



RatSWD

Working Paper Series

Working Paper

No. 146

Experiments, Surveys and the Use of Representative Samples as Reference Data

Thomas Siedler and Bettina Sonnenberg

May 2010

Working Paper Series of the German Data Forum (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/2008 Heike Solga; since 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Experiments, Surveys and the Use of Representative Samples as Reference Data

Thomas Siedler and Bettina Sonnenberg

DIW Berlin (tsiedler[at]diw.de, bsonnenberg[at]diw.de)

Abstract

During the last two decades, laboratory experiments have come into increasing prominence and constitute a popular method of research to examine behavioral outcomes and social preferences. However, it has been debated whether results from these experiments can be extrapolated to the real world and whether, for example, sample selection into the experiment might constitute a major shortcoming of this methodology. This note discusses potential benefits of combining experimental methods and representative datasets as a means to overcome some of the limitations of lab experiments. We also outline how large representative surveys can serve as reference data for researchers collecting their own datasets in order to explore potential sample selection biases.

Keywords: experiments, survey, representativity

JEL-Classification: C01, C52, C8, C9, D0, D6, D81, D84

1. Introduction

During the last two decades, laboratory experiments have come into increasing prominence and now constitute a popular method of research to examine behavioral outcomes and social preferences. There are obvious advantages of laboratory experiments. First, researchers can control the environment under which individuals make their decisions and allow causal inferences by exogenously varying one parameter while holding all others constant. Second, the simplicity of many such experiments makes it easy to explain the findings to non-academics and policy makers. However, major limitations of most experiments are that they are administered to students, who usually self-select themselves into the study and are therefore not representative of the entire adult population. In fact, due to self-selection, experimental studies with student subjects might not even be representative of the entire student population. For example, Eckel and Grossman (2000) investigate the impact of recruitment methods on behavior in a series of dictator experiments with a charitable organisation as a recipient in laboratory sessions. The authors compare altruistic behavior among student subjects recruited voluntarily through announcements in graduate and undergraduate courses (“voluntary sample”) with students in which the experiment was conducted during class time (“pseudo-voluntary sample”). They find that pseudo-volunteers are significantly more generous on average than their volunteer counterparts, and that socio-economic characteristics such as religion or survey measures of altruistic preferences have a larger effect on giving behavior among students recruited pseudo-voluntarily. Similarly, Harrison et al. (2007) examine potential self-selection bias in both a field experiment and a laboratory experiment with student subjects. The authors start with the observation that samples observed in the experiment might suffer from randomization bias (Heckman and Smith 1995). Being interested in individuals’ risk attitudes, the authors note that the likelihood to participate in the experiment might be higher for individuals with on average higher risk attitudes than among the general population. On the other hand, the researchers offer participants a fixed show-up fee that might encourage individuals that are more risk-averse to participate in the experiment, potentially outweighing sample selection into the experiment in their study due to randomization bias. The authors report significant self-selection into both the field experiment and the laboratory experiment with adult subjects drawn from the general Dutch population, arguing that their sample is on average more risk-averse than the general population (see also Roe et al. 2009). In addition, most laboratory experiments are conducted on very homogenous samples (typically students studying the same subject at the same university) and often information on potentially important socio-

economic background characteristics is missing or lacks sufficient variation. Another shortcoming of laboratory experiments is the lack of anonymity. In most laboratory studies, students play against each other and know that the other player is a student. Hence, the degree of anonymity is rather low. Both the degree of homogeneity and anonymity in the subject pool might influence revealed social preferences (Sapienza et al. 2007). The question has also been raised whether laboratory experiments are externally valid and to what extent laboratory findings can be extrapolated to the general population (Levitt and List 2007). A branch of the recent literature examines the external validity of laboratory experiments by comparing behavior in laboratory sessions with experimental outcomes in more heterogeneous and representative samples (Bosch-Domenech et al. 2002; Haigh and List 2005; Benz and Meier 2006). The majority of these studies report that the behavior in the lab differs from that observed in other contexts. For a detailed discussion of potential limitations of laboratory experiments measuring social preferences, see Levitt and List (2007). For a recent discussion regarding potential improvements and future challenges in the field of experimental economics, see Gächter (2009).

Another strand of research in economics and the social sciences makes use of survey questions from large representative cross-sectional or household panel datasets. One criticism of using attitudinal questions from these surveys concerns the lack of behavioral underpinnings and the absence of meaningful survey questions in certain contexts. For example, Glaeser et al. (2000) and Ermisch et al. (2009) discuss the difficulties of measuring respondents' trustworthiness by means of survey questions. Combining attitudinal survey questions with behavioral experiments that include monetary rewards can potentially provide a fuller understanding of economic behavior and help to overcome some of these shortcomings. This note briefly discusses potential benefits of combining experimental methods and representative datasets when studying economic outcomes and social behavior. We also provide a short overview about the recent literature combining the experimental approach with survey methods. Finally, we discuss potential benefits of using large representative surveys as reference data for researchers collecting their own datasets. An overview of recent selected studies combining behavioral experiments with survey questions or using representative surveys as reference datasets is provided in table 1.

2. Combining behavioral experiments and survey methods

2.1 Trust and trustworthiness

A new research strand combines behavioral experiments and survey methods. Fehr et al. (2002) incorporate the standard trust-game experiment (Berg et al. 1995) into a representative survey of the German population and asked respondents several survey measures of trust. Fehr et al. (2002) find a positive association between attitudinal survey measures of trust and sender's behavior, but no significant correlation between survey-based measures of trust and trustworthiness in the experiment. In addition, the authors report that individuals aged 65 and above, highly skilled workers, and those living in larger households exhibit less trusting behavior in the experiment.

Using nationally representative data for Germany, Naef and Schupp (2009) compare survey and behavioral measures of trust. The authors create a new survey measure of trust and find that it is significantly correlated with the experimental trust measure. Moreover, they report that their experimental measure of trust is not subject to a social desirability bias and is robust to variations in stakes and the use of strategy method. This study demonstrates how survey measures can be tested by combining the experimental approach with survey methods.

In a representative sample of the Dutch population, Bellemare and Kröger (2007) measure levels of trust and trustworthiness elicited through an experiment similar to those presented by Berg et al. (1995) in a representative sample of the Dutch population. The authors also compare their representative trust experiment with a sample of college students in an equivalent laboratory experiment. They find that college students have considerably lower levels of trust and trustworthiness than individuals in the representative sample and that these differences can be explained mainly by differences in socio-economic and background characteristics, in particular age, gender, and education. For example, the authors find that women have higher levels of trust than men, but display lower levels of trustworthiness. In line with Fehr et al. (2002), Bellemare and Kröger (2007) find a positive, inverted U-shaped association between age and trust. The authors do not find evidence of a participation bias in their trust experiment with student subjects, and therefore argue that trust and trustworthiness as measured in the laboratory are informative about the behavior in the general population.

Ermisch et al. (2009) integrate a new experimental trust design into a sample of the British population. The authors' rationale for using an alternative trust design is based on observations that the sender's behavior in the standard trust-game experiment (Berg et al. 1995) is not only influenced by trust but also depends on other motivations such as sender's reciprocity, risk aversion, altruism, or inequality aversion (Cox 2004; Karlan 2005; Ermisch

and Gambetta 2006 and Sapienza et al. 2007). In their “one-shot” trust experiment, the sender faces the decision as to whether or not to pass on a fixed amount of money (e.g., whether or not to send £10. If £10 are sent, the experimenter increases it by £30 so that the second person receives £40) and the receiver must decide whether or not to pay back a fixed amount of money (e.g., the sender has the choice of either paying back £22 or keeping all £40). Thus, the players cannot choose whether or not to transfer a certain amount of money between, say, £1-£10; rather they face the decision whether to transfer the entire amount or nothing. The authors argue that this binary trust game is more likely to measure revealed trust and trustworthiness than the standard trust game experiment, in which the possibility of sending “any amount favours the intrusion of other motives such as ‘gift giving’, ‘let’s risk part of it’, ‘I like to gamble’.” Ermisch et al. (2009) find that the experiment is more likely to reveal trust if people are older, if they are homeowners, if their financial situation is “comfortable,” or if they are divorced or separated. Trustworthiness is lower if a person’s financial situation is perceived by them as difficult or as “just getting by.”

2.2 Risk attitudes

Another recent example demonstrating the benefits of combining incentive-compatible experimental measures with survey methods is the study by Dohmen et al. (2009). In a previous related study, Dohmen et al. (2007) examine the relationship between individual’s risk aversion, impatience, and cognitive abilities. They find that lower cognitive abilities are significantly associated with greater risk aversion and more pronounced impatience. These relationships are found to be robust to controlling for a broad set of socio-economic characteristics, such as age, gender, education, and income, which are measured through standard survey questions. In their study, both risk aversion and impatience are measured by choice experiments that involve real monetary choices and relatively large stakes. Respondents were told in advance that the experiment was about financial decisions, that they would have the chance to win money, and that the earned amount would depend on their choices in the experiment. Subjects were also informed that every seventh participant would win. For instance, in the lottery experiment, a financial decision is represented by the choice between a certain payoff (*Option A*) and a risky lottery (*Option B*). Participants were also informed that, for each paired lottery, *Option B* always implies a 50 percent chance of winning €300 and a 50 percent chance of winning nothing. The experiment starts with the following lottery choice: respondents can choose between a certain payoff of €0 (*Option A*) and *Option B*. If participants choose *Option B*, the amount of *Option A* is increased by €10 in

the next decision round. Thus, the second lottery choice is between the “safe” payoff of €10 and Option *B*. Similarly, conditional on prior decisions, a third lottery choice is between a certain payoff of €20 and Option *B*. The experiment ends when subjects choose Option *A* for the first time, or when the maximum amount of €190 for Option *A* is reached. This study is another example demonstrating the potential benefits of combining experimental and survey measures in a representative sample of the population.

3. Using representative surveys as reference data

In this section, we briefly discuss potential benefits of using large representative surveys as reference datasets for researchers collecting their own data. Household panels might offer a useful reference point for experimental studies, thanks to their longitudinal character and the sampling of all household members – for example, the British Household Panel Study (BHPS), the new household panel *Understanding Society* in the United Kingdom, and the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*). Register data can constitute another fruitful source of reference data (Harrison et al. 2007). The basic idea here is that large representative surveys can serve as reference data for researchers collecting datasets that do not represent the full universe of the population of interest (e.g., through clinical trials, intervention studies, laboratory and behavioral experiments, and cohort studies). An important issue when investigators collect their own data is whether the sample represents the general population, or conversely, whether it is selective (for example, by design or through choice-based sampling). This approach might offer several benefits. First, by asking participants similar questions to those in representative surveys, researchers can compare their sample with either a sub-sample or the whole representative survey. Second, in contrast to many of the scales and questionnaire instruments developed by psychologists, for instance, questions in household panel surveys like the SOEP or BHPS are not copyrighted and can be used by other researchers free of charge. Thus, these datasets can be a valuable point of reference for designing new questionnaires. Combining experimental sessions with a questionnaire collecting basic individuals’ socio-demographic characteristics used in representative surveys gives researchers valid information as to the representativeness of their sample with respect to the individual characteristics surveyed.

Two recent studies exemplify the potential for using questions from a panel survey when researchers collect their own data. In Germany, Geyer et al. (2009) examine whether individuals aged 17-45 with operated congenital heart disease have adverse employment

chances compared to people without heart problems. The authors compare their sample of patients (N=314; treatment group) with a sample drawn from the SOEP, which serves as a comparison group. The treatment group consisted of women and men who had a congenital heart disease and were operated on at the University Hospital of Göttingen. The authors conducted a face-to-face interview with patients using several SOEP questions. Comparing their hospital sample with the SOEP as reference data they found considerable differences between the two samples with respect to gender, age, and employment status.

Two recent projects that also follow the idea of using a representative household panel study (SOEP) as reference data are the Berlin Aging Study II and the Brain Gene Behavior Project. The Berlin Aging Study II, collecting data on objective socio-economic and biological characteristics like objective health, functional capacity, subjective health, and well-being, draws on SOEP questions with regard to health and life satisfaction to enable comparisons with the SOEP data (Max Planck Institute for Human Development 2009). Likewise, the Brain Gene Behavior Project, a large-scale study on the molecular genetic basis for personality, cognitive, and individual behavioral differences, makes use of the SOEP questionnaire to exploit comparable reference data (Neuroeconomics Lab Bonn and Socio-Economic Panel 2009).

In the United Kingdom the study by Ermisch et al. (2009) demonstrates how a panel survey can help in determining the extent to which a particular sample is representative of the general population. The authors integrate a new experimental trust design into a former sample of the British population and compare their trust sample with a sample from the BHPS. By using a questionnaire similar to the BHPS, the authors are able to determine that their trust sample over-represents women, people who are retired, older, divorced, or separated. Together, these two studies show that household panel studies can serve as useful reference data for researchers collecting their own samples and can help to reveal the representativeness of their own collected data.

4. Conclusion

The studies reviewed demonstrate that enormous academic benefits can be derived from combining experimental studies with representative surveys.¹ First, experiments based on representative samples help to assess potential biases of studies based on student subjects who self-select themselves into the sample. This advances our knowledge on whether and to what

¹ See also Falk et al. (2009).

extent experimental studies on student samples can be generalized. Second, research measuring both revealed preferences and stated preferences allows researchers to validate their measures. For example, Fehr et al. (2002), Ermisch et al. (2009), and Naef and Schupp (2009) report that answers to attitudinal questions on trust toward strangers do predict real trusting behavior in the experiment.

The recent studies by Eckel and Grossman (2000) and Roe et al. (2009) demonstrate the importance of self-selection into experimental studies, and their studies suggest that results from laboratory experiments might not be generalized to the entire population. In this note, we briefly discussed potential benefits of using large representative survey as reference data for researchers who are collecting their own datasets and point readers to two recent examples in the literature.

Table 1: Studies Combining an Experimental Design with Survey Methods

Author(s)	Topic	Method	Data	Finding
Bellemare and Kröger 2007	Measure levels of trust and trustworthiness elicited through an experiment similar to those presented by Berg et al. (1995) in a representative sample of the Dutch population.	Trust and trustworthiness measured by an invest- and-reward experiment.	Representative sample of the Dutch population and a laboratory sample with college students.	The smaller amount of students' investments predominantly demonstrates differences in socio-economic and background characteristics. While these characteristics can explain different revealed behavior, they have almost no impact on stated trust. Return ratios are significantly lower in the lab sample as well.
Benz and Meier 2006	Explore the correlation between individual behavior in laboratory experiments and in a similar situation in the field.	Donation lab experiments with college students.	Secret use of the real donation spending behavior of the students.	The authors find a rather moderate or weak relationship between lab and field behavior.
Dohmen et al. 2009	Investigate the relevance of survey questions on risk-taking behavior in field experiments and actual behavior in the real world.	Risk-taking measured by a lottery game in a field experiment and SOEP survey questions with a representative sample of 450 participants.	Comparison with representative data of the whole SOEP sample on seven different survey questions with regard to risk attitudes.	The general risk attitude survey questions is significantly correlated with behavior in the lottery game as well as with actual behavior in the real world, e.g., with regard to financial, sports, and health-related behavior. Simultaneously, specific behavior is best predicted by context-specific risk survey measures in the respective domain.
Eckel and Grossman 2000	Compare the effect of recruitment method in dictator experiments with student subjects.	Altruism measured by means of dictator games.	Laboratory experiment with self-recruited students (voluntary sample) and in classroom recruited college students during the class period (pseudo-voluntary sample).	Volunteers are less generous in distributing endowments and are more motivated by incentives than classroom-recruited students. Respondents' characteristics such as sex, religion, and altruism influence the behavior of pseudo-volunteers more than that of volunteers. The authors conclude that self-selection into the sample matters.

Author(s)	Topic	Method	Data	Finding
Ermisch et al. 2009	Measure trust and trustworthiness in Britain using an experimental and survey design.	One-shot trust experiment with former respondents of the BHPS in combination with survey questionnaires.	Comparison with representative BHPS sample allows the authors to examine whether their experimental sample is representative of the general population	For example, the authors report that their experimental sample over-represents women, people who are retired, divorced, or separated. Individual behavior in experiments is found to be a reliable and superior measure compared to standard common trust survey questions.
Fehr et al. 2002	Investigate trust and trustworthiness by comparing behavioral experimental outcomes and representative survey data.	Implementation of a trust experiment in a representative survey of the German population in 2002.	—	Trust in strangers and past trusting behavior correlate with trust behavior in the experiment, but no survey measure predicts trustworthiness.
Gächter et al. 2004	The authors present survey and experimental evidence on trust and voluntary cooperation in Russia using both a student and a non-student sample.	One-shot public goods experiment.	Not fully representative survey data of Russian non-students and a student subject pool.	Non-students display higher levels of trust than students, and also contribute more to the public good as long as socio-economic background is not controlled for. Individuals who believe that most other people are fair contribute significantly more to the public good game than those without such beliefs. Likewise, optimists make higher contributions than pessimists.
Geyer et al. 2009	Examine the effect of congenital heart disease on employment status.	Sample of 628 patients surveyed in clinic combined with medical check-up (treatment group).	Their comparison group is 10 percent sample drawn from the German SOEP	The authors find significant differences between male patients and male control subjects. Those with congenital heart disease are less likely to be employed full-time, more likely to be employed part-time, and in marginal employment. The differences between treatment and control group depend on the severity of the disease.
Glaeser et al. 2000	Examine the validity of trust survey questions with a behavioral trust experiment.	Laboratory experiment with Harvard undergraduates.	Survey measures on trust (self-reported attitudes and behavior) of 258 Harvard undergraduates.	Ten out of 12 GSS trust questions do not predict trust, but are related to trustworthiness as measured in the experiment. Trust in the experiment is associated with past trusting behavior.

Author(s)	Topic	Method	Data	Finding
				Trust and trustworthiness rise with closer social distance.
Harrison et al. 2007	Investigate whether experiment samples are biased because of the risk of randomization. The authors undertake both a laboratory experiment and a field experiment to examine whether selection into the experiment influences measures of risk attitudes.	Eliciting individual risk attitudes through an experimental lottery game in both a field experiment and a laboratory experiment.	First, the authors collect information on subjects' socio-economic characteristics by means of questionnaires and use this information to correct for potential self-selection into the field experiment. Second, in their laboratory experiment, they investigate the impact of variation in recruitment information on individual risk attitudes.	The authors find that the use of show-up fees generates a more risk-averse sample. Participants in both the field and laboratory experiment are found to be more risk-averse than the general population once they control for selection into the experiment.
Levitt and List 2007	Discuss whether estimates on pro-social behavior from laboratory experiments can be extrapolated to the real world.	Literature review.		The authors argue that pro-social behavior in experiments depends on a number of experimental situation and design factors, e.g., stakes, sample recruitment, anonymity, as well as unobserved respondents' characteristics. They caution against generalizing results from laboratory to real-world situations.
Naef and Schupp 2009	Test the correlation and validity of trust survey questions with experimental measures of trust.	Trust experiment with survey respondents, representative for Germany.	Self-reported trust and trustworthiness by different measurements with an representative survey sample	GSS Survey question do not measure trust in the experiment. However, the authors find a significant correlation between self-reported SOEP trust measures and experimental measures of trust. Students are found to be slightly more trustful than non-students.

References:

- Bosch-Domenech, A./Montalvo, J.G./Nagel, R. and Satorra, A. (2002): One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review* 92 (5), 1687–1701.
- Bellemare, C. and Kroeger, S. (2007): On representative social capital. *European Economic Review* 51, 183-202.
- Benz, M. and Meier, S. (2006): Do people behave in experiments as in the field? Evidence from donations. Working Paper No. 248, Institute for Empirical Research in Economics, University of Zurich.
- Berg, J./Dickhaut, J. and McCabe, K. (1995): Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122-142.
- Cox, J.C. (2004): How to identify trust and reciprocity. *Games and Economic Behavior* 46, 260-281.
- Dhali, M.K./Hertwig, R. and Hoffrage, U. (2004): The Role of Representative Design in an Ecological Approach to Cognition. *Psychological Bulletin* 130 (6), 959-988.
- Dohmen, T./Falk, A./Huffman, D. and Sunde, U. (2007): Are risk aversion and impatience related to cognitive ability? IZA discussion paper No. 2735. [Forthcoming in the *American Economic Review*].
- Dohmen, T./Falk, A./Huffman, D./Sunde, U./Schupp, J. and Wagner, G.G. (2009): Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. [Forthcoming in *Journal of the European Economic Association*].
- Eckel, C.E. and Grossman, P.J. (2000): Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics* 3, 107-120.
- Ermisch, J. and Gambetta, D. (2006): People's trust: the design of a survey-based experiment. ISER Working Paper 34/2006.
- Ermisch, J./Gambetta, D./Laurie, H./Siedler, T. and Uhrig, N. (2009): Measuring people's trust. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (4), 749-769.
- Falk, A./Dohmen, T. and Sunde, U. (2009): Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten. *Perspektiven der Wirtschaftspolitik* 10 (1), 54-74.
- Fehr, E./Fischerbach, U./von Rosenblatt, B./Schupp, J. and Wagner, G.G. (2002): A Nation-Wide Laboratory. *Schmollers Jahrbuch* 122 (4), 519-542.
- Gächter, S./Herrmann, B. and Thöni, C. (2004): Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *Journal of Behavior & Organisation* 55, 505-531.
- Gächter, S. (2009): *Experimental Economics*. [In this volume].
- Geyer, S./Norozi, K./Buchhorn, R. and Wessel, A. (2009): Chances of employment in a population of women and men after surgery of congenital heart disease: Gender-specific comparisons between patients and the general population. *Congenital Heart Disease* 4 (1), 25-33.
- Glaeser E.L./Laibson, D.I./Scheinkman, J.A. and Soutter, C.L. (2000): Measuring trust. *Quarterly Journal of Economics* 115, 811-846.