Austrian Journal of Statistics
*April 2015, Volume 44, 13–30.*
*http://www.ajs.or.at/*
doi:10.17713/ajs.v44i2.60

# Measuring Nonresponse Bias in a Cross-Country Enterprise Survey

**Katarzyna Bańkowska, Małgorzata Osiewicz*, Sébastien Pérez-Duarte**
European Central Bank, Frankfurt am Main

**Abstract**

Nonresponse is a common issue affecting the vast majority of surveys. Efforts to convince those unwilling to participate in a survey might not necessary result in a better picture of the target population and can lead to higher, not lower, nonresponse bias.

We investigate the impact of nonresponse in the European Commission & European Central Bank Survey on the Access to Finance of Enterprises (SAFE), which collects evidence on the financing conditions faced by European SMEs compared with those of large firms. This survey, conducted by telephone bi-annually since 2009 by the ECB and the European Commission, provides a valuable means to search 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, which is however partly limited in the case of the SAFE. We also assess the representativeness of the SAFE particular subsample created by linking the survey responses with the companies' financial information from a business register; this sub-sampling is another potential source of bias which we also attempt to quantify. Finally, we suggest possible ways how to improve monitoring of the possible nonresponse bias in the future rounds of the survey.

*Keywords*: business survey, representativeness, bias, nonresponse, R-indicators.

## 1. Nonresponse bias and its measurement

Nonresponse bias occurs when the survey estimates for the respondents are different from those who did not answer to the survey. While initially the nonresponse was treated as a fixed characteristic of a respondent, the more currently popular stochastic approach assumes that people have a certain probability $\rho_i$ of participating, which varies depending on circumstances. In this sense, the bias of the respondents' mean $\bar{y}_r$ is approximated by $\frac{\sigma_{y\rho}}{\bar{\rho}}$, where $\sigma_{y\rho}$ is the population covariance between the survey variable, $y$, and the response propensity, $\rho$, and $\bar{\rho}$ is the mean propensity in the target population over sample realisations (Groves 2006).

*Corresponding author: malgorzata.osiewicz@ecb.europa.eu

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

Dependent on the available information, various approaches are applied to analyse the nonresponse (Montaquila & Olson 2012). First, the survey estimates can be compared to the external sources, like administrative records. In this case, highly accurate benchmark and consistent measurement of analysed indicators between both datasets are prerequisite to the meaningful evaluation.

A second set of methods compares the survey estimates under alternative weighting schemes using additional characteristics associated with the key survey estimates or response propensities. Sensitivity of the results to different weighting would indicate the presence of nonresponse bias. On the other hand, no or insignificant differences might stem rather from lack of good predictors than absence of bias.

A third approach relies on the information from the sampling frame and observations collected during the fieldwork for the whole sample. Such data are the basis for the calculation of different statistics (e.g. sample means, proportions) separately for respondents and nonrespondents or various reasons for nonparticipation (noncontact, refusal). Additionally for longitudinal studies, past information on the initial respondents, who turned nonrespondents in the subsequent rounds, help to detect response patterns and possible causes of attrition (National Research Council 2013). Furthermore, the auxiliary sample information allows computing response rates by characteristics. Within the respondent set, the survey estimates can be presented for cooperative and more reluctant respondents, measured by variables like number of call attempts, early versus late respondents, provided incentives and techniques used for refusal conversion. Large variation between specific subgroups would point to the potential bias and its source. R-indicators, which are the focus of this paper, fall also into this set of methods for nonresponse analysis.

Fourth, follow-up surveys, aimed at collecting information on the initial nonrespondents, are another possibility to investigate how distinct they are from the respondents. Such studies usually apply enhanced recruitment techniques, different survey modes, and shorter questionnaires targeted on the main variables. Apart from the drawbacks of the extra cost and the extended fieldwork, achieving high response rate in the follow-up survey is essential, which might prove a difficult objective[1].

In this paper, we apply the third approach based on the sample information to the Survey on Access to Finance of Enterprises (SAFE), with the main focus on the R-indicators developed within the Representativity Indicators of Survey Quality (RISQ) project[2].

SAFE is a qualitative telephone survey conducted with the purpose of providing regular information on the financing conditions of micro, small and medium-sized enterprises (SMEs). 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 of the financing conditions of firms in the euro area countries. A more comprehensive version of the survey with an extended questionnaire is run every two years, in cooperation with the European Commission. The survey is conducted by an external survey company. The sample is a probability sample based on quo-

[1] Additional data collection can also take the form of randomised nonresponse experiments, where different design features (e.g. "warm-up" questions, mode) are assigned to different random subsamples. The results and the response rates of the treatment groups are then compared and effective design identified, although it might be challenging to find one treatment which performs well in terms of reducing nonresponse bias, not only for a particular group, but for the full sample (Kruskal & Mosteller 1979).

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

tas by country and size. The SAFE has also a rotating opt-in panel component – at the end of the interview the respondents are asked whether they would like to participate in the future survey rounds. Around 80% of firms agree, however, afterwards only a part is successfully re-contacted. As a result, panel constitutes currently around 50% of the respondents.

Given the restricted length of phone interview and respondent's difficulties in answering questions related to quantitative accounting elements, to obtain balance sheet information of the interviewed companies, the survey data are matched with the quantitative financial information from the Bureau van Dijk's Amadeus database.

The objective of this study is to examine the representativity of the SAFE sample, as well as the subsample containing the matched financial information. This paper gives first the rationale for applying the R-indicators to the probability sample based on quotas. Secondly, we present an overview of the nonresponse in SAFE. In the following sections, we describe briefly the methodology of various types of R-indicators and present the implementation of the indicators in SAFE and the matched dataset of SAFE and Amadeus. In final section, we conclude and give the recommendation for fieldwork monitoring.

## 2. Probability sampling based on quotas in the SAFE

A word is warranted on the nature of the sample in the SAFE, as "quota sample" carries a negative connotation among survey statisticians, and indeed, when improperly done, data collected through such a sample offer no guarantee of representativity and do not allow any sort of probabilistic analysis. However, the SAFE sample is very far from the quota samples of the 1950s where interviewers had to choose a convenience sample respecting quotas. The SAFE sample follows the work of Sudman (1966) in order to confer probabilistic properties to quota sampling.

We describe the selection of the sample of first-time participants in the survey; panel firms are not considered here[3]. The sample is drawn from the Dun & Bradstreet company database, which has the benefit of adequately, if not perfectly, covering the universe of enterprises in the euro area. From Dun & Bradstreet, a stratified random sample is drawn, with strata composed of country (11 in the euro area surveys) and size class (4 such classes). In line with other cold-call business surveys, response rates are very low. Consequently, the initial sample is 10 to 15 times larger than the desired sample, to account for nonresponse.

As in other surveys working with firm data in a multinational setting, we assume that the Dun & Bradstreet population is a good image of the 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, conditional on 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 (the same as above). The initial sample is randomly sorted, and the firms are dialled from this sample. Up to ten calling attempts are made to each address, at different times or even outside normal office hours. Call-back appointments are not subject to the limit of ten attempts. From 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 contacted but they refuse to participate ("refusal") and others 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).

[3]For the description of the panel selection, see section 1.

In order to analyse response behaviour and response rates across countries, those "fresh" firms are dropped from the initial sample. The initial number of records drawn from the register is a decision of the survey company based on the 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 some countries with lower quality of the contact information (e.g. incorrect telephone numbers, out-of-date records) or lower than expected cooperation rates, there is a need of topping-up the initial sample with additional fresh records. Thus, the amount of unused sample would not be comparable across countries as the ratio of the initial sample to targeted interviews varies. However, even if this ratio was the same in each quota, the amount of the fresh sample is an arbitrary decision and should not be taken into account in the analysis of the response indicators[4]. 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 only because it is more probable to conclude the interview than trying to contact again the firm, for which the previous contacts were unsuccessful. If this is the case, then the quota sampling is not less probabilistic than a probability sample where nonresponse 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 not enough contact attempts for the firms in the calling roster are carried out. We study this phenomenon in section 5.2 below, when we consider the representativity of the sample through the length of the fieldwork.

The final estimation weight is then obtained by calibrating on official counts by country, size class, and sector (4 main sector groupings), correcting in this manner differential response rates as long as the nonresponse can indeed be considered conditionally random by country, sector and size class.

One interesting theoretical aspect that would need to be further explored 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 sample but random final one. We consider this issue to be of a secondary nature in the measure 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. Nonresponse in the Survey on Access to Finance of Enterprises (SAFE)

A common problem across nearly all types of surveys is low response rates, which in fact have dropped substantially over the last decades (see e.g. National Research Council 2013, p. 12-30). A low response rate is also a concern for the SAFE. The overall response rate reached around 14% in the last survey rounds[5], below 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. As this may be a source of uncertainty about the quality of the results, in this paper we apply R-indicators to analyse from several angles possible nonresponse bias and its origin.

[4] 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 responses rate would be very different for those two scenarios, although the response behaviour is the same.

[5] 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.

In the first step, we present the outcome rates for the SAFE by main characteristics of enterprises: country of residence, size and sector. In addition, we split firms into those which participate for the first time in the survey (non-panel firms) and those which took part at least in one of the earlier survey rounds (panel).. Those results will be later cross-checked with the findings coming from the R-indicators. We focus on the three latest survey rounds ($8^{th}$ to $10^{th}$) as detailed information on the full sample including nonrespondents, was not available in the earlier rounds. When computing response and cooperation rates, break-off interviews are treated as nonresponse. In case of unknown eligibility, the proportion of cases of unknown eligibility that are eligible is estimated[6] and increased from 0.6 in $8^{th}$ survey round to 0.8 in the $10^{th}$ 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 companies' sector nor 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 Figure 1). The largest divergence shows between panel and non-panel enterprises with relatively high response rate of 40% for panellist in $8^{th}$ survey round, either through a positive image of the survey acquired through previous participation or a higher propensity to participate (see Figure 2).

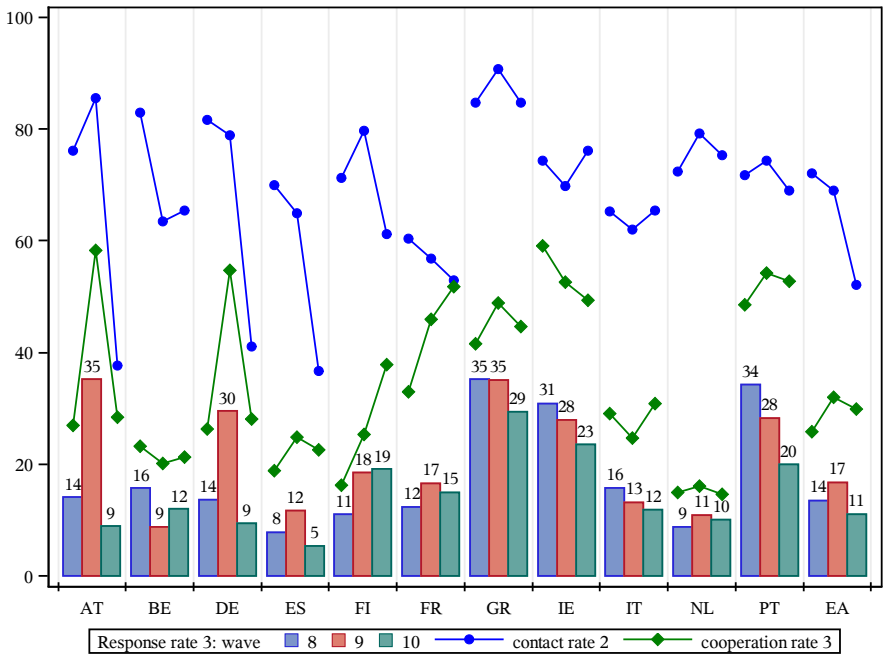



Figure 1: Outcome rates for SAFE from $8^{th}$ to $10^{th}$ survey round by country.
Note: Residency of a firm is indicated by country ISO-3166 code (AT – Austria, BE – Belgium, DE – Germany, ES – Spain, FI – Finland, FR – France, GR – Greece, IE – Ireland, IT – Italy, NL – the Netherlands, PT – Portugal). EA stands for aggregated figure for all presented euro area countries combined.

Country variation 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

[6]Following the definitions:

- 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.

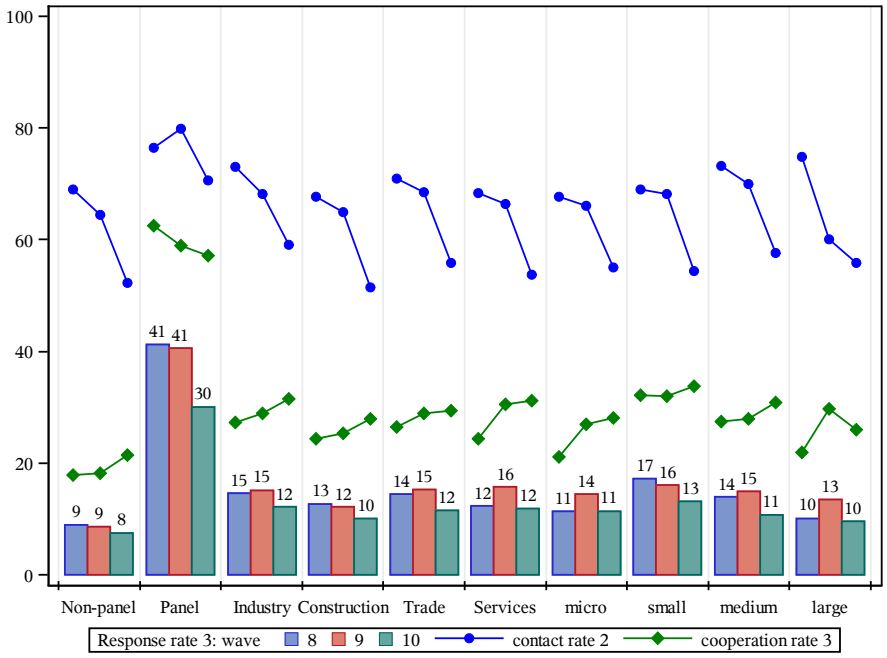



Figure 2: Outcome rates for SAFE from 8$^{th}$ to 10$^{th}$ survey round by panel dummy, sector and size (excluding Austria and Germany).

surveys conducted by the survey company, while in other countries, where the refusals are softer, good interviewers can more easily convince initial nonrespondents to eventually take part in the study. Second, the quality of the sampling frame differs across countries. The low quality of the enterprises' contact information, number of employees or sector will result in unsuccessful phone calls (in case of wrong company's number) or necessity to exclude a respondent after the screener questions (in case of 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). Third, the situation in the local offices of the survey company, such as the experience and training of the interviewers, work load at the time of conducting the survey can also have an impact on the response rate. In case of SAFE, additional factor which can explain the divergences is different CATI system used by the survey company in Germany and Austria and it is apparent that the outcome codes are not fully harmonised with offices in other locations. For that reason, we excluded those two countries from the subsequent analysis.

## 4. R-indicators as a measure of representativity

The concept of "representativeness" does not have single clear interpretation. Kruskal & Mosteller (1979) review the statistical and other scientific literature and divides the meaning of term "representative" into no less that nine different groups, varying from "general acclaim for data", through "miniature of the population" to "representative sampling as permitting good estimation".

Representativity indicators (R-indicators) are based on 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, the R-indicators attempt to capture the overall impact of the nonresponse for the whole survey, and not only at the level of a particular estimate.

Although it is not the point of this paper to describe in details the theoretical properties of the R-indicators, which is much better done in Shlomo & Schouten (2013) or in Schouten, Cobben & Bethlehem (2009), we present their definition and main features.

The R-indicator is based on the standard deviation of the response propensities transformed to lie between 0 and 1, where 1 is 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 $R$:

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

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 & Bethlehem 2009, p. 104) depends on the response rate: $R \geq 1 - 2\sqrt{\bar{\rho}(1-\bar{\rho})}$. Notably, it reaches its minimum of 0 for response rate of 0.5, i.e. when the individual response propensities can have largest variation, while it increases when the response rate decreases from 0.5 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 variances are the basis for the conditional partial indicators (Schouten, Bethlehem, et al. 2012). Those indicators can be further decomposed into the category level R-indicators showing the contributions to the variation of the respective categories (de Heij, Schouten, Shlomo 2010).

| | **Unconditional** | **Conditional** |
|---|---|---|
| $S^2\left(\rho\right) =$ | $S^2_{between}\left(\rho\right)$ | $S^2_{within}\left(\rho\right)$ |
| **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 $\underline{X}$.<br>$n_h = \sum_{i=1}^{n} d_i\Delta_{h,i}$ 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 $h$.<br>$U_l$ 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, if the nonresponse correlates maximally with the variable of interest is $B_m\left(X\right) = \frac{1-R(\rho)}{2\bar{\rho}} \leq 1 - \bar{\rho}$ and it can be shown that it cannot be larger than the nonresponse rate (see Schouten, Morren, et al. 2009).

## 5. R-indicators for SAFE survey

For the computation of R-indicators and associated statistics, we used the SAS code available at the website of the RISQ project[7] (see also de Heij, Schouten, Shlomo 2010) for the methods of bias adjustment and computation of confidence intervals of the R-indicators).

[7] http://www.risq-project.eu/tools.html; We would like to thank Natalie Shlomo for providing additional SAS code for stratified simple random samples and useful suggestions.

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

Although the methods to estimate representativity were not designed for quota samples, and consistent with 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 nonrespondent. Since the objective of the paper is to assess the influence of the firm characteristics on the 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 we computed the R-indicators for stratified samples[8]. As expected, the overall R-indicator improves; however, the effect of the remaining variables (sector and panel) is similar to the presented results without stratification.

All R-indicators were computed using four above mentioned variables, i.e. country (9 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 (8 to 10)

We start the examination from the R-indicators for each survey round looking at the overall response and contact rates. It would be possible to split the response process into successive sub-processes of contact, cooperation and final response, as it was done in Schouten, Bethlehem, et al. (2012). However, being unsure to which extent the outcome codes are harmonised among countries, we limit this initial analysis to two processes mentioned.

Interestingly, the R-indicator for overall response is the lowest for the 9$^{th}$ round, although the highest response rate was achieved in that round (see Table 1). Notably, it was the time when longer questionnaire was used. We cannot draw conclusion from this one observation, but it would be recommended to monitor in the future the development of the nonresponse bias in the rounds with the extended questionnaire.

Table 1: R-indicators and other associated information for the survey rounds 8 to 10.

| | **Response** | | | **Contact** | | |
|---|---|---|---|---|---|---|
| **Round** | **8** | **9** | **10** | **8** | **9** | **10** |
| Total sample | 70,432 | 58,689 | 62,090 | 70,432 | 58,689 | 62,090 |
| 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 |
| Ave 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 |

A higher response rate does not guarantee better representativeness. For instance, the R-indicator for the response is the highest and maximal bias is the lowest for round 10, although

[8] Available upon request.

the response rate was higher for the previous round (see Table 1 and Figure 3). It is also useful in the analysis of the overall representativeness to look at the maximal bias, especially since it is not sensitive to the level of the response rates. Figure 3 illustrates that R-indicators are higher for the overall response process than for the contact, but the maximal bias is much lower for the contact. It seems that other sub-processes of the overall response behaviour (such as cooperation of the respondents) may play a bigger role and contribute to the potential loss of representativeness.

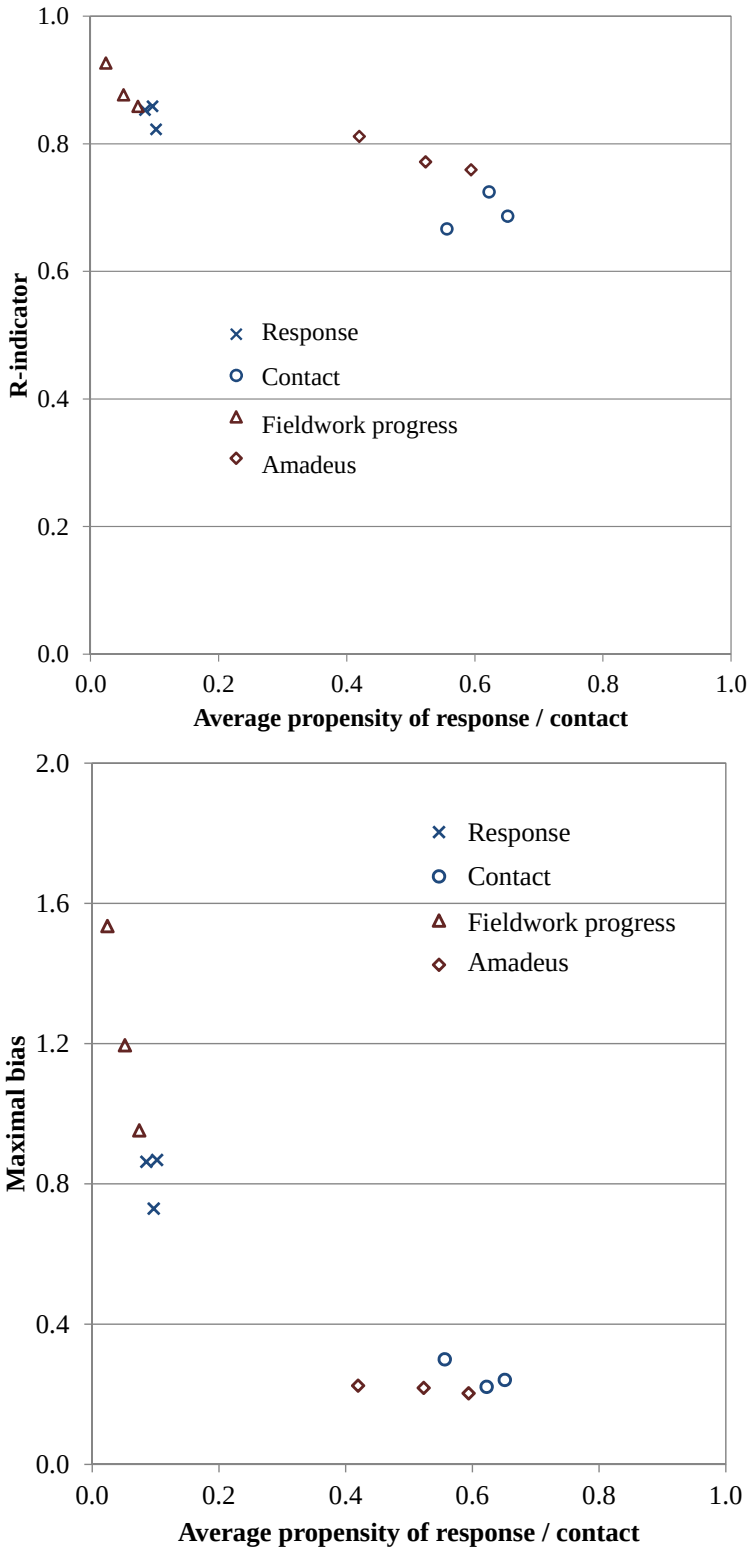
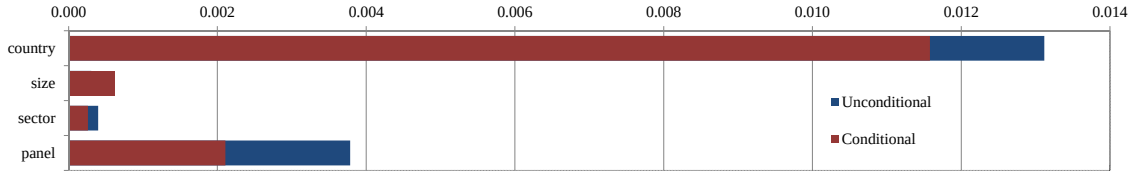
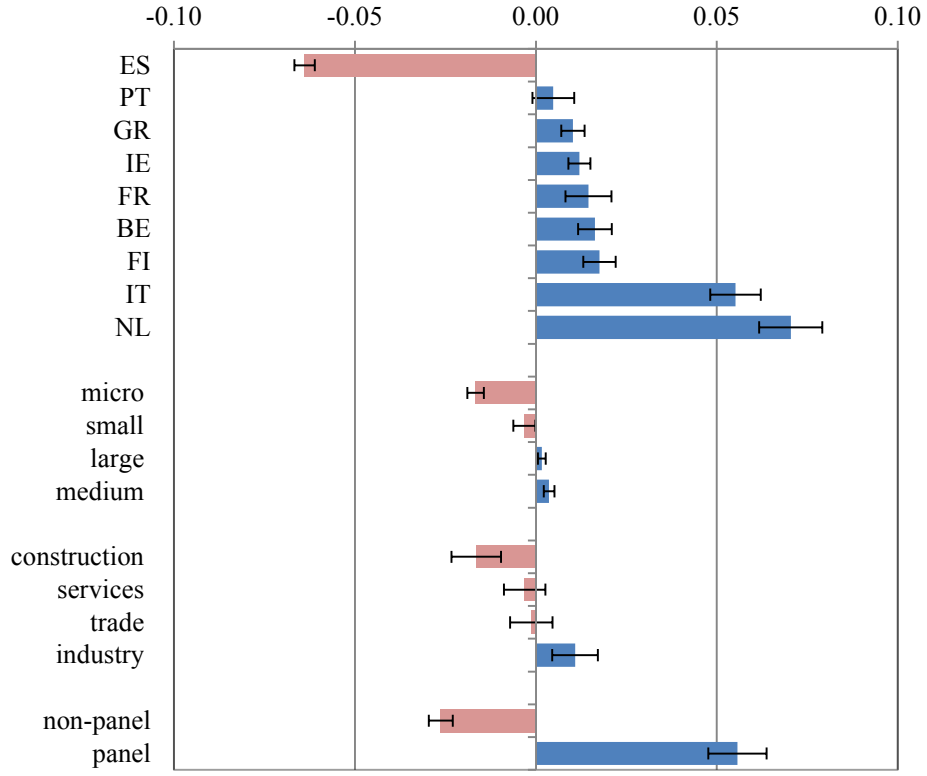



Figure 3: R-indicators and maximal bias as a function of the average propensity of response/contact for the survey rounds 8 to 10.

Looking at the R-indicator corresponding to contact propensities, the 10$^{th}$ survey round scores the worst. It was already visible from the investigations of outcome rates, where the contact rate dropped dramatically from round 9 to 10, particularly in three countries: Austria, Germany (both excluded from the analysis) and Spain[9]. In this case, low contact rate is also associated with higher bias – the large negative unconditional values for 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 in comparison overrepresented (see Table 4 and Figure 4).




(a) Variable level: Conditional and unconditional partial indicators



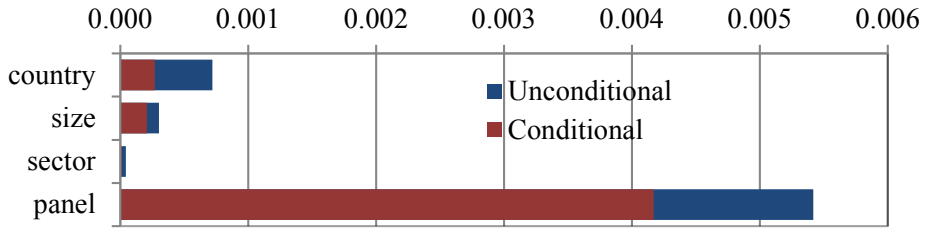

(b) Category level: Unconditional partial indicators with 95% confidence bands

Figure 4: Partial indicators for contact in 10$^{th}$ survey round.

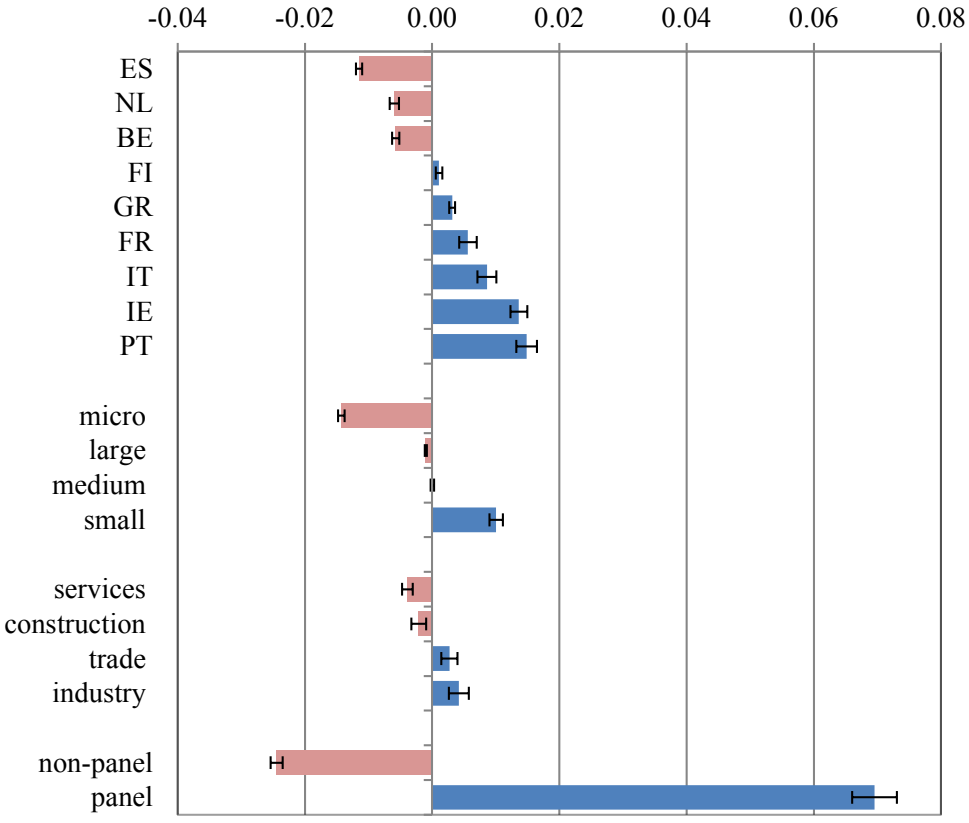
More generally, with respect to contact the unconditional and conditional partial R-indicators are the highest for the variable country and the country variation contributes the most to the loss of representativeness in all examined survey rounds. It seems that enterprises in some countries are more difficult to contact than in other regions, which points out also to the issues with the quality of the sampling frame. For SAFE, enterprises are all sampled from Dun & Bradstreet; however, the availability and accuracy of the contact information is not homogenous, given that the underlying sources of information differ by country. Consequently, it would be recommended to increase the efforts in the improvement of the sampling frame.

[9]The disproportionately high non-contact rate in drop in the 10$^{th}$ wave was a result of approaching relatively many enterprises at the beginning of the fieldwork. Enterprises, which 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 the respondents.

If we turn to the overall response, unsurprisingly, the fact whether the enterprise belongs to the panel or not 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 4 and Figure 5). This is consistent with the earlier finding about much higher response propensities of the panel firms. It is also comforting that the firm's characteristics available in the registers do not play a role in the response patterns. This is confirmed when the R-indicators were calculated separately for the firms which in a given round participated for the first time in the survey – also in this case the unconditional and conditional indicators at the variable level were not statistically different from zero[10].




(a) Variable level: Conditional and unconditional partial indicators




(b) Category level: Unconditional partial indicators with 95% confidence bands

Figure 5: Partial indicators for response in 8$^{th}$ survey round.

## 5.2. R-indicators during the SAFE fieldwork

The R-indicators can be implemented as a tool for monitoring the representativeness during the data collection. They can be computed for different amount of efforts, e.g. number of attempts, level of interviewer's experience. In SAFE such fieldwork information is limited and we analyse the development of the R-indicators during fieldwork progress.

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

The SAFE is conducted usually within one month, however, the start and end of the fieldwork can slightly vary by country. To account for these differences, we divide 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 round are presented in Table 2.

Table 2: R-indicators for the response and other associated information for each quartile on the fieldwork (8th survey round)

| | 1st quartile | 2nd quartile | 3rd quartile | Full fieldwork |
|---|---|---|---|---|
| Total sample | 70,432 | 70,432 | 70,432 | 70,432 |
| R-indicator | 0.926 | 0.877 | 0.859 | 0.853 |
| Standard error | 0.003 | 0.004 | 0.003 | 0.003 |
| Ave 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 the data collection, the representativity is the highest with R-indicator reaching 0.93. It drops slightly in the second quartile to 0.88 and remains broadly stable till the end of the fieldwork. In this case, the split of the sample into the enterprises which are part of the panel and those participating for the first time plays the major role as indicated by increasing partial R-indicator as the fieldwork progresses (see Table 5). However, a positive impact of each additional week of the fieldwork is visible when looking at maximal bias – it decreases steadily from 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 Figure 3).

## 6. R-indicators for SAFE data matched with Amadeus database

In this section, first we describe briefly the matching methodology of the SAFE dataset with the Bureau van Dijk's Amadeus database and comment on the quality of the matching. Second, with the dataset, containing both qualitative and quantitative firm-level information, we analyse the R-indicators looking at the availability of the financial information among respondents.

To link the companies from SAFE and Amadeus the information on tax identification number, company name, street, postcode, city and country are used. In the 8th round, 86% of SAFE respondent[11] were successfully matched with Amadeus business register. The quality of matching varies substantially between countries, with success rates over 90% in Belgium, Spain, France and the Netherlands and the lowest in Greece of 67%. There is also a significant difference between the size classes, with the large companies being successfully matched in 98% of cases, whereas the micro firms only in 72%. The difference on the 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; a record may have missing financial information. For that reason, we examine separately the representativeness of the SAFE subsamples containing the respondents with the available information on loans, value added and turnover in 8th survey round (in short, "Amadeus sample").

The R-indicators were computed using the same auxiliary variables as in earlier analysis (i.e. country, size, sector and panel dummy), and amount to 0.81 for the value added and are a bit lower for loans and turnover (0.76 and 0.77 respectively; see Table 3). In all three cases, the

[11] As in the previous section, Austria and Germany were excluded from the analysis.

lack of representativity, measured by both partial conditional and unconditional R-indicators, comes from the country variable, similarly to results for the contact (see section 5.1). However, given the smaller sample size, the unconditional partial indicator is statistically significant at 0.1 level only for value added (for turnover p-value equals to 0.12 for country and 0.11 for size variable; see Table 6). Estimated negative values for the category level partial indicators, suggest that the enterprises in the Netherlands and to a lesser extent in Greece are underrepresented in the set of companies with available financial information. Looking at value added and turnover this applies also to Belgium and Ireland. On the other hand, France and Spain are strongly overrepresented with respect to all the three variables considered.

Table 3: R-indicators and other associated information for the availability of information on loans, value added and turnover ($8^{th}$ survey round, respondents)

| | Loans | Value added | Turnover |
|---|---|---|---|
| Total sample | 6,008 | 6,008 | 6,008 |
| R-indicator* | 0.759 | 0.812 | 0.772 |
| Standard error | 0.003 | 0.002 | 0.003 |
| Ave 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 |

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

Similarly to the analysis of the whole SAFE sample with respect to the contact, the size class breakdown also contributes to the loss of representativity in the dataset matched with quantitative financial variables[12]. As expected, micro companies, for which financial information are scarce, are strongly underrepresented also in the matched SAFE subsample (see Table 6). The findings are also reflected in the overall matching rates at the enterprise level, as mentioned above.

It is also worth noting that in the $8^{th}$ round the maximal bias for the SAFE respondents among the whole sample is higher than for the subsample of the respondents with financial information (0.86 for the SAFE sample in comparison to 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.

## 7. Conclusions and outlook

In this paper we present R-indicators for SAFE and show that the level of representativity is comparable to other surveys (e.g. see Schouten, Bethlehem, et al. 2012). We found that for the SAFE sample, the country variation contributes mostly to the loss in representativity, while for the Amadeus subsample also size class plays some role with the evident underrepresentation of micro firms.

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

Since September 2014 (corresponding to $11^{th}$ 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 will have the

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

opportunity to disentangle the country variation from the differences in the organisation of local offices. Since the introduction of the online questionnaire in September 2014, it will be important to investigate and monitor the representativity of different survey modes.

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

Bańkowska K, Osiewicz M, 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, 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, 646-675.

Heij V de, Schouten B, 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, F. Mosteller (1979). *Representative Sampling, III: The Current Statistical Literature.* International Statistical Review 47, 245-265.

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

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

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

Schouten B, Cobben F & Bethlehem J (2009). *Indicators of Representativeness of Survey Response.* Survey Methodology Vol. 35, 101-113.

Schouten B, Morren M, Bethlehem J, Shlomo N & Skinner C (2009). *How to Use R-Indicators?.* URL http://www.risq-project.eu/publications.html.

Shlomo N & 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 (Sep 1966), 749-771.

Table 4: Unconditional and conditional partial R-indicators for contact and response in 8 to 10 survey round.

| | ***Unconditional*** | | | | | | ***Conditional*** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **response** | | | **contact** | | | **response** | | | **contact** | | |
| *Round* | **8** | **9** | **10** | **8** | **9** | **10** | **8** | **9** | **10** | **8** | **9** | **10** |
| *Variable level* | | | | | | | | | | | | |
| 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 0.01 level, ** indicates significance at 0.05 level and * indicates significance at 0.1 level.

Table 5: Unconditional and conditional partial R-indicators for response during fieldwork progress in round 8.

| | ***Unconditional*** | | | | ***Conditional*** | | | |
|---|---|---|---|---|---|---|---|---|
| | **1st quartile** | **2nd quartile** | **3rd quartile** | **Full fieldwork** | **1st quartile** | **2nd quartile** | **3rd quartile** | **Full fieldwork** |
| *Variable level* | | | | | | | | |
| 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 0.01 level, ** indicates significance at 0.05 level and * indicates significance at 0.1 level.

Table 6: Unconditional and conditional partial R-indicators for SAFE respondents matched with Amadeus database ($8^{\text{th}}$ survey round).

| | ***Unconditional*** | | | ***Conditional*** | | |
|---|---|---|---|---|---|---|
| | **Loans** | **Value added** | **Turnover** | **Loans** | **Value added** | **Turnover** |
| *Variable level* | | | | | | |
| 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 0.01 level, ** indicates significance at 0.05 level and * indicates significance at 0.1 level.

**Affiliation:**

Katarzyna Bańkowska
European Central Bank – Statistics Development/Coordination Division
60640 Frankfurt am Main, Germany
E-mail: katarzyna.bankowska@ecb.europa.eu

Małgorzata Osiewicz
European Central Bank – Statistics Development/Coordination Division
60640 Frankfurt am Main, Germany
E-mail: malgorzata.osiewicz@ecb.europa.eu

Sébastien Pérez-Duarte
European Central Bank – Statistics Development/Coordination Division
60640 Frankfurt am Main, Germany
E-mail: sebastien.perez_duarte@ecb.europa.eu

Austrian Journal of Statistics
published by the Austrian Society of Statistics
Volume 44
April 2015

http://www.ajs.or.at/
http://www.osg.or.at/

*Submitted:* 2014-09-03
*Accepted:* 2015-01-08