Mirosław Szreder
Jerzy P. Gwizdała

POSSIBLE FUTURE DEVELOPMENTS OF SAMPLE SURVEYS IN FINANCE

Abstract

Although sample surveys do not play a primary role in research, their impact has increased recently in many areas of finance and economics. At the same time, however, new difficulties have arisen in conducting surveys and making inference on the basis of sample results. These difficulties concern not only up-to-date sampling frames and proper modes of interviewing respondents, but also the rapid increase of non-response rates, which affects the quality of outcomes. In these circumstances it seems reasonable to search for external data in order to reduce the consequences of non-sampling errors. The article presents the authors' view on the possible use of big data as supplementary information. It also discusses some of the main challenges for survey-based research in the near future. The main conclusion is that big data will support rather than overtake surveys in finance in the years ahead.

Keywords: sample survey, statistical inference, survey-based research, big data.
JEL Classification: C10.

1. Introduction

For the last several decades sample surveys have been successfully applied in various research areas, including empirical economics and finance. Their main objective is to describe certain characteristics of the population under study on the basis of a small representative fraction of it – usually a random sample. The increasing impact of survey-based research in modern times is mainly a result of continuous progress in the theory of statistical inference,

Mirosław Szreder, University of Gdańsk, Faculty of Management, Department of Statistics, Armii Krajowej 101, 81-824 Sopot, Poland, e-mail: miroslaw.szreder@ug.edu.pl

Jerzy P. Gwizdała, University of Gdańsk, Faculty of Management, Department of Finance and Financial Risk, Armii Krajowej 101, 81-824 Sopot, Poland, e-mail: jgwizdala1@wp.pl

and the rapid development of new technological tools for collecting and processing data.

Statistical inference – estimation and hypothesis testing – is used in finance in a large number of research problems which include inference about unknown parameters that describe various financial phenomena. Together with the theory of probability, inferential statistics constitute a basic set of methods and techniques for analysing stochastic processes and other models popular in financial research and practice. Finance is one of the sciences where a large amount of statistical data is generated every day (or every second, if one looks at it from the global perspective). Therefore, it provides on the one hand good opportunities for testing new inferential approaches and methods, and on the other creates new challenges connected with the increasing complexity of financial problems[1].

Along with research that includes statistical or econometric modelling, other methods are also used in finance where statistical inference plays a vital role, one of these being survey-based research. This kind of research has gained more attention in recent years not only in finance but in empirical economics as a whole. Inference based on incomplete information (sample information) is used in various practical problems, including decision--making in enterprises. Business activity surveys and consumer sentiment surveys are examples of research that tends to be carefully monitored by many players on the market. Survey-based research is the main concept we are focusing on in this paper. The main features of survey-based research are: quantitative characteristics of the variables of interest in the population, sampling techniques which allow the researcher to generalise about the population being investigated, and soliciting information from people using a questionnaire.

The main objective of our paper is to attempt to answer the following two questions:

1) Will the growing popularity of sample surveys in finance, accompanied by time pressure to obtain quick results and the unwillingness of many respondents to cooperate, allow the surveys to remain reliable and accurate sources of information?

2) How likely is it that survey-based research will be overtaken in the near future by investigations based on data-rich information resources, such as metadata, paradata, and big data?

---

[1] See, for instance, Gwizdała (2013) about the role of reliable information in solving multi-faceted risk problems in banking.

2. The Role of Sample Surveys in Finance

Although survey-based research does not play a primary role in finance, its impact should not be neglected. Moreover, this kind of research has recently become more popular than ever before. In our opinion, there are three main reasons to account for this.

First, quantitative survey research is quite frequently the only way to empirically test theories describing the activities and behaviours of people in financial markets. Additionally, such surveys can be helpful in discovering new phenomena or trends in finance.

Second, the surveys provide data which form the basis for a number of widely recognised indicators, indices, and rankings. All of these are designed to describe certain aspects of financial markets in a condensed, synthetic form. Examples of such measures are: the Global Findex[2], Consumer Confidence Index (CCI)[3], Stock Market Confidence Index[4], PENGAB[5].

Third, the increasing proportion of internet surveys involves lower costs and a shorter period for obtaining information from respondents compared with the pre-internet era. This relates to both nationwide and worldwide surveys, which in the past were particularly difficult to conduct. As a consequence, the whole process of collecting and analysing data is nowadays less costly and less time-consuming than in the past.

The internet and other technological advances have led to growing interest among researchers in sample surveys, not only in finance. Survey--based research tends to drive out other research approaches, including qualitative studies. This tendency may raise some concern, as some researchers seem to ignore the fact that quantitative description essentially simplifies reality. Numbers themselves reflect the real world in a very imperfect manner. Therefore, if a study focuses on the motivations behind the choices of consumers or investors, a quantitative approach may not be appropriate. Moreover, taking into account that in sample surveys respondents' answers are typically confined to a few close-ended statements, and the scales used in particular questions tend to be very simple, it becomes

[2] An index which measures the financial inclusion of people and is based on interviews with about 150,000 adults in over 140 countries.

[3] An indicator, formed from survey results of more than 5,000 US households, gauging financial health, spending power and consumer confidence.

[4] An index based on sample surveys of individual investors and institutions in the United States.

[5] A popular banking sentiment indicator in Poland based on surveys carried out by TNS Polska among bank managers.

clear that the final picture may be far from accurate. There are many examples of such oversimplified surveys in finance[6].

Apart from these constraints, the practical difficulties in performing a survey may also affect the quality of its findings. This includes problems with designing the sampling frame, the refusal of some respondents to cooperate, the misunderstanding of questions (question wording), and in the case of enterprises or institutions – uncertainty about the competence of the respondent representing such an entity. All this creates scepticism among some researchers about the validity and accuracy of scientific findings based on results obtained from sample surveys. H. K. Baker and T. K. Mukherjee (2007) have attempted to answer the question about the current status of survey-based research in finance. They asked editors of 50 "core" and "non--core" financial journals about their views on the role which survey-based research should play in finance literature[7].

Table 1. Role which Survey-based Research Should Play in the Finance Literature

Which of the following statements best describes your view on the role that survey-based research should play in the finance literature?	"Core" journal editors	"Non-core" journal editors	Total (*n*)	Total (%)
A. Survey-based research should be considered equal to other types of original research	-	10	10	43.5
B. Survey-based research should play a complementary role to other types of original research	4	6	10	43.5
C. There is a limited (or no) role for survey-based research relative to other types of original research	2	1	3	13.0
D. The role of survey-based research should be as follows	–	–	–	–

Source: Baker and Mukherjee (2007, p. 21).

The precise questions and answers given by 23 editors who agreed to cooperate are presented in Table 1. Although the majority of editors argue that survey-based research should be considered equal to other types of research or play a complementary role, there are clear differences among

---

[6] One can be found in Sokołowska (2014, chapter 5).

[7] Classification of journals into "core" and "non-core" was based on perceived quality. The response rate did not exceed 50% in this survey.

them. Those representing "core" journals seem unwilling to consider survey-
-based research equal to other original forms of research. They indicate that
the role of the former should be considered either complementary or limited
(or nil). At the same time, "non-core" journal editors view the role of the
survey-based research as complementary or equal to other types of original
research. None of the editors admitted that survey-based manuscripts are
generally discouraged in their journals.

In our opinion, the future status and impact of survey-based research in
finance will be equally determined by advances in the theory and practice of
sample designs, and by an integrated approach to the combination of sample
and non-sample (or external) information. We look at some of the main
challenges below. The majority of these are common to all survey-based
research and are not limited to applications in finance.

3. Main Challenges for Survey-based Research

The increasing variety and efficiency of ways in which interviewers
communicate with respondents seems to be one of the important incentives
for using sample surveys in contemporary research. Instead of time-
-consuming mail surveys which are rarely used these days, telephone surveys
and internet (Web-based) surveys have become more popular. It is worth
noting, however, that interviewers' relatively easy access to communication
channels with respondents does not necessarily imply high sample quality.
First, in the era of mobile phones, random digit dialling (RDD), as was
applied for years with regard to landline phones, does not unambiguously
identify a household or adult. Moreover, population units (respondents) who
have more than one mobile phone stand an unequal chance of being selected
for the sample. The problem of an appropriate and good quality sampling
frame when the survey is based on telephone interviews is still open[8]. In its
surveys, the Gallup Organisation uses a 50% listed landline sample and 50%
RDD mobile phone sample. This is not, however, a perfect solution, as was
proved by the authors of a report aimed at identifying factors responsible for
the failure of opinion polls before the US presidential election in 2012 (see
Gallup 2013). They argue that there were significant differences, including
in the field of demographics, in both these subsamples (selected by mobile
and landline phones). The weighting adjustment procedures applied were

---

[8] In Britain, for instance, different demographic groups have substantially different patterns of
mobile/landline use. Most notably, nearly 30% aged between 18 and 34 use a mobile but do not
have a landline, while the corresponding rate for those aged 55 and over is just 5%. See Sturgis *at
al.* (2016).

able to reduce some differences, although their impact on the final survey results was not without significance.

An alternative could be an internet (online) survey which is as relatively easy for respondents to handle as it is convenient for researchers to process the data recorded. Also, many respondents feel more anonymous and consequently are not tempted to give answers which would be in line with expectations (socially desirable) rather than reflecting their own attitudes or opinions. One of the challenges connected with using internet surveys is the difficulty of constructing sampling frames and designing procedures for recruiting respondents. If these are not properly solved, it is likely that they will generate a so-called coverage error[9]. More challenging, however, is identifying and explaining the differences that tend to crop up between results obtained by telephone and online surveys. Such differences occurred, for instance, in pre-election opinion polls in Britain in 2015, and also in opinion polls before the European Union membership referendum in the UK a year later[10]. There has been little evidence that one of the two ways of surveying respondents is superior to the other. Nor is there any firm knowledge about the circumstances in which one of these modes should be preferred over the other. For this reason, the use of mixed-mode designs or different modes is recommended for different segments of the population.

In finance, where enterprises, households and individual consumers regard a lot of information as confidential, a serious challenge for researchers using surveys is the high rate of non-response. One general reason for people being unwilling to take part in surveys or opinion polls is that many are fed up with them. It is presumably the price we all pay for the increasing number of surveys with which we bother others. Respondents' doubts as to whether the survey is legitimate, declining human interaction in surveys over the decades[11], and concerns about privacy and confidentiality among respondents (disclosure risk) are also responsible for lowering response rates. Nowadays, regardless of the subject of the survey, it is difficult to achieve response rates above 50%[12]. From a statistical perspective, high non-response rates should be considered a serious problem. A high proportion of non-responses can potentially render the whole inference invalid, as

---

[9] This error arises if not all population units are represented in the sampling frame.

[10] For details, see Sturgis *et al.* (2016).

[11] "The sample survey has been transformed from being a comfortable face-to-face conversation to a highly impersonal experience" (Dillman, Smyth & Christan 2009, p. 1).

[12] Interestingly, or perhaps surprisingly, a mere 50% rate of return was achieved in the previously mentioned survey among editors of financial journals. See Baker and Mukherjee (2007).

estimates are likely to be biased. The consequences of non-response are less harmful only if they can be regarded as missing at random observations. In this particular case, although the estimates suffer from lack of precision, they remain unbiased. However, in practice such fortunate situations are rare. More frequently, non-response is associated or correlated with some other population characteristics. This, in turn, results in underestimating or overestimating the population characteristics of interest. Additionally, in these circumstances the bias does not diminish with an increase in sample size[13]. R. M. Groves and E. Peytcheva (2008) show that at any given level of non-response the actual bias is unknown and may vary considerably. One of the most efficient ways to reduce the non-response error involves enriching sample information with other relevant information that could be used in weighting schemes or calibration procedures. This could be either sample information about other characteristics of the units or non--sample information, including external secondary data, e.g. commercial databases and administrative records. Weights are used in surveys not only for adjusting for non-response, but also for correcting the sample structure in order to improve the precision of inference. The effectiveness of all such measures aimed at substituting non-responses depends on the relevance and quality of the auxiliary data (see Särndal & Lundström 2006).

Apart from these general rules, there exist some practical recommendations for optimising response rates in a particular research area. In our opinion, as regards the field of finance, these recommendations should include generating confidence in respondents as to the anonymity of their answers, assuring them that the survey is valid, and choosing the most appropriate mode for conducting the survey. Because financial issues are commonly regarded as sensitive, it may be useful, for instance, to avoid those interviewing techniques which allow for the presence of other people.

Another challenge worth considering is the problem of sample section. Two questions are particularly important. First, will increasing the volume of information on various populations under investigation be capable of providing sufficient information to reduce the impact of random selection and eventually substitute probability sampling techniques with non--probability equivalents? Second, what is the actual scientific value of increasingly popular opt-in Web surveys?

Let us remember that the classical theory of statistical inference is based on the assumption of having a (simple) random sample. All basic estimation

---

[13] For details, see Bethlehem (2002) or Szreder (2010, chapter 6.3).

procedures and tests of significance rely on this assumption. However, in practical applications of sample surveys in economics, including finance, or in social science, simple random sampling is used only occasionally[14]. This discrepancy between theory and practice has one main reason. The researcher is not satisfied with using the random selection mechanism, which generates more or less representative samples. He or she seeks a highly representative sample in the particular survey and, as a consequence, is ready to employ other sources of information or knowledge that can improve a purely random selection mechanism. For example, if the population of interest is strongly diversified, stratified sampling which uses prior information about the structure of the population is usually more efficient than simple random sampling. If the amount of external (non-sample) information increases, and ultimately provides substantial knowledge about population units, one can imagine that random sampling can be replaced by non-probability techniques, such as quota sampling. Quota samples are popular in many commercial surveys and opinion polls in Britain[15]. Whether this tendency develops in the near future will depend on new methods and techniques enabling researchers to combine various sources of relevant information, including sample data, for the purpose of inference. The external sources also include administrative records and other official registers which can help verify the reliability of sample data. In Poland, for instance, information given by households and individuals on their incomes requires correction based on other sources of information. It has been shown that the data obtained from representative samples of households in surveys conducted by public statistical bodies are not consistent with the corresponding data records in tax offices[16]. More reliable information, and as a consequence more accurate inference, can be obtained if a combination of both these data sources is employed. It should be noted, however, that the theory of statistical inference does not yet give substantive support for this kind of combined information[17]. Further progress in developing a theoretical

---

[14] Among his seven original maxims of sampling, C. O'Muircheartaigh (2008, p. 296) includes the following: "Never use simple random sampling!".

[15] MORI-Ipsos claims that "most polls in Britain use quota sampling which pre-determines the demographic make-up of the sample to ensure it matches the profile of the whole adult population" and points out that "historically in Britain, the record of quota samples in predicting elections has been better than that of random samples". For details, see Worcester (1991).

[16] See Kośny and Mazurek (2009), Brzeziński (2015), and Brzeziński and Kostro (2010).

[17] The lack of proper theory also applies to register-based inference. A. Wallgren and B. Wallgren (2007, p. IX) write: "Although register-based statistics are the most common form of statistics, no well-established theory in the field has existed up to now".

framework for estimation and hypothesis testing based on a combination of various information sources seems to be one of the main challenges.

The other problem mentioned above relates to the validity of so-called opt-in (or self-identified) surveys in which respondents are self-selecting. This kind of survey is popular in market research and in some social sciences, mainly because it is inexpensive and convenient, as it attracts people who are really interested in completing the questionnaire. In response to the allegation that the selection bias can seriously affect the representativeness of such samples, proponents argue that it is likely to be compensated for by the large sample size. In fact, it is not. Those who take part in the survey may be systematically different from the remaining units of the population. In practice, people from high-income families refuse to cooperate in all surveys focused on financial issues. Similarly, representatives of large companies tend to be reluctant to complete surveys about their activities, particularly those related to the financial aspects of their activities. The self-selection mechanism and other techniques of convenient sampling will very rarely cover all segments of the population. Ch. Wheelan (2013, p. 118) seems to be correct when he writes: "If you ask 100 people in a public place to complete a short survey, and 60 are willing to answer your questions, those 60 are likely to be different in significant ways from the 40 who walked by without making eye contact". In these circumstances, inference based on such samples can be heavily biased. The amount of bias may be reduced if proper weight-adjustment procedures or other bias-correction techniques are applied. This, however, requires reliable external information about the subject population. Unfortunately, opt-in surveys are frequently used when the researcher has little or no knowledge about the structure of the population in question. Therefore, in research activities, applications of this sample construction technique should be confined to pilot surveys.

4. Big Data – Hopes and Opportunities

In contemporary data analyses a great deal of attention is paid to external sources of data that aim to extend the sample information on which the inference is based. Several categories of external information can be distinguished, depending on the nature and content, including metadata and paradata.

Metadata is commonly understood as "data about the data", i.e. it summarises basic information about the data collected – its structure, content, and context. Examples of metadata are survey instruments,

interviewer instructions, software used for processing data, and documentation of the survey process. This kind of data was probably first recognised and used by official statistics, and then gradually became common in commercial surveys. Collection of various kinds of metadata has been facilitated in recent years thanks to technological progress. Given the problems with declining response rates in sample surveys, the impact of metadata is on the increase. However, this area has witnessed a lack of theoretical background and common standards for using metadata in survey--based research.

Paradata, on the other hand, refers to more detailed information that is more difficult to record. The term "paradata" describes all types of data about the process and context of data collection. According to F. Kreuter (2015), it was first used in the survey research context by M. Couper in 1998 to describe automatically generated process data, such as data from computerised interviewing software. This kind of data includes contact data (day, time, outcome), keystrokes (response times, edits), interviewer personal observations (respondents' interest). Proper use of this sort of data can help reduce the total error in a survey. For example, an analysis of time the respondent took to "click" and answer questions may suggest how interested he or she was in the survey, and consequently what quality of data they provided. Similarly, in telephone or face-to-face interviews, information about the availability of respondents can prove helpful for the researcher[18]. Despite growing opportunities to collect various kinds of paradata in recent years, its future impact will depend on the anticipated progress in working out a systematic approach to handling this kind of data.

Both metadata and paradata constitute elements of a broader concept known as "big data", a term to describe ways of acquiring new knowledge and learning about reality which can be achieved on a large scale by using new opportunities for capturing and processing large-size data files. Some authors who deal with this subject expect revolutionary changes not only in survey-based research but also in our lives. V. Mayer-Schönberger and K. Cukier in their famous book (2013) argue that "the concept of sampling no longer makes as much sense when we can harness large amounts of data" (p. 26), and that "reaching for a random sample in the age of big data is like clutching at a horse whip in the era of the motor car" (page 31). These

---

[18] For example, in one of the pre-election face-to-face British Social Attitudes surveys in 2015, Labour was six points ahead among respondents who answered the door at the first visit, whereas the Conservatives enjoyed an 11-point advantage among interviewees that required between three and six home visits (see Clark & Parraudin 2016).

statements are, in our opinion, debatable. The authors, who are clearly impressed by the huge amounts of data available, mainly Internet-derived, seem to suggest that quantity can compensate for lack of quality. They are of course aware of the nature of big data, and realise that it may often be unstructured, messy, and possibly of poor quality. Anyone who argues that in such circumstances big data is superior to a well-designed and well conducted sample survey, is paying too much attention to sampling errors and is probably ignoring non-sampling errors. In modern sample surveys the latter are much more difficult to combat than sampling errors. If a survey fails, the principal reasons are likely to be connected with non--sampling errors, which can occur both in sample surveys and in surveys designed to cover the whole population under study. Therefore, one of the main challenges is to seek efficient tools for reducing non-sampling errors. If "sacrificing a little accuracy", to quote (Mayer-Schönberger & Cukier 2013), means ignoring some categories of non-sampling errors prior to analysis, then a well-designed and carefully performed sample survey will presumably provide better quality information about the population of interest.

Substituting a random sample by gathering as much data as possible – "N = all", according to V. Mayer-Schönberger and K. Cukier's concept (2013) – is, in our opinion, also controversial. There are many populations where small fractions may seriously affect the inference if they are not represented in the sample. Even a figure as high as 99% of the population covered in the survey may not be sufficient to make reliable statements about some variables. The slogan "We are the 99%" used by the Occupy Wall Street movement refers to one such variable – global wealth (and the distribution of family wealth in the USA). According to Credit Suisse, the top 1% of adult wealth holders worldwide own more than half of all global wealth. If a researcher fails to cover this one percent of the population, then his or her estimates about the world's wealth will certainly be skewed. Similar examples can be found in Poland, where failure to cover the top 3% of high-income earners in a survey will result in a drop of 25% in the personal income tax they account for (2015).

The use of big data in areas previously occupied by statistical surveys and censuses does not have to be viewed in terms of mutual competition. They are complementary rather than competitive. Big data, and administrative records in particular, may effectively complement sample data. It would be particularly valuable in surveys that are subject to large non-sampling errors, including coverage error, non-response error, and processing error.

In other words, the additional information about the population commonly used in weighting adjustment procedures or calibration techniques may have its origin in big data. We believe that sample surveys supported by big data sources and instruments will play an important role in finance in the near future.

5. Conclusions

Survey-based research has gained acceptance in many sciences, including finance. It is particularly useful if no other possibilities exist for obtaining data that could be employed to generalise findings to a larger (pre-defined) population. However, one can simultaneously observe growing scepticism among some researchers towards the validity and accuracy of surveys which nowadays have to deal with sampling problems and the unwillingness of respondents to cooperate. These and other practical difficulties in conducting surveys mean that auxiliary information has become increasingly desirable. This information can be implemented in weighting schemes and various error correction procedures. Big data is expected to play an important role in handling non-sampling errors. If understood as a broad concept, it involves, among others, metadata and paradata. Both can provide relevant and valuable information about the process and context of data collection. Big data is, in addition, likely to play a supplementary role for sample surveys. Unlike V. Mayer-Schönberger and K. Cukier, we are convinced that big data, despite its huge potential, will not substitute sample surveys in finance and other sciences. Surveys are likely to survive at least into the near future. One of the main challenges in this area, however, is to work out a coherent theoretical framework for applications of combined sources of information in statistical inference.

Bibliography

Baker, H. K. and T. K. Mukherjee (2007) "Survey Research in Finance: Views from Journal Editors". *International Journal of Managerial Finance* 3 (1): 11–25, https://doi.org/10.1108/17439130710721635.

Bethlehem, J. G. (2002) "Weighting Nonresponse Adjustments Based on Auxiliary Information" in R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. Little (eds) *Survey Nonresponse*. New York: Wiley Series in Probability and Statistics.

Brzeziński, M. (2015) "Trzeba patrzeć w PIT-y" [We have to look at personal income tax forms (PITs)]. *Gazeta Wyborcza*, 24 July.

Brzeziński, M. and Kostro, K. (2010) "Income and Consumption Inequality in Poland, 1998–2008". *Bank i Kredyt* 41 (4): 45–72.

Clark, T. and Parraudin F. (2016) "General Election Opinion Poll Failure Down to Not Reaching Tory Voters". *The Guardian*, 19 January.

Couper, M. (1998) "Measuring Survey Quality in a CASIC Environment" in *Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 41–49.

Dillman, D. A., Smyth, J. D. and Christan, L. M. (2009) *Internet, Mail and Mixed-mode Surveys: The Tailored Design Method.* J. Wiley.

Gallup (2013) *Gallup 2012 Presidential Election Polling Review* (2013). The Gallup Organization, 4 June.

Groves, R. M. and Peytcheva, E. (2008) "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis". *Public Opinion Quarterly* 72 (2): 167–89, https://doi.org/10.1093/poq/nfn011.

Gwizdała, J. P. (2013) "The Role of Information in Bank's Credit Risk Management". *Zarządzanie i Finanse / Journal of Management and Finance* 3 (1): 5–17.

Kośny, M. and Mazurek, E. (2009) "Redistribution and Equity of Polish Personal Income Tax: Measurement Using Micro Data from Tax Returns". *Statistica & Applicazioni* 7 (2): 211–21.

Kreuter, F. (2015) "Paradata" in J. A. Krosnick, S. Presser, K. Husbands Fealing, and S. Ruggles (eds) *The Future of Survey Research: Challenges and Opportunities*. Arlington, VA: The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live*, *Work, and Think.* Houghton Mifflin Harcourt.

O'Muircheartaigh, C. (2008) "Sampling" in W. Donsbach and M. W. Traugott (eds) *Public Opinion Research*. Thusands Oaks: Sage.

Särndal, C. E. and Lundström, S. (2006) *Estimation in Surveys with Nonresponse*. New York: John Wiley.

Sokołowska, E. (2014) *Alternative Investments in Wealth Management: A Comprehensive Study of the Central and East European Market*. Springer.

Sturgis, P., Baker, N., Callegaro, M., Fisher, S., Green, J., Jennings, W., Kuha, J., Lauderdale, B. and Smith, P. (eds) (2016) *Report of the Inquiry into the 2015 British General Election Opinion Polls*. London: Market Research Society and British Polling Council.

Szreder, M. (2010) *Metody i techniki sondażowych badań opinii* [Methods and techniques of public opinion polls]. Warsaw: PWE.

Wallgren, A. and Wallgren, B. (2007) *Register-based Statistics: Administrative Data for Statistical Purposes*. New York: John Wiley.

Wheelan, Ch. (2013) *Naked Statistics: Stripping the Dread from the Data*. New York: W. W. Norton.

Worcester, R. M. (1991) *British Public Opinion: A Guide to the History and Methodology of Political Opinion Polling.* London: Basil Blackwell.

Abstract

O możliwym dalszym rozwoju badań próbkowych w finansach

Mimo że badania próbkowe nie stanowią najczęściej podstawowego podejścia w naukach nieeksperymentalnych, to ich znaczenie w ostatnich latach rośnie zarówno w ekonomii, jak i w finansach. Jednocześnie jednak pojawiają się nieznane dotychczas trudności w prowadzeniu badań próbkowych i we wnioskowaniu, którego są podstawą. Są to nie tylko problemy z właściwym operatem losowania i skutecznym sposobem komunikowania się ankietera z respondentem, ale także szybko rosnące wskaźniki odmowy respondentów udziału w tego typu badaniach. W tych warunkach szczególnego znaczenia nabierają informacje spoza próby, mogące poprawić jakość wnioskowania. Artykuł przedstawia punkt widzenia autorów na kwestię wykorzystania możliwości *big data* jako zewnętrznego, wspomagającego próbę źródła informacji. Autorzy prezentują ponadto najważniejsze wyzwania, jakie stawiane są przed badaniami próbkowymi w najbliższej przyszłości. Głównym wnioskiem z rozważań jest przekonanie autorów, że w najbliższej przyszłości *big data* nie zdoła zastąpić badań próbkowych, ale może przyczynić się do poprawy ich jakości.

Słowa kluczowe: badania próbkowe, wnioskowanie statystyczne, badania ankietowe, *big data*.