EUROPEAN CENTRAL BANK

EUROSYSTEM

Statistics Paper Series

Katarzyna Bańkowska
Małgorzata Osiewicz
Sébastien Pérez-Duarte

Measuring non-response bias in a cross-country enterprise survey

No 12 / December 2015

Contents

Abstract 2

Non-technical summary 3

1 Non-response bias and its measurement 5

2 Probability sampling based on quotas in the SAFE 7

3 Non-response in the Survey on Access to Finance of Enterprises (SAFE) 10

4 R-indicators as a measure of representativity 13

5 R-indicators for the SAFE survey 15

5.1 R-indicators across survey rounds (8th to 10th) 15

5.2 R-indicators during the SAFE fieldwork 19

6 R-indicators for SAFE data matched with the Amadeus database 20

7 Conclusions and outlook 22

References 23

Annex 25

Abstract

Non-response is a common issue affecting the vast majority of surveys, and low non-response is usually associated with higher quality. However, efforts to convince unwilling respondents to participate in a survey might not necessarily result in a better picture of the target population. It can lead to higher, rather than lower, non-response bias, for example if incentives are effective only for particular groups, e.g. in a business survey, if the incentives tend to attract mainly larger companies or enterprises encountering financial difficulties.

We investigate the impact of non-response in the European Commission and European Central Bank Survey on the Access to Finance of Enterprises (SAFE), which collects evidence on the financing conditions faced by European small and medium-sized enterprises compared with those of large firms. This survey, which has been conducted by telephone biannually since 2009 by the European Central Bank and the European Commission, provides a valuable means of searching for this kind of bias, given the high heterogeneity of response propensities across countries.

The study relies on so-called “Representativity Indicators” developed within the Representativity Indicators of Survey Quality (RISQ) project, which measure the distance to a fully representative response. On this basis, we examine the quality of the SAFE at different stages of the fieldwork as well as across different survey waves and countries. The RISQ methodology relies on rich sampling frame information, although this is partially limited in the case of the SAFE. To enrich this frame information, we also assess the representativity of a SAFE subsample created by linking the survey responses with the companies’ financial information as obtained from a business register. This subsampling is another potential source of bias which we also attempt to quantify. Finally, we suggest possible ways of improving the monitoring of the non-response bias in future rounds of the survey.

Keywords: business survey, representativity, bias, non-response, R-indicators

JEL codes: C81 (Methodology for Collecting, Estimating, and Organizing Microeconomic Data • Data Access), C83 (Survey Methods • Sampling Methods), D22 (Firm Behavior: Empirical Analysis)

Non-technical summary

Non-response bias occurs when the survey results obtained for the respondents differ from the results that would have been obtained from those who did not answer the survey, even though they were in the target sample. The reasons for the non-response might be diverse and can include: i) the information in the register used to draw up the sample, such as the telephone number or address, is not up-to-date and the respondents cannot be reached; ii) the target person is not available in the survey time frame; or iii) the person explicitly refuses to participate in the study. Low non-response rates have often been used as a proxy for higher survey quality, and survey practitioners have investigated methods of increasing response rates. However, somewhat paradoxically, efforts to convince those unwilling to participate in a survey to do so might not necessarily result in a better picture of the target population and can lead to a higher, rather than lower, non-response bias. For instance, monetary incentives may increase the response rate in a household survey but could result in a higher proportion of poorer respondents; if income correlates with the topic of the survey, this could lead to biased estimates. In business surveys, a pre-announcement letter containing information about the study can serve as an incentive to participate. However, if the survey pre-notification mostly attracts those who are interested in the topic or those more knowledgeable about the subject (in the case of the SAFE, this would be more financially sophisticated companies), such respondents will not be representative of the overall population of firms (Groves and Peytcheva, 2008).

The Survey on Access to Finance of Enterprises (SAFE) is a qualitative telephone survey conducted to provide regular information on the financing conditions of micro, small and medium-sized enterprises (SMEs) in the European Union. A sample of large firms (250 employees or more) is also included in order to be able to compare developments for SMEs with those for large firms. A subset of the survey is run by the ECB every six months to assess the latest developments in the financing conditions of firms in euro area countries. A more comprehensive version of the survey with an extended questionnaire is conducted in cooperation with the European Commission. Initially, it was run every two years but since 2013 it has been conducted yearly. The survey is run by an external survey company using the Dun & Bradstreet business register as a sampling frame. The sample is a probability sample based on quotas by country and size. The SAFE also has a rotating opt-in panel component – at the end of the interview the respondents are asked whether they would like to participate in future survey rounds. Even though around 80% of firms agree, only a portion of these are subsequently re-contacted successfully. As a result, the panel currently consists of around 50% of the respondents.

The motivation for this paper stems from the relatively low response rate for the SAFE, which stood at about 14% in the last rounds of the survey. The biggest differences are between panel and non-panel enterprises – while the response rate among those contacted for the first time is below 10%, it can be over 40% for panellists. There are also significant differences across countries,[1] with response rates varying from 5% in Spain in the 10th wave to over 20% in Ireland, Greece and Portugal.

Given the restricted length of the telephone interviews and the difficulties encountered by respondents in answering questions related to quantitative accounting elements, the survey data are matched with the quantitative financial information from the Bureau van Dijk's Amadeus database in order to obtain balance sheet information for the interviewed firms. The portion of the SAFE dataset with successfully matched records constitutes the Amadeus subsample.

The objective of this study is to examine the representativity of the SAFE sample and the subsample containing the matched financial information using so-called "Representativity Indicators" developed within the Representativity Indicators of Survey Quality (RISQ) project. R-indicators are based on the standard deviation of response propensities for various sub-populations. They attempt to capture the overall impact of non-response for the whole survey, unlike comparisons with external sources, which only capture representativity at the level of a particular estimate.

Our study is pioneering in three respects: first in terms of its wide geographical scope as it covers several euro area countries in a harmonised manner; second, while most of the academic statistical research refers to probability sampling, to the best of our knowledge, this is the first time that R-indicators have been applied to probability sampling based on quotas; and third, we investigate not only the possible bias of the standard survey sample, but also the representativity of the SAFE subsample linked to balance sheet information from the business register – an area where formal academic research is still lacking.

Our findings shed light on the use of survey results in the presence of non-response – a crucial issue for the SAFE which is frequently used in policy-relevant studies. First, the computed R-indicators show that the level of representativity of the SAFE is comparable to that of other surveys. For the SAFE sample, we find that the different response patterns across countries make the largest contribution to the loss in representativity, while for the Amadeus subsample size class also plays a role due to the evident under-representation of micro firms. We also see a positive impact from fieldwork length - the potential maximal bias decreases steadily with each additional week that the survey is conducted. Finally, we propose further improvements to the survey data collection to enhance the monitoring of the potential bias.

[1] The country-by-country variation can stem partially from differences in the coding of non-response in the local offices. Please see Section 3 for more details.

1 Non-response bias and its measurement

Non-response bias occurs when the survey estimates for the respondents differ from the results that would have been obtained from those who did not answer the survey. While initially, non-response was treated as a fixed characteristic of a respondent, the more current popular stochastic approach assumes that people have a certain probability ρ_i of participating, which varies depending on circumstances. In this sense, the bias of the respondents' means $\bar{y}_r$ is approximated by $\sigma_{y\rho} / \bar{\rho}$, where $\sigma_{y\rho}$ is the population covariance between the survey variable, y, and the response propensity, ρ, and $\bar{\rho}$ is the mean propensity in the target population over sample realisations (Groves, 2006).

However, the relationship between the response propensities and the non-response biases is not straightforward and higher response rates do not necessarily lead to a lower bias if increased efforts to convert non-respondents are only effective for particular groups, e.g. in a business survey, if they attract mainly larger companies or enterprises encountering financial difficulties. Groves (2006) presents the absolute relative bias together with the corresponding response rate for over 200 estimates from 30 different methodological studies and shows a weak correlation between the two. Interestingly, most of the variation comes from the estimates within the same survey.

Depending on the available information, various approaches are applied to analyse the non-response (Montaquila and Olson, 2012). First, follow-up surveys, aimed at collecting information from the initial non-respondents, are a possible way of investigating their characteristics and how they differ from the respondents. Such studies usually apply enhanced recruitment techniques, different survey modes and shorter questionnaires targeting the main variables. Apart from the drawbacks of the extra cost and the extended fieldwork, achieving a high response rate in the follow-up survey is essential, which might prove a difficult objective.[2]

Second, the survey estimates can be compared with external sources such as administrative records. In this case, highly accurate benchmarks and consistent measurement of analysed indicators between both datasets are prerequisites for a meaningful evaluation.

A third set of methods compares the survey estimates under alternative weighting schemes using additional characteristics associated with the key survey estimates or response propensities. On the one hand, sensitivity of the results to different

[2] Additional data collection can also take the form of randomised non-response experiments, where different design features (e.g. "warm-up" questions, mode) are assigned to different random subsamples. The aim is to identify the most effective design by comparing the response rates of the treatment groups. However, it might be challenging to find a single treatment which performs well in reducing the non-response bias for the full sample and ,not just for a particular group (Kruskal and Mosteller, 1979).

weightings would indicate the presence of a non-response bias. On the other, an absence of or insignificant differences might indicate a lack of good predictors rather than an absence of bias.

A fourth approach relies on the information from the sampling frame and observations collected during the fieldwork for the whole sample. These data are used to calculate different statistics (e.g. sample means, proportions) for various subgroups, such as: i) respondents and non-respondents; or ii) those who could not be contacted, those who refused and those who participated. Additionally, for longitudinal studies, past information on initial respondents who became non-respondents in subsequent rounds helps to detect response patterns and possible causes of attrition (National Research Council, 2013). Furthermore, the auxiliary sample information enables the calculation of response rates by characteristics. Within the respondent set, the survey estimates can be presented separately for cooperative and for more reluctant respondents, distinguishing between both groups using variables such as the number of call attempts, early versus late responses, incentives provided and techniques used for refusal conversion. A significant variation between specific subgroups would point to a potential bias and its source. R-indicators, which are the focus of this paper, also fall into this set of methods for non-response analysis.

In this paper, we apply this last approach to the SAFE sample information, focusing primarily on the R-indicators developed as part of the Representativity Indicators of Survey Quality (RISQ) project.[3] This paper first gives the rationale for applying the R-indicators to the probability sample based on quotas. It then gives an overview of the non-response in the SAFE. In the subsequent sections, we briefly present the methodology used for various types of R-indicators and describe the implementation of the indicators in the SAFE and in the matched SAFE–Amadeus dataset. In the final section, we present our conclusions and provide recommendations for fieldwork monitoring.

[3] http://www.risq-project.eu/

2 Probability sampling based on quotas in the SAFE

In this section, we describe the SAFE sample which is the outcome of a random selection process involving probabilistic sampling based on quotas.

A probability sample applies various random selection methods which ensure that all units in the population have a known, non-zero probability of being chosen. Unlike non-probabilistic sampling methods, it gives a single theoretical framework for making statistical inferences about the whole population and for that reason is the preferred method, particularly among statistical agencies (American Association for Public Opinion Research, 2013).

The term “quota sample” covers different techniques and usually carries a negative connotation among survey statisticians. Indeed, when improperly done, data collected through this type of sampling offer no guarantee of representativity and do not allow any form of probabilistic analysis. However, the SAFE sample is very different from the quota samples of the 1950s where interviewers had to choose a convenience sample which respected set quotas. The SAFE sample follows the work of Sudman (1966) conferring probabilistic properties on quota sampling.

We describe the selection of the sample of first-time participants in the survey; panel firms are not considered here.[4] The sample is drawn from the Dun & Bradstreet business register, which has the advantage of adequately, if not perfectly, covering the universe of enterprises in the euro area. A stratified random sample is drawn from Dun & Bradstreet, where the strata consist of country (11 in the euro area surveys) and size class (four such classes). In line with other cold-call business surveys, response rates are unfortunately low. Consequently, the initial sample is 10 to 15 times larger than the desired final sample, to account for non-response.

As in other surveys working with firm data in a multinational setting, we assume that the Dun & Bradstreet population is a good reflection of the actual population of firms. The total number of firms in the target population is known from Eurostat’s Structural Business Statistics, by country, sector, and size class. If, for each country, sector and size class, firms have the same probability of being included in Dun & Bradstreet, then firms not in that register can be considered to be missing at random (MAR in Donald Rubin’s terminology). Hence, the initial sampling probability can be estimated for all firms in the population and thus in the initial sample.

The interviews are based on this initial sample, with targets or quotas for the number of interviews conducted by country and size class (as above). The initial sample is randomly sorted and the firms are dialled from this sample. Up to ten call attempts are made to each telephone number at different times or even outside normal office

[4] For the description of the panel selection, see Section 1.

hours. Call-back appointments are not subject to the limit of ten attempts. Under this interviewing strategy, a certain number of firms will not have been called at all ("fresh" sample), some firms will have been called and not contacted ("non-contact"), others will have been contacted but refused to participate ("refusal") and others will have been successfully interviewed ("respondents"). At the end of the fieldwork, some firms will still be fresh and will be so at random (conditional on the quota cell).

In order to analyse response behaviour and response rates across countries, the fresh firms are dropped from the initial sample. The initial number of records drawn from the register is decided by the survey company based on past response rates in the SAFE and similar studies. Usually, a sample ten times larger than the targeted number of interviews is sufficient. However, in countries with lower quality contact information (e.g. incorrect telephone numbers, out-of-date records) or lower-than-expected cooperation rates, the initial sample will need to be topped-up with additional fresh records. Thus, the size of the unused sample is not comparable across countries as the ratio of the initial sample to targeted interviews varies. However, even if this ratio were the same in each quota, the size of the fresh sample is an arbitrary decision and should not be taken into account in the analysis of the response indicators.[5] Consequently, the unused records are removed and only the records where at least one contact attempt was made are taken into account in the analysis.

During fieldwork, however, the way the fresh sample is integrated into the calling roster is crucial for the probabilistic nature of the sample. A firm in the fresh sample should not be called simply because the interviewer feels he/she is more likely to secure an interview in this way rather than by re-contacting a firm that had previously been called unsuccessfully. If fresh firms are contacted in this way, then the quota sampling is not less probabilistic than a probability sample where non-response causes randomness in the firms that are interviewed. Of course, since the survey has a tight deadline and priority is given to the timeliness of the results, towards the end of the fieldwork it is more likely that insufficient attempts will be made to contact the firms in the calling roster. We study this phenomenon in Section 5.2 below, where we consider the representativity of the sample according to the length of the fieldwork.

The final estimation weights for the survey are then obtained by calibrating the results according to official figures for the country, size class and sector breakdown of the enterprise population (four main sector groupings), and thus correcting differences in the response rates (as long as the non-response can indeed be considered conditionally random by country, sector and size class).

One interesting theoretical aspect that would need to be explored further in connection with the R-indicators is the randomness of the effective initial sample (excluding the fresh firms) and the fixed number of firms in the final respondent sample, which is the converse of the standard probabilistic setup of fixed initial

5 To illustrate it, we can consider two initial samples: one ten times larger and another one hundred times larger than the number of targeted interviews. Computed response rates would be very different for these two scenarios, although the response behaviour is the same.

sample but random final one. We consider this issue to be of a secondary nature in the measurement of the representativity of the final sample, and will hence take the effective initial sample as the true initial sample and the final sample as the result of the interviewing process of all the firms in the initial sample.

3 Non-response in the Survey on Access to Finance of Enterprises (SAFE)

A common problem for nearly all types of survey is low response rates, which have in fact dropped substantially over the past decades (see, for example, National Research Council, 2013, pp. 12-30)). A low response rate is also a concern for the SAFE. The overall response rate was around 14% in the last survey rounds,[6] which is lower than those of other business surveys run by central banks. While these other surveys are not directly comparable given the differences in how they are conducted, in absolute terms the response rates for the SAFE can nevertheless be objectively deemed low.

In this paper we focus on non-response at the level of the unit (respondent), which should be distinguished from item non-response. In the SAFE, most questions offer two silent answers “Don’t know” (DK) and “Not applicable” (NA). Overall, less than 2% of the answers to the qualitative question are missing as DK, which is sufficiently low to consider that item non-response is not a serious concern. With respect to the quantitative information, the frequency of DK is similar (e.g. annual turnover, share of exports in annual turnover); this is in part because the SAFE allows responses to be provided in brackets.[7]

Given that non-response at the respondent level, and not item non-response, is the main source of uncertainty over the quality of the results, in this paper we apply R-indicators to analyse unit non-response bias and its causes from several possible angles.

In the first step, we present various outcome rates for the SAFE, i.e. response, contact and cooperation rates, by main enterprise characteristics: country of residence, size and sector. In addition, we split firms into those participating in the survey for the first time (non-panel firms) and those which took part in at least in one of the earlier survey rounds (panel). These results will be cross-checked later with the findings obtained from the R-indicators. We focus on three survey rounds (8th to 10th) as detailed information on the full sample including non-respondents was not available in the earlier rounds. When computing response and cooperation rates, break-off interviews are treated as non-responses as they are excluded from the survey respondents. In the case of unknown eligibility, the proportion of cases of

[6] Response rate 3, following the definition of outcome rates advocated by AAPOR (see American Association for Public Opinion Research 2011). Since the original AAPOR definitions refer to household surveys, they were adapted to the features of a business survey.

[7] The only exception is a new question, introduced on an experimental basis, on the interest rates charged for the credit line or bank overdraft, where the respondents are asked to give the exact value. In this case around 30% of respondents do not provide an answer. This is understandable given that the survey is conducted mostly over the telephone and the respondents do not have an opportunity to check such detailed information. This is also the reason why the SAFE questionnaire contains mostly qualitative questions which are much easier to answer by telephone.

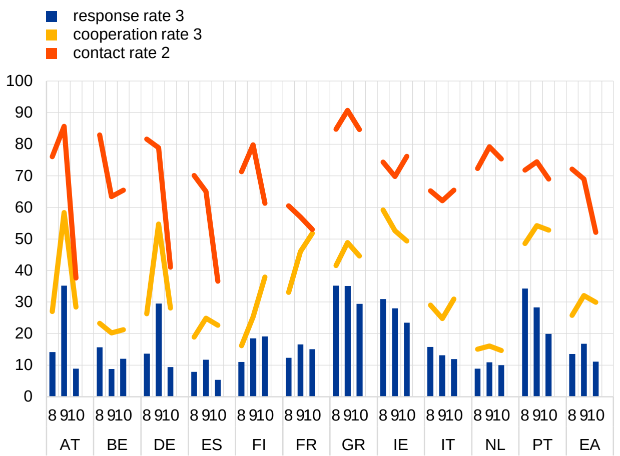
unknown eligibility that are eligible is estimated[8] and increased from 60% in the 8th survey round to 80% in the 10th round, which is rather conservative, since the higher this proportion, the lower the response rate.

While contact, cooperation and response rates vary considerably across countries, neither the companies' sector nor their size class have a large impact on the response rates (small firms have a slightly higher propensity to participate, while construction firms have a lower one; see Chart 1). The largest divergence is between panel and non-panel enterprises, with a relatively high response rate of 40% for panellist in the 8th and 9th survey rounds. This can stem from various factors: i) a positive image of the survey acquired through previous participation; ii) up-to-date contact information for those companies (reflected in a higher contact rate than for non-panel firms, see Chart 2); or iii) a higher propensity to participate.

Chart 1

Response, contact and cooperation rates for the SAFE

(from 8th to 10th survey round by country)

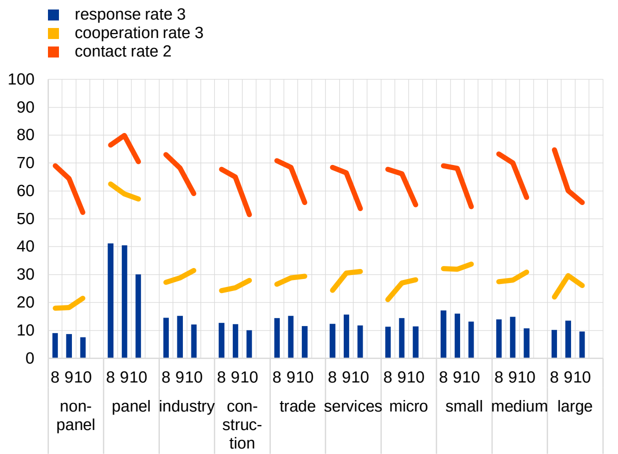


Sources: ECB, SAFE.
Notes: Residency of a firm is indicated by country ISO-3166 code (AT – Austria, BE – Belgium, DE – Germany, ES- Spain, FI – Finland, FR – France, GR – Greece, IE- Ireland, IT – Italy, NL – the Netherlands, PT- Portugal). EA stands for aggregated figure for all presented euro area countries combined.

Chart 2

Response, contact and cooperation rates for the SAFE

(from 8th to 10th survey round by panel dummy, sector and size; excluding Austria and Germany)



Sources: ECB, SAFE.

Differences in response patterns across countries can stem from many factors. First, cultural differences play a role. In some countries, the respondents strongly refuse to participate, asking to be excluded from any future surveys conducted by the survey company, while in other countries, where the refusals are softer, good interviewers can more easily convince initial non-respondents to eventually take part in the study.

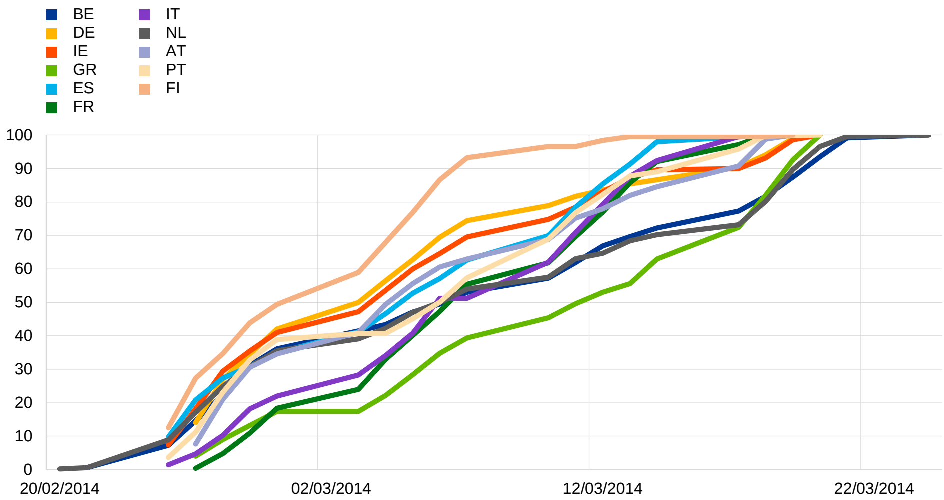
[8] The full definitions of outcome rates applied in this paper are as follows (see also footnote 6):
- response rate 3: I/((I+P) + (R+NC+O)+ e*U),
- cooperation rate 3: I/(I+P+R),
- refusal rate 2: R/((I+P)+(R+NC+O) + e*U),
- contact rate 2: ((I+P)+R+O) / ((I+P)+R+O+NC+ e*U),
- e: (I+P+R+NC+O)/(I+P+R+NC+O+NE),

where I – Interview, P – Partial interview, R – Refusal, NC – Non-contact, O – Other contact (non-refusals), U – Unknown if firm, NE – Non-eligible, e- the estimated proportion of cases of unknown eligibility that are eligible.

Second, the quality of the sampling frame differs across countries. Low quality company contact details can result in unsuccessful telephone calls (e.g. if the number is wrong); similarly, poor information on the number of employees or sector of activity can lead to a respondent being excluded after the screener questions (in the case of the SAFE, if the firm is non-profit, has no employees other than the owner or belongs to a sector which is out of the scope of the SAFE). As a result, a lower quality frame increases the cost of conducting the survey in that particular country and may require longer fieldwork owing to a high non-contact rate. This could have a negative impact on the comparability of the results between firms interviewed at the beginning and the end of the fieldwork period. However, this is not the case for the SAFE, given the relatively short fieldwork period and the generally stable progress achieved in carrying out successful interviews, in all countries and in all analysed survey rounds (see Chart 3).

Chart 3

Fieldwork progress by country in the 10th survey round



Sources: ECB, SAFE.

Third, the situation in the local offices of the survey company, such as the experience and training of the interviewers, and their workload at the time of conducting the survey can also have an impact on the response rate. In the case of the SAFE, one additional factor which can explain the divergences is the different CATI (Computer-Assisted Telephone Interviewing) system used by the survey company in Germany and Austria and the fact that the outcome codes are not fully harmonised with offices in other locations. For this reason, we excluded these two countries from the subsequent analysis.

4 R-indicators as a measure of representativity

The concept of “representativity” or “representativeness” does not have a single clear interpretation. Kruskal and Mosteller (1979) review the statistical and other scientific literature and divide the meaning of the term “representative” into no less than nine different groups, varying from “general acclaim for data”, through “miniature of the population” to “representative sampling as permitting good estimation”. Alongside these vague uses of the term, they also briefly present some mathematical methods of comparing two probability distributions and give examples of distance functions measuring the extent of “disagreement”. Representativity indicators (R-indicators) would fall into this category.

R-indicators are based on a definition linked to the mechanism of Missing Completely at Random (MCAR) and individual response propensities. Following Schouten, Bethlehem et al. (2012), p. 384, “response is called representative with respect to [the vector of auxiliary variables] *X* when the response propensities of all subpopulations formed by the auxiliary variables are constant and equal to the overall response rate”, in other words “when the respondents form a random subsample of the survey sample”. In this sense, R-indicators attempt to capture the overall impact of non-response for the whole survey and not only at the level of a particular estimate.

Although it is not the objective of this paper to describe in detail the theoretical properties of R-indicators, which is much better done in Shlomo and Schouten (2013) or in Schouten, Cobben and Bethlehem (2009), we provide a brief definition and a description of their main features.

R-indicators are based on the standard deviation of response propensities transformed to lie between 0 and 1, where 1 is a fully representative response: $R = 1 - 2S(\rho)$. The response propensities and then the variance of the response propensities are estimated, leading to the following estimator of R:

$$\hat{R} = 1 - 2\hat{S}(\hat{\rho}) = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i=1}^{n} d_i(\hat{\rho}_i - \hat{\bar{\rho}}\,)^2}\ ,$$

where d_i are the design weights, $\hat{\bar{\rho}} = \frac{1}{N}\sum_{i=1}^{n} d_i\hat{\rho}_i$ is the weighted sample mean of the estimated response propensities and *N* is the size of the population (see Shlomo & Schouten (2013), p. 4).

It can be shown that the lower bound of the R-indicator (see Schouten, Cobben and Bethlehem, 2009, p.104) depends on the response rate: $R \geq 1 - 2\sqrt{\bar{\rho}(1-\bar{\rho})}$. In particular, it reaches its minimum level of 0 for a response rate of 50%, i.e. when individual response propensities can have the largest variation, while it increases when the response rate decreases from 50% to 0%.

The decomposition of the variance $S^2(\rho)$ into between- and within-components of the response propensities for the sample subgroups is the foundation of the partial R-indicators at variable level. The unconditional partial R-indicator corresponds to the between-subgroup variance, while the within-subgroup variances are the basis for the conditional partial indicators (Schouten, Bethlehem et al. (2012)). These indicators can be further decomposed into category-level R-indicators showing the contributions to the variation of the respective categories (de Heij, Schouten and Shlomo, 2010).

Table 1

Variance decomposition and partial R-indicators at variable and category level

	Unconditional	Conditional
$S^2(\rho)$ =	$S^2_{between}(\rho)$	$S^2_{within}(\rho)$
Variable level	$P_U(X_k) = \sqrt{\frac{1}{N}\sum_{h=1}^{H} n_h(\bar{\rho}_h - \bar{\rho})^2}$	$P_C(X_k) = \sqrt{\frac{1}{N}\sum_{l=1}^{L}\sum_{i \in U_l} d_i(\rho_i - \bar{\rho})^2}$
Category level	$P_U(X_k, h) = \sqrt{\frac{n_h}{N}}(\bar{\rho}_h - \bar{\rho})$	$P_C(X_k, h) = \sqrt{\frac{1}{N}\sum_{l=1}^{L}\sum_{i \in U_l} d_i\Delta_{h,i}(\rho_i - \bar{\rho}_l)^2}$
Notation	X_k is a categorical variable with H categories and it is a component of the vector X. $n_h = \sum_{i=1}^{n} d_i\Delta_{h,i}$ is the weighted sample size in the category h, where $\Delta_{h,i}$ is a 0-1 dummy variable for sample unit i being a member of stratum h U_l is a cell in the cross-classification of all model variables except X_k	

Standardised maximal absolute bias (in short "maximal bias") in the worst case scenario, i.e. if the non-response correlates maximally with the variable of interest, is $B\_m\ (X) = (1 - R(\rho))/(2\bar{\rho}) \leq 1 - \bar{\rho}$ and it can be shown that it cannot be larger than the non-response rate (see Schouten, Morren et al., 2009).

5 R-indicators for the SAFE survey

For the computation of R-indicators and associated statistics, we used the SAS codes available on the RISQ project[9] website (see also de Heij, Schouten and Shlomo, 2010) for the methods of bias adjustment and computation of confidence intervals for the R-indicators).

The main requirement for computing the R-indicators is the availability of the auxiliary information from the sampling frame. The microdata for the whole SAFE sample have only been provided from the 7th survey round onwards- although they have not yet been fully harmonised - and contain detailed outcome codes for each telephone call (interview, refusal, answering machine, etc.), size class and sector from the Dun & Bradstreet business register, and a dummy for panel firms (only from the 8th survey round onwards). We also have the date of the last contact attempt or actual contact which, in the case of respondents, is the time of the interview.

Although the methods for estimating representativity were not designed for quota samples, in keeping the description of the probability sample based on quotas used in the SAFE, we will neglect this issue in this paper and assume that the respondents were obtained through a simple random sample. We will consider that every firm for which a contact was attempted (the "non-fresh" sample) is to be included in the sample as a non-respondent. Since the objective of this paper is to assess the influence of firm characteristics on response behaviour, we do not use the R-indicators for stratified samples as this would mask the impact of stratification variables (country and size). However, for comparison purposes we computed the R-indicators for stratified samples.[10] As expected, the overall R-indicator improves; however, the effect of the remaining variables (sector and panel) is similar to the results presented without stratification.

All R-indicators were computed using the four abovementioned variables, i.e. country (nine euro area countries), size class (micro, small, medium and large), sector (industry, construction, trade and services) and panel dummy. The response propensities were estimated by a logistic regression with all mentioned variables as predictors, without interactions.

5.1 R-indicators across survey rounds (8th to 10th)

We start the examination of the R-indicators for each survey round by looking at the overall response and contact rates. The response process could be split further into successive sub-processes of contact, cooperation and final response, as in Schouten, Bethlehem, et al. (2012). However, as we are unsure of the extent to

[9] http://www.risq-project.eu/tools.html. We thank Natalie Shlomo for providing additional SAS codes for stratified simple random samples and useful suggestions.

[10] Available upon request.

which outcome codes have been harmonised between countries, we limit this initial analysis to the two processes mentioned above.

Table 2

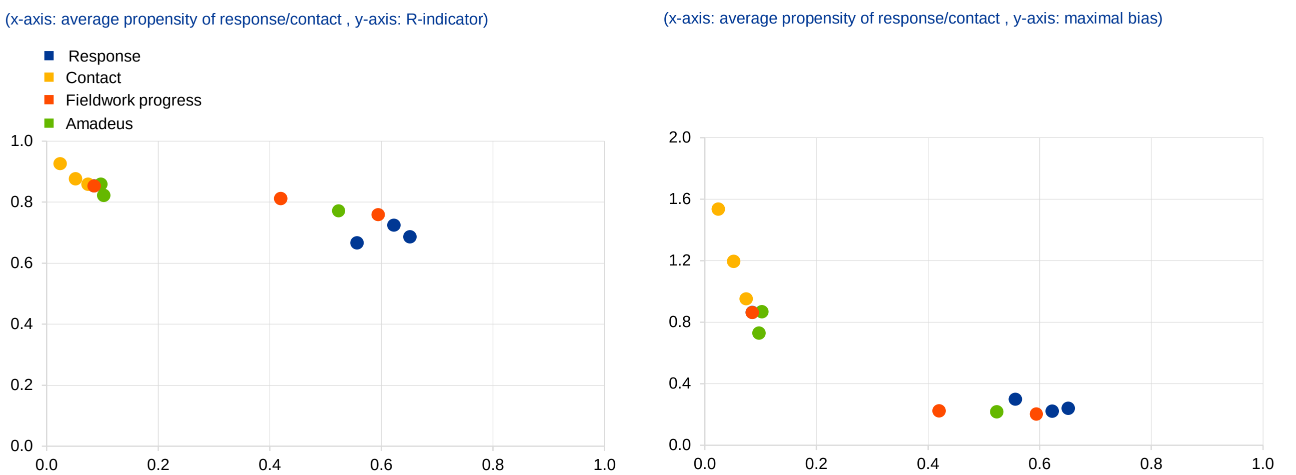
R-indicators and other associated information for the survey rounds 8 to 10.

Round	8	9	10	8	9	10
	Response			Contact		
Total sample	70432	58689	62090	70432	58689	62090
Response rate 3/contact 2	13.4%	15.0%	11.6%	70.0%	67.3%	55.5%
R-indicator	0.853	0.822	0.859	0.725	0.686	0.666
Standard error	0.003	0.004	0.003	0.003	0.003	0.003
Avg propensity	0.085	0.102	0.097	0.622	0.651	0.556
Maximal bias	0.863	0.868	0.729	0.221	0.241	0.300
Lower bound for R	0.441	0.394	0.408	0.030	0.047	0.006

Interestingly, the R-indicator for overall response is lowest for the 9th survey round, even though the highest response rate was achieved in that round (see Table 2). In particular, it was also the round where the longer questionnaire was used. We cannot draw conclusions from this single observation; however, it would be useful to monitor the future evolution of the non-response bias in the rounds with the extended questionnaire.

Chart 4

R-indicators and maximal bias as a function of the average propensity of response/contact for the 8th to 10th survey rounds

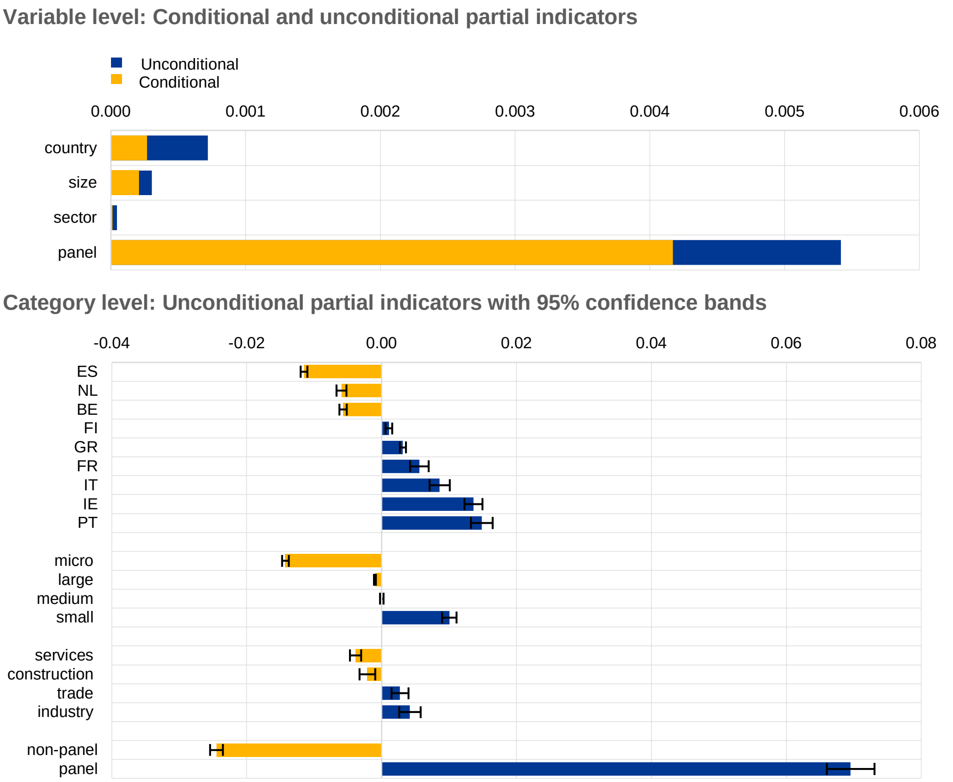


A higher response rate does not guarantee greater representativity. For instance, the R-indicator for the overall response is highest and the maximal bias lowest in round 10, even though the response rate was higher for the previous round (see Table 2 and 4). It is also useful to look at the maximal bias in the analysis of overall representativity, especially as it is not sensitive to the response rate. Chart 4 shows that R-indicators are higher for the overall response process than for the contact process, but the maximal bias is much lower for contact. It would appear that other sub-processes of the overall response behaviour (such as the cooperation of the respondents) may play a bigger role and contribute to the potential loss of representativity.

In the case of the R-indicator corresponding to contact propensities, the 10th survey round scores the worst. This was already visible from the outcome rates, as the contact rate dropped dramatically between the 9th and 10th rounds, particularly in three countries: Austria, Germany (both excluded from the analysis) and Spain.[11] In this case, a low contact rate is also associated with a higher bias – the large negative unconditional values for the R-indicator point to the underrepresentation of Spanish businesses in the pool of contacted enterprises, while the Netherlands and Italy, with high positive unconditional values, are, by comparison, overrepresented (see Chart 5 and 6).

Chart 5

Partial indicators for response in the 8th survey round



More generally, with respect to contact, the unconditional and conditional partial R-indicators are highest for the country variable, and country variation contributes most to the loss of representativity in all examined survey rounds. It seems that enterprises in some countries are more difficult to contact than in others, which also

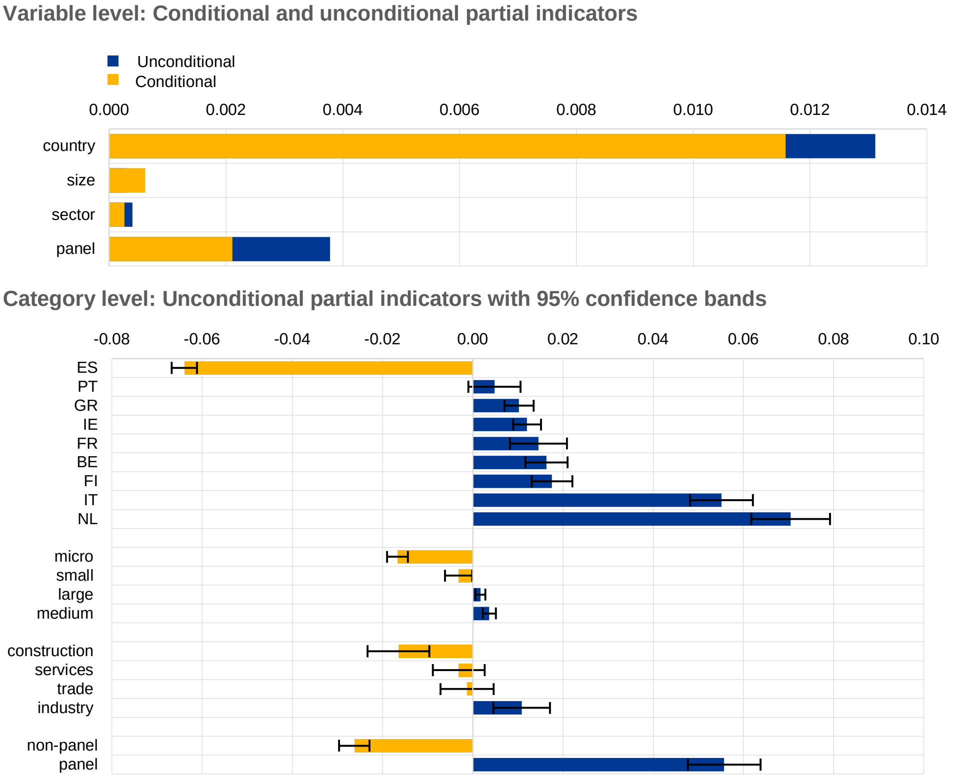
[11] The disproportionately large drop in the non-contact rate in the 10th wave was a result of approaching more enterprises at the beginning of the fieldwork. Enterprises that were not contacted successfully were not re-approached since the quotas were already filled. In other countries, such companies would be re-contacted and possibly converted into respondents.

points to issues with the quality of the sampling frame. For the SAFE, enterprises are all sampled from Dun & Bradstreet; however, the availability and accuracy of the contact information is not homogenous as the underlying sources of information differ from country to country. Consequently, it would be advisable to increase efforts to improve the sampling frame.

With regard to the overall response, unsurprisingly, whether or not the enterprise belongs to the panel plays the biggest role, while the company's characteristics, such as country, size and sector are not statistically significant at the variable level (see Table 5 and Chart 5). This is consistent with the earlier finding regarding much higher response propensities for panel firms. It also backs up the finding that firm characteristics, as available in the registers, do not play a role in response patterns. This is confirmed when the R-indicators are calculated separately for the firms which participated for the first time in the survey – in this case also, the unconditional and conditional indicators at the variable level are not statistically different from zero.[12]

Chart 6

Partial indicators for contact in the 10th survey round



[12] These results are not presented in the paper but are available upon request.

5.2 R-indicators during the SAFE fieldwork

R-indicators can be used as a tool for monitoring representativity during the data collection, and can be computed for different amounts of effort, e.g. number of attempts or level of interviewer's experience. In the SAFE, however, this information on fieldwork is limited; therefore we analyse the development of the R-indicators as the fieldwork progresses.

The SAFE is usually conducted over a period of one month; however, the start and end of the fieldwork can vary slightly from country to country. To account for these differences, we divide the fieldwork into four periods based on the quartiles of the total number of fieldwork days, calculated separately for each country. The results for the 8th survey round are presented in Table 3.

Table 3

R-indicators for the response and other associated information for each quartile of the fieldwork (8th survey round)

	Up to 1st quartile	Up to 2nd quartile	Up to 3rd quartile	Full fieldwork
Total sample	70432	70432	70432	70432
R-indicator	0.926	0.877	0.859	0.853
Standard error	0.003	0.004	0.003	0.003
Average response propensity	0.024	0.052	0.074	0.085
Maximal bias	1.535	1.195	0.953	0.863
Lower bound for R	0.694	0.558	0.476	0.441

For the first fieldwork quartile, which corresponds to approximately the first week of data collection, the representativity is highest, with the R-indicator reaching a level of 0.93. It drops slightly in the second quartile to 0.88 and remains broadly stable until the end of the fieldwork. In this case, splitting the sample into enterprises which are part of the panel and those participating for the first time plays the biggest role, as indicated by the increase in the partial R-indicator as the fieldwork progresses (see Table 6). However, a positive impact from each additional week of fieldwork is also visible when looking at the maximal bias – it decreases steadily from a maximum of 1.54 standard deviation of a survey estimate of interest in the first part of the fieldwork to 0.86 at the end of the fieldwork (see also Chart 4).

6 R-indicators for SAFE data matched with the Amadeus database

In this section, we first describe briefly the methodology used to match the SAFE dataset with that of Bureau van Dijk's Amadeus database, and then comment on the quality of the matching. Second, we analyse the R-indicators for this dataset containing both qualitative and quantitative firm-level information, taking into account variations in the availability of financial information among respondents.

To link the companies from the SAFE and Amadeus databases, information on tax identification number, company name, street, postcode, city and country are used. In the 8th round, 86% of SAFE respondents[13] were successfully matched with the Amadeus business register. The quality of matching varies substantially between countries, with success rates of over 90% in Belgium, Spain, France and the Netherlands, and the lowest rate obtained for Greece at 67%. There is also a significant difference between size classes, with large companies successfully matched in 98% of cases, and micro firms only in 72% of cases. The difference at sector level is much less pronounced (see also Bańkowska, Osiewicz and Pérez-Duarte, 2014 for more information on matching results).

Being in Amadeus is not enough as some records may have missing financial information. For this reason, we separately examine the representativity of three different SAFE-Amadeus subsamples in the 8th survey round: one containing respondents with available information on loans, one containing those with information on value added and one containing those with information on turnover (in short, the "Amadeus subsample").

Table 4

R-indicators and other associated information for respondents with available details on loans, value added and turnover (8th survey round)

	Loans	Value added	Turnover
Total sample	6008	6008	6008
R-indicator*	0.759	0.812	0.772
Standard error	0.003	0.002	0.003
Average propensity	0.594	0.420	0.523
Maximal bias	0.203	0.225	0.218
Lower bound for R	0.018	0.013	0.001

* Owing to the smaller sample size the R-indicator adjusted for bias is used as in de Heij, Schouten and Shlomo (2010).

The R-indicators were computed using the same auxiliary variables as in the earlier analysis (i.e. country, size, sector and panel dummy). The R-indicator for value added amounts to 0.81, while for loans and turnover it is slightly lower at 0.76 and 0.77 respectively (see Table 4). In all three cases, the lack of representativity, measured by both partial conditional and unconditional R-indicators, stems from the

13 As in the previous section, Austria and Germany were excluded from the analysis.

country variable, which is similar to the results obtained for the contact process (see Section 5.1). However, given the smaller sample size, the unconditional partial indicator is only statistically significant at the 0.1 level only for value added (for turnover, p-value equals 0.12 for the country variable and 0.11 for the size variable; see Table 7). The estimated negative values for the category-level partial indicators suggest that firms in the Netherlands and, to a lesser extent, in Greece are underrepresented in the set of companies with available financial information. In the case of value added and turnover, this also applies to Belgium and Ireland. On the other hand, France and Spain are strongly overrepresented with respect to all three variables considered.

As in the analysis of the contact process for the whole SAFE sample, size class also contributes to the loss of representativity in the dataset matched with quantitative financial variables.[14] As expected, micro companies, for which financial information is scarce, are also strongly underrepresented in the matched SAFE subsample. These findings are also reflected in the overall matching rates at enterprise level, as mentioned above.

It is also worth noting that in the 8th survey round the maximal bias for the whole sample of SAFE respondents is higher than for the subsample of respondents with financial information (0.86 for the SAFE sample compared with 0.23 for value added in the Amadeus subsample). However, it should be borne in mind that this is an additional potential bias since the matched SAFE-Amadeus dataset is already a subsample of the SAFE respondents.

[14] The levels of the partial indicators for the size variable are comparable to the partial R-indicators for the contact process. However, given the smaller sample size, they turn out to be statistically not significant at the 0.1 level (p-value is 0.15 for value added and 0.11 for turnover).

7 Conclusions and outlook

In this paper we present R-indicators for the SAFE and show that the level of representativity is comparable to that of other surveys (see Schouten, Bethlehem, et al. (2012)). As with other surveys, the panel firms participating in the SAFE have much higher response propensities. However, our findings also show that the firm characteristics available in business registers, such as size and sector, do not play a role in determining response patterns. Further investigations into other potential sources of bias stemming from the panel are left for future studies.

In addition, we find that, for the SAFE sample, country variation in contact propensities makes the largest contribution to the loss in representativity, whereas, for the Amadeus subsample, size class also plays a role, with the clear underrepresentation of micro firms.

Based on these findings, we make the following recommendations: i) increase efforts to enhance the quality of sample contact information; ii) fully harmonise the use of outcome codes across countries and interviewers; and iii) collect more detailed information on the fieldwork which is useful for monitoring data collection, i.e. outcome codes for each attempt and possibly interviewers' performance and experience.

Since September 2014 (corresponding to the 11th survey round), a new survey company has been in charge of the SAFE fieldwork. Given that this new supplier conducts interviews from one central call centre, as opposed to having local agencies in each region, we can disentangle the differences in response patterns across countries from differences in the organisation of local offices. At the same time, the option of completing the questionnaire online was introduced and it will be important to investigate and monitor the effects of different survey modes on representativity.

This paper could be extended in three directions. First, the representativity of the sample frame can be assessed with respect to official statistics on the population of enterprises. Second, the sensitivity of the survey results can be tested using different weighting schemes. Finally, as mentioned above, the analysis presented in this paper can be extended using newly available information on the fieldwork, and by splitting the response process into several sub-processes (such as contact, cooperation and response) to identify the main causes of potential non-response bias.

References

The American Association for Public Opinion Research (2011), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 7th edition, AAPOR.

The American Association for Public Opinion Research (2013), *Report of the AAPOR Task Force on Non-Probability Sampling*, AAPOR, June.

Bańkowska, K., Osiewicz, M. and Pérez-Duarte, S. (2014), "Linking qualitative survey responses with quantitative data. Methodology, quality and data analysis from the matching of the ECB/EC Survey on Access to Finance of Enterprises and the Amadeus database", available at URL http://www.bis.org/ifc/events/7ifcconf_bankowska.pdf

Groves, R. M. (2006), "Nonresponse Rates and Nonresponse Bias in Household Surveys", *Public Opinion Quarterly*, Vol. 70, pp. 646-675.

Groves, R. M. and Peytcheva, E. (2008), "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis", *Public Opinion Quarterly*, Vol. 72, pp. 167-189.

Heij, V. de, Schouten, B. and Shlomo, N. (2010), "RISQ manual, Tools in SAS and R for the computation of R-indicators and partial R-indicators", URL http://www.risq-project.eu/publications.html

Kruskal, W. and Mosteller, F. (1979), "Representative sampling, III: The current statistical literature", *International Statistical Review*, Vol. 47, pp. 245-265.

Montaquila, J. M. and Olson, K. M. (2012), "Practical Tools for Nonresponse Bias Studies", available at URL http://www.amstat.org/sections/srms/webinarfiles/NRBiasWebinarApril2012.pdf.

National Research Council (2013), *Nonresponse in Social Science Surveys: A Research Agenda*, The National Academies Press, Washington DC.

Schouten, B., Bethlehem, J., Beullens. K., Kleven, O., Loosveldt, G., Luiten, A., Rutar, K., Shlomo, N. and Skinner, C. (2012), "Evaluating, Comparing, Monitoring and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators", *International Statistical Review*, Vol. 80, pp. 382-399.

Schouten, B., Cobben, F. and Bethlehem, J. (2009), "Indicators for the Representativeness of Survey Response", *Survey Methodology*, Vol. 35, pp. 101-113.

Schouten, B., Morren, M., Bethlehem, J., Shlomo, N. and Skinner, C. (2009), "How to use R-indicators?", available at URL http://www.risq-project.eu/publications.html.

Shlomo, N. and Schouten, B. (2013), *Theoretical Properties of Partial Indicators for Representative Response*, Technical Report, University of Southampton.

Sudman, S. (1966), “Probability Sampling with Quotas”, *Journal of the American Statistical Association*, Vol. 61, No. 315 (September), pp. 749-771.

Annex

Table 5

Unconditional and conditional partial R-indicators for contact and response in the 8th to 10th survey rounds.

	Unconditional						Conditional					
Round	**8**	**9**	**10**	**8**	**9**	**10**	**8**	**9**	**10**	**8**	**9**	**10**
Variable level		**response**			**contact**			**response**			**contact**	
country	0.001	0.001	0.001	0.003***	0.005***	0.013***	0.000	0.000	0.001	0.005***	0.005***	0.012***
size	0.000	0.000	0.000	0.002***	0.002**	0.000	0.000	0.000	0.000	0.003***	0.001	0.001
sector	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
panel	0.005***	0.007***	0.004***	0.001	0.002***	0.004***	0.004***	0.006***	0.004***	0.001	0.002	0.002*
Category level												
BE	-0.006***	-0.010***	-0.002***	0.031***	-0.023***	0.016***	0.005***	0.008***	0.003***	0.037***	0.018***	0.017***
ES	-0.011***	-0.009***	-0.016***	-0.037***	-0.032***	-0.064***	0.007***	0.007***	0.014***	0.040***	0.033***	0.067***
FI	0.001***	0.006***	0.010***	0.011***	0.023***	0.018***	0.002***	0.004***	0.011***	0.013***	0.025***	0.018***
FR	0.006***	0.001	0.007***	-0.012***	0.026***	0.015***	0.007***	0.004***	0.007***	0.023***	0.029***	0.008**
GR	0.003***	0.009***	0.012***	0.021***	0.014***	0.010***	0.001**	0.007***	0.012***	0.021***	0.013***	0.010***
IE	0.014***	0.004***	0.008***	-0.004***	0.000	0.012***	0.006***	0.001	0.003***	0.007***	0.002***	0.008***
IT	0.009***	-0.002***	0.002***	0.006*	-0.026***	0.055***	0.004***	0.003***	0.008***	0.014***	0.023***	0.041***
NL	-0.006***	-0.007***	-0.003***	-0.002	0.040***	0.070***	0.006***	0.008***	0.004***	0.004***	0.043***	0.069***
PT	0.015***	0.022***	0.011***	0.014***	0.012***	0.005	0.008***	0.011***	0.008***	0.012***	0.008***	0.003**
micro	-0.014***	-0.006***	-0.007***	-0.044***	-0.039***	-0.017***	0.012***	0.002***	0.004***	0.050***	0.029***	0.024***
small	0.010***	0.004***	0.004***	0.000	0.000	-0.003**	0.007***	0.003	0.003	0.010***	0.005**	0.005
medium	0.000	-0.001***	0.000	0.009***	0.008***	0.004***	0.000	0.000	0.000	0.010**	0.008	0.005
large	-0.001***	0.000	-0.001***	0.001***	-0.002***	0.002***	0.001	0.001	0.001	0.003	0.002	0.002
industry	0.004***	0.002***	0.003***	0.012***	0.008**	0.011***	0.002	0.003*	0.003	0.004	0.007***	0.007***
construction	-0.002***	-0.008***	-0.006***	-0.014***	-0.018***	-0.016***	0.001	0.006***	0.004***	0.009***	0.014***	0.014***
trade	0.003***	0.001	0.000	0.008***	0.001	-0.001	0.002***	0.001**	0.001	0.008***	0.004***	0.004***
services	-0.004***	0.001	0.000	-0.008***	0.000	-0.003	0.002	0.001	0.001	0.003**	0.001	0.001
non-panel	-0.024***	-0.037***	-0.028***	-0.010***	-0.021***	-0.026***	0.027***	0.035***	0.026***	0.010***	0.018***	0.019***
panel	0.070***	0.077***	0.060***	0.028***	0.043***	0.056***	0.059***	0.073***	0.056***	0.024***	0.041***	0.042***

Note: *** indicates significance at the 0.01 level, ** indicates significance at the 0.05 level and * indicates significance at the 0.1 level.

Table 6

Unconditional and conditional partial R-indicators for response during fieldwork progress in the 8th round.

	Unconditional				Conditional			
Variable level	**Up to 1st quartile**	**Up to 2nd quartile**	**Up to 3rd quartile**	**Full fieldwork**	**Up to 1st quartile**	**Up to 2nd quartile**	**Up to 3rd quartile**	**Full fieldwork**
country	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
size	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sector	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
panel	0.001	0.004**	0.005***	0.005***	0.001	0.003***	0.004***	0.004***
Category level								
BE	-0.002***	-0.003***	-0.005***	-0.006***	0.002***	0.003***	0.005***	0.005***
ES	-0.003***	-0.006***	-0.010***	-0.011***	0.001	0.003***	0.006***	0.007***
FI	-0.001***	0.000**	0.001***	0.001***	0.001*	0.000	0.002***	0.002***
FR	0.001***	0.000	0.003***	0.006***	0.003***	0.001*	0.004***	0.007***
GR	0.001***	0.002***	0.003***	0.003***	0.001**	0.001	0.001*	0.001**
IE	0.004***	0.011***	0.014***	0.014***	0.000	0.006***	0.007***	0.006***
IT	0.000	0.006***	0.007***	0.009***	0.005***	0.002	0.003***	0.004***
NL	-0.002***	-0.005***	-0.006***	-0.006***	0.002**	0.006***	0.007***	0.006***
PT	0.010***	0.012***	0.014***	0.015***	0.009***	0.007***	0.007***	0.008***
micro	-0.002***	-0.007***	-0.012***	-0.014***	0.001	0.005***	0.010***	0.012***
small	0.003***	0.007***	0.010***	0.010***	0.003	0.005**	0.008***	0.007***
medium	0.000***	-0.001***	-0.001***	0.000	0.001	0.001	0.001	0.000
large	0.000***	-0.001***	-0.001***	-0.001***	0.001	0.000	0.001	0.001
industry	0.001**	0.003***	0.004***	0.004***	0.001	0.002	0.002	0.002
construction	0.000	-0.001**	-0.002***	-0.002***	0.001**	0.000	0.001	0.001
trade	0.001**	0.002***	0.003***	0.003***	0.001**	0.001	0.001*	0.002***
services	-0.001***	-0.003***	-0.003***	-0.004***	0.000	0.001	0.001	0.002
non-panel	-0.011***	-0.020***	-0.024***	-0.024***	0.014***	0.023***	0.026***	0.027***
panel	0.032***	0.058***	0.067***	0.070***	0.031***	0.051***	0.057***	0.059***

Note: *** indicates significance at the 0.01 level, ** indicates significance at the 0.05 level and * indicates significance at the 0.1 level.

Table 7

Unconditional and conditional partial R-indicators for SAFE respondents matched with Amadeus database (8th survey round).

	Unconditional			Conditional		
Variable level	**Loans**	**Value added**	**Turnover**	**Loans**	**Value added**	**Turnover**
country	0.002	0.004*	0.003	0.001	0.002	0.002*
size	0.001	0.002	0.002	0.001	0.002	0.002
sector	0.000	0.001	0.001	0.000	0.000	0.000
panel	0.000	0.000	0.000	0.000	0.000	0.000
Category level						
BE	0.012***	-0.015***	-0.017***	0.015***	0.010***	0.014***
ES	0.007***	0.018***	0.011***	0.006**	0.016***	0.009***
FI	0.004***	0.000	0.007***	0.004**	0.001	0.007***
FR	0.006**	0.016***	0.029***	0.004***	0.011***	0.028***
GR	-0.007***	-0.013***	-0.005***	0.006	0.009**	0.004
IE	-0.002	-0.017***	-0.022***	0.002***	0.011***	0.016***
IT	0.008**	0.031***	0.020***	0.004***	0.021***	0.012***
NL	-0.038***	-0.031***	-0.033***	0.031***	0.020***	0.025***
PT	0.004*	0.019***	0.011***	0.004***	0.019***	0.012***
micro	-0.036***	-0.044***	-0.045***	0.034***	0.046***	0.042***
small	0.005***	0.002	0.003	0.007***	0.008***	0.007***
medium	0.003***	0.006***	0.006***	0.003	0.004*	0.004
large	0.002***	0.003***	0.003***	0.002	0.002	0.002
industry	0.010***	0.015***	0.015***	0.004***	0.004***	0.006***
construction	0.002	0.002	0.001	0.002**	0.003***	0.001
trade	-0.015***	-0.020***	-0.018***	0.007***	0.006***	0.007***
services	-0.001	-0.002	-0.002	0.002**	0.001*	0.002**
non-panel	-0.002	-0.001	0.002	0.001	0.000	0.001**
panel	0.003	0.001	-0.003	0.001*	0.000	0.001***

Note: *** indicates significance at the 0.01 level, ** indicates significance at the 0.05 level and * indicates significance at the 0.1 level.

Acknowledgements
We thank Natalie Shlomo for providing additional SAS codes for stratified simple random samples and useful suggestions.

Note
An earlier version of this paper was published in the *Austrian Journal of Statistics*, Vol 44, No 2 (2015).

Katarzyna Bańkowska
European Central Bank; e-mail: Katarzyna.Bankowska@ecb.europa.eu

Małgorzata Osiewicz
European Central Bank; e-mail: Malgorzata.Osiewicz@ecb.europa.eu

Sébastien Pérez-Duarte
European Central Bank; e-mail: Sebastien.Perez_Duarte@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

ISSN 2314-9248 (online)
ISBN 978-92-899-1908-1
DOI 10.2866/936989
EU catalogue No QB-BF-15-004-EN-N