DE GRUYTER OPEN

Journal of Official Statistics, Vol. 30, No. 4, 2014, pp. 623–649, http://dx.doi.org/10.2478/JOS-2014-0041

Measuring Representativeness of Short-Term Business Statistics

Pim Ouwehand[1] *and Barry Schouten*[2]

Short-term statistics (STS) are important early indicators of economic activity. The statistics are obligatory for all EU countries and also serve as input to national accounts. In most countries, short-term Statistics are based on business surveys. However, in recent years a number of countries have gradually replaced their business surveys with business VAT registry data. An important question is whether these surveys and registries are representative of the populations and whether representativity is stable in time. We apply R-indicators and partial R-indicators to measure the representativity of both kinds of data sources. We find large differences between different months of the year and between the two data sources. We discuss dual frame approaches that optimize the accuracy of STS statistics.

Key words: Business surveys; registry data; survey nonresponse.

1. Introduction

Short-term business statistics (STS) provide early indicators of economic activity in the EU countries. These statistics are produced on a monthly basis and represent estimated total revenue for various business sizes (in terms of number of employees) and types of economic activity (according to NACE classification of business activities, an abbreviation of 'Nomenclature statistique des activités économiques dans la Communauté européenne'). The STS estimates are mostly based on business surveys. However, an increasing number of countries are starting to include Value Added Tax (VAT) registry data or even to use registry data to replace business surveys entirely. In the Netherlands, registry data is used because legislation prohibits surveying economic indicators that can be derived from registry data with sufficient accuracy. The prerequisite for the use of registry data, hence, is a constraint on quality, which is to some extent left ambiguous. In this article, we investigate an important aspect of quality: the representativeness of the business data that form the input to the STS. We do so by applying a new set of indicators that has recently been proposed and that can supplement more traditional measures such as the unit and quantity response rates.

Both business surveys and business registry data suffer from nonresponse and thus may not be completely representative of the population. Although participation in STS business

[1] Department of Methodology, Statistics Netherlands, PO Box 4000, 2270JM Den Haag, The Netherlands. Email: powd@cbs.nl
[2] Department of Methodology, Statistics Netherlands, PO Box 4000, 2270JM Den Haag, The Netherlands. Email: bstn@cbs.nl

surveys is obligatory by law, some of the businesses do not respond or respond too late. Reporting business data to the VAT register is also obligatory, but the VAT register was not set up to serve statistical needs and, as a consequence, reporting deadlines do not meet the STS deadlines. Some of the reports are still missing when the STS is produced. Furthermore, the Tax Authorities allows smaller businesses to report at a lower frequency than larger businesses. Smaller businesses have to apply for permission to do so, but it can be presumed that this option has a considerable impact on the accuracy of STS statistics.

The nonresponse error is an influential component of the total estimation error. Nonresponse leads to missing data, which in turn may lead to biased estimators of population parameters. The response to business surveys and business registry data should therefore be representative of the population. Although the feature of representativeness is often discussed and debated, it is seldom defined with mathematical rigor. Little and Rubin (2002) provide a clear definition of three missing data mechanisms that underlie inferences about a population parameter of a certain variable. Nonresponse is Missing-Completely-at-Random (MCAR) for a certain variable, say revenue, when the nonresponse is independent of that variable. Nonresponse is Missing-at-Random (MAR) for a variable conditional on a specified set of covariates, when the nonresponse is independent of the variable given the covariates. All other nonresponse is called Not-Missing-at-Random (NMAR). Most business statistics implicitly assume a Missing-at-Random mechanism conditional on business size and type of activity.

Schouten et al. (2009) gave explicit definitions for representative response and for conditionally representative response and introduced quality indicators that measure deviations from these two properties. They have labelled the indicators generally as representativeness indicators, or R-indicators. Response is termed representative for a set of covariates when the propensities to respond are equal over the classes formed by these covariates. Response is termed conditionally representative for one set of covariates conditionally on another set of covariates, when the response propensities for the first set are equal within the classes formed by the second set. The two definitions are closely related to the missing data mechanisms: When response is representative for a set of covariates X, then it is MCAR for all variables in X. When response is conditionally representative for X given Z, then it is MAR for all variables in X given Z. The indicators are based on the estimated variation in response probabilities and have been extensively tested on social survey data. The indicators serve four purposes: comparison of representativeness over surveys, comparison of representativeness of a survey in time, monitoring of representativeness during data collection, and optimization of data collection designs. The choice of covariates depends on the purpose of the indicators, but clearly always excludes the survey variables themselves. Therefore, indicators cannot be used to extrapolate conclusions about MCAR, MAR or NMAR mechanisms beyond those of the selected covariates and one should always mention the selected covariates in order to avoid such conclusions. The rationale behind the indicators is, however, that they measure process quality: The stronger the deviation from representative response on relevant covariates, the more one should worry about nonrepresentative response on survey variables. In our case study for the STS, the available covariates are strongly related to the main survey variables. An extensive exposition and discussion of representativity is given in the papers by Kruskal and Mosteller (1979 a, b and c).

To produce statistics more efficiently, less labour intensively, and of higher quality, Statistics Netherlands is replacing part of its surveys with registry data, particularly for small and medium-sized enterprises. However, there is a clear difference in missing data mechanisms between these two sources of data, which is based on the reporting schedule (monthly, quarterly, annually) and lateness of the VAT data.

The main underlying question of this article is whether VAT data can lead to the same accuracy of STS statistics as survey data or whether a dual frame approach is required. Of course, this question has many angles, of which representativeness of response is just one, but an important one in our view. In order to investigate representativeness, we focus on two purposes of the indicators: comparison of STS over time and monitoring during data collection. This is done for both business survey data and business VAT registry data. The monitoring and adjustment of the collection process based on R-indicators is clear for STS, but less clear for VAT data, since these latter data are not collected via a survey. However, the collection process can be influenced in a less direct way, by agreeing with the Tax Authorities when data is sent. The detailed research questions are:

- How representative are survey and registry data with respect to relevant business characteristics?
- How does representativeness evolve in time, that is, over months and during data collection?
- What groups need to be targeted to improve representativeness of survey and registry data?
- How can survey and registry data be optimally combined in a dual frame approach?

The answer to the fourth question is dependent on the answer to the third question; only if both data sources attract different respondents can they complement each other. We will show that VAT and the STS survey indeed have different underrepresentations of businesses.

To evaluate the representativeness of response and be able to compute the R-indicators, we linked various registries to the business survey and VAT registry of 2007. The VAT registry data for 2006 were linked to both data sets. The Tax Authorities registry of wages and the type of economic activity as derived by the Chamber of Commerce were linked to the VAT data as well. We linked similar variables from the business population register as maintained by Statistics Netherlands to the business survey.

In Section 2, we provide a short background with respect to representativeness and representativeness indicators. In Section 3, we describe the STS data sources and the available business characteristics. We answer the four research questions in Section 4 and end with a discussion in Section 5.

2. How to Measure Representativeness?

In this section, we briefly revisit the definitions of representative response and of so-called representativeness indicators or R-indicators. These measures were introduced by Schouten et al. (2009) and Schouten et al. (2011). We do not give a detailed statistical account of their statistical properties but refer to Shlomo et al. (2012) for details.

In this section, we also link R-indicators and unit response rates to quantity response rates, which are more common measures of nonresponse error in business surveys (see the recent review by Thompson and Oliver 2012).

Throughout this section, we illustrate the concepts using a simplified example. Consider a simulated business population stratified into four disjoint subpopulations defined by crossing two characteristics: type of economic activity (NACE) in two categories and activity status in previous calendar year (yes or no reported VAT). The sizes of the four groups in the population are: NACE Type 1 business and not active in previous year = 33%, NACE Type 1 business and active in previous year = 17%, NACE Type 2 business and not active in previous year = 17%, and NACE Type 2 business and active in previous year = 33%. Table 1 contains the unit response rates over the first six months for the four subpopulations. Also given are the average monthly revenues of businesses in the four subpopulations, which we take as constant over the six months for the sake of simplicity. The unit response rates for the four subpopulations are consistently different, with each response pattern remaining fairly consistent over the observed months. In the following sections, we evaluate the representativeness of the response over time in the example.

2.1. *Overall Representativeness – R-indicators*

In daily survey practice, the term ‘representativeness’ is often used as a desirable property of response, but without a rigorous definition. Schouten et al. (2009) therefore propose a definition of representative response. They call a response representative when response probabilities are equal for all population units, or, in other words, when the population units all show exactly the same response behaviour. A natural measure of deviation from representative response given the definition is the standard deviation of response probabilities. Schouten et al. (2009) transform the standard deviation, so that it takes values between 0 (fully nonrepresentative) and 1 (fully representative), and call it a representativeness indicator or R-indicator. The rationale behind R-indicators is that they are a relevant measure that can be monitored, evaluated and compared over different surveys or registry data, and that are complementary to the unit response rate. In Subsection 2.4, we show how the unit response rate and the R-indicator relate to the quantity response rate.

We introduce some notation. Let X be a vector consisting of auxiliary variables, for example, number of employees, reported VAT in a previous year and economic activity

Table 1. Monthly unit response rates and average monthly revenue for subpopulations based on type of economic activity (NACE) and activity status in the previous year (reported VAT > 0)

Type	Status	Jan	Feb	Mar	Apr	May	Jun	Average revenue
1	Not active	65%	72%	71%	69%	71%	65%	100
1	Active	92%	88%	92%	92%	89%	85%	300
2	Not active	62%	66%	65%	66%	66%	60%	200
2	Active	91%	89%	90%	89%	88%	85%	400

Table 2. Monthly R-indicators with respect to economic activity and status

	Jan	Feb	Mar	Apr	May	Jun
R(X)	0.726	0.809	0.779	0.778	0.807	0.781

(NACE). Let the response propensity function $\rho_X(x)$ be defined as the probability of response given that $X = x$. A response to a survey is called representative with respect to X when response propensities are constant for X, that is, when $\rho_X(x)$ is a constant function.

The R-indicator for X is defined as the standard deviation $S(\rho_X)$ of the response propensities transformed to the [0,1] interval by

$$R(X) = 1 - 2S(\rho_X). \tag{1}$$

When all propensities are equal, the standard deviation is zero and hence fully representative response is represented by a value of 1 for the indicator. A value of 0 indicates the largest possible deviation from representative response.

Table 2 provides the R-indicator values for the example of Table 1 based on the two auxiliary variables ‘type of economic activity’ and ‘activity status’. It shows that the indicator for January is considerably lower than for the other months. Hence, in January the variation in the subpopulation response propensities is largest and the businesses show the most diffuse response behaviour.

2.2. Disentangling Nonrepresentative Response – Partial R-indicators

In order to locate the sources of deviations from representative response, Schouten et al. (2011) introduce partial R-indicators. Partial R-indicators perform an analysis of variance decomposition of the total variance of response probabilities into between and within variances. The between and within variance components help to identify variables that are responsible for a large proportion of the variance. The partial R-indicators are linked to a second definition called conditional representative response, defined as a lack of within variance. The resulting between and within components are termed unconditional and conditional partial R-indicators.

Again we introduce some notation. Let Z be an auxiliary variable not included in X, for example, the region in which a business is located. Let $\rho_{X,Z}(x, z)$ be the probability of response given that $X = x$ and $Z = z$. The response to a survey is called conditionally representative with respect to Z given X when conditional response propensities given X are constant for Z, that is, when $\rho_{X,Z}(x, z) = \rho_X(x)$ for all z. Hence, when the response propensities over country regions are the same for businesses employing the same type of economic activity, then response for region is conditionally representative given economic activity.

The square root of the between variance $S_B(\rho_{X,Z})$ for a stratification based on Z is called the unconditional partial R-indicator. It is denoted by $P_u(Z)$ and it holds that $P_u(Z) \in [0, 0.5]$. So values of $P_u(Z)$ close to 0 indicate that Z does not produce variation in response propensities, while values close to 0.5 represent a variable with maximal impact on representativeness.

For categorical variables the between variance can be further decomposed to the category level in order to detect which categories contribute most. Let Z be a categorical variable with categories $k = 1, 2, \ldots, K$ and let Z_k be the 0–1 variable that indicates whether $Z = k$ or not. For example, Z represents the region of a country and Z_k is the indicator for area k. The partial R-indicator for category k is defined as

$$P_u(Z,k) = \sqrt{\frac{N_k}{N}}\left(\frac{1}{N_k}\sum_U Z_k\rho_{X,Z}(x_i,z_i) - \frac{1}{N}\sum_U \rho_{X,Z}(x_i,z_i)\right) \tag{2}$$

with N_k the number of population units in category k. It follows that $P_u(Z,k) \in [-0.5, 0.5]$. So a value close to 0 implies that the category subpopulation shows no deviation from average response behaviour, while values close to -0.5 and 0.5 indicate maximal underrepresentation and overrepresentation respectively. The category-level indicators are the category components in the total between variance.

The logical counterpart to the unconditional partial R-indicator is the conditional partial R-indicator. It considers the other variance component: the within variance. The conditional partial R-indicator for Z given X, denoted by $P_c(Z|X)$, is defined as the square root of the within variance $S_W(\rho_{X,Z})$ for a stratification based on X. Again it can be shown that $P_c(Z|X) \in [0, 0.5]$, but now the interpretation is conditional on X. A value close to 0 means that the variable does not contribute to variation in response propensities in addition to X, while large values indicate that the variable brings in new variation. When X is type of economic activity and Z is region, then $P_c(Z|X) = 0$ means that one should focus on economic activity when improving response representativeness, as region does not add any variation.

Again for categorical variables Z, the within variance can be broken down to the category level. The category-level conditional partial R-indicator for category k is

$$\boldsymbol{P}_c(\boldsymbol{Z},\boldsymbol{k}|\boldsymbol{X}) = \sqrt{\frac{1}{N-1}\sum_U Z_k(\rho_{X,Z}(x_i,z_i) - \rho_X(x_i))^2}. \tag{3}$$

Unlike the unconditional indicators, the conditional indicators do not have a sign. A sign would have no meaning as the representation may be different for each category of X. For instance, in some categories a certain economic activity may have a positive effect on response while in others it may have a negative effect. The conditional partial R-indicator for Z is always smaller than the unconditional partial R-indicator for that variable; the impact on response behaviour is to some extent removed by accounting for other characteristics of the population unit.

Table 3 shows the partial R-indicators for the two variables in the example of Table 1; type of economic activity and activity status. As expected, January shows larger values for the partial indicators. However, after conditioning it follows that the extra contribution to selective response in January comes mostly from activity status. In all months, the activity status is the dominant source of selective response.

Table 3. Monthly unconditional and conditional partial R-indicators for type of economic activity (NACE) and activity status (reported VAT in the previous year >0)

		Jan	Feb	Mar	Apr	May	Jun
Type of activity	P_u	0.04	0.02	0.02	0.02	0.02	0.03
	P_c	0.01	0.02	0.02	0.01	0.02	0.01
Activity status	P_u	0.14	0.09	0.11	0.11	0.10	0.11
	P_c	0.13	0.09	0.11	0.11	0.10	0.11

2.3. Representativeness and Nonresponse Bias

R-indicators can be interpreted in terms of nonresponse bias through the variance of response propensities. Consider the standardized bias of the design-weighted, unadjusted response mean $\hat{\bar{y}}_r$ of an arbitrary variable y, say total revenue. The standardized bias of the mean can be bounded from above by

$$\frac{|B(\hat{\bar{y}}_r)|}{S(y)} = \frac{|Cov(y, \rho_Y)|}{\rho_U S(y)} = \frac{|Cov(y, \rho_{\aleph})|}{\rho_U S(y)} \leq \frac{S(\rho_{\aleph})}{\rho_U} = \frac{1 - R(\aleph)}{2\rho_U}, \tag{4}$$

with ρ_U the unit response rate (or average response propensity) and $\aleph$ some 'super' vector of auxiliary variables providing full explanation of nonresponse behaviour.

Clearly, the propensity function $\rho_{\aleph}$ is unknown. Since R-indicators are used for the comparison of the representativeness of response in different surveys or the same survey over time, the interest lies in the general representativeness of a survey, that is, not the representativeness with respect to single variables. Therefore, as an approximation for (4) is used:

$$B_m(X) = \frac{1 - R(X)}{2\rho_U}. \tag{5}$$

B_m is the maximal (standardized) bias for all variables that are linear combinations of the components of X. For other variables, (5) does not provide an upper bound to the bias. The choice of X, therefore, is very important, but even for relevant X, (5) cannot be extrapolated to all survey target variables. If the selected set of variables in X is correlated with the survey variables, then (5) is informative as a quality indicator. If it is not correlated with the survey variables, then it has limited utility.

A useful graphic display of unit response rates and response representativeness is given by so-called response-representativity functions. Ideally, one would like to bound the R-indicator from below, that is, to derive values of the R-indicator that are acceptable and values that are not. If the R-indicator takes a value below some lower bound, then measures to improve response are paramount. Response-representativity functions can be used for deriving such lower bounds for the R-indicator. They are a function of a threshold γ and the unit response rate ρ_U. The threshold γ represents a quality level. The functions are defined as

$$RR(\gamma, \rho) = 1 - 2\rho_U \gamma, \tag{6}$$

and follow by demanding that the maximal bias given by (5) is not allowed to exceed the prescribed threshold γ, that is, from taking $B_m(X) = \gamma$. For STS, a reasonable threshold γ

can be set by considering the final response obtained at the end of data collection when unit response rates are very high.

Figure 1 presents an RR-plot for the example given in Table 1. The pair of values for January is the only set that is above the 15% level. All other months are between the 10% and 15% levels.

2.4. R-Indicators, Unit Response Rates and Quantity Response Rates

A measure commonly used in business statistics is the quantity response rate. It is the ratio between the quantity reported by the respondents and the quantity that would be reported if all sample units were respondents. The quantity response rate is different for each study variable Y. We denote it by ρ_Q and suppress the dependence on Y. The application to business statistics is natural; businesses have diverse revenues and often a small number of businesses make up most of the total revenue. For a useful and recent discussion we refer to Thompson and Oliver (2012).

The quantity response rate is defined as

$$\rho_Q = \frac{y_r}{y_n} = \frac{\sum_{i=1}^{n} d_i r_i y_i}{\sum_{i=1}^{n} d_i y_i}, \tag{7}$$

with d_i the design weight for business i, r_i the $0-1$ response indicator for business i, and y_i the value of the study variable for business i. Hence, y_r and y_n denote, respectively, the design-weighted response and sample totals. In Appendix A we show that for large sample sizes, the expected value of (7) is approximately equal to

$$E\rho_Q = \rho_U + \frac{\text{cov}\,(y, \rho_Y)}{\bar{y}_N}, \tag{8}$$

with $\bar{y}_N$ the population mean. We may view (8) as the population representation of the quantity response rate which is estimated by (7). From (8) we can conclude that the quantity response rate is equal to the unit response rate whenever there is no linear relation

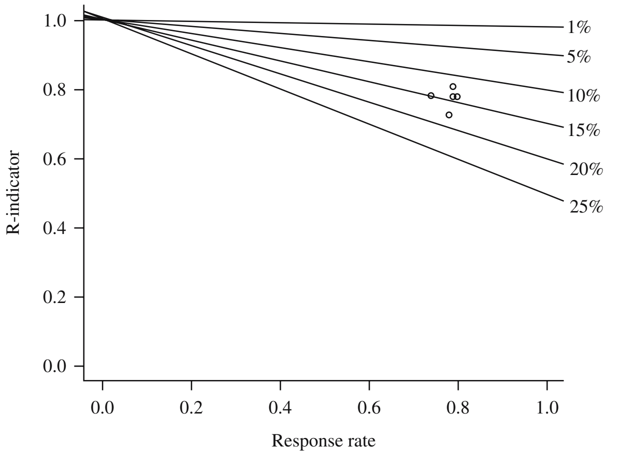

Fig. 1. RR-plot for six months. The thresholds γ are 1%, 5%, 10%, 15%, 20% and 25% (from top to bottom)

between the quantity under study and response propensities. With similar arguments to (4) it can be shown that

$$E\rho_Q \in \rho_U \pm \frac{(1 - R(\aleph))S(y)}{2\bar{y}_N}, \tag{9}$$

so that the R-indicator appears as a component in lower and upper limits to the quantity response rate for auxiliary variables.

The quantity response rate in (7) is an (unbiased) estimator for (8) but can only be computed for variables that are not subject to nonresponse themselves, that is, variables that are auxiliary and can be linked to the sample. For survey variables, the denominator in (7) is unknown and needs to be estimated. It is usually estimated by imputing the nonrespondents or weighting the respondents. The denominator in (7) is then replaced by an estimator that employs auxiliary information, usually taken from the same set of available auxiliary variables that are input to R-indicators. As a result the estimated quantity response rate may be biased itself. Furthermore, when new response comes in during data collection this bias may change and the estimator must be updated retrospectively. Consequently, quantity response rate patterns that are computed when data collection is completed may look different from quantity response rate patterns that are computed in real time during data collection. As a result, and somewhat confusingly, the quantity response rate is not necessarily monotone increasing and may decrease through some periods of data collection; the estimated sample total may become larger when new response comes in.

In this article, we estimate the denominator of (7) using a poststratification estimator. The population is stratified into H subpopulations, $h = 1, 2, \ldots, H$, based on an auxiliary variable, say Z, and the sample mean per stratum is estimated by the design-weighted response mean per stratum. The quantity response rate estimator is then defined as

$$\rho_Q = \frac{y_r}{y_{post}} = \frac{y_r}{\sum_{h=1}^{H} N_h \bar{y}_{r,h}}, \tag{10}$$

with N_h the size of stratum h in the population and

$$\bar{y}_{r,h} = \frac{\sum_{i \in h} d_i r_i y_i}{\sum_{i \in h} d_i r_i} \tag{11}$$

the design-weighted response mean in stratum h.

Appendix A shows that the expected value of (10) can be approximated by

$$E\rho_Q = \frac{\rho_U \bar{y}_N + \operatorname{cov}(y, \rho_Y)}{\bar{y}_N + \sum_{h=1}^{H} \frac{N_h}{N} \frac{\operatorname{cov}_h(\varepsilon, \rho_Y)}{\rho_{U,h}}}, \tag{12}$$

where ε represents the residuals in the poststratification.

When the residuals show no correlation to the response propensities, that is, when the poststratification provides unbiased estimators of the stratum means, then (12) equals (8). If there is a nonzero correlation, then the denominator is biased and (12) and (8) are different. Assuming that the study variable only takes non-negative values, it is possible to derive lower and upper limits to (12) that are expressed in terms of unit response rates and unconditional partial R-indicators

$$E\rho_Q \in \frac{\rho_U \bar{y}_N}{\bar{y}_N - \sum_{h=1}^{H} \frac{N_h}{N} \frac{S_h(\varepsilon)|P_U(Z,h)|}{\rho_{U,h}}} \pm \frac{(1 - R(\aleph))S(y)/2}{\bar{y}_N - \sum_{h=1}^{H} \frac{N_h}{N} \frac{S_h(\varepsilon)|P_U(Z,h)|}{\rho_{U,h}}}. \tag{13}$$

In Table 4, we show the two response rates for the example of Table 1. As expected, the quantity response rate is always higher as businesses with larger revenues have higher response probabilities (see Table 1). Both rates are relatively stable over the months, except for June that has smaller response rates. The simultaneous drop of the rates for June indicates that this drop is not strongly related to revenue.

2.5. The Utility and Limitations of R-Indicators

R-indicators and partial R-indicators can be useful tools to supplement unit and quantity response rates, but they also have limitations. We discuss both here.

In the setting of STS, the quantity response rate would be computed for total business revenue, the key variable. As a single indicator, the complement of the quantity response rate represents the total revenue that is still missing. In conjunction with the unit response rate however, it allows for more elaborate conclusions. The height of the quantity response rate relative to the unit response rate tells whether larger or smaller businesses are overrepresented. A difference in slope between the two rates can provide information on the evolution of these representations; for example, when the quantity response rate grows faster than the unit response rate, then it is likely that bigger businesses have responded better over that time window. The utility of the R-indicator, in addition to unit and quantity response rates, is that it quantifies over- and underrepresentation, it allows for a multivariate view on multiple business characteristics, and it can in theory be estimated without bias both after and during data collection. The R-indicator and partial R-indicators are designed to have a multivariate view. The R-indicator measures the simultaneous deviation from representative response for a range of variables and allows any particular variable to be zoomed in on. The unconditional partial R-indicators do just what quantity response rates are doing: show the impact on single variables. The conditional partial R-indicators allow for a search for the strongest variables in a multivariate context, which is what quantity response rates are lacking; they do not account for multicollinearity.

It is important to stress that the R-indicator values depend on the vector of auxiliary variables X. For different selections of X, the R-indicator attains different values and the (partial) R-indicators do not allow for statements about NMAR nonresponse outside the selected vector of variables. Therefore the selection is a crucial and influential part of

Table 4. Monthly unit and quantity response rates

	Jan	Feb	Mar	Apr	May	Jun
ρ_U	78%	79%	80%	79%	79%	74%
ρ_Q	84%	83%	85%	84%	83%	79%

the analysis. The purpose of the indicator determines the selection of the auxiliary variables that are used. When multiple surveys are compared, it is essential that representativeness is evaluated in terms of generally available and relevant characteristics, such as type of economic activity or business size. For the other three purposes mentioned in Section 1, it is important to select characteristics that are closer to the survey topics and key variables. In the case of short-term statistics, it is paramount to have variables that relate to the revenue of a business. We return to this issue in Section 3.

The response propensity function ρ_X is unknown, and needs to be estimated from the survey response data. A consequence of the estimation of the propensities is that R, P_u, P_c and B_m need to be estimated as well. Schouten et al. (2011) and Shlomo et al. (2012) propose estimators for these population parameters and derive analytic approximations to their standard errors and bias. The estimators replace population means with design-weighted sample and response means and response propensities with estimated propensities. Propensities are estimated by means of general linear models such as linear regression, logistic regression, or probit regression. The resulting estimators have a standard error and indicator values need to be evaluated along with their precision. On the website www.risq-project.eu code in SAS and R is available to compute indicators and their standard errors. To allow for comparison it is crucial that the set of auxiliary variables and the link function, for example, linear or logistic, are kept fixed. Hence, variables are selected beforehand based on their relevance to the survey variables and are always included in the models when monitoring or comparing nonresponse.

Since only response propensities for X need to be estimated, the models for nonresponse cannot be misspecified in terms of omitted variables and in theory response propensities can be estimated without bias. However, since sample sizes are always limited in practice, some interactions between the variables may have to be omitted and/or some categories of variables may have to be merged. In order to enable comparison over surveys and over time, such adaptations need to be applied beforehand to all data sets under study. As a result, the models for nonresponse may be viewed as misspecified for the selected variables and leading to biased estimators for response propensities. It is therefore not enough to provide the variable names when presenting indicators; their classification also needs to be specified.

The R-indicator, variable-level and category-level partial R-indicators together form a set of tools that can be used to search effectively for population subgroups that need to be targeted in data collection. A strategy is given by Schouten et al. (2012):

1. Compute the R-indicator for different time periods.
2. When strong differences are found in Step 1, assess the unconditional variable-level partial R-indicators for all auxiliary variables; the variables that have the highest scores have the strongest single impact on representativity of response. They are also the strongest candidates to be monitored and analysed more closely and subsequently to be involved in design changes and data collection interventions.
3. Assess the conditional variable-level partial R-indicators for all auxiliary variables; the conditional values are needed in order to check whether some of the variables are strongly collinear. If indicator scores remain high, then the strongest variables are selected. If indicator scores vanish by conditioning, then it is sufficient to focus only

on a subset of the variables. A low conditional indicator value implies that the corresponding variable is conditionally representative.
4. Repeat Steps 2 and 3 but now for the category-level partial R-indicators and for the selected auxiliary variables only; the subgroups that need to be targeted in design changes are those categories that have large negative unconditional scores and large conditional scores.

This strategy is used in Subsection 4.3, where we identify the business groups that influence representativeness the most.

3. Short-Term Statistics

3.1. Survey and Registry Data

The traditional way of collecting data for business statistics is to send questionnaires to a sample of enterprises. To produce statistics more efficiently, less labour intensively, and with higher quality, Statistics Netherlands is replacing part of its surveys with registry data, particularly for small and medium-sized enterprises. Apart from costs, a strong incentive for the use of registry data is business response burden. The use of VAT data reported to the Tax Authorities would reduce the burden to enterprises as they have to provide data only once. Yet another advantage of registry data is their sheer size. Registry data aim at a full enumeration of the population. As a consequence, the number of observations is much larger than for regular business surveys.

However, in both surveys and registers part of the data is missing at the time when statistics need to be produced. For the VAT registry data this is particularly the case for monthly statistics (Vlag and Van den Bergen 2010). Although both sources of data are subject to missing data, the missing data mechanisms are very different. In a survey, typically some of the enterprises in the sample do not respond to the questionnaire, or have to be prompted several times. At Statistics Netherlands, however, the enterprises are not targeted in a specific way and data collection is therefore uniform. Registers, on the other hand, may not be complete due to regulations about reporting of enterprises to register holders and time delays in reporting.

The data sets used represent turnover data for both Retail trade and Manufacturing industries for 2007. Turnover refers to the invoice value of sales to third parties of goods and services produced within a company. The VAT register is linked to the employment register containing wages, so that these can be used as auxiliary information. About 75% of the VAT units could be linked to wages from the employment register. For the smallest enterprises (<€2,500 VAT) this was about 60%; for the larger enterprises this was at least 80%. For VAT, we selected all VAT units that were obliged to report their VAT. The number of records is given in Table 5.

The VAT data includes records for companies reporting on a monthly, quarterly or annual basis. The reporting frequency depends on the amount of VAT a company is expected to report, or is based on individual requirements made by the Tax Authorities. If the VAT of a company lies below €1,883 per year, they can report on an annual basis. If it exceeds €15,000 per quarter, they should report on a monthly basis. Most companies

Table 5. Sample and register size

	Retail trade		Manufacturing	
	VAT	*Survey*	*VAT*	*Survey*
January	124,602	7,852	59,346	5,393
June	126,158	7,871	60,229	5,381
July	127,568	7,727	61,023	5,355
December	128,212	7,864	61,521	5,078

report on a quarterly basis (Van Delden and Aelen 2008, and Slootbeek and Van Bemmel 2010).

In the case of the VAT register, companies are required to report 25 days after a reporting period has ended, and statistics are produced 30 days after that period. However, some companies do not report within 30 days. For the STS survey, companies are given the same deadline for responding. The STS sample is a stratified random sample of all enterprises where strata are business size classes. The design weights also depend on the NACE category at the highest level, that is, between the Retail and Manufacturing industries but not within these business types.

3.2. Auxiliary Variables in the Computation of the R-indicators

The comparison was made for four different months with very distinct characteristics of VAT data: January, June, July, and December. The data for January includes only companies reporting on a monthly basis. In June, we have companies reporting on a monthly and on a quarterly basis. July, again, only includes companies reporting on a monthly basis, but unlike January is not at the beginning of the year. December includes companies reporting on a monthly, quarterly, and annual basis.

Ideally, we would like to compare the R-indicators for both types of data using the same auxiliary variables. However, the VAT and survey data sets do not share the exact same set of auxiliary variables. This is caused by the difference in population frames as used by Statistics Netherlands and the Tax Authorities. For VAT data, we can use the current year's monthly wages records, the previous year's VAT records (for the same month), and a business classification (enterprise groups according to NACE classification of 1974). For survey data we can use business size, business classification (economic activity according to NACE classification of 1993) and VAT of the previous year (for the same month). Table 6 presents an overview.

Table 6. Available variables and their number of categories

Variable	# categories
VAT($t - 12$)	9
Wages(t)	10
Business size	9
NACE 2-digit (1974) Manufacturing	20
NACE 3-digit (1974) Retail trade	18
NACE subsection (1993) Manufacturing	12
NACE 3-digit (1993) Retail trade	7

The current year's wages resemble business size. Business size is a classification of the number of employees which can be expected to be proportional to the wages. It is, however, not the *same* variable so that a direct comparison is hampered. The two business classifications also show a clear resemblance but are not exactly the same. This leads to a problem when we want to combine these specific VAT and STS data sets. Tables, however, are available that link the codes of both classifications. For the majority of codes there is a direct translation between the classifications. However, some codes in one classification may be divided into two or more codes in the other system. For this, heuristic solutions are available.

A second difference between the data sets is the units for which turnover is recorded. Tax units do not completely match survey units, especially when larger businesses are concerned. The differences in variables and units imply that some care is needed in the comparison of absolute values of indicators. However, what can be compared is the patterns of representativeness over months and in time.

For both Retail trade and Manufacturing, we tested a model based on the VAT register and a model based on STS survey data. Table 7 presents an overview of the models. For the moment we ignore the type of economic activity.

Since VAT data should replace surveys, we compute the representativity of the response through time, as additional survey or VAT data becomes available. We compare the representativity for both types of data and compare representativity to the unit response rate. Since companies are required to report 25 days after a reporting period has ended, and statistics are produced 30 days after that period, we computed both unit response rate and R-indicator 25, 26, 27, 28, 29, 30, and 60 days after a reporting period had finished.

4. Results

4.1. *What is the Representativeness of STS Based on Survey and Registry Data?*

We first computed the representativity at 25 days after the end of the reporting period has finished. This is currently the deadline for companies to report their VAT, and the moment at which the production of statistics commences.

Table 8 and Table 9 show unit response rates, quantity response rates, and R-indicators for both industries and all months. For VAT, the unit response rate is the number of units that have reported VAT as a proportion of all units in the register (i.e., units that should report their VAT on either a monthly, quarterly or annual basis). For survey data, the unit response rate is the proportion of units in the sample that have responded. The quantity response rate is the proportion of total turnover available at a certain time point. For VAT, these proportions are calculated as the sum of turnovers of reporting units divided by turnover of all units in the register. For survey data, the proportions are calculated as the

Table 7. Models used for the estimation of response propensities

Model	Data set used	Specifications
VAT	VAT register	VAT($t-12$) + Wages(t)
STS	STS survey data	VAT($t-12$) + Business size

Table 8. VAT data: Unit response rates, quantity response rates and R-indicators for four months, 25 days after the reporting period

		Response rate		
Industry	Month	Unit	Quantity	R-indicator
Retail trade	January	0.20	0.57	0.68
	June	0.64	0.74	0.74
	July	0.15	0.41	0.76
	December	0.48	0.42	0.85
Manufacturing	January	0.26	0.65	0.54
	June	0.67	0.57	0.61
	July	0.18	0.46	0.68
	December	0.49	0.34	0.80

sum of turnovers of responding units divided by turnover of all units in the sample. For survey data, the quantity response rate is thus calculated retrospectively.

When we look at the results for VAT data for both Retail trade and Manufacturing, the unit and quantity response rates clearly vary from month to month due to the types of businesses that respond. January and July have lower response rates than June and December. For STS data, response rates show less variation since there is less variation in business types reporting than for VAT data. Despite variation in response rates, the representativity shows less variation. Apparently, the additional enterprises that respond in some months do not make response more representative.

4.2. *How Does Representativeness Evolve in Time?*

The results in the previous subsection focus on a single time lag only. In this subsection we will discuss how response and representativity change during data collection.

In Figures 2 and 3, we present graphs of the unit response rate, the quantity response rate and R-indicator for Retail trade using the VAT model and the STS model, respectively. The graphs show results for January, June, July, and December 2007. In all graphs, indicators are computed after 25, 26, 27, 28, 29, 30, and 60 days of data collection.

Table 9. STS data: Unit response rates, quantity response rates and R-indicators for four months, 25 days after the reporting period

		Response rate		
Industry	Month	Unit	Quantity	R-indicator
Retail	January	0.67	0.71	0.89
	June	0.71	0.70	0.90
	July	0.64	0.71	0.93
	December	0.73	0.65	0.91
Manufacturing	January	0.63	0.67	0.93
	June	0.67	0.72	0.94
	July	0.64	0.69	0.93
	December	0.72	0.76	0.92

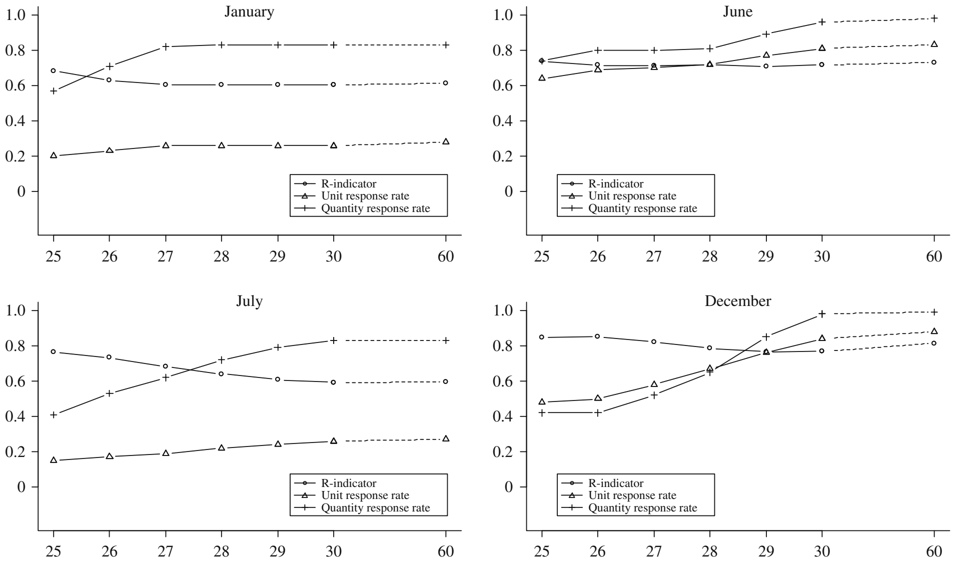

Fig. 2. Unit response rates, quantity response rates and R-indicators based on VAT data for Retail trade (VAT model), for four months

The figures show that for both the VAT and STS model, and all four months under investigation, both the unit response rate and quantity response rate increase as the data collection period progresses. For VAT data, the quantity response rate for July and December approaches 100% (since all companies must report), while for survey data it is relatively stable after 25 days. The response patterns for VAT in January and July are quite

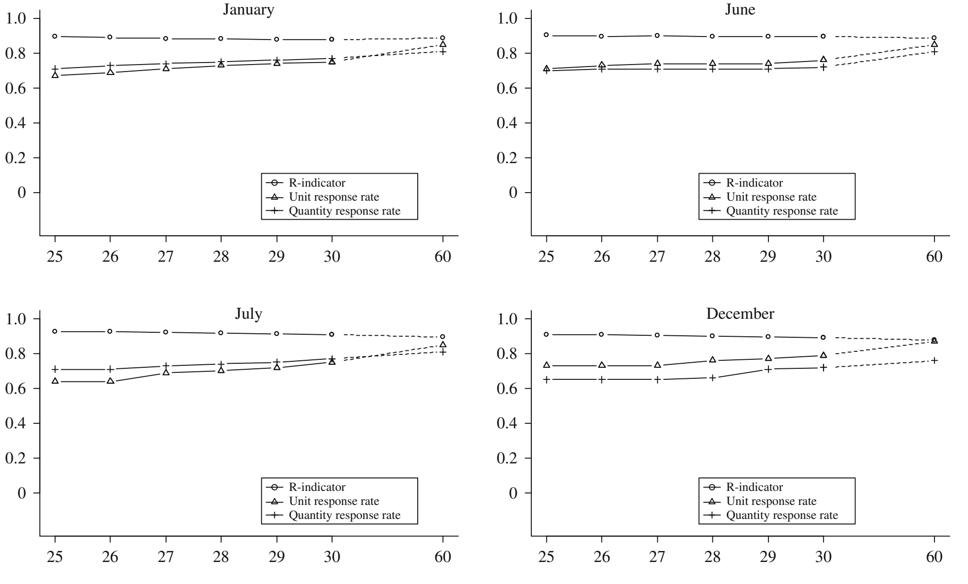

Fig. 3. Unit response rates, quantity response rates and R-indicators based on survey data for Retail trade (STS model), for four months

similar, since these consist of monthly reporters only. Likewise, the patterns for July and December are similar, since these also include quarterly reporters.

In some cases there is a clear difference between the development of the unit response rate and the quantity response rate. This is an indication of a bias in the response. For January and for July (in case of VAT data), these two lines are far apart, meaning that only a relatively small number of companies have reported a large portion of total turnover. This indicates that large companies are overrepresented. At the same time, at the beginning of data collection, the slope of the quantity response rate of January (between 25 and 27 days) and July (between 25 and 30 days) is steeper than that of the unit response rate. In these periods, the number of companies reporting increases only slightly, while the amount of turnover reported increases significantly. This shows that the composition of the response is changing, and this is reflected in the change in the representativity indicators as well. They may change only slightly as the unit response rate increases, or may even decline. Generally, the R-indicators drop as data collection proceeds and there is only a slight increase after 30 days of data collection.

We conclude that the contrast between reporting and nonreporting units increases. The additional response between 25 and 30 days is thus not as representative of the population as the initial response. Hence, waiting longer than 25 days before producing statistics based on VAT data does not make the data more representative.

For the survey data, the difference between the four months is only small. As was mentioned above, in our dataset we only have companies taking part in surveys on a monthly basis. It is only in July that the unit response rate is slightly lower than in other months, which may be due to seasonal effects in Retail trade, such as holidays. Representativity, however, is not different from other months.

In Figures 4 and 5, we present RR-plots of the unit response rate and the R-indicator for VAT and survey data for Retail trade using the VAT model and STS model, respectively.

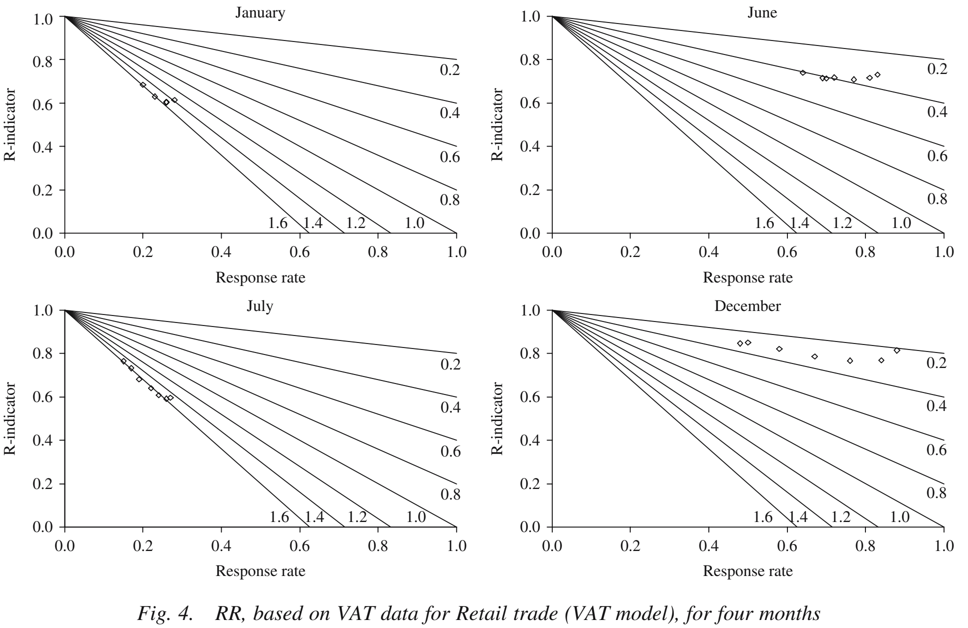

Fig. 4. RR, based on VAT data for Retail trade (VAT model), for four months

The straight lines in the plots represent the maximal bias levels of 0.2, 0.4, . . . 1.6. The plots confirm the previous analyses. The STS survey data show stable patterns over the months. During data collection the maximal bias level remains almost constant. For the VAT data, however, the maximal bias levels vary considerably over the months. Periods with only monthly reporters have a higher maximum bias than other periods.

In summary, the main difference between representativeness of response to surveys and to register holders is the stability over time and during data collection. We conclude that for VAT there is no improvement in the R-indicator and no improvement in the maximal bias when data collection is continued between 25 and 30 days. Since it is crucial for the editing and imputation of business data to start as early as possible, we recommend starting these activities at 25 days after the end of the reference month. For VAT data one must, however, rely much more strongly on nonresponse adjustment methods in months with only monthly reporters and, equally important, be aware that comparability over months is weaker.

4.3. *What Groups of Businesses to Target?*

In this section we deal with the important question of how we can improve the representativeness of STS and VAT. To answer this question, we first need to identify the subpopulations that impact representativeness most. Second, the data collection design needs to be adapted in such a way that these subpopulations receive more attention.

In the previous sections we restricted ourselves to two auxiliary variables: VAT of the previous reporting year and business size. For the VAT data we used total wages as a proxy for business size. In addressing subpopulations, we now add the type of economic activity, see Table 10, as a variable to the assessment of representativeness.

With the exception of Manufacturing in STS, the R-indicator values decrease only marginally when type of economic activity is added. However, for the Manufacturing

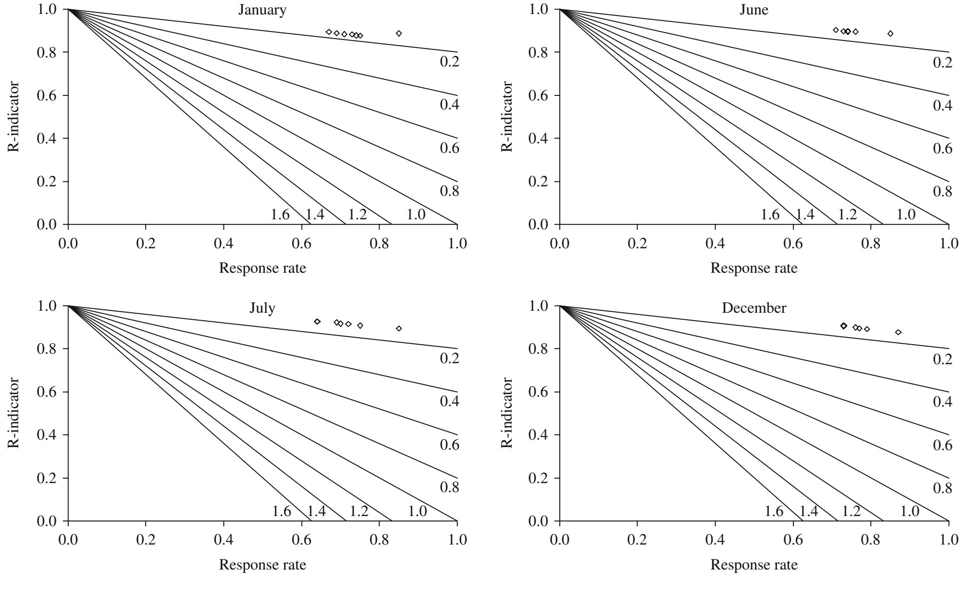

Fig. 5. RR, based on survey data for Retail trade (STS model), for four months

Table 10. Extended models used for the estimation of response propensities

Model	Specifications
Extended VAT model	
Manufacturing	VAT ($t-12$) + wages (t) + NACE-2 digit (1974)
Retail trade	VAT ($t-12$) + wages (t) + NACE-3 digit (1974)
Extended STS model	
Manufacturing	VAT($t-12$) + Business size + NACE subsection (1993)
Retail trade	VAT($t-12$) + Business size + NACE -3 digit (1993)

industry the drop of the STS R-indicator is almost 0.1 and hence the variable provides additional deviation from representative response. Table 11 presents an overview of the variable level partial R-indicators for the auxiliary vector with and without type of economic activity.

From Table 11 we conclude that the extended models do not alter the impact of VAT ($t-12$) and business size or wages (t). We can conclude that type of economic activity plays an almost separate, independent role in representativeness. For this reason, in the remainder of this section, we shall consider the extended models only.

Next, let us explore the dependence of the partial impact of the variables on the data collection month. Table 12 contains the partial R-indicator values for January, June, July and December.

From Table 12 we conclude that the STS representativeness is relatively stable over months and over variables. For VAT, however, the months present quite different pictures. The table also demonstrates that in December, generally, the impact of all variables has reduced considerably. One exception is the impact of wages (t) for Manufacturing, which is strongest in June and comparable to January in December. Furthermore, Table 12 shows that for STS the strongest impact comes from VAT ($t-12$) for Retail trade, and from type of economic activity for Manufacturing. For VAT the strongest impact comes from VAT

Table 11. Unconditional and conditional partial R-indicators without (small) and with (extended) type of economic activity for January after 25 days of data collection

			Unconditional		Conditional	
	Type	Variable	Small	Extended	Small	Extended
STS	Retail	VAT ($t-12$)	0.051	0.051	0.048	0.046
		Business size	0.022	0.022	0.013	0.013
		Activity	–	0.029	–	0.025
	Manufacturing	VAT ($t-12$)	0.023	0.023	0.024	0.023
		Business size	0.028	0.029	0.029	0.028
		Activity		0.038		0.036
VAT	Retail	VAT ($t-12$)	0.152	0.152	0.114	0.117
		Wages (t)	0.110	0.109	0.043	0.051
		Activity	–	0.074	–	0.088
	Manufacturing	VAT ($t-12$)	0.224	0.225	0.213	0.207
		Wages (t)	0.081	0.081	0.039	0.038
		Activity	–	0.057	–	0.028

Table 12. Unconditional and conditional partial R-indicators for the extended model for January, June, July and December

			Unconditional				Conditional			
			Jan	Jun	Jul	Dec	Jan	Jun	Jul	Dec
STS	Retail	VAT ($t-12$)	0.051	0.045	0.034	0.038	0.046	0.039	0.032	0.031
		Business size	0.022	0.026	0.016	0.032	0.012	0.016	0.013	0.023
		Activity	0.029	0.031	0.024	0.028	0.024	0.026	0.022	0.023
	Manufacturing	VAT ($t-12$)	0.023	0.014	0.017	0.022	0.022	0.015	0.019	0.022
		Business size	0.029	0.026	0.030	0.032	0.028	0.025	0.029	0.030
		Activity	0.038	0.039	0.039	0.040	0.036	0.037	0.036	0.039
VAT	Retail	VAT ($t-12$)	0.152	0.129	0.114	0.074	0.117	0.105	0.090	0.061
		Wages (t)	0.109	0.075	0.076	0.045	0.051	0.020	0.036	0.018
		Activity	0.074	0.043	0.053	0.026	0.088	0.033	0.061	0.020
	Manufacturing	VAT ($t-12$)	0.225	0.132	0.159	0.063	0.207	0.084	0.148	0.039
		Wages (t)	0.081	0.173	0.051	0.094	0.038	0.139	0.027	0.078
		Activity	0.057	0.052	0.040	0.032	0.028	0.024	0.016	0.021

($t - 12$) for Retail trade only. For Manufacturing, the same is true with the exception of June and December, where wages (t) is strongest. We added more detail to the evaluation, in line with the proposed guidelines in Subsection 2.4, for variable VAT ($t - 12$) in STS Retail trade, VAT Retail trade and VAT Manufacturing and for variable type of economic activity in STS Manufacturing. For reasons of brevity, we here omit the detailed analysis of variable wages (t) in VAT Manufacturing, but refer the reader to Ouwehand and Schouten (2011).

Table 13 shows that the lack of availability of VAT ($t - 12$) has a negative impact on representativeness in all cases. It also shows that the impact does not decrease after conditioning on the other variables. When VAT of the previous year is not available, then in most cases it concerns newcomers, that is, businesses that launched at some point during the year under consideration. It is not surprising that these businesses are bad responders as they are still starting up and may not have all reporting procedures in order. For VAT, the smaller businesses in terms of revenue also perform worse. This effect was anticipated as small businesses report VAT annually. The values for VAT are smoothed when they are conditioned on wages (t) and type of economic activity; part of the impact of revenue is compensated for by these variables. Surprisingly, for STS Retail trade there is little difference between businesses given that they were active one year ago; the values over the different wage categories are almost constant.

Figure 6 plots the category-level partial R-indicators for type of economic activity in STS Manufacturing. As expected, the unconditional and conditional values are almost identical in an absolute sense: The variable has an orthogonal impact on the other variables. It must be noted here that one group of businesses stands out negatively: the businesses that manufacture food products (NACE 15 and 16). The businesses that manufacture chemicals and chemical products (NACE 23 and 24) perform best.

In sum, with respect to VAT, small businesses and newcomers deserve more attention, while for STS Retail trade it is the newcomers that should be targeted in the data collection. Finally, for STS Manufacturing more effort is needed for specific NACE categories.

Table 13. Categorical partial R-indicators for VAT (t − 12) in STS Retail trade, VAT Retail trade and VAT Manufacturing for January

	STS retail		VAT retail		VAT manufacturing	
	Pu	Pc	Pu	Pc	Pu	Pc
<2.5k	0.012	0.009	−0.045	0.034	−0.079	0.073
2.5k–10k	0.011	0.010	−0.046	0.034	−0.071	0.079
10k–20k	0.013	0.010	−0.008	0.015	−0.007	0.018
20k–30k	0.016	0.013	0.020	0.019	0.030	0.027
30k–50k	0.005	0.003	0.044	0.031	0.059	0.053
50k–100k	0.012	0.010	0.070	0.047	0.092	0.085
100k–200k	0.005	0.006	0.064	0.044	0.089	0.080
>200k	−0.005	0.005	0.081	0.062	0.126	0.108
Not available	−0.041	0.038	−0.034	0.039	−0.047	0.040

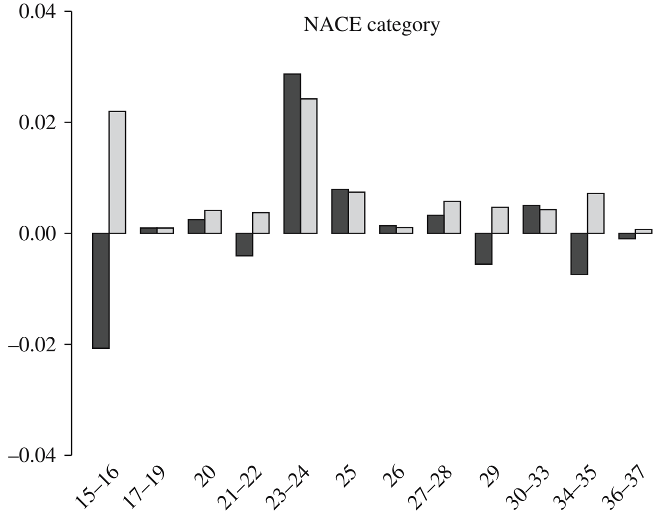

Fig. 6. Categorical partial R-indicators for type of economic activity in STS Manufacturing for January. Black columns represent unconditional and grey columns conditional values. A description of the categories can be found in Appendix B.

4.4. How to Combine the STS Survey and VAT into a Dual Frame Approach?

An important next step is a change of design to obtain higher unit response rates for the underrepresented groups. In this case, three dual-frame approaches can be adopted: VAT-based statistics, STS-based statistics and a combination of STS and VAT. The first and second approaches assume that VAT or STS, respectively, is the primary input to statistics and the other source is used only to supplement types of businesses that are strongly underrepresented. The third approach is a hybrid design, in which both sources are treated as equal. This approach is pragmatic and uses the source per type of business that performs best. For all approaches, however, the explicit targeting of data collection to business units needs to take into account costs and the response burden of data collection. Therefore, representativeness should be optimized subject to constraints on costs and the number of requests for revenue data. Such designs are termed adaptive survey designs (Wagner 2008; Schouten et al. 2013). These designs have begun to emerge in social surveys and may also be applicable to business data collection.

There are three complications to a dual-frame approach that need mentioning first. The first complication is formed by the population frames of the STS survey and the VAT register. Although they are essentially based on the same underlying frame that is maintained by the Dutch Chamber of Commerce, the frames used by Statistics Netherlands and the Tax Authorities are different. This difference applies mostly to larger businesses, where the two offices use different criteria to cluster economic activity. These criteria are logical from their respective operations and perspectives, but a nuisance to any method that combines the two frames. For smaller businesses there is a one-to-one correspondence for virtually all business units, but for larger business units there could be n to 1, 1 to m or even n to m correspondences. As a result, linkage of the two frames cannot be performed without dividing or combining business units. Clearly, when a dual frame approach is applied, complex decision rules are needed for the larger businesses.

The second complication is a conceptual difference in the STS variables themselves. The definition of total revenue and its components is not fully harmonised across the two sources, again for the same operational reasons. This difference is more severe again for the larger businesses. The third complication lies in the classification of businesses. The survey and VAT population frames have different sets of additional, auxiliary variables, as mentioned above. These variables are used to classify businesses. Since there is no one-to-one correspondence between the two frames, transformation rules need to be applied in order to link auxiliary variables from one frame to the other. In summary, it can be concluded that any dual frame approach will need to find methodological solutions for the larger businesses.

We first look at STS Retail trade. These businesses are mostly smaller and both frames have a strong correspondence. Here, it is anticipated that the above-mentioned complications play only a minor role. In Subsection 4.3, we concluded that the smallest businesses are underrepresented in the VAT and that both STS survey and VAT have an underrepresentation of newcomers. Hence, for newcomers no approach will be satisfactory and there is no suitable hybrid approach. The only solution is the development of special invitation letters and instructions and guidance to raise response rates of newcomers in the STS survey. For the small businesses, the STS survey can be conducted to supplement or replace VAT. In STS-based statistics, there is no reason to employ VAT. In VAT-based statistics, the STS survey can be conducted to supplement response for small businesses and, if successful nonresponse reduction methods can be developed, also for newcomers. A hybrid approach would employ STS for small businesses and newcomers and VAT for all other businesses.

For STS Manufacturing, the picture is very different as it consists of larger businesses. Here, frame differences and conceptual differences may complicate a dual-frame approach. The conceptual differences imply that the three approaches are likely to cause method effects. We concluded in Subsection 4.3 that specific NACE categories have a lower representation in the STS survey, while for VAT no specific types of businesses are underrepresented. Hence, VAT-based statistics and a hybrid approach do not employ STS survey data and coincide, but STS-based statistics may employ VAT for these NACE categories. Because of the method effects, STS-based statistics should use a stable design in order to maintain comparability in time.

When adopting a dual frame approach, the focus is on design. Even when more effort is made to raise the response rates of underrepresented businesses, it is likely that some businesses will have lower response rates than others. Apart from a change of design, one may therefore in addition use the VAT records of the previous reporting period to adjust for nonresponse in either the VAT or the STS survey of the current reporting period. Such adjustment is termed nowcasting in economic studies. In nowcasting, the frame differences again pose problems but conceptual differences are not an issue; VAT of the previous reporting period is merely used as a predictor.

5. Discussion

This article compared the unit response rate, quantity response rate and representativity of the STS survey and VAT data over several months and during data collection. Both data sources can be used to produce monthly short-term business statistics. However, Statistics Netherlands intends to replace part of its survey efforts with data from administrative registers. To this end, the available data should, of course, lead to accurate statistics. An important data quality aspect that is assumed to be a good predictor of accuracy is the representativity of the data. In this article, we therefore compared the two data sources with respect to representativity, as measured by the R-indicator.

In our comparison, we focused completely on nonresponse error and ignored measurement and sampling errors. Clearly, the STS survey response has a bigger sampling error than the VAT data as the Tax Authorities records are a full enumeration of enterprises in the Netherlands. Measurement errors were conjectured to play an important role as well. However, there is little empirical evidence in favour of survey or administrative data. A complete comparison of both data sources should also account for these errors.

In our comparison, we answered three research questions. They regard the representativeness of survey and registry data per industry, per month, and through time, but also regard the enterprise groups that need to be targeted to improve representativeness of response. The main question underlying these investigations is the more general issue of whether STS statistics should be based on a dual-frame approach using both register and survey data.

The representativeness of survey data and register data is quite different over the months. The results indicate that the unit response rate for both Retail trade and Manufacturing is substantially lower for VAT than for STS, due to the nature of the collection method. However, the R-indicator for VAT can still be relatively high even in months of low response rates. This shows that the unit response rate alone is not sufficient for assessing data quality.

During data collection, and more specifically between 25 and 30 days after the end of the reference month, the unit response rates increased, as could be expected. Representativity, however, is not in line with the unit response rate patterns: It may change only slightly as the unit response rate increases, or it may even decline. From this we conclude that the contrast between reporting and nonreporting units may increase as data reporting proceeds. Hence, waiting longer before producing statistics based on VAT data does not make the data more representative.

The findings for the R-indicators are in line with the combined patterns of unit response rates and quantity response rates. Whenever quantity response rates showed a different increase from unit response rates, the R-indicator also changed. The strong feature of the R-indicator is that it quantifies over- and underrepresentation, it allows for a simultaneous assessment on multiple auxiliary variables and it can be estimated without bias after and during data collection. The quantity response rates in this article were computed retrospectively, but could normally not be estimated during or shortly after data collection without bias.

In summary, the main difference between representativeness of response to surveys and to register holders is the stability over time and during data collection. The survey data are more stable in time and during data collection.

Representativity patterns may differ from subpopulation to subpopulation. We found that in VAT small businesses and newcomers deserve more attention, while for STS Retail trade it is the newcomers that should be targeted in the data collection. Finally, for STS Manufacturing more effort is needed for specific NACE categories.

Future research is required. Our study had some limitations with respect to the data set used. It used a specific set of auxiliary variables, only focused on a period of two years and on two industries. The auxiliary variables were not the same for the two data sources (caused by the difference in population frames), so the absolute values of the R-indicator could not be compared. It is important, therefore, that our results are replicated on other years and industries and in other countries.

Appendix A: Approximations to the Expected Quantity Response Rate

We restrict ourselves to a first-order Taylor approximation of (7) and (10). For the expected value of a ratio of two random variables, this leads to the ratio of the expected values of the two random variables. This is a crude approximation, but we merely want to show how the various indicators relate to each other for large sample sizes. The STS survey sample sizes are indeed large and VAT is a full enumeration of the population.

Assuming that the population is large and $(N-1)/N \approx 1$, for the numerator and denominator of (7), respectively, we arrive at

$$E\sum_{i=1}^{n} d_i r_i y_i = \sum_{i=1}^{N} \rho_i y_i = N\operatorname{cov}(y, \rho_Y) + N\rho_U \bar{y}_N, \tag{A.1}$$

$$E\sum_{i=1}^{n} d_i y_i = \sum_{i=1}^{N} y_i = N\bar{y}_N. \tag{A.2}$$

For the denominator of (10), we first rewrite as

$$\sum_{h=1}^{H} N_h \frac{\sum_{i\in h} d_i r_i y_i}{\sum_{i\in h} d_i r_i} = \sum_{h=1}^{H} N_h \frac{\sum_{i\in h} d_i r_i (y_i - \bar{y}_h)}{\sum_{i\in h} d_i r_i} + \sum_{h=1}^{H} N_h \bar{y}_h, \tag{A.3}$$

and define residual $\varepsilon_i = y_i - \bar{y}_h$ for unit i. The expectation of a weighted stratum response mean can be approximated (again using a first-order Taylor expansion) by

$$E\frac{\sum_{i\in h} d_i r_i \varepsilon_i}{\sum_{i\in h} d_i r_i} = \frac{\operatorname{cov}_h(\varepsilon, \rho_Y)}{\rho_{U,h}}, \tag{A.4}$$

since the stratum residual means $\bar{\varepsilon}_h$ are equal to zero. In (A.4) $\rho_{U,h}$ is the unit response rate in stratum h and $\operatorname{cov}_h(y, \rho_Y)$ is the covariance between response propensities and residuals within stratum h.

Using (A.3) and (A.4) the expectation of the denominator of (10) is approximated as

$$E\left(\sum_{h=1}^{H} N_h \frac{\sum_{i\in h} d_i r_i \varepsilon_i}{\sum_{i\in h} d_i r_i} + \sum_{h=1}^{H} N_h \bar{y}_h\right) = \sum_{h=1}^{H} N_h \frac{\operatorname{cov}_h(\varepsilon, \rho_Y)}{\rho_{U,h}} + N\bar{y}_N. \tag{A.5}$$

Appendix B: NACE categories

15–16 : manufacture of food products
17–19 : manufacture of apparel, leather, leather products, and footwear
20 : manufacture of wood, and wood and cork products, except furniture
21–22 : manufacture of paper and paper products, and printing and reproduction of recorded media
23–24 : manufacture of chemicals and chemical products
25 : manufacture of rubber and plastic products
26 : manufacture of other nonmetallic mineral products
27–28 : manufacture of basic metals and manufacture of fabricated metal products, except machinery and equipment
29 : manufacture of machinery and equipment
30–33 : manufacture of computers, electronic and optical products
34–35 : manufacture of transport equipment
36–37 : manufacture of furniture

6. References

Kruskal, W. and F. Mosteller. 1979a. "Representative Sampling I: Non-Scientific Literature." *International Statistical Review* 47: 13–24. DOI: http://dx.doi.org/10.2307/1403202.

Kruskal, W. and F. Mosteller. 1979b. "Representative Sampling II: Scientific Literature Excluding Statistics." *International Statistical Review* 47: 111–123. DOI: http://dx.doi.org/10.2307/1402564.

Kruskal, W. and F. Mosteller. 1979c. "Representative Sampling III: Current Statistical Literature." *International Statistical Review* 47: 245–265. DOI: http://dx.doi.org/10.2307/1402647.

Little, R.J.A. and D.B. Rubin. 2002. *Statistical Analysis with Missing Data, Wiley Series in Probability and Statistics*. New York: John Wiley & Sons.

Ouwehand, P. and B. Schouten. 2011. Representativity of VAT and Survey Data for Short Term Business Statistics. CBS discussion paper 201106. Available at: http://www.cbs.nl/nl-NL/menu/methoden/onderzoek-methoden/discussionpapers/archief (accessed date).

Schouten, B., F. Cobben, and J. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35: 101–113.

Schouten, B., N. Shlomo, and C. Skinner. 2011. "Indicators for Monitoring and Improving Survey Response." *Journal of Official Statistics* 27: 231–253.

Schouten, B., J. Bethlehem, K. Beulens, Ø. Kleven, G. Loosveldt, K. Rutar, N. Shlomo, and C. Skinner. 2012. "Evaluating, Comparing, Monitoring and Improving Representativeness of Survey Response Through R-indicators and Partial R-indicators." *International Statistical Review* 80: 382–399. DOI: http://dx.doi.org/10.1111/j.1751-5823.2012.00189.x.

Schouten, B., M. Calinescu, and A. Luiten. 2013. "Optimizing Quality of Response Through Adaptive Survey Designs." *Survey Methodology* 39: 29–58.

Shlomo, N., C. Skinner, and B. Schouten. 2012. "Estimation of an Indicator of the Representativeness of Survey Response." *Journal of Statistical Planning and Inference* 142: 201–211. DOI: http://dx.doi.org/10.1016/j.jspi.2011.07.008.

Slootbeek, M. and K. van Bemmel. 2010. *Onderzoek Naar Overstappers/Blijvers*, Internal report, Statistics Netherlands.

Thompson, K.J. and B.E. Oliver. 2012. "Response Rates in Business Surveys: Going Beyond the Usual Performance Measure." *Journal of Official Statistics* 28: 221–237.

Van Delden, A. and F. W. L. Aelen. 2008. Redesigning the Chain of Economic Statistics at Statistics Netherlands: STS Statistics as an Example. IAOS Conference 'Reshaping Official Statistics', 14–16 October 2008, Shanghai. Available at: http://www.stats.gov.cn/english/specialtopics/IAOSconference/200811/w020130912531695358471.doc (accessed September 30, 2014).

Vlag, P. and D. van den Bergen. 2010. The Use of VAT for Short Term Statistics: Some Quality Aspects, Statistics Netherlands. In Proceedings ESSnet seminar on administrative data – Rome, 18–19 March 2010. Available at: http://essnet.admindata.eu/Document/Getfile?objectid=5147 (accessed September 30, 2014).

Wagner, J. 2008. *Adaptive Survey Design to Reduce Nonresponse Bias*. Ann Arbor, MI: University of Michigan, Ph.D. thesis.

Received November 2012
Revised April 2014
Accepted June 2014