

Emilia Tomczyk[‡]

To be (accounted for) or not to be: Factors influencing probability of non-response in economic tendency surveys

Abstract

Non-response constitutes the major challenge for empirical studies based on survey data as it introduces the element of self-selection and damages representativeness of surveys. So far, formal analyses of causes of survey non-response have not resulted in supplying satisfactory remedial measures. The aim of this paper is to establish key factors affecting probability that a firm will respond to an economic tendency survey, and to propose methods to increase response rates. Empirical results lead to conclusion that information collected in the RIED (Research Institute for Economic Development of the Warsaw School of Economics) questionnaires does not allow to identify factors that influence non-response. The only statistically significant finding is a tendency of petroleum, chemical, pharmaceutical, rubber and plastic producers to be more responsive than other companies. Any additional factors that determine probability of responding to the RIED economic tendency survey remain unknown. Short-term solutions to the non-response problem include establishing direct contact with non-respondents and use of incentives; in the long-run, requests for additional information from the respondents may facilitate further studies on causes of survey non-response.

Keywords: economic tendency surveys, non-response, survey data

JEL classification: C83

[‡] SGH Warsaw School of Economics, Collegium of Economic Analysis, Institute of Econometrics.

Czynniki prawdopodobieństwa wystąpienia obserwacji brakujących w badaniach koniunktury metodą testu

Streszczenie

Brakujące obserwacje stanowią poważny problem w badaniach koniunktury gospodarczej metodą testu, ponieważ wprowadzają element autoselekcji i podważają ich reprezentatywność. Nie udało się jak dotąd wypracować skutecznych metod zapobiegania występowaniu brakujących obserwacji. Celem niniejszej pracy jest identyfikacja czynników, które decydują o tym, czy uczestnik badania koniunktury odpowie na pytania ankiety, i zaproponowanie działań, które zwiększą zwrotność ankiet. Z analizy danych pochodzących z badań koniunktury w przemyśle przetwórczym, prowadzonych przez Instytut Rozwoju Gospodarczego SGH, nie wynika jednak, by zawierały one informacje umożliwiające wskazanie takich czynników. Jedyny pozytywny i statystycznie istotny rezultat analizy to taki, że wytwórcy produktów naftowych, chemikaliów i tworzyw sztucznych udzielają pełniejszych informacji niż inni producenci.

Słowa kluczowe: brakujące obserwacje, dane ankietowe, test koniunktury

Kod klasyfikacji JEL: C83

1. Sources and consequences of non-response in tendency surveys

Survey data is subject to two general types of errors: sampling and non-sampling errors. Sampling errors result from weaknesses of the statistical sampling procedures and occur in every quantitative analysis based on sampled data. Non-sampling errors are usually described as all the remaining errors; such a wide designation makes them difficult to define and to measure. They may be caused by many factors, among them:

- gaps in the sampling frame due to incomplete or out of date business registers,
- faulty selection of the units to be sampled,
- non-response (total or partial): refusal of some of the selected units to provide information, either intentionally or unintentionally (for example, when a potential respondent cannot be contacted),
- honest mistakes and deliberately incorrect answers of survey respondents,
- inaccuracies introduced during collecting, editing, coding and entering responses into the data base,
- errors caused by further processing of survey data: data revisions, imputation of missing responses, quantification of qualitative data, etc.

Non-sampling errors are sometimes further classified (see Groves, 1989) into non-observation errors, that is, those resulting in loss of data (i.e., coverage errors, non-response, errors in applying statistical sampling techniques) and observational errors which lead to collection of faulty data.

In this paper, a subset of non-sampling non-observation errors is considered: those resulting from non-response. Non-response is acknowledged to be the main source of non-sampling errors in survey data in general, and economic tendency surveys in particular (Särndal *et al.*, 1992; Bethlehem, 2009). The main cause of concern for the analyst of survey data is that non-response, generally, is not random. It introduces the element of self-selection and consequently reduces reliability of surveys in terms of representativeness of results. Rasmussen & Thimm (2009) further classify non-respondents into two groups:

- passive, who may have planned to provide answers but circumstances prevented them from doing so,
- active, who deliberately chose not to respond.

Authors stress that active non-response requires more attention as it can lead to biased (self-selected) survey sample whereas passive non-response may be considered random. Since both types of non-response reduce sample size, and active non-response does so in non-random way, representativeness

of the sample is damaged. Significant non-response may introduce doubts as to validity of analysis of survey data or lead to incorrect results.

The extent of non-response varies from one type of survey to another, but it is usually substantial due to voluntary character of most economic tendency surveys. Let us define unit response rate (URR) following Thompson & Oliver (2012) as a percentage of total number of responding sample units (or net sample; see Schwarz, 2013) to the total number of sample units drawn (gross sample). In March 2017, out of 1204 industrial enterprises which received the RIED monthly economic tendency questionnaire (administered by the Research Institute for Economic Development of the Warsaw School of Economics), only 274 responded. That is, URR is equal 22.8%, and non-response rate – to 77.2%. Such non-response rates are not uncommon. Rasmussen & Thimm (2009) indicate that typical survey response rates deteriorated from 64.4% in 1975 to 50% in 1995 and are further declining, reaching about 20% in 2009. Tomczyk & Kowalczyk (2010) show that URR rates in the RIED economic tendency surveys in industry range between 20.4% and 41.5% (but do not show the tendency to deteriorate with time). Response rates commonly obtained in economic surveys can be therefore judged unsatisfying, and since this phenomenon negatively affects the quality of research based on survey data, research into causes and solutions for non-response is warranted. In this paper I attempt to determine factors which influence the probability of an individual respondent to take part in an economic tendency survey in order to find ways to limit the negative influence of non-response on empirical results drawn from survey data.

2. Description of data

In this paper, empirical analysis is based on enterprise-level data collected by the Research Institute for Economic Development (RIED) of the Warsaw School of Economics in four separate periods in time: September 2015, March 2016, September 2016 and March 2017 (to avoid drawing conclusions that might be specific to a unique moment in time). The models used are binary logit models with a dependent variable RESP coding whether an industrial enterprise has responded to the RIED economic tendency survey in a given period or not. RESP is equal to 1 if questionnaire has been returned with at least one question answered, and 0 otherwise.

The following explanatory variables are available in the RIED database:

- type of ownership (private or public),
- size of enterprise as measured by the number of people employed, in the following categories: up to 50 employees; 51-250 employees; 251-500 employees; 501-2000 employees; more than 2001 employees,

- province (voivodeship),
- sector of business activity according to the Code List of Business Activities in Poland (see GUS, 2008).

Sample sizes vary between 831 (March 2016) and 1285 (September 2016). Further limitations are imposed on the models by the necessity to remove observations with item non-response which in this particular case means missing information on firm characteristics.

Ownership

The sample is dominated by private enterprises: only between 3.3% and 6.2% of the firms belong to the public sector. Ownership is coded in the binary X_SECTOR variable equal to 0 for private sector (base category) and 1 for public sector.

Size

To code enterprise size as measured by its employment, five binary variables have been constructed. Definitions are summarized in Table 1.

Table 1. Classification of enterprises according to size (sample: March 2017).

Variable name	Number of employees	Percentage of observations
X_EMPL_01	up to 50	46.0%
X_EMPL_02	51 – 250	35.2%
X_EMPL_03	251 – 500	13.9%
X_EMPL_04	501 – 2000	4.4%
X_EMPL_05	more than 2001	0.5%

Employment in majority of the firms does not exceed 250 persons. For example, in the March 2017 sample, larger companies (that is, the last three categories in Table 1) constitute only 18.8% of the sample. For the purposes of empirical analysis, these three categories are grouped into one variable: X_EMPL_L, equal to 1 for large enterprises which employ more than 251 persons and 0 otherwise. Employment variables are subject to the worst case of missing data in the dataset: only 52% observations are available in September 2016 and March 2017, 68% in September 2015, and 89% in March 2016.

Provinces

There are 16 provinces (voivodeships) in Poland; see Appendix I for their listing and corresponding number of respondents in the RIED economic

tendency survey in March 2017. To analyze influence of geographical region on probability of a respondent answering the RIED questionnaire, provinces are further classified according to two sets of criteria:

- growth potential (see Nazarczuk, 2013), presented in Table 2,
- competitive power (see Golejewska & Gajda, 2012), presented in Table 3.

Table 2. Classification of provinces according to their growth potential.

Variable name	Classified as	Provinces (voivodeships)
VOI_NZ_I	leader	mazowieckie
VOI_NZ_II	above average	śląskie, dolnośląskie, wielkopolskie, łódzkie
VOI_NZ_III	average	pomorskie, zachodniopomorskie, małopolskie
VOI_NZ_IV	below average	opolskie, lubuskie, kujawsko-pomorskie, podlaskie, warmińsko-mazurskie, podkarpackie, lubelskie, świętokrzyskie

Source: Nazarczuk (2013), p. 121.

Table 3. Classification of provinces according to their competitive power.

Variable name	Classified as	Provinces
VOI_GG_I	leader	mazowieckie
VOI_GG_II	high	małopolskie, śląskie, dolnośląskie, pomorskie
VOI_GG_III	average	podkarpackie, lubuskie, kujawsko-pomorskie, wielkopolskie, opolskie, warmińsko-mazurskie, łódzkie
VOI_GG_IV	low	lubelskie, podlaskie, świętokrzyskie, zachodniopomorskie

Source: Golejewska & Gajda (2012), p. 14.

In order to avoid arbitrary assumptions whether it is growth potential or competitive power of their geographical region of operation that influences respondents' decision to answer the questionnaire, both classifications are included in the empirical analysis.

Sectors of business activity

Enterprises surveyed by the RIED economic tendency questionnaires can be classified into 25 categories out of 99 defined by the Code List of Classification of Business Activities in Poland – PKD 2007. Since the number

of PKD categories is high and some of them include only a few companies, basic categories are aggregated into 11 clusters for the purposes of the empirical analysis, and 11 binary variables are created (see Appendix II). Category VIII (manufacture of basic metals and fabricated metal products, except machinery and equipment) is the most numerous and will constitute a base variable for the analysis of influence of sector or business activity on probability that an industrial company responds to the economic tendency survey.

Certainly there may be factors other than basic and measurable characteristics of an enterprise that influence probability of responding to an economic tendency survey: for example, organizational structure, work culture, company-specific attitude towards questionnaires, etc. These factors, however, remain unobservable and highly subjective. They may even be of psychological nature (relating to individuals who fill out the questionnaires); for example, negative response style of households has been found to influence inflation expectations (Białowolski, 2016). Still, these psychological factors should play a minor role in more formalized environment of business enterprises. The scope of this article is limited to information provided by the industrial enterprises in the descriptive label of the questionnaire.

3. Empirical results

Initially, for each of the samples a binary logit model with all the explanatory variables was estimated to evaluate size and statistical significance of all the potential causes of non-response, and then successive models were estimated by maximum likelihood method according to the from-general-to specific principle in two versions: with provinces classified according to their growth potential and to their competitive power. In the course of the analysis, almost all these factors proved to be statistically insignificant. Some patterns can be discerned, however, and will be discussed on the basis of estimation results presented in Tables 4-7. Some of the insignificant variables were retained in order to enable comparisons between the four periods and highlight general results that seem to emerge from the analysis.

From the summary of estimation results presented in Tables 4-7 it is clear that the only statistically significant factors that influence the probability of obtaining a response are two specific sectors of business activity: manufacture of coke and refined petroleum products, chemicals and chemical products, basic pharmaceutical products and pharmaceutical preparations (variable PKD_VI) and manufacture of rubber and plastic products and other non-metallic mineral products (variable PKD_VII). Companies in these

sectors are characterized by significantly higher probability of providing response to the RIED economic tendency survey than firms in other sectors in recent samples (September 2016 and March 2017); in earlier samples, the effect is limited to the PKD_VII category (March 2016) or disappears (September 2015). Also, large enterprises and private companies tend to be more likely to respond to the survey but these effects are not statistically significant.

Table 4. Estimation results: logit model of variable RESP (September 2015).

Model: Logit, using observations 1-1158 (n = 779)
Missing or incomplete observations dropped: 379
Dependent variable: RESP
Standard errors based on Hessian

	<i>coefficient</i>	<i>standard error</i>	<i>z</i>	<i>p-value</i>	<i>slope at mean</i>
const	-0.877639	0.116824	-7.5125	<0.0001	
X_SECTOR	-0.257223	0.333885	-0.7704	0.4411	-0.055243
PKD_VII	0.33587	0.214841	1.5633	0.1180	0.077715
X_EMPL_02	0.33515	0.171274	1.9568	0.0504	0.076171
X_EMPL_L	0.37268	0.219839	1.6953	0.0900	0.086250

Table 5. Estimation results: logit model of variable RESP (March 2016).

Model: Logit, using observations 1-831 (n = 819)
Missing or incomplete observations dropped: 12
Dependent variable: RESP
Standard errors based on Hessian

	<i>coefficient</i>	<i>standard error</i>	<i>z</i>	<i>p-value</i>	<i>slope at mean</i>
const	-0.648328	0.083243	-7.7883	<0.0001	
X_SECTOR	-0.205203	0.321022	-0.6392	0.5227	-0.0455539
PKD_VI	-0.011636	0.364504	-0.0319	0.9745	-0.0026547
PKD_VII	0.420301	0.207884	2.0218	0.0432	0.0995965

Table 6. Estimation results: logit model of variable RESP (September 2016).

Model: Logit, using observations 1-1285 (n = 1269)
Missing or incomplete observations dropped: 16
Dependent variable: RESP
Standard errors based on Hessian

	<i>coefficient</i>	<i>standard error</i>	<i>z</i>	<i>p-value</i>	<i>slope at mean</i>
const	-1.31452	0.075503	-17.4103	<0.0001	
X_SECTOR	0.51102	0.340587	1.5004	0.1335	0.101173
PKD_VI	0.66585	0.344990	1.9300	0.0536	0.135855
PKD_VII	0.46772	0.188488	2.4814	0.0131	0.089958

Table 7. Estimation results: logit model of variable RESP (March 2017)

Model: Logit, using observations 1-1204 (n = 619)

Missing or incomplete observations dropped: 585

Dependent variable: RESP

Standard errors based on Hessian

	<i>coefficient</i>	<i>standard error</i>	<i>z</i>	<i>p-value</i>	<i>slope at mean</i>
const	-0.66268	0.102159	-6.4868	< 0.0001	
X_SECTOR	-0.55220	0.375882	-1.4691	0.1418	-0.119079
X_EMPL_L	0.26196	0.220225	1.1895	0.2342	0.062275
PKD_VI	1.07679	0.425846	2.5286	0.0115	0.262811
PKD_VII	0.55478	0.233603	2.3749	0.0176	0.134213

The following patterns emerge from the empirical analysis:

- Geographical location (province) does not influence the probability of responding to the questionnaire, either for growth potential classification or competitive power classification. Province variables are not statistically significant in any of the models considered.
- Private companies are slightly more likely to respond; this effect is present in most models considered but generally statistically insignificant.
- Large companies are slightly more likely to respond; this effect is present in most models considered but generally statistically insignificant.
- Of all the sectors of business activity, two stand out: namely, manufacture of coke and refined petroleum products, chemicals and chemical products, basic pharmaceutical products and pharmaceutical preparations, rubber and plastic products and other non-metallic mineral products. Companies in these sectors are more likely to provide answers to the economic tendency survey. This effect is present in most models considered and statistically significant in recent samples.

The following general conclusion can be drawn on the basis of the empirical analysis: information gathered by the RIED on prospective respondents (that is, type of ownership, sector of operation, size measured by employment, and geographical location) does not suffice to identify factors influencing probability of obtaining an actual response from an addressee. Apart from a tendency of petroleum, chemical, pharmaceutical, rubber and plastic producers to respond to the survey with higher probability than companies belonging to other sectors, any additional factors that determine whether a firm answers a questionnaire or not are still to be isolated and explained. Further research is clearly needed to identify statistically significant variables that systematically influence the decision of an industrial

enterprise to comply or to refuse when presented with a business tendency survey.

4. Summary and conclusions

This paper presents an attempt to determine factors which influence the probability of the individual respondent to take part in the RIED economic tendency survey, and to identify methods to reduce non-response. All the respondent characteristics available in the RIED database (namely, type of ownership, sector of operation, size, and geographical location) have been taken under consideration by the means of explanatory variables in binary logit models. Empirical results allow to conclude that these characteristics fail to explain probability of obtaining a response in a satisfactory way. The only factor that remains statistically significant across the models is a tendency of petroleum, chemical, pharmaceutical, rubber and plastic producers to be more responsive than other companies. Any additional factors that determine probability of responding to the RIED economic tendency survey remain unknown.

This result makes it considerably more difficult to identify methods of reducing non-response; its causes do not appear to be reflected in companies' characteristics such as size, geographical location or sector of operation. In this light, what can be done to improve response rates of the RIED economic tendency survey?

Since information gathered in the questionnaire does not allow to identify factors that influence non-response, efforts should be made to obtain more detailed information on prospective respondents. This could be done via requesting additional information from the companies (for example, other measures of companies' size, their financial standing, etc.) or employing mixed methods of surveys – that is, postal questionnaires followed by internet or telephone contact. Of course, in order to receive more detailed information from respondents, they must first be convicted to reply to a business tendency survey. Literature suggests that establishing direct rapport with the addressees and gaining their support and understanding of the aims of the survey may induce them to participate in the survey more actively. Re-contacting non-respondents in particular may persuade them to become involved in the survey. Apart from various forms of contact with the respondents, several authors (see Curtin *et al.*, 2005; Toepoel & Schonlau, 2017) point to the use of incentives; in literature on economic surveys, incentives vary from simple thank-you notes to hard cash. While most authors agree that incentives promote response, high incentives may consume a significant portion of a project's budget, and skew responses when provided only in order to obtain

a payoff. Curtin *et al.* (2005) report that reminders, tokens of appreciation, the wording of the cover letter and the option to choose the frequency of being surveyed are effective means of increasing response rates. However, the survey mode used by RIED (that is, postal) reduces response rates, according to this study.

Schwarz (2013) identifies two significant determinants of response rates: type of institution conducting the survey (with statistical offices and national banks response rates higher than private institutions' by, on average, 23 percentage points) and use of quota samples. However, for quota samples, in which non-responding firms are replaced with similar companies to fulfill the quota plan, no meaningful "response rates" can be calculated (see Curtin *et al.*, 2005). As far as professional standing of the agency administering the survey is concerned, however, RIED is very well positioned in the academia: it is a part of one of the best Polish economic universities, and is widely recognized as a leading force in Polish economic tendency survey analysis.

Other methods, which can be classified as remedies to non-response rather than prevention measures, include post-survey adjustments: post-stratification weighting schemes, filling of missing data by imputation and extrapolation, and combination of these methods (see Groves, 2006; Rasmussen & Thimm, 2009; Toepoel & Schonlau, 2017). Various aspects of imputation and weighting have been previously analyzed for RIED survey data (see Tomczyk & Kowalczyk, 2010; Kowalczyk, 2010) but no general solution to the non-response problem has been found.

Continuation of the current study is planned with a panel model in order to verify whether a dynamic component would help to identify sources of non-response in economic tendency surveys. However, since samples in panel surveys tend to deteriorate with time as panel members cease to participate in multiple-wave study, sample attrition, together with item non-response, may substantially limit the number of observations available.

To summarize, even though results of the empirical analysis do not provide ready answers as to which factors determine probability of obtaining responses in an economic tendency survey, some conclusions follow. In the short term, efforts to contact non-respondents and encourage them to take part in the survey – either by direct incentives or non-monetary inducements like personalized letters – should lead to higher response rates. In the long run, requesting additional information that could be used in subsequent econometric studies to determine causes of non-response, and possibly changing the mode of the survey from postal to internet-based, should also limit the extent of non-response. These measures require additional expenditures in terms of money and commitment on the part of the survey

administrators but should be rewarded with higher response rates, and improving response rates would go a long way towards convincing economists to use survey data on a larger scale.

References

- Bethlehem, J. (2009). *Applied Survey Methods: A Statistical Perspective*. New Jersey: John Wiley & Sons.
- Białowolski, P. (2016). The influence of negative response style on survey-based household inflation expectations. *Quality & Quantity. International Journal of Methodology*, 50: 509-528.
- Curtin, R., Presser, S., Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *The Public Opinion Quarterly*, 69: 87-98.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *The Public Opinion Quarterly*, 70: 646-675.
- GUS (2008). *Zmiany strukturalne grup podmiotów gospodarki narodowej w 2007 r.* Warszawa: GUS.
- Kowalczyk, B. (2010). On selected issues of non-response in the case of business tendency surveys. *Acta Universitatis Lodziensis. Folia Economica*, 235: 263-271.
- Rassmusen, K., Thimm, H. (2009). Fact-based understanding of business survey non-response. *The Electronic Journal of Business Research Methods*, 7(1): 83-92.
- RIED (2017). *Koniunktura w przemyśle*. 344.
- Särndal, C., Swensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Schwarz, G. (2013). *Response Rates in European Business Tendency Surveys*. Vienna: Austrian Institute of Economic Research (WIFO).
- Thompson, K., Oliver, B. (2012). Response rates in business surveys: Going beyond the usual performance measure. *Journal of Official Statistics*, 28(2): 221-237.
- Toepoel, V., Schonlau, M. (2017). Dealing with nonresponse: Strategies to increase participation and methods for postsurvey adjustments. *Mathematical Population Studies*, 24(2): 79-83.
- Tomczyk, E., Kowalczyk, B. (2010). Influence of non-response in business tendency surveys on properties of expectations. *Statistic in Transition New Series*, 11: 403-422.