h=1}^{H} N_{H+1,h}$ denote the ratio between the total number of "missed units" in the true industries $\{1,\ldots,H\}$ and the number of "wrong units" in the observed industries $\{1,\ldots,H\}$. As noted above, the sum $\sum_{h=1}^{H} N_{H+1,h}$ can be estimated from the '2014 audit sample'. Given an value for B, we can estimate the sum $\sum_{g=1}^{H} N_{g,H+1}$ as $B \sum_{h=1}^{H} N_{H+1,h}$. We propose to approximate B by the corresponding ratio of observed yearly transitions in the GBR, that is the ratio between the numbers of units that enter and leave the target industries. Given an estimate for $\sum_{g=1}^{H} N_{g,H+1}$, we can estimate the probabilities $p_{g,H+1}$ using the log-linear model of Section 3.2 (see van Delden et al., 2016).

3.4 Consequence of model estimation for bootstrapping

For the bootstrap simulations, the probability $p_{g,H+1} = P(\hat{s}_i \in \mathcal{H}_{\text{full}} \backslash \mathcal{H} | s_i = g)$ refers to the event that a unit from a given target industry g is observed in an unspecified industry outside the target set ("missed turnover"). Therefore, there is no need for further refinement as those units do not contribute to the target industries. Likewise, the probability $p_{H+1,h} = P(\hat{s}_i = h | s_i \in \mathcal{H}_{\text{full}} \backslash \mathcal{H})$, refers to the event that a unit from an unspecified industry outside the target set is observed in a given target industry h. For this "excess turnover" we do need a further refinement, because the properties of an erroneously included unit may depend on its actual industry. For the car trade case study, to be discussed below, Van Delden et al. (2016) found that erroneously included units originated from a wide range of non-target industries with different turnover distributions.

In that paper, we assumed that the relative number of units from each non-target industry $g \in \mathcal{H}_{\text{full}} \backslash \mathcal{H}$ that are erroneously observed in a given target industry $h \in \mathcal{H}$ is proportional to the corresponding yearly transitions in the GBR. The associated excess turnover values were obtained from a simple log-normal distribution. The latter step is not easily extended to time-related classification errors because in the current study we need a time series of turnover for each unit. Therefore, we introduce an alternative approach for the turnover values.

We propose to make use of the actual units in a given year that according to the GBR move from outside the target set to an industry within the target set. These units encompass the empirical

distribution of the erroneously observed units within the target industries. In other words we extend our population of observed units by drawing a bootstrap sample (with replacement) from the missed units, such that the number of erroneously observed units is $1/B$ times the number of missed units of that industry. Note that we use a non-parametric bootstrap for these units.

The procedure can in principle be repeated for multiple years. However, for practical reasons, we will limit the procedure to sets of two subsequent years. The reason is that the simplifications we used to handle the missed units and the erroneously included units become less realistic over time.

4. Modelling classification errors for the change transition matrix

4.1 The model

The probabilities $\mathbf{P}_i^{OC} = [\, p_{jklhi}^{OC}]$ can be grouped into four situations (A–D), given the values for s_i^t, s_i^{t-1} and $\hat{s}_i^{t-1}$. We take the true industry code in current situation, thus s_i^t, as the starting point. Next we consider whether the true industry code is the same as the one in previous year ($s_i^t = s_i^{t-1}$, situations A and B) or not ($s_i^t \neq s_i^{t-1}$, situations C and D). Further, we regard whether last year's observed industry code is now correct $\hat{s}_i^{t-1} = s_i^t$ (situation A and D) or not $\hat{s}_i^{t-1} \neq s_i^t$ (situation B and C). The logic behind this approach is that the GBR aims (from the viewpoint of the current situation) to obtain the correct industry code during the actualisation of its industry codes from December to January. Thus when the previously observed industry code is now correct, thus when $\hat{s}_i^{t-1} = s_i^t$, there is in fact no need to change the observed industry code, so the correct transition would be $\hat{s}_i^t = \hat{s}_i^{t-1}$. However, when $\hat{s}_i^{t-1} \neq s_i^t$ the GBR should change its observed industry code into the true value, thus $\hat{s}_i^t = s_i^t$ and $\hat{s}_i^t \neq \hat{s}_i^{t-1}$.

Table 2: Four situations for s_i^{t-1}, s_i^t and $\hat{s}_i^{t-1}$.[1]

Sit.	s^{t-1}	s^t	$\hat{s}^{t-1}$	Possibility			Audit substr	Probability
				Nr.	$\hat{s}^t$	Event		
A	j	j	j	1	j	U	AS 3	$1 - p_S$
				2	k	S	AS 1	$p_S \rho(j,k)$
B	j	j	k	1	k	U	AS 2	$\frac{(1-p_R)(1-p_S)}{1-p_R p_S}$
				2	j	R or S	AS 1	$\frac{p_R(1-p_S) + (1-p_R)p_S\rho(k,j)}{1-p_R p_S}$
				3	l	S	AS 1	$\frac{(1-p_R)p_S\rho(k,l)}{1-p_R p_S}$
C	j	l	k or j	1	k or j	U	AS 2	$\frac{(1-p_N)(1-p_S)}{1-p_N p_S}$
				2	l	N or S	AS 1	$\frac{p_N(1-p_S) + (1-p_N)p_S\rho(k,l)}{1-p_N p_S}$
				3	r	S	AS 1	$\frac{(1-p_N)p_S\rho(k,r)}{1-p_N p_S}$
D	j	k	k	1	k	U	AS 2	$1 - p_S$
				2	l	S	AS 1	$p_S \rho(k,l)$

Legend: grey = no change in true industry, blue = change in true industry, green = correct observed industry, red = incorrect observed industry. The colouring of $\hat{s}^{t-1}$ and of $\hat{s}^t$ is relative to the value of s^t.

In summary, we have the following four situations (see also Table 2):

A. ($s_i^t = s_i^{t-1}$) and ($\hat{s}_i^{t-1} = s_i^t$): "no change in true industry, the previously observed code is now correct";
B. ($s_i^t = s_i^{t-1}$) and ($\hat{s}_i^{t-1} \neq s_i^t$): "no change in true industry, the previously observed code is now incorrect";
C. ($s_i^t \neq s_i^{t-1}$) and ($\hat{s}_i^{t-1} \neq s_i^t$): "change in true industry, the previously observed code is now incorrect";
D. ($s_i^t \neq s_i^{t-1}$) and ($\hat{s}_i^{t-1} = s_i^t$): "change in true industry, the previously observed code is now correct".

Within these four situations different events may occur. When $\hat{s}^t = \hat{s}^{t-1}$ we say that the observed industry is UNCHANGING (event U). In case that the observed industry code changes ($\hat{s}^t \neq \hat{s}^{t-1}$) there are a number of possibilities for the transition $\hat{s}^{t-1} \rightarrow \hat{s}^t$. Each of these has a certain probability of occurrence, depending on the situation. We model the probability for $\hat{s}^t \neq \hat{s}^{t-1}$ for three events:

- NOTICE (N) a true change in industry. We denote the probability of this event—given $s^t \neq s^{t-1}$—by p_N. When this event occurs then $\hat{s}^t = s^t$.
- RESTORE (R) an industry error that was present in year $t-1$ (for instance when the true industry code had changed in the past but that change was not noticed in the GBR). The probability that this event occurs—given $s^t = s^{t-1}$ AND $\hat{s}^{t-1} \neq s^{t-1}$—is p_R. When the event occurs then $\hat{s}^t = s^t$.
- SPURIOUS CHANGE (S) of the observed industry. This event can occur under any condition, with probability p_S. The newly observed industry code $\hat{s}^t$ is drawn from a transition matrix with elements $\rho(j,k) = P(\hat{s}_i^t = k | \hat{s}_i^{t-1} = j, \text{S occurs})$ where $\rho(j,j) = 0$. This event concerns all changes without a clear explanation.

For each of the four situations, the different events that may occur are shown in Table 2. For instance, in situation A, after the transition $\hat{s}^{t-1} \rightarrow \hat{s}^t$ the observed industry code may be correct ($\hat{s}^t = s^t$) corresponding to event U, or incorrect ($\hat{s}^t \neq s^t$), corresponding to event S. Notice that event R can occur only for enterprises in situation B. Event N only applies to units in situation C[2]. Event S can occur for any unit. We further assume that events N, R and S occur independently, with the restriction that at most one event can occur in the transition from year $t-1$ to t. The probability that two (or more) events occurs simultaneously is small anyway, but it is simpler to exclude that option. We assume initially that p_N, p_R, p_S and $\rho(j,k)$ are the same for all units (but see Section 6).

We have worked out the probabilities of different possibilities within the situations A–D, in terms of the three parameters p_N, p_R and p_S. The result is shown in the final column of Table 2. Note that within each of the situations A, B, C, and D, the probabilities sum to 1.

4.2 Estimating the parameters

The model for $\mathbf{P}_i^{OC}$ can be estimated from an audit sample where for each sampled unit the values for $\hat{s}_i^t$, $\hat{s}_i^{t-1}$ and s_i^t, s_i^{t-1} were obtained (see Section 5). We propose to estimate only the parameters p_N, p_R and p_S from the audit sample, and this only for "simple" units (see Section 5). Since the events are rare, a small audit sample is insufficient to estimate all of the conditional transition probabilities $\rho(j,k)$. Instead, we approximate all $\rho(j,k)$ by the relative frequencies of observed changes within the GBR. So we assume that the $\rho(j,k)$ for the true industry codes are close to those of the observed industry codes.

We estimate the parameters p_N, p_R and p_S by maximum likelihood (ML). Maximising the likelihood function of the observed data directly is complicated, because there are two cases for which it is not clear which event occurs: "Situation B – possibility 2" (see Table 2) and "Situation C – possibility 2",

[2] For units in situation D, doing nothing actually has the same effect as event N. Somewhat arbitrarily, we restrict the event N to apply only to units in situation C. It turns out that this greatly simplifies the estimation of the model.

namely either R or S (first case) or N or S (second case) occurred. By introducing two latent binary variables that indicate which event occurred in these situations, we obtain a complete-data likelihood function that is easy to maximise. An EM algorithm (Little and Rubin, 2002) can then be used to obtain ML estimates for p_N, p_R and p_S.

In this case, the EM algorithm works as follows. Denote the number of sampled units for situation $X \in \{A, B, C, D\}$ as n_X. Further, let $n = n_A + n_B + n_C + n_D$ and let n_0 be the number of units where $(\hat{s}_i^{t-1} \neq \hat{s}_i^t)$. The estimated population equivalents (after multiplying by the sampling weights w_i) are $\widehat{N} = \widehat{N}_A + \widehat{N}_B + \widehat{N}_C + \widehat{N}_D$ and $\widehat{N}_0$. To initialise the EM algorithm we use starting values $p_N^{(0)} = \widehat{N}_{C2}/\widehat{N}_C$, $p_R^{(0)} = \widehat{N}_{B2}/\widehat{N}_B$ and $p_S^{(0)} = \left(\widehat{N}_{A2} + \widehat{N}_{D2}\right)/\left(\widehat{N}_A + \widehat{N}_D\right)$.

E-step. Given the current parameter estimates $p_N^{(g)}$, $p_R^{(g)}$ and $p_S^{(g)}$ compute

$$M_B^{(g)} \equiv M_B\left(p_R^{(g)}, p_S^{(g)}\right) = \sum_{\substack{i \in s_B: \\ \hat{s}_i^t = s_i^t}} w_i \left(\frac{p_R^{(g)}\left[1 - p_S^{(g)}\right]}{p_R^{(g)}\left[1 - p_S^{(g)}\right] + \left[1 - p_R^{(g)}\right] p_S^{(g)} \rho\left(\hat{s}_i^{t-1}, \hat{s}_i^t\right)} \right) \tag{8}$$

and

$$M_C^{(g)} \equiv M_C\left(p_N^{(g)}, p_S^{(g)}\right) = \sum_{\substack{i \in s_C: \\ \hat{s}_i^t = s_i^t}} w_i \left(\frac{p_N^{(g)}\left[1 - p_S^{(g)}\right]}{p_N^{(g)}\left[1 - p_S^{(g)}\right] + \left[1 - p_N^{(g)}\right] p_S^{(g)} \rho\left(\hat{s}_i^{t-1}, \hat{s}_i^t\right)} \right). \tag{9}$$

where $M_B^{(g)}$ and $M_C^{(g)}$ denote the expected numbers within situation B2 and C2 (given the current parameter estimates) where the first event occurs (R or N).

M-step. Given the expected numbers $M_B^{(g)}$ and $M_C^{(g)}$ from the E-step compute

$$\begin{aligned} p_N^{(g+1)} &= \frac{M_C^{(g)}\left[\widehat{N} - M_B^{(g)} - M_C^{(g)}\right]}{M_C^{(g)}\left[\widehat{N} - M_B^{(g)} - M_C^{(g)}\right] + \left[\widehat{N}_C - M_C^{(g)}\right]\left(\widehat{N} - \widehat{N}_0\right)}, \\ p_R^{(g+1)} &= \frac{M_B^{(g)}\left[\widehat{N} - M_B^{(g)} - M_C^{(g)}\right]}{M_B^{(g)}\left[\widehat{N} - M_B^{(g)} - M_C^{(g)}\right] + \left[\widehat{N}_B - M_B^{(g)}\right]\left(\widehat{N} - \widehat{N}_0\right)}, \\ p_S^{(g+1)} &= \frac{\widehat{N}_0 - M_B^{(g)} - M_C^{(g)}}{\widehat{N} - M_B^{(g)} - M_C^{(g)}}. \end{aligned} \tag{10}$$

The E- and the M-step are repeated until the estimated parameters have converged.

5. Case study: data and audit samples

5.1 Case study

We applied the effect of NACE classification errors to the estimation of quarterly growth rates for the short-term business statistics. We derived turnover from value added tax (VAT) data for the smaller and simple statistical units. By simple units we mean statistical units with a $1{:}m$ relationship to VAT units. The complex and most complex units concern units for which the enterprise group is split up into multiple enterprises (for the simple units the enterprise group consists of one enterprise, as an approximation). The most complex units concern units that belong to an enterprise group with complicated, international structures, that are treated by a special team at Statistics Netherlands (SN).

This special team ensures that all variables that are collected across different outputs for those units are consistent with each other. VAT units cannot be uniquely related to the complex and most complex units, and for the latter units we use census survey data. Altogether we have turnover data for all units in the target population.

We limit our accuracy estimates to the economic sector *car trade* (NACE G45). Within car trade, there are six publications cells for which STS estimates are published and there are nine industries. Based on those nine industries all publications that use turnover (STS, SBS, National accounts) can be produced. Using the set of industries we tried to develop a method to quantify the effect of time-related classification errors. We used quarterly turnover data of eight quarters: the first quarter of 2014 through to the fourth quarter of 2015.

5.2 2015 Audit sample

We aim to find instances of all four situations and of all events mentioned in Table 2. It might require a large sample to find instances where the true industry code changed whereas the observed industry code did not change. To solve this issue we attempted to divide the units that are present in both t and $t-1$ into three "audit strata" (AS):

AS 1: the observed industry in the GBR has changed ($\hat{s}_i^t \neq \hat{s}_i^{t-1}$);

AS 2: the observed industry in the GBR has not changed ($\hat{s}_i^t = \hat{s}_i^{t-1}$) and there is a large probability that the observed industry contains an error in either period ($\hat{s}_i^t \neq s_i^t$ and/or $\hat{s}_i^{t-1} \neq s_i^{t-1}$);

AS 3: the observed industry in the GBR has not changed ($\hat{s}_i^t = \hat{s}_i^{t-1}$) and there is a small probability that the observed industry contains an error in either period ($\hat{s}_i^t \neq s_i^t$ and/or $\hat{s}_i^{t-1} \neq s_i^{t-1}$).

The demarcation of AS 1 follows directly from the observed GBR data. The criteria by which the remaining units were assigned to either AS 2 or AS 3 will be discussed in Section 5.3.

Next, we divided the population of panel units at 1 July 2015 into 27 strata (nine industries times three audit strata) and sampled 10 units from each stratum. We refer to this as the '2015 audit sample'. Since the vast majority of units is part of AS 3, this sample allocation implies that units with changed observed industry codes (AS 1) or a large probability of erroneous observed industry codes (AS 2) are oversampled. In addition we divided the population of units that were present at 1 July 2015 but not at 1 July 2014 (births) into 9 industry strata and sampled 3 units from each stratum. The total sample size was 270 + 27 = 297 units.

We gave the IDs of the sampled enterprises—without the AS information—to an expert in industry classification at CBS. For each unit this expert (aimed to) determine the true actual value of the industry at time of judgement and the true industry 12 months earlier. He made use of ownership relations of the unit, of current and past internet information and he contacted the enterprise when needed. Past internet information was obtained by using the internet archive "waybackmachine.org" that saves snapshots of internet pages at regular intervals (a few times per year). For practical reasons the judgement was done from November 2015 – January 2016. The expert also looked up the observed industry codes in the GBR at the same time points. Note that the obtained information does not aim to give information on the state of the codes at 1 July 2015, but gives insight into the difference between true and observed industry codes at two time points with 12 months difference.

5.3 Determining the audit strata

In order to distinguish AS 2 from AS 3 we re-used the '2014 audit sample'. This sample was drawn at 1 July 2014 and consisted of 25 units per car trade industry, for which the true and the observed industry codes were determined at that time (see Van Delden et al., 2016). We added background variables to the '2014 audit sample' to compute eight indicators that potentially indicate the presence of an error in the observed industry code:

1. the so-called EMP fraction—the relative contribution of the legal units (measured as the relative number of employees) that have the same industry code as the enterprise as a whole—is ≤ 0.4 in both years (t and $t-1$);
2. a clear change in the EMP fraction from ≤ 0.4 at $t-1$ to ≥ 0.6 at t or vice versa;

3. number of legal units per enterprise in t is ≥ 4;
4. name of the enterprise changes from $t-1$ to t;
5. change in set of names of the legal units underlying the enterprise from $t-1$ to t;
6. large change in turnover per employee (for an enterprise) from $t-1$ to t;
7. change in industry code within VAT data set;
8. change in activity code within a commercial data set (www.locatus.com) of a company that visits shops in the Netherlands and applies its own classification of economic activity.

Denote the indicators by I_{ki} $(k = 1, \cdots, 8)$, with $I_{ki} \in \{\text{TRUE (T)}, \text{FALSE(F)}\}$. We analysed the effectiveness of the eight indicators by computing $E_k = \sum_i (I_{ki} = \text{T} \;\&\; \hat{s}_i^{2014} \neq s_i^{2014}) \,/\, \sum_i (I_{ki} = \text{T})$, i.e., the fraction of units selected by the indicator for which the observed industry code is erroneous. We also combined the scores of two indicators by $I_{k,\ell i}^{\text{MAX}} = \text{T}$ if ($I_{ki} = \text{T}$ or $I_{\ell i} = \text{T}$), and similarly for three or more indicators.

Based on E_k, indicator 2 was the most effective, followed by 1, 8, 4, 5 and 3 (Table 3). Indicators 6 and 7 were not effective at all. For the indicator I_2 and subsequent combined indicators $I_{2,1}^{\text{MAX}}$, $I_{2,1,8}^{\text{MAX}}$ etc. in the order of their effectiveness according to Table 3, we computed an ROC curve (receiver operating characteristic) (Figure 2). The vertical axis shows the probability that true positives (=classification errors) are found and the horizontal axis the probability that false positives are found. Figure 2 shows that the probability for true positives is larger than that for false positives. The curve bends near $I_{2,1,8}^{\text{MAX}}$. We selected $I_{2,1,8i}^{\text{MAX}}$ to define the stratum AS 2 (TRUE) and versus AS 3 (FALSE).

Table 3: Effectiveness of the indicators.

k	1	2	3	4	5	6	7	8
E_k	0.333	0.714	0.108	0.278	0.200	0.000	0.000	0.333

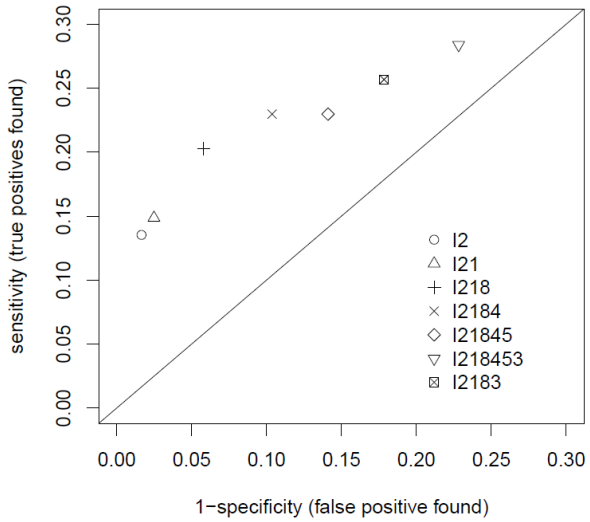

Figure 2: ROC of the (combined) indicators.

6. Results

6.1 Audit sample

The results of the '2015 audit sample' of the continuing units are shown in Table 4. The net sample size was 252, because 18 units had ceased to exist between 1 July 2015 and the moment of auditing. The vast majority of the enterprises in the audit sample—after applying the design weights based on industry × audit stratum—fell under "Situation A – possibility 1" (no change in true and observed industry and observed = true code). The second most frequent occurring case is "Situation B – possibility 1" (no change in true and observed industry and observed ≠ true code). We also clearly observed cases where an error was restored (part of "Situation B – possibility 2") but this was a

limited part of the total number of cases in situation B. Likewise, we found situations where an observed change in industry codes corresponded with a true change (part of "Situation C – possibility 2"), but this is a limited fraction compared to the total number of cases in C. Note that B2 and C2 had much smaller design weights that B1 and C1 so the ratios as directly computed from the sample are much larger than those computed using the design weights.

Table 4: Results of the audit sample (Sample = unweighted counts; Pop = weighted counts).

type	subgroup	45111	45112	45191X	45194	45200	45310	45320	45401	45402
Sample	A1	6	26	13	21	18	20	14	23	22
	A2	0	0	0	0	0	0	3	0	0
	B1	18	0	10	2	6	7	7	4	5
	B2	2	0	3	4	1	0	2	1	1
	B3	2	0	1	0	0	0	0	0	1
	C1	0	0	0	1	1	0	0	0	0
	C2	0	1	1	1	1	0	0	0	0
	C3	1	0	0	0	0	1	0	0	0
	D1	0	0	0	0	0	1	0	0	0
	D2	0	0	0	0	0	0	0	0	0
Pop	A1	44	16551	539	262	3605	1357	607	329	891
	A2	0	0	0	0	0	0	18	0	0
	B1	53	0	624	32	1202	352	103	78	211
	B2	4	0	18	10	20	0	3	2	4
	B3	3	0	9	0	0	0	0	0	4
	C1	0	0	0	29	521	0	0	0	0
	C2	0	32	9	2	20	0	0	0	0
	C3	1	0	0	0	0	9	0	0	0
	D1	0	0	0	0	0	16	0	0	0
	D2	0	0	0	0	0	0	0	0	0

Estimated from the sampled data (using the design weights), 0.608% of the units in the population had a change in the observed industry and 0.369% had a change in the true industry code. That exemplifies that it is very difficult to draw a small sample in which all the different situations are found. To increase the efficiency of the sampling in that respect, we used the audit stratum. We analysed whether the use of the audit stratum was effective. First we estimated the relative proportion of population units per situation from the results in Table 4. Next, we computed the expected number of sampling units per situation (left column in Table 5) that would have been obtained when we would have sampled randomly the same number of active units (as in the audit sample) from each industry but now without using the audit stratum. We compared those numbers with the actual sampling numbers per situation. (right column of in Table 5). Both are aggregated over the industries within car trade. Using the audit stratum proved to be effective in finding cases for Situation B2 and C2.

We computed the parameter estimates by the EM-algorithm in two ways. First we used the sampling weight according to industry × audit stratum. That resulted in the estimates $\hat{p}_N = 0.614$, $\hat{p}_R = 0.026$ and $\hat{p}_S = 0.00173$. Next, we used only the audit stratum weights, which resulted in $\hat{p}_N = 0.159$, $\hat{p}_R = 0.043$ and $\hat{p}_S = 0.00054$. Under the assumption that our model is correct—thus that the parameters p_N, p_R and p_S do not vary by industry—the latter estimates are the best, because including the industry weights will increase the variance of the estimates (Kish, 1992). In fact, the weights according to industry × audit stratum varied much more than those by the audit stratum only. The parameters varied considerably by the two different sets of weights.

Table 5: Expected versus realised number of sampling units per situation (explanation in text).

Situation	Expected (without audit stratum)	Realised (with audit stratum)
A1	182.4	163
A2	0.7	3
B1	58.4	59
B2	2.8	14
B3	1.2	4
C1	5.1	2
C2	0.5	4
C3	0.5	2
D1	0.3	1
Total	252	252

The estimated parameters refer to the simple units that belong to size class 10−40. We expect that the other units have higher probabilities p_N and p_R and a lower probability p_S because more manual effort is put into them in daily production. For the largest, most complex units we expect that p_N and p_R are close to 1.0 and that p_S is close to 0.0 since they are thoroughly checked by the special team at SN (see Section 5.1). Especially for the most complex units it is nearly impossible to estimate those parameters from an audit sample. Instead, we used a linear interpolation in error probability from the most complex units to the simple units (with size class), similarly to the model that we used for the diagonal elements of $\mathbf{P}_i^{OL}$ in Van Delden et al. (2016, figure 2). This is shown in Table 6. In this table $\hat{p}_{audit}$ is an estimated probability from the audit sample, $\hat{p}_{limit}$ is an expert guess of the corresponding probability for largest, most complex units, and $\hat{p}_{1rd} = \frac{2}{3}\hat{p}_{audit} + \frac{1}{3}\hat{p}_{limit}$ and $\hat{p}_{2rd} = \frac{1}{3}\hat{p}_{audit} + \frac{2}{3}\hat{p}_{limit}$. For p_N and p_R we chose $\hat{p}_{limit} = 1.0$ and for p_S we chose $\hat{p}_{limit} = 0.0$.

Table 6: Relative values of the parameters p_N, p_R and p_S per complexity class.

Complexity class	10−30	40	50	60−90
Simple	$\hat{p}_{audit}$	$\hat{p}_{audit}$	$\hat{p}_{1rd}$	$\hat{p}_{2rd}$
Complex	$\hat{p}_{1rd}$	$\hat{p}_{1rd}$	$\hat{p}_{2rd}$	$\hat{p}_{limit}$
Most Complex	$\hat{p}_{1rd}$	$\hat{p}_{2rd}$	$\hat{p}_{limit}$	$\hat{p}_{limit}$

6.2 Accuracy

We limit ourselves to presenting the results for three of the nine car industries, as an example. The other six industries showed similar outcomes. Those three are industry (NACE code) 45112 (sale and repair of passenger cars), 45191X (trade and repair of goods vehicles) and 45402 (retail trade in maintenance and repair of motorcycles). They had quarterly turnover levels of about 8.0 billion euro, 1.3 billion euro and 0.13 billion euro (not shown). Their quarter-on-quarter (qoq) turnover growth rates varied considerably with the quarter of the year (Figure 3, left panel). For instance the sale of cars (45112) decreased in the third quarter of 2014 relative to the second one and increased from the third quarter to the fourth quarter in both years. The large qoq growth rates in the fourth quarter can be explained by special tax regulations that stimulated the sale of cars at the end of both 2014 and 2015.

Provisional results on the accuracy of those qoq changes in turnover are shown in the right panel of Figure 3. Results are also provisional because we have used only 100 bootstrap replicates since the computation time for the bootstrap replicates is rather long. We ultimately wish to use 10.000

replicates because classification errors of large enterprises occur with a small probability but they may have a large impact on the outcomes.

The statistical division at SN considers a root mean squared error of about 1.5 per cent points as an acceptable accuracy level for publication. That means that the accuracy of the qoq changes for industry 45112 and of 45191X are acceptable. However, qoq turnover changes for industry 45402 are too inaccurate. Notice that the RMSE, bias and standard errors of the qoq changes in turnover increased with decreasing turnover levels. This was also found for the six car trade industries that are not shown here.

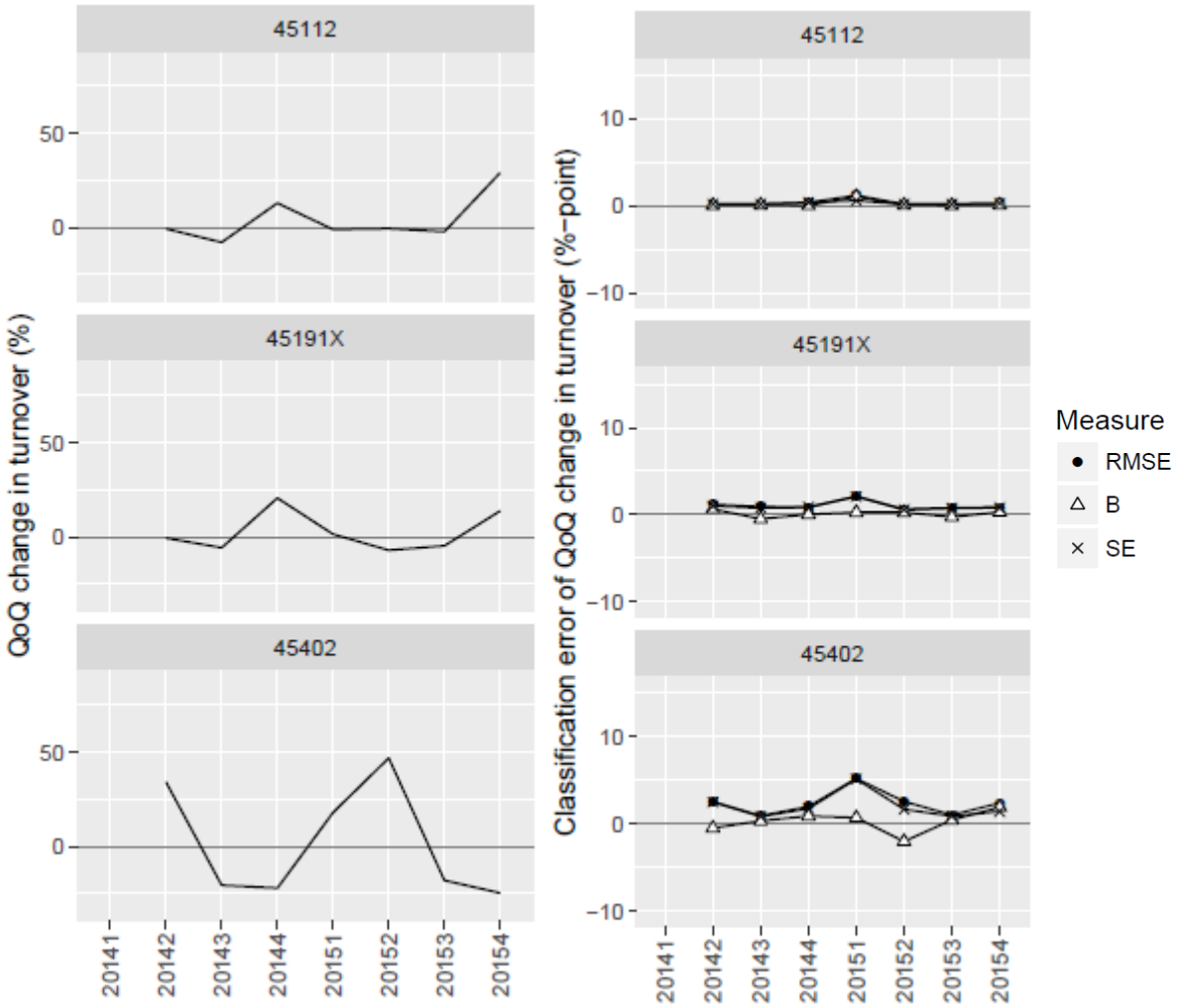

Figure 3: Quarter-on-quarter changes and their corresponding accuracy for three car trade industries; accuracy is measured as root mean squared error (RMSE), bias (B) and standard error (SE).

For industry 45191X and 45402 the RMSE is mainly dominated by the standard error. Furthermore, the RMSE is larger in the first quarter of the year than in the other three quarters of the year. Possibly the peak in the first quarter of 2015 is because units may then change their industry code (according to the change matrix). A further analysis into the data revealed that the number of erroneously included units in the target industry was larger in 2015 than in 2014 (not shown). Maybe this increase in the number of units also caused an increase in standard error. When the erroneously included units have other growth rates than the missing units, that will lead to inaccuracy. It is a point for future research to verify whether this explanation is correct.

7. Discussion

The industry code is a key characteristic of enterprises and it is used to differentiate economic characteristics such as production, use and economic growth into economic sectors. European regulations prescribe that this characteristic is maintained in a central business register that can thus be used to unify different economic outputs of NSIs. In practice, the industry code is often derived automatically from different administrative sources because it is simply too time consuming to determine this characteristic manually for all enterprises. In addition to that, budget cuts at NSIs restrict possibilities to find and correct errors in industry code. Given these conditions, it is relevant to investigate the size of the classification errors and their effect on accuracy of statistical outcomes. In this paper, we focused in particular on the effect of time-related errors in observed industry codes on estimated growth rates by industry.

Despite the fact that (apparently) less than 1% of units within the target population change their true industry code, we were able to collect some information on errors in the observed changes in industry code, using an audit sample of limited size. The background variables that we linked to the population of units were effective in differentiating between units with high and low probabilities to have a true change in industry code. Nonetheless, because the estimates of the parameters p_N, p_R and p_S are based on a limited number of audit cases they are not very precise. It is a point for future research how we can improve the estimation of classification error probabilities. A first idea is to estimate the industry code of enterprises from their web sites using text mining methods (independent from the GBR codes). We could then obtain two independent sets of industry code estimates and use those to estimate classification errors by means of a latent class model (Biemer, 2011).

Manually determining the industry code of a small audit sample (of nine industry codes) required a considerable amount of time. It is not feasible to apply this approach to all industry codes in the STS domain, which encompasses more than 300 industries. An important question for further research is therefore to investigate which simplifications can be introduced to reduce the number of parameters to estimate, while still obtaining acceptable accuracy estimates (compared to the more elaborate approach that we currently use). Notice that the application that we are interested in is a complicated one in that respect because any information on non-sampling errors can only be obtained by collecting *additional* information. In many other applications, different overlapping data sources are combined so that there are two or more measurements per variable per unit. This overlap can provide information on the amount of error in each source, without the need for collecting additional data.

Our bootstrap procedure will lead to some bias in the estimated accuracy, because we start our procedure from the observed data (including the observed industry codes) and then draw new industry codes, see Figure 1. For the case of level estimates, we derived formulas for the size of this bias and we also derived a correction for this bias in Van Delden et al. (2016). For the case of the accuracy of the growth estimates we still have to look into this bias and try to find a correction for it; this a point of future research. So far we ignored this bias, because in Van Delden et al. (2016) we found that the bias-corrected level accuracy estimates were close to the uncorrected estimates.

Finally, we would like to remark that the ultimate practical aim is not just to *determine* the accuracy of the growth rate estimates but also to *improve* the accuracy for those industries for which we find that the current estimates are too inaccurate (e.g. industry 45402 according to the provisional results). It is still an open question how to do this in an efficient manner. In Van Delden et al. (2016) we showed that it is not sufficient to focus on the industries with the lowest accuracies because there are also transitions (between observed and true codes) from other industry strata.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands. We thank Arjen de Boer and Danny van Elswijk for providing the raw data and

Lei Dirrix, Marian Immerzeel, Ivonne Valent, and René Wevers for providing their help with the audit sample. We thank Harm Jan Boonstra for his useful comments.

References

Biemer, P.P., 2011. *Latent Class Analysis of Survey Error*. Hoboken, New Jersey: John Wiley & Sons.

Burger, J., A. van Delden, and S. Scholtus, 2015. Sensitivity of mixed-source statistics to classification errors. *Journal of Official Statistics* 31: 489–506.

Christensen, J.L., 2008. "Questioning the precision of statistical classification of industries." Paper presented at the DRUID Conference on Entrepreneurship and Innovation, June 17–20, Copenhagen.

Costanzo, L. (ed.), 2011. Main Findings of the Information Collection on the Use of Admin Data for Business Statistics in EU and EFTA Countries. Admin Data ESSnet, Work Package 1. Deliverable 1.1

Efron, B. and R.J. Tibshirani, 1993. *An Introduction to the Bootstrap*. London: Chapman & Hall/CRC.

Kish, L. (1992), Weighting for Unequal P_i. *Journal of Official Statistics* 8: 183–200.

Kuha, J. and C. Skinner, 1997. Categorical Data Analysis and Misclassification. In: Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin (eds.), *Survey Measurement and Process Quality*. New York: John Wiley & Sons, pp. 633–670.

Little R.J.A. and D.B. Rubin, 2002. *Statistical Analysis with Missing Data* (second edition). Hoboken, New Jersey: John Wiley & Sons.

McCullagh, P. and J.A. Nelder, 1989. *Generalized Linear Models* (2nd Edition). London: Chapman & Hall.

Struijs, P., 2015. Statistical Units Delineation and the Quality of Business Statistics. ENBES Conference, 7-9 September 2015, Poznań, Poland.

Van Delden, A., S. Scholtus, and J. Burger, 2016. Accuracy of mixed-source statistics as affected by classification errors. Accepted for publication in the *Journal of Official Statistics*.

Van Delden, A., S. Scholtus, and J. Burger, 2015. Quantifying the effect of classification errors on the accuracy of mixed-source statistics. Discussion Paper 2015-10. Available at www.cbs.nl/NR/rdonlyres/B6004222-3760-4E8B-A3A7 3F3FC30939B1/0/2015accuracyclassificationerrors.pdf