Respondent Behavior and Data Quality Aspects in Panel Surveys

Four Empirical Contributions

DISSERTATION

zur Erlangung der Würde

eines Doktors der Staatswissenschaften

vorgelegt der
Wirtschaftswissenschaftlichen Fakultät
der Universität Basel

von

Oliver Serfling
aus
Deutschland

Eigenverlag, Wiesbaden 2006

Genehmigt von der Wirtschaftswissenschaftlichen Fakultät der Universität Basel auf Antrag von Prof. Regina T. Riphahn, Ph.D. und Prof. Dr. George W. Sheldon.

Basel, den 28. Oktober 2005

Der Dekan

Prof. Dr. Heinz Zimmermann

Contents

I. Introduction

Household surveys have become a widespread source of data for empirical research across various disciplines. With the dispersion of household survey data, the growing utilization of statistical methods across scientific disciplines, and the increase in fields of application for data on the micro-social level, questions on data quality have risen. Correspondingly, the reliability of the results of empirical studies based upon survey data came into the focus of empirical research. A large and growing literature on measurement errors in surveys emerged. One part of this literature discusses the reliability of survey data and develops techniques to reduce the impact of data deficiencies for a given dataset, e.g. through imputation of missing values. Since the collected data and their quality are the result of an interview process, another part of the literature focuses on the behavior of interview participants and the conditions under which an interview takes place. A better understanding of the data collection process may help to circumvent data deficiencies before they occur.

In a similar vein, this doctoral thesis aims to further the understanding of respondent's behavior in survey interviews and considers various aspects of the interaction of respondents and interviewers in panel surveys with respect to data quality, in particular missings due to item nonresponse, panel attrition and rounding in self-reported income figures.

In most survey studies, respondents act on a voluntary base. When they are asked to participate in a survey interview they are free in the decision to accept or reject this request. If they decide to participate, they are free to decide whether to answer the survey questions. Should they decide to answer, they are free on how to answer, truthfully or bogusly, exactly or imprecisely. In reiterated survey interviews, they are free to decide to be re-interviewed or not. These decisions are taken on various levels of survey cooperation and by different and – with the exception of the initial participation request - self-selected groups of persons. The four studies presented here aim to deepen the understanding of the influences on such decisions at the various stages. Which influences can be attributed to the observable characteristics of the respondent, the interviewer, the interactions between interviewer and respondent, the interview mode, and the interview situation? To what extent can the implications of theories on human behavior from various disciplines of the social sciences be corroborated statistically? The motivation of research on this field of human behavior stems lastly from the insight that this behavior has a large influence on the quality and reliability of survey data, and has the potential to interfere with analyses based upon such data. Furthermore, the understanding of the interaction processes during the interview request by the survey organization, and during the interview itself may help to improve the data collection process. From the viewpoint of the survey organization, this may on the one hand improve the cost-efficiency of data collection, and may on the other hand increase returns, since data which meet high quality standards are likely to have more success in the market.

Within the range of data deficiency concerns, some attracted more attention than others across the social sciences. Sociologists and psychologists are mainly concerned with the influence of the interviewer on subjective statements of the respondent. In the political sciences, the role of filter questions, which may have an impact on the opinion statement of the respondent, garners more attention.

For economists, the most important item in the above mentioned household panel databases is income. Be it in the form of yearly or monthly payments, net or gross income, earnings or subsidies, at the household or individual level, this item suffers strongly from survey measurement errors, and its reliability is in the focus of quality endeavors of data collectors and a permanent concern of data users. Asset and wealth items share both, the attention of economists as well as data quality concerns, and suffer even more from misreporting errors. This may also be one reason why they are less often surveyed than income. Income and wealth items are therefore the subject of investigation of the following chapters. The empirical contributions within this dissertation provide evidence of the influential effects of respondent- and interviewer-characteristics and their interactions on various outcomes of the interview: the first essay studies the determinants of item nonresponse. The second study introduces questionnaire nonresponse as a new category and examines its determinants and interactions with item and unit nonresponse. The third study is motivated by the findings of the previous short study and investigates the interplay between item and unit nonresponse in greater detail. The fourth paper lastly deals with the rounding behavior of respondents concerning their income statements.

Organization of this dissertation

The first essay “Item Nonresponse on Income and Wealth Questions” investigates the mechanisms determining the item refusal of survey participants. Three issues are addressed: the first asks if there are significant differences in the patterns of nonresponse across financial items. The results show large item fixed-effects and systematic differences in the response patterns between income and wealth questions.

The second issue evaluates whether the social distance between interviewer and respondent may have an effect on the occurrence of item nonresponse. If so, the survey organization could improve the informational value of surveys by matching interviewers and respondents based on their observable characteristics. The results show that the interaction between the gender of the interviewer and the respondent impacts on the probability of item nonresponse, in the way that women have a higher tendency not to respond to income questions and female interviewers induce more item nonresponses on income and wealth questions. Also significant effects of the age-difference are found, but the rest of the observable interviewer-interviewee interaction was found to be negligible.

The third issue examines whether offering a "don't know" answer option affects respondent behavior. The results show that "don't know" statements result from response mechanisms that differ from those yielding informative responses and those yielding nonresponse. This finding is unsatisfactory from the view of survey design, since guidance as to whether the questionnaire should include a "don't know" option, could not be provided by our results.

The second study cursorily presents some empirical findings on respondents' behavior with respect to the interaction of item-, questionnaire-, and unit-nonresponse. Questionnaire nonresponse, i.e. respondents' selective response to single questionnaires of a multi-questionnaire survey, is introduced as a new outcome of respondents' behavior. This analysis focuses on income and wealth items from the household questionnaires of two subsequent panel waves of the German Socio-Economic Panel (GSOEP). In the panel wave of 1988 households wealth was surveyed by the GSOEP as a separate questionnaire. The findings show that neither item nor questionnaire nonresponse are significantly correlated with subsequent unit nonresponse. In turn, a negative correlation between item and questionnaire nonresponse is found. This result is contradictory to the general wisdom of survey research that the several nonresponse types stem from the same decision process and should be positively related. The next research question addressed is, whether unit nonresponse is selective with regard to item nonresponse. There is some indication for endogenous sample selection with regard to item nonresponse behavior of wealth questions, but none for the income questions.

Motivated by these unexpected findings of this short study, I re-examine the interplay of unit and item nonresponse in surveys in greater detail and with a larger sample in the third contribution of this doctoral thesis. While the second paper focuses on respondent behavior in the household questionnaires of only two panel waves, the third study uses information on response behavior of all original households and participants of the first GSOEP wave in 1984, observed over all 19 panel waves until 2003. The question whether item and unit nonresponse may result from the same decision process is pursued. In that case, both types should be positively correlated and the sample selection due to unit nonresponse (or panel attrition) should be endogenous with respect to item nonresponse. The study introduces the hypothesis that respondents may behave according to two opposing types of a cooperation continuum, which may lead to the balancing of the selectivity of such drop-outs. Understanding the relationship between several types of nonresponse may permit the development of techniques that jointly reduce item and unit nonresponse. And it furthers the understanding of motivation and cooperation processes of respondents and the interaction with their environmental situation.

It is found that item nonresponse, measured by item nonresponse on the income question (income INR) and by the score of total item nonresponses during the interview (INR rate) is weakly, but mostly significantly, associated with subsequent unit nonresponse. Nonetheless, the iteration of such drop-out processes after each panel wave does not increase the selectivity of the sample with respect to item nonresponse, which may be explicable by the above outlined hypothesis. The preliminary results of the previous short study are partly confirmed.

Furthermore the determinants of natural panel mortality, e.g. death, sickness, or removal, are investigated, which should - by definition - not be related to the same decision process as interview refusal. Nonetheless, the same response patterns are found, which in turn are interpreted that a better part of "natural panel mortality" is active interview refusal, communicated to the interviewer or survey organization by a "policy of closed doors".

The last study within this dissertation provides a novel perspective on the quality of data provided within a survey: it concerns the rounding behavior of respondents by the provision of income information. The rounding error of data may have an impact on the estimates within empirical analyses relying on such data. Moreover, rounding may reflect the motivation of the respondent towards survey participation and may be a precursor of subsequent nonresponse.

Even if this research is limited by the possibilities of the used dataset, i.e. raw survey data from the Swiss Household Panel Study (SHP), the results show that rounding does not occur at random, but is explicable by cost/benefit considerations of the respondent. Respondents' sex, age, and health status are found to be influential regarding the magnitude of rounding. Also the experience of the interviewer determines the precision of the income statement of the respondent. Furthermore, it is shown that several common assumptions about the measurement error of survey data are likely to be violated: the rounding error is correlated with the provided income amount, autocorrelated with itself over time, and correlated with observable characteristics of the respondent. Additionally, the rounding behavior is found to be weakly correlated with subsequent "don't know" statements and item nonresponse.

From a methodological point of view, this study develops informally two approaches to test for the ordinality of the outcomes of categorical variables with respect to a given behavioral model. If the categories of a dependent variable are spuriously ordered, this may impact the reliability of regression coefficients within ordinal regression models, like ordered probit or ordered logit. The hitherto existing literature has not yet discussed the problem of spurious ordered categories, since in most econometric applications the ordinality of categories is obvious, or seems to be obvious. The two test approaches developed in this study, are based on the stereotype regression model and on the predicted latent variable in the ordered probit framework. As an indication of robustness, it is shown that both test procedures lead mostly to the same results with respect of the correct ordering of categories.

Authorship and Publications

The first article is co-authored with Regina T. Riphahn. A short conference-paper version of it has been published as "Heterogeneity in Item Nonresponse on Income and Wealth Questions" in *Schmollers Jahrbuch - Journal of the Applied Social Science Studies*, 123 (1), 2003. A longer version is forthcoming as "Item Nonresponse on Income and Wealth Questions" in *Empirical Economics*, 2005.

The second, third and fourth article are single authored. The second paper was published as "The interaction between Item, Questionnaire and Unit Nonresponse in the German SOEP" in *Schmollers Jahrbuch - Journal of the Applied Social Science Studies*, 125 (1), 2005. The third paper has been revised and resubmitted to the *Journal of the Royal Statistical Society*, London.

Acknowledgements

I acknowledge with gratitude generous help from my supervisor Regina T. Riphahn, as well as George Sheldon, Jörg-Peter Schräpler, a number of anonymous referees, participants of the Workshop on Item Nonresponse and Data Quality in Large Social Surveys in Basel, October 2003, participants at the GSOEP Users Conferences 2002 and 2004 in Berlin, participants of the Workshop on Data Quality the Royal Statistical Society in London for very helpful comments, Erwin Zimmermann with his team from the Swiss Household

Panel Study in Neuchâtel, especially Alexis Gabadinho and Annette Scherpenzeel, my dear colleagues at the Faculty of Economics and Business Administration at the University of Basel, especially Iva Cecchin and Reto Scherrer for helping me with the manuscript, Fabienne Romanens, Bret Kragel, Mirjam Schiffer and Stephan Mayer for proofreading, my parents for their lifelong support, and my dear friend Nicole.

II. Item Nonresponse on Income and Wealth Questions

Regina T. Riphahn and Oliver Serfling

Abstract

This study investigates the mechanisms determining item nonresponse focusing on three issues: First, is there significant heterogeneity in item nonresponse across financial questions and in the association of covariates with item nonresponse across outcomes? Second, can the informational value of surveys be improved by matching interviewers and respondents based on their characteristics? Third, how does offering a "don't know" answer option affect respondent behavior? The questions are answered based on detailed survey and interviewer data from the German Socioeconomic Panel using a broad set of income and wealth outcomes. We find considerable heterogeneity in nonresponse across financial items, little explanatory power of interviewer-respondent matches and strong evidence that "don't know" answers result from mechanisms that differ from those yielding valid responses and outright refusals to respond.

Acknowledgements

We thank Jörg-Peter Schräpler for generous support regarding the interviewer data used in this study and numerous participants of the GSOEP 2002 conference, of the 2003 meeting of the Ausschuss für Ökonometrie / Verein für Socialpolitik, and George Sheldon for very helpful comments.

"... the subject of item nonresponse
is badly in need of investigation."
Ferber (1966, p.415)

1 Introduction

Survey data form the basis of much empirical economic research. Accordingly its quality and the various determinants thereof, as well as the implications of data deficiencies deserve the attention of researchers.

Within the range of data problems and quality concerns some garnered more attention in the social sciences than others: the disciplines of sociology and psychology, where interest often focuses on subjective statements, are mainly concerned with effects of the interview situation or interviewer influences:[1] if respondents seek interviewers' respect, their answers may deviate from the truth. Unit nonresponse and sample representativeness are issues raised in the economic literature (cf. Hill and Willis 2001 or Horowitz and Manski 1998). Here also measurement error and recall bias find attention.[2] In contrast, the problem of item nonresponse is largely neglected. This is astounding as the loss of information due to item nonresponse could be even more problematic than respondent dropout from a survey.

Given the typically high rates of item nonresponse on sensitive issues such as income and wealth it is important to learn about the determinants of nonresponse behavior. An understanding of the mechanisms driving item nonresponse may permit the development of techniques to reduce it and thus to substantively increase the value of interviews. It might improve researchers' ability to rigorously deal with nonresponse in their own analyses and finally it may yield important insights to improve imputation procedures for missing data.

This study investigates such mechanisms and addresses the following questions: First, we ask whether the matching of interviewers to respondents affects respondents' willingness to provide information. This could be the case if 'observational closeness' between the interview partners aids in building up the level of trust required to reveal rather private information. If this were the case, survey administrators might be able to improve data quality by carefully pairing interviewers and respondents. Second, we analyze whether offering the option of "don't know" answers in questionnaires helps increase the amount of information provided, and third we investigate whether there is measurable heterogeneity in

[1] For careful discussions of these problems see Esser (1984) or Reinecke (1991).
[2] See the special issues of the Journal of Human Resources (1998.2, 2001.3) and sources cited there.

the response propensity for different types of financial questions. Little evidence exists on these last two issues. If item nonresponse propensities depend on the way the question is posed and differ for different types of financial questions it might be possible to utilize the findings to optimize survey strategies and to improve informational outcomes.[3]

This study adds to the literature in various ways. First, it briefly summarizes theoretical models of nonresponse behavior and surveys the empirical literature on item nonresponse. Second, it exploits excellent data from the German Socioeconomic Panel (GSOEP), which contains information on respondent and interviewer characteristics, thus permitting research on the relevance of interviewer-respondent matches. Finally, we extend the scarce literature on item nonresponse which focused on income measures, by considering item nonresponse for a variety of income and wealth outcomes.

We find significant heterogeneity in item nonresponse across financial questions. Our data reveal that there is not much to be gained for the informational value of surveys by matching interviewers and respondents, once age and gender effects are controlled for. Our third result with respect to "don't know" answer options is that the observed and unobserved characteristics of "don't know" respondents are neither close to those providing informative answers nor to those refusing to respond. Therefore the "don't know" responses should be considered as a separate group. Simple statements as to whether offering a "don't know" answer option takes away from valid answers or from nonresponses are not feasible. This also implies that missing data imputations should use different procedures depending on whether missing values derive from "don't know" answers or nonresponses.

The paper next reviews theoretical approaches, hypotheses, and empirical findings regarding item nonresponse behavior. Section 3 discusses our hypotheses before we describe data and empirical strategy in section 4. Section 5 presents the findings and section 6 concludes.

2 The Phenomenon of Item Nonresponse

2.1 Theoretical Frameworks for the Analysis of Item Nonresponse

Respondent behavior in interviews and surveys has been addressed in an interdisciplinary literature. The most prominent explanations of response behavior in the literature are the cognitive model and the rational choice framework.

[3] For a survey of possible procedures see Juster and Smith (1997).

The cognitive model of respondent behavior extends earlier psychological models of thought processes (cf. Lachman et al. 1979) by also taking social aspects of the survey situation into consideration (Sudman et al. 1996). It separates several stages in the process of answering a question: After having heard or read a question, the respondent must interpret it: The issue addressed has to be recognized and understood by the respondent. At this stage problems may arise based on the content of the question or the definitions used by respondent and interviewer. Here face-to-face interviews could be helpful, easing communication and understanding.

In the second cognitive stage the respondent gathers the information. Here the familiarity of the issue matters: Being asked about ones' age imposes less of a cognitive effort than providing the amount of interest earned last year. Complex issues require more of respondents' knowledge, cognitive ability, and willingness to cooperate. When the respondent successfully gathered the information, the survey may impose an answer format (e.g. categories or subjective intensity statements), which may require additional formatting of the answer. In a final stage the respondent adjusts the answer to objectives such as self representation or social desirability. Only after the intended response is "filtered" through these "mental screens", it is provided.

This last stage is at the focus of rational choice theory (Esser 1984), which suggests that the respondent evaluates behavioral alternatives based on their expected consequences and chooses to maximize expected utility. Responding to a question consists of understanding the question, evaluating behavioral alternatives, and choosing the preferred behavior.

The rational choice approach predominates the literature: Hill and Willis (2001, p.418) state that an individual responds if "the act of participation is expected to bring rewards that exceed the cost of participation." The *rewards* may be pecuniary or non-pecuniary such as social acknowledgment or being appreciated by others. Benefits consist of supporting a potentially appreciated cause (e.g. scientific value, public interest) and of avoiding the negative effects of a refusal, such as breaking social norms or violating courtesy towards the interviewer (Schräpler 2001). The *costs* of participation consist of the time it takes to respond, the effort of recalling information, and the emotional experience of going through potentially embarrassing, painful, intrusive, or cognitively difficult interviews. In addition, there might be negative consequences of providing private information from tax authorities, through data abuse, or breach of privacy.

A connected aspect concerns the relevance of trust in the interview situation. If a respondent distrusts an interviewer he is less ready to expend effort to recall the information or to reveal information at all. Hill and Willis (2001) refer to Dillman (1978) who emphasized the relevance of trust, and describe steps to render interviewers more trustworthy. Schräpler (2001) discusses the importance of "confidence building" which involves reducing the social distance between interviewer and respondent over time.

2.2 Detailed Hypotheses

The theoretical frameworks describe individual item nonresponse behavior as determined by the relationship between respondent and interviewer as well as by the costs and benefits of providing an answer. The literature operationalized these aspects by interpreting respondent and interviewer characteristics in the light of their effect on trust, and cost-benefit considerations.

If trust affects response behavior, sending a new interviewer to a given household should generate fewer informative responses than sending a well known interviewer. Thus an interviewer change in a panel survey is hypothesized to increase item nonresponse. Further, the match between interviewer and respondent characteristics may affect respondents' perception of the interviewers' trust-worthiness: We hypothesize that a match in relevant characteristics of interviewer and respondent, e.g. in age, gender, or schooling increases response propensities.[4]

In general we assume that nonresponse propensities increase with the cost and decline with perceived benefits of answering. These costs and benefits will vary with the type of question, the characteristics of the interview partners and the general setting of the interview. As the cognitive ability of a respondent may determine the effort involved in answering a question and as the cognitive ability may be correlated to education, we expect a negative correlation between high education and nonresponse. A factor that might be correlated with the perceived benefit derived from survey participation is the appreciation of public service. Existing studies suggest that this is particularly high among public sector employees, who in turn seem to be more ready to participate and provide information in surveys (e.g. Biewen 2001).

The costs and benefits of an interview might also be affected by the characteristics of the interview situation. One might e.g. take the size of a person's town of residence as an indicator of a general attitude of openness and trust. This is based on evidence showing that individuals refuse to participate in surveys because of fear of crimes and that larger cities often entail a sense of anonymity, where the limits of privacy are guarded more carefully than in rural

[4] While the literature presents evidence regarding characteristics of the interview partner (see e.g. Schräpler 2001, Sousa-Poza and Henneberger 2000) the investigation of the matching effects has been surprisingly neglected.

areas.[5] Similarly, it might be easier for individuals to communicate with an interviewer if they are used to such exchanges. Therefore residents in large households might be more at ease answering questions and providing information. Another characteristic of the interview situation is whether a respondent answers a questionnaire partly in writing as opposed to an oral interview. As it should be easier and less costly to refuse an answer if this does not have to be communicated to the interviewer it is plausible to expect higher item nonresponse in this situation.

2.3 Prior Evidence on Item Nonresponse

Given the focus in the social sciences on social desirability effects and on *unit* nonresponse, evidence on the determinants of *item* nonresponse is scarce. Lillard et al. (1986) investigate the distribution of item nonresponse in the United States' Current Population Survey. The authors distinguish between individuals who refused to respond only to the income question and those who did not answer a number of questions. While the latter represent the lower part of the income distribution, exclusive income nonresponses are more likely at high incomes.

In an interesting analysis of financial information provided in the Health and Retirement Survey (HRS) and the Asset and Health Dynamics Among the Oldest Old Survey (AHEAD) Juster and Smith (1997) investigate follow-up bracket responses. They show that missing data involve nonignorable response bias which in part can be remedied by follow-up brackets.

Similarly, Sousa-Poza and Henneberger (2000) focus on the income question in Swiss telephone interviews. They find that nonresponse probabilities are significantly higher for respondents with low education, among the self-employed and home owners. They investigate the relevance of interviewer - respondent matches and show that similarity in age increases the response probability, that education differences do not affect item nonresponse, and that male interviewers are more successful in eliciting income information than females.

Biewen (2001) compares alternative methods to address item nonresponse based on GSOEP data, and shows that nonresponse of income is highest in the tails of the income distribution. He points out that nonresponse is only weakly associated with personal characteristics and mainly driven by unobservables. The study by Zweimüller (1992) on Austrian data is similar to Biewen (2001). Based on wage equations for women Zweimüller (1992) concludes that selection due to item nonresponse is more important than sample selection bias.

[5] De Maio (1980) found significantly more survey cooperation among rural than among urban dwellers.

Schräpler (2001) focuses on the longitudinal development of item nonresponse for earnings and measures of individual concerns. He shows that those in low social positions tend to withhold income information. Again respondents seem to be much more uncooperative in front of females than males. Schräpler concludes that "with increasing trust the item-nonresponse rate falls off over time." (2001, p.22)

Hill and Willis (2001) evaluate the effectiveness of paying respondents for their time and of enhancing the psychic value of participation: Reassigning the same interviewer to a given respondent strongly increases the propensity to respond. The authors emphasize the importance of respondent engagement and cognitive ease as predictors of survey participation.

The reported evidence yields four main results: (i) item nonresponse on income questions is concentrated in the tails of the distribution, certainly in the lower tail; (ii) there seems to be only little systematic variation in item nonresponse behavior and considerable randomness; (iii) the predictions based on the "cognitive process" model and the "trust" framework find support: Interviewer-respondent matching seems to affect survey success and certain matched characteristics, such as age or sex may improve responses; (iv) the cognitive requirement and the sensitivity of an issue seem to affect respondents' willingness to answer: The cost of a response seems to be higher when difficult, sensitive, or threatening issues are considered.

One limitation of this literature is that almost all studies of item nonresponse investigate merely the income question. If there is heterogeneity in the level of cognitive challenges and item-specific sensitivities across financial outcomes this has been neglected in prior analyses.[6] Also the literature does not investigate the role of framing: If individuals show different responses depending on how a question is formulated, this information is relevant for the design of future surveys.[7] We address these issues below.

3 Empirical Approach

In our model of response behavior we follow a rational choice framework and consider factors discussed in models of item nonresponse. When asked a survey question individual i may respond in J different ways (e.g. provide a valid answer, not respond at all, or -if possible- answer "don't know").[8] In the

[6] Exceptions are Schräpler (2001) who also investigates subjective concerns and Loosveldt et al. (1999) who look at political preferences.

[7] Framing is much discussed in the literature on attitude surveys. Trometer (1996) summarizes the evidence which suggests that offering respondents who are queried about their opinions the option of a "don't know" answer affects responses in important ways.

[8] We consider the event of the interview, the selection of the respondent, and the fact that the individual is in principle willing to respond to the survey as being exogenously given.

framework of a random utility model we can describe utility u resulting for individual i from behavior option j as follows

$$u_{ij} = c_{ij}\ \alpha_{1j} + b_{ij}\ \alpha_{2j} + X_i\ \beta_{1j}\ + W_m\ \beta_{2j} + (X_i * W_m)\ \beta_{3j} + \mu_{ij} \tag{1}$$

where c_{ij} and b_{ij} represent the costs and benefits of answer option j, α and β are coefficients, and μ is random noise. Also, respondent (i) and interviewer (m) characteristics (X, W), and their interactions may affect the utility connected to a given behavioral response. Summarizing the right hand side variables in vector z and the coefficients in vector γ, our random utility model is

$$u_{ij} = z_{ij}\ \gamma_j\ + \mu_{ij} \tag{2}$$

The probability that individual i chooses option j is then

$$\begin{aligned} &Pr\ (option\ j\ is\ chosen \mid c, b, X, W) = Pr\ (u_{ij} > u_{ik} \mid c, b, X, W) \\ &= Pr\ (z_{ij}\ \gamma_j - z_{ik}\ \gamma_k > \mu_{ik} - \mu_{ij} \mid c, b, X, W) \\ &for\ all\ k \neq j,\ k = 1, 2, \ldots J \end{aligned} \tag{3}$$

which must hold jointly for all J-1 alternative options k. Assuming a distribution for $\mu_{ik} - \mu_{ij}$ the resulting cumulative distribution function can be estimated by maximum likelihood.

Within this framework we investigate three issues: First we describe whether item nonresponse rates differ across questions, and study whether such differences are associated with observable and unobservable determinants of item nonresponse.[9] For an intuitive indication of outcome-specific heterogeneity we pool item nonresponse outcomes across questions and test the significance of question specific covariate effects in addition to question specific fixed effects.

Second, we investigate whether the match between interviewer and respondent affects response behavior, by controlling for interactions between interviewer and respondent characteristics. Many authors confirm the relevance of trust and confidence building between interviewer and respondent. We hypothesize that individuals feel more confident reporting financial information to someone of their own characteristics as suggested by Sousa-Poza and Henneberger: "One potential source of nonresponse is the existence of a mis-match between the characteristics of the interviewer and the characteristics of the respondent." (2000, p.83)

Finally, response probabilities might be affected by the way questions are posed. This has been looked at before (cf. Trometer 1996 and sources cited

[9] If e.g. wealth items are considered a more private issue than income, the cost of revealing wealth may exceed that of income and we expect higher nonresponse for wealth. Similarly, if information on wealth is less familiar and difficult to obtain, we expect differences in response based on cognitive ease.

there). However, these studies typically do not focus on financial measures and an analysis of the effect of alternative answer options for financial questions is missing in the literature.[10] Our data contain some questions with the option of answering "don't know", and others without this option. We first describe item nonresponse rates for both and then perform two tests to find out whether response processes yielding "don't know" answers differ from those yielding nonresponses or informative answers.

Both tests are applied within the framework of the multinomial logit estimator. The first tests the assumption underlying this estimator that the disturbances of alternative (answer-) outcomes are uncorrelated. This 'independence of irrelevant alternatives' (IIA) property states that the set of outcome alternatives is correctly specified only if the estimates do not vary with the set of considered outcomes (Hausman and McFadden 1984).[11] This will be investigated applying a Hausman test. If the Hausman test yields that the *unobservable determinants* of the three outcomes - valid response, "don't know" answer, and item nonresponse - are uncorrelated, this provides a first piece of evidence for the independence of the "don't know" answer alternative. If uncorrelatedness is rejected, "don't know" answers are not truly independent outcomes. In that case further investigation into the similarity to response or nonresponse alternatives is required.

The second test looks at whether the *observable determinants* of the three possible outcomes are correlated. The test was suggested by Cramer and Ridder (1991) but had been performed before by Hill (1983). Cramer and Ridder (1991) describe the condition under which a subset of multinomial logit outcomes may be treated as a single state. They assume uncorrelated unobservables and describe a criterion by which one may choose the most parsimonious set of outcomes: If the slope coefficients of two outcome options do not differ significantly, the two options may be combined. We test whether the coefficients for the "don't know" answer option differ from those for the two alternative response behaviors.[12] If the mechanisms determining the choice of a "don't know" answer do not differ significantly from those determining valid answers, these processes are very

[10] Juster and Smith (1997) concentrate on financial measures but focus on the impact of adding follow-up bracket answer options for "don't know" and nonresponse outcomes and for imputation results.

[11] The classic illustration of the IIA property looks at alternative means of public transport. While taxi, train, and bus constitute valid alternatives, a split between red and blue buses most likely violates the IIA assumption: One would assume that the unobservable determinants of the choice between red and blue buses are correlated. We test whether don't know answers are a "red bus" as opposed to being an independent alternative such as the train.

[12] Hill (1983) investigated whether females consider the decision to enter the labor force as an employee as being distinct from the choice to enter the labor force as a family worker. Similar tests were performed by Flinn and Heckman (1983), and Riphahn (1997).

similar and it might well be that offering a "don't know" answer option takes away from valid answers. Similarly we can test explicitly whether the mechanisms leading to "don't know" answers and nonresponses are similar. These tests provide a second indicator as to whether the availability of a "don't know" answer option takes away from valid answers or whether this reply is a substitute for nonresponses. In the former case offering "don't know" answers reduces the informative value of the survey.

4 Data Description

4.1 Dataset and Sample

Our data are taken from the German Socioeconomic Panel (GSOEP). The GSOEP gathers information on German households and individuals periodically adding topical modules to the survey (SOEP Group 2001). Since the 1988 module was devoted to household wealth we evaluate item nonresponse for that year, when 4,814 households with 10,023 individuals were interviewed. Our data are taken from three questionnaires. The individual survey was administered to everybody aged 16 or older, whereas the household and wealth questionnaires were answered by heads of households.[13] We also take advantage of data describing GSOEP interviewers (cf. Schräpler and Wagner 2001), which is matched to respondent records.

The GSOEP applies various interview methods: Individuals can answer questions orally, they can fill in the questionnaire themselves with or without interviewer support, questionnaires may be sent out by mail, or answered via telephone (see Table 1). Generally interviewers are required to perform oral interviews but they may use different formats depending on the situation.

To circumvent language problems, we select German respondents from the GSOEP's nationally representative subsample "A".[14] We disregard observations where the survey was administered other than by meeting the interviewer in person, because our research interest concerns the interaction between interviewer and respondent. We also drop observations where information on the interviewer is missing. Table 2 presents the sample sizes after each selection step. Clearly, conditioning on an interviewer being present is the most stringent sample requirement and induces a loss of between 35 and 25 percent of observations.[15]

---

[13] The GSOEP has no strict definition of the "head of household". Instead it surveys a knowledgeable person for every household and tries to re-interview that same person in subsequent surveys. (Hanefeld 1987)

[14] In addition the GSOEP covered a subsample with an oversample of guestworkers.

[15] Preliminary results (not presented) confirm prior studies in that the presence of an interviewer strongly affects item nonresponse (cf. Lillard et al. 1986, Schräpler 2001). A comparison of

Table 1: Distribution of Interview Formats for three Questionnaires of 1988

Interview Format	Individual Questionnaire	Household Questionnaire	Wealth Module Questionnaire
1 Oral Interview	55.1	70.5	64.6
2 Self administered without interviewer	23.3	13.0	14.4
3 Self administered with interviewer	6.1	2.6	3.4
4 Partly oral, partly self	3.7	2.0	1.7
5 By telephone	0.2	0.5	0.4
6 Through mail	9.3	10.0	10.6
7 Proxy interview	0.1	0.0	0.0
8 Information missing	2.3	1.5	5.0
Total (in percent)	100	100	100
Number of interviews conducted	7,360	3,691	3,483

Source: Own calculations based on GSOEP. Only German respondents from subsample A are considered.

Table 2: Criteria for Sample Selection

	Individual Questionnaire	Household Questionnaire	Wealth Module Questionnaire
Full GSOEP data	10,023	4,814	4,606
... thereof in subsample A	7,481	3,743	3,535
... thereof with German respondent	7,360	3,691	3,483
... thereof with non-missing information on interview type	7,194	3,637	3,310
... thereof with interviewer present during interview (formats 1, 3, 4)	4,775	2,769	2,427
... thereof with non-missing interviewer information	4,744	2,769	2,427

Source: Own calculations based on GSOEP.

4.2 Dependent and Explanatory Variables

Dependent Variables: The financial variables of interest are taken from the individual, household, and wealth questionnaires. Table 3 describes the measures gathered in the individual survey. Due to filtering mechanisms in the questionnaire the sample sizes vary by question.[16] The last column of Table 3 describes the item nonresponse rate for each measure. The rates vary markedly

item nonresponse rates for the pool of all outcome measures yielded a rate of 6.03 percent before selecting on the basis of interviewer presence and of 4.52 percent when conditioning on interviewer presence.

[16] E.g. only those who had indicated employment were asked about labor incomes, or those who were retired could indicate retirement benefits.

between 15 percent for income from self-employment and less than 3 percent for the "13. monthly salary", a common employment benefit in Germany. Averaging across all outcomes, we obtain a nonresponse rate of 5.2 percent for individual income variables.

Table 3: Item Nonresponse Rates for Individual Income Questions

Number of Question [1)]	Type of Income	Number of cases [2)]	Nonresponses N	Share
53.02	Income from self employment [3)]	274	42	15.3%
54.09	Bonus / profit sharing [3)]	106	13	12.3%
54.11	Other benefits [3)]	30	3	10.0%
53.08	General unemployment transfer [3)]	149	14	9.4%
53.09	Means tested unemployment transfer [3)]	47	4	8.5%
44.01	Gross earnings last month	2,546	211	8.3%
53.03	Earnings from other employment [3)]	140	9	6.4%
44.02	Net earnings last month	2,546	135	5.3%
54.03	End of year payment: 14. monthly salary [3)]	52	2	3.8%
53.01	Gross wage [3)]	2,454	91	3.7%
54.01	End of year payment: 13. monthly salary [3)]	676	24	3.6%
54.07	Vacation benefits [3)]	1,501	47	3.1%
54.05	Christmas bonus [3)]	1,149	33	2.9%
53.04	Retirement benefits[3)]	983	26	2.6%
	Total all:	12,653	654	5.2%
	Total for 53.02, 44.01, 44.02, 53.01, 54.07:	9,321	536	5.6%

Notes:
[1)] Question number in individual questionnaire.
[2)] Number of cases indicating receipt of income.
[3)] Average gross monthly amount in the last calendar year. If the respondent was unable to provide exact figure the questionnaire prompted for an approximation.

Based on cognitive ease one might assume that providing last month's earnings should require less effort than last year's average monthly income. However, item nonresponse on the former (questions 44.01 and 44.02 in Table 3) is about twice that for the latter (question 53.01). If it is the sense of privacy that determines the cost of reporting earnings, this outcome may indicate that current earnings are more sensitive than those of the past. Generally regulated payments, such as vacation or retirement transfers seem to involve lower reporting costs - possibly because they are considered as less private - than those that entail information on labor market success (e.g. unemployment benefits, or earnings).

Table 4 describes financial indicators from the household questionnaire. It combines measures as to whether a household has incomes or expenditures of a given type at all, with those specifying amounts. Nonresponse rates are highest and at over thirty percent with respect to interest payment and annuity payments.

Table 4: Item Nonresponse Rates Household Questionnaire

Number of Question [1]	Type of income / expenditure	Number of cases[2]	Nonresponse N	Share
39c	Interest payments (amount last year)	342	126	36.8%
39b	Annuity and interest payments (amount last year)	342	110	32.2%
41	Interest and dividend income (last year)	2,149	312	14.5%
39a	Maintenance expenditures on property (amount last year)	342	43	12.6%
33	General welfare benefits (amount)	70	8	11.4%
34	Special welfare benefits (amount)	70	8	11.4%
42	Monthly household net income (amount)	2,769	84	3.0%
37	Rental or lease incomes (yes / no)	2,769	7	0.3%
38	Rental or lease incomes (amount)	342	6	1.8%
36a	Child benefits (yes / no)	1,721	5	0.3%
36c	Child benefits (amount)	1,048	2	0.2%
32	Welfare receipt (yes / no)	2,769	5	0.2%
	Total all:	14,733	716	4.9%
	Total for 41, 42:	4,918	396	8.1%

Notes:
[1] Question number in household questionnaire
[2] Number of cases eligible to respond to the question.

The wealth questionnaire typically asked whether the household holds a given asset and if so at which value. If the respondent indicated possession of a given item but could not provide the exact amount, the person was first asked to guess and if that failed in most cases answer categories or a "don't know" reply were offered. Column 1 in Table 5 shows that the number of cases for each measure varies depending on the number of households owning the asset. The rates of nonresponse and "don't know" answers differ strongly across items. The highest refusal rates of about 30 percent are observed for questions on stock, bond, and equity ownership, which agrees with the findings of Juster and Smith (1997) for the U.S. Health and Retirement Survey. The "don't know" responses are distributed differently: The highest rates appear for equity (15 percent) and inheritances (16.6 percent). As the value of these items seems difficult to determine "don't know" likely reflects lack of knowledge. This seems less plausible in the case of monthly life insurance payments, where the respondent should be familiar with a figure showing up regularly on bank statements. Here an 11 percent "don't know" rate seems high.

While the nonresponse rates in Table 5 do not differ markedly from those in Tables 3 and 4, the joint share of nonresponse and don't know answers more than doubles these rates. Two factors might explain this difference: Either, offering an answer option "don't know" induces individuals who may have otherwise provided an answer to indicate ignorance. Alternatively, wealth is either more sensitive than income or it is more difficult to know the correct answer.

Explanatory Variables: Equation 1 describes individual response behavior as determined by the costs and benefits of providing an answer, the characteristics of respondent and interviewer, as well as their interactions. Clearly it is not possible to actually measure individual costs and benefits in answering a given question. Therefore the characteristics of respondents and interviewers are interpreted in the light of their effect on cost and benefit considerations. More detailed hypotheses were discussed in section 2.2 above. The indicators considered in our item nonresponse model are described in Table 6[17]. We control for characteristics of the respondent-interviewer match, such as equal labor market status and schooling, for the age difference, as well as for the gender combination between interviewer and respondent. We also measure whether a household's interviewer has changed since the last survey, which should increase item nonresponse. The remaining covariates were chosen as indicators of relevant costs and benefits in an interview situation. Education, as indicator of a respondent's cognitive ability, is measured using three categorical indicators. We consider an indicator for whether respondents work in the public sector, control for household size, the size of a person's town of residence and for whether a respondent answers a questionnaire partly in writing as opposed to orally.

[17] We do not consider interview duration, discussed as a potential cost factor in the literature (see Hill and Willis 2001), since interview duration might be a function of nonresponse behaviour, it may not be the same person answering different parts of the interview, and it is difficult to determine the relevant duration indicator, as different questionnaires could be considered separately.

Table 5: Item Nonresponse Rates for Household Wealth Questions

Number of Question[1]	Type of Asset[2]	Number of cases	Nonresponse N	Nonresponse Share	"Don´t know" N	"Don´t know" Share	Total N	Total Share
4	Equity in a business	164	43	26.2%	25	15.2%	68	41.5%
5c	Stocks and bonds	636	217	34.1%	27	4.2%	244	38.4%
5b	Home loan savings certificates (*Bausparvertrag*)	1,001	150	15.0%	82	8.2%	232	23.2%
11	Inheritances since 1960	392	23	5.9%	65	16.6%	88	22.4%
3	Farm[3]	62	12	19.4%	-	-	12	19.4%
8d	Life Insurance: Current monthly payment	1,330	34	2.6%	149	11.2%	183	13.8%
8c	Life Insurance: Originally insured amount	1,330	14	1.1%	141	10.6%	155	11.7%
2	Property other than occupied flat or home	306	6	2.0%	20	6.5%	26	8.5%
5a	Savings account	2,064	70	3.4%	97	4.7%	167	8.1%
1	Owned home: Market value	1,065	8	0.8%	74	6.9%	82	7.7%
10	Total household wealth	2,427	32	1.3%	124	5.1%	156	6.4%
9	Household debt	771	7	0.9%	25	3.2%	32	4.2%
	Total all:	12,613	631	5.0%	1,049	8.3%	1,680	13.3%
	Total for 1, 5a, 5b, 5c, 10:	7,193	477	6.6%	404	5.6%	881	12.2%

Notes:
[1] Question number in wealth questionnaire
[2] The survey first posed yes / no questions as to whether the household owns a given asset. Then the respondent was prompted for the exact amount held in this type of asset, or for an estimate. If that was not provided, response categories including the "don't know" option were provided. Nonresponse is coded if the asset type is available, but the amount was not provided. "Don't know" is coded if the first yes / no answer was positive and the respondent replied that the exact amount is unknown.
[3] The "don't know" category was not offered for this question.

Source: Own calculations based on GSOEP.

Table 6: Descriptive Statistics: Explanatory Variables

Variable	Individual Questionnaire		Wealth Module	
	Mean	Std. Dev.	Mean	Std. Dev.
Demographic Indicators				
respondent female interviewer male	0.294	0.456	0.239	0.427
respondent male interviewer female	0.205	0.403	0.232	0.422
respondent female interviewer female	0.232	0.422	0.211	0.408
respondent male interview. male (ref.)	0.269	0.443	0.318	0.466
respondent part time employed	0.089	0.285	0.074	0.261
respondent not employed	0.464	0.499	0.451	0.498
interviewer part time employed	0.132	0.338	0.136	0.343
interviewer not employed	0.464	0.499	0.492	0.500
same employment status	0.419	0.493	0.428	0.495
respondent medium level schooling	0.201	0.400	0.187	0.390
respondent high schooling	0.127	0.333	0.149	0.356
interviewer medium level schooling	0.469	0.499	0.464	0.499
interviewer high schooling	0.207	0.405	0.204	0.403
same schooling	0.347	0.476	0.349	0.477
respondent age	46.461	18.579	50.903	17.130
Age difference: respond. - interviewer	-3.873	21.484	0.685	20.053
Other indicators				
change of interviewer	0.102	0.303	0.115	0.319
public sector employee	0.134	0.340	0.156	0.363
Self administered survey	0.150	0.357	0.073	0.261
Lives in small town	0.577	0.494	0.539	0.499
household size	2.828	1.310	2.431	1.264
respondent schooling missing	0.010	0.100	0.009	0.095
Number of observations		4,744		2,427

Notes: Low schooling is coded for mandatory schooling, medium schooling for the German *Realschule,* and high schooling for degrees preparing for academic studies.

Source: Own calculations based on GSOEP.

5 Results and Discussion

5.1 Heterogeneity in Item Nonresponse Behavior and its Determinants

Covariates of item nonresponse

For the individual and household measures we estimate bivariate logit models and present marginal effects in Table 7a. For wealth measures the dependent variable contains the additional outcome category "don't know". To impose the least restrictive model we estimate multinomial logits and calculate the marginal effects for the probability of nonresponse (Table 7b). By comparing the estimation results across outcomes we can evaluate the robustness of the estimated covariate effects - an aspect that has been neglected in this literature.

The estimations yield only few statistically significant coefficients. The first group of variables describes the gender combination of respondent and interviewer with two males as the reference. All marginal effects that are based on significant coefficients indicate positive associations between female interviewers and item nonresponse. Sizeable effects on nonresponse are found e.g. for stock and bond ownership of plus 26.4 percentage points relative to an average of 34.1 percent. If we assume that it is easier to avoid an answer in front of a female than a male the pattern fits the rational choice model's predictions.

The next set of indicators describes the employment status of the participants. Overall there seems to be a weak tendency for non full time employed respondents to refuse an answer. Again the finding can be explained within the rational choice model: If the earnings of part-time workers are comparatively low, and these respondents prefer to indicate personal labor market success to the interviewer they may choose nonresponse. The evidence on the role of the interviewers' employment status is mixed and there is no indication of matching effects based on interviewer and respondent employment status.

Similarly, the evidence on schooling effects does not suggest clear patterns. Having respondents and interviewers with similar schooling does not affect the results. The marginal effects of higher respondent schooling on income nonresponses are almost all positive, yet insignificant. High interviewer education does not seem to improve outcomes. The reduction in item nonresponse when interviewers have medium schooling is difficult to rationalize.

Table 7: (a) Logit Estimates of Item Nonresponse Across Financial Measures at the Individual and Household level

Variable	Self-employment		Gross earnings		Net earnings		Vacation Benefits		Hh. Interest and Dividend Inc.		Hh. Net Income	
	ME	t	ME	t	ME	t	ME	t	ME	t	ME	t
respondent female interviewer male	0.033	0.53	-0.002	-0.09	-0.014	-1.00	-0.010	-0.68	0.033	1.46	-0.001	-0.10
respondent male interviewer female	0.128	2.19	0.024	1.44	0.002	0.16	-0.007	-0.55	0.062	2.57	0.018	1.80
respondent female interviewer female	0.066	0.91	0.038	2.05	-0.011	-0.71	-0.003	-0.20	0.079	3.34	0.015	1.41
respondent part time employed	-0.033	-0.50	0.063	4.02	0.035	2.53	-0.016	-0.72	0.043	1.43	0.017	1.39
respondent not employed	0.291	3.95	x	x	x	x	-0.061	-1.50	0.089	4.18	-0.004	-0.43
Interviewer part time employed	-0.036	-0.44	-0.075	-3.07	-0.030	-1.67	0.046	1.52	0.020	0.84	0.023	2.23
Interviewer not employed	-0.049	-0.69	-0.020	-0.97	-0.013	-0.76	0.057	1.77	-0.042	-1.90	0.008	0.81
same employment status	0.012	0.21	-0.007	-0.33	-0.001	-0.06	0.039	1.34	-0.003	-0.18	0.011	1.46
respondent medium level schooling	0.120	2.30	0.019	1.39	0.024	2.16	0.007	0.57	-0.108	-4.53	0.008	0.91
respondent high schooling	0.001	0.01	0.007	0.42	0.029	2.31	0.009	0.68	-0.022	-0.98	0.016	1.82
Interviewer medium level schooling	-0.123	-2.43	-0.030	-2.27	-0.003	-0.26	-0.032	-2.51	0.023	1.08	-0.017	-2.02
Interviewer high schooling	-0.040	-0.66	-0.003	-0.18	0.010	0.74	0.005	0.44	-0.006	-0.24	-0.002	-0.17
same schooling	-0.020	-0.43	-0.006	-0.49	0.011	1.11	0.008	0.79	0.038	1.87	0.010	1.32
respondent age	0.004	1.23	0.002	3.08	0.002	2.44	0.000	-0.61	0.000	-0.17	0.000	-0.39
age difference: respondent - interviewer	0.001	-0.44	-0.001	-1.20	0.000	-0.48	0.001	1.05	-0.002	-1.95	0.000	1.01
change of interviewer	0.064	1.09	0.022	1.41	0.027	2.23	0.018	1.37	-0.020	-0.82	0.007	0.71
public sector employee	x	x	-0.061	-3.96	-0.045	-3.46	-0.030	-2.42	-0.032	-1.21	-0.031	-2.36
self administered survey	0.001	0.20	0.028	2.13	0.010	0.93	0.004	0.35	0.034	1.11	-0.014	-0.80
lives in small town	-0.032	-0.67	0.021	1.79	0.012	1.22	0.007	0.73	-0.073	-4.74	-0.002	-0.35
household size	0.019	1.09	0.003	0.76	0.002	0.47	-0.009	-2.09	-0.031	-4.36	-0.001	-0.31
respondent schooling missing	0.218	2.07	0.048	1.03	0.058	1.86	x	x	-0.093	-1.04	0.013	0.42
Number of observations	274		2,546		2,546		1,501		2,149		2,769	

Table 7: (b) Multinomial Logit Estimates of Item Nonresponse Across Wealth Measures at the Household Level

Variable	Stocks / Bonds		Owned home		Home loan savings		Savings		Total wealth	
	ME	t	ME	t	ME	t	ME	t	ME	t
Respondent female interviewer male	0.074	0.50	-0.009	-0.93	0.004	0.01	-0.014	-0.99	0.002	0.19
Respondent male interviewer female	-0.050	-0.41	-0.008	-0.68	0.034	0.62	0.017	1.21	0.010	1.21
Respondent female interviewer female	0.264	1.64	-0.008	-0.72	0.123	2.09	0.002	0.10	0.017	2.12
Respondent part time employed	0.213	1.24	0.021	1.45	-0.097	-1.10	0.044	2.47	-0.003	-0.21
Respondent not employed	0.123	0.97	0.177	4.96	0.007	0.13	0.014	0.98	0.005	0.66
interviewer part time employed	-0.638	-3.36	-0.007	-0.75	-0.051	-0.90	-0.012	-0.72	-0.019	-1.31
interviewer not employed	-0.192	-1.49	0.000	-0.09	0.026	0.33	0.007	0.50	-0.003	-0.65
same employment status	-0.033	-0.30	0.000	0.04	0.025	0.55	0.002	0.24	-0.009	-1.49
Respondent medium level schooling	-0.296	-2.63	-0.004	-0.42	0.003	-0.04	-0.004	-0.38	0.000	0.04
Respondent high schooling	-0.703	-5.27	-0.275	0.00	-0.146	-2.50	-0.034	-1.87	-0.022	-1.53
interviewer medium level schooling	-0.092	-0.88	0.002	0.17	-0.006	-0.15	-0.022	-2.01	-0.003	-0.41
interviewer high schooling	0.127	0.76	0.003	0.28	0.031	0.51	-0.025	-1.72	-0.012	-1.16
same schooling	0.094	1.08	0.002	0.20	-0.071	-1.54	0.003	0.23	0.002	0.38
Respondent age	0.007	1.27	-0.001	-1.62	0.004	1.57	-0.001	-0.88	0.001	1.52
age difference: respondent – interviewer	-0.015	-2.82	0.000	0.95	0.000	-0.29	0.001	0.95	0.000	-1.60
change of interviewer	-0.213	-1.42	0.000	0.03	-0.167	-2.40	0.016	1.26	0.006	0.86
public sector employee	0.004	0.11	0.172	5.29	-0.005	-0.06	0.009	0.67	-0.011	-1.01
self administered survey	-0.318	-1.73	-0.267	0.00	-0.109	-1.37	0.009	0.52	0.017	2.26
lives in small town	0.063	0.52	-0.020	-2.87	-0.066	-1.96	0.015	1.37	0.008	1.37
Household size	-0.083	-1.93	-0.002	-0.47	-0.080	-4.69	-0.013	-2.82	-0.003	-1.21
Respondent schooling missing	-16.166	0.00	-0.272	0.00	-5.828	0.00	-1.112	0.00	-0.010	1.04
Number of observations	636		1,065		1,001		2,064		2,427	

Notes for Table 7:
The columns labeled ME present marginal effects, which were calculated in Table 7(a) on the basis of logit estimates for item nonresponse for each outcome separately, and in Table 7(b) on the basis of multinomial logit estimates for each outcome separately. Here the dependent variable was coded to indicate response, don't know, and nonresponse. In Table 7(b) the marginal effects describe the impact of the indicator on the probability of item nonresponse. The columns labeled t present the asymptotic t statistics for the coefficient estimates on the relevant variables in the estimations.
Cells containing an x do not indicate coefficient or marginal effects, because the variables had to be dropped from the model estimation due to collinearity.
All estimations controlled for constants which are not presented to save space.

Source: Own calculations based on GSOEP.

Older respondents seem to be more prone to item nonresponse than younger individuals. We also find some evidence that having interviewers who are younger than the respondents reduces nonresponse. For income measures we find that a ten years age difference among respondents is correlated with 2 percentage points higher nonresponse rate, holding interviewer age constant. - There are only few consistent patterns in the remaining control variables. In contrast to the literature we find a significant nonresponse effect of an interviewer change only for one outcome. Possibly the change of an interviewer has strong effects on unit nonresponse (cf. Rendtel 1995) such that item nonresponse cannot even be observed. Public sector employees seem to be significantly less likely to refuse an answer on income but not so on wealth items.

Based on the lower disutility involved one might expect more nonresponses among those who completed the questionnaire without an interviewer. This is confirmed only for gross earnings and total wealth. The evidence is similarly mixed with respect to the effect of rural residence where significant effects go in both directions. - The household size effect yields as expected that individuals living in larger households have significantly lower nonresponse rates.

5.2 Heterogeneity in Item Nonresponse across Outcomes

Outcome-specific heterogeneity is addressed in two steps. First, we pool the outcome data described in Tables 3 to 5 and add fixed outcome specific effects to the specification.[18] These fixed effect controls (see Table 8) are jointly highly significant and reflect heterogeneity across outcomes even after controlling for

[18] To render the bivariate nonresponse outcome measure of the income variables comparable to the multivariate outcome measure of the wealth indicators we dropped those wealth observations with "don't know" answers from the sample. The results presented below justify this procedure.

covariates. Adding outcome specific fixed effects to the model increases the pseudo (McFadden) R^2 from about 1.8 (not presented) to almost 14 percent. This result holds in smaller subsamples as well, when we pool outcomes at the individual, household, or wealth level only (not presented).

Since nonresponse rates differ between wealth and income outcomes, we investigate in a second step whether this is a level effect or whether the covariate effects differ across the two outcome groups. We reestimate the fixed effects model, now adding a full set of interaction terms (I) which indicate whether a wealth or income measure is observed.[19] The model is thus:

$$\begin{aligned} u_{ij} = \; & c_{ij}\ \alpha_1 + b_{ij}\ \alpha_2 + X_i\ \beta_1 + W_m\ \beta_2 + (X_i * W_m)\,\beta_3 \\ & + c_{ij} * I_j\ \alpha_1' + b_{ij} * I_j\ \alpha_2' + X_i * I_j\,\beta_1' \\ & + W_m * I_j\,\beta_2' + (X_i * W_m) * I_j\,\beta_3' + \mu_{ij} \end{aligned} \qquad (4)$$

The explanatory power of the model increases significantly with the full set of interaction terms added, of which a number are statistically significant (see last columns of Table 8).

The results suggest that the increase in nonresponse for female interviewers is somewhat more pronounced for wealth outcomes. A clear pattern appears for the schooling indicators, confirming the results from Table 7: whereas item nonresponse on income measures increases with higher respondent education, we find the opposite result for wealth outcomes. The differences are significant and difficult to interpret. If education is correlated with a respondent's level of information about wealth, then the high response propensity might be explained by cognitive ability. However, given that the same individuals should also be well informed on their income one can only speculate that they consider income as more private information.

The age effects seem to differ between income and wealth outcomes. The nonresponse probability on income measures increases with respondent age. The effect disappears for wealth questions. The negative correlation between the respondent-interviewer age difference and item nonresponse pointed out above seems to be based mostly on wealth outcomes.

[19] In these estimations we treat "income from interest and dividends" as an indicator of wealth holdings and group it with the outcomes listed in Table 5.

Table 8: Logit Estimates on Pooled Outcomes

Variable	Fixed Effects		Fixed Effects with Wealth Interactions			
			Main Effects		Interaction	
	Coeff.	t	Coeff.	t	Coeff.	t
respondent female interviewer male	-0.007	-0.08	-0.010	-0.08	0.000	0.00
respondent male interviewer female	0.140	1.72	0.149	1.30	0.004	0.02
respondent female interviewer female	0.357	4.16	0.196	1.51	0.291	1.67
respondent part time employed	0.422	4.59	0.520	4.34	-0.230	-1.19
respondent not employed	0.353	4.19	0.405	2.54	0.020	0.10
interviewer part time employed	-0.215	-2.20	-0.242	-1.68	0.036	0.18
interviewer not employed	-0.071	-0.92	-0.001	-0.01	-0.137	-0.85
same employment status	-0.054	-0.91	-0.035	-0.36	-0.035	-0.28
respondent medium level schooling	-0.126	-1.79	0.261	2.68	-0.754	-5.29
respondent high schooling	-0.192	-2.33	0.259	2.25	-0.880	-5.32
interviewer medium level schooling	-0.235	-3.45	-0.411	-4.34	0.331	2.42
interviewer high schooling	-0.041	-0.50	-0.056	-0.49	0.018	0.11
same schooling	0.094	1.47	0.086	0.98	0.045	0.35
respondent age	0.009	2.30	0.016	3.00	-0.013	-1.77
age difference: respond. - intvwr.	-0.004	-1.30	-0.000	-0.05	-0.008	-1.33
change of interviewer	0.068	0.82	0.302	2.68	-0.457	-2.75
public sector employee	-0.341	-4.20	-0.632	-5.47	0.582	3.53
self administered survey	0.109	1.32	0.137	1.33	-0.037	-0.21
lives in small town	-0.081	-1.45	0.099	1.20	-0.365	-3.22
household size	-0.126	-5.16	-0.016	-0.46	-0.240	-4.87
respondent schooling missing	-0.036	-0.13	0.659	2.10	-1.958	-2.85
Significance test fixed effects (χ^2,p)	90.75	0.00		53.54	0.00	
Log Likelihood		-5,444.8			-5,376.8	
Pseudo R^2 (McFadden)		0.138			0.149	
Number of observations		28,531			28,531	

Notes: The estimations combine the following outcome measures: All income categories listed in Table 3, the outcome measures No. 41 and 42 from Table 4, and all measures from Table 5 except for farm value.
Since income from interest and dividends (reported at the household level) is an indicator of wealth we considered this outcome as a wealth outcome.
Fixed effect coefficients are not presented to save space.
The figures in the row on fixed effect significance tests provide the test statistic of a chi^2 test with 26 degrees of freedom. The p-value is given behind the test statistics.

Source: Own calculations based on GSOEP.

Differences in covariate associations with nonresponse probabilities by outcome are observable also for the remaining variables: While the change of an interviewer increased nonresponses for incomes, it reduces them for wealth. The beneficial effect of public sector employment on the propensity to provide financial information seems to be limited to incomes: Since the earnings of public sector workers in Germany typically follow publicly available pay scales, it is possible that these workers are more open about their income, as these may be public knowledge anyway. When it comes to wealth, however, their privacy protection instincts seem to be the same as for anyone else. Living in a small town is correlated with significantly lower nonresponse on wealth while the effect on income is insignificant. Also the negative effect of household size differs significantly for the two outcomes. Thus, nonresponse is heterogeneous in frequency and correlation patterns across outcomes.[20]

5.3 A Closer Look at "don't know" Answers

In this section we investigate whether answering "don't know" is an independent outcome, or whether this response can be grouped with valid responses or with nonresponses. As described above we first apply a Hausman test to determine whether the unobserved determinants of item nonresponse are correlated with those of valid answers or outright nonresponses. We start with a sample that pools all of the outcomes presented in Table 5, combining the 12,613 observations of the "total" row. Then we consider some of the wealth outcomes separately to determine whether the results for the pooled sample are robust.

Table 9: Summary of Hausman Test Results

Dependent Variable	Number of obs.	Test-Statistic	p-value
Pooled wealth measures	12,613	0.20	1.00
Stocks and Bonds	636	0.00	1.00
Home Loan Savings	1,001	0.00	1)
Savings Account	2,064	0.00	1.00
Owned home: Market value	1,065	0.00	1.00
Total household wealth	2,427	-3.23	1)

Note:
1) The test statistic takes on a negative value, which can be interpreted as strong evidence against rejecting the null hypothesis that the IIA assumption holds (Hausman and McFadden 1984, p. 1226 footnote 4, or Stata 7 Manual volume 2 p.13).

Source: Own calculations based on GSOEP

[20] Since the combination of outcomes considered in the sample used in Table 8 is somewhat arbitrary, we performed robustness tests by reestimating the same models for alternative outcome subsets. There most coefficients have the same sign, but their statistical significance is not always robust to modifications of the sample.

The evidence presented in Table 9 seems to be strong and clear: The null hypothesis that the IIA assumption holds cannot be rejected in any of the tests. The unobservables do not seem to be correlated and therefore "don't know" answers are "relevant" and "independent" alternatives to informative responses and to nonresponses in the IIA sense.

Next we perform the (Hill-) Cramer-Ridder test of whether the observable covariates have significantly different effects on the propensity to provide a "don't know" answer relative to either valid answers or to outright nonresponse.

The test compares the multinomial logit slope coefficients to those of a model where the slope parameters for don't know and one alternative answer option are restricted to be identical. Again, we perform it first for the sample of pooled wealth measures and then for some of the wealth items separately.

Again the evidence (see Table 10) is clear: In all cases we can reject the hypotheses that the coefficients of the "don't know" answer are identical to those of valid responses at high levels of statistical significance. For all outcomes but ‘market value of an owned home' we reject that the coefficients of the "don't know" answer are identical to those of valid responses mostly at the one percent level. We read this evidence as indicative of the independence of "don't know" answers: Typically observable determinants of response behavior have significantly different impacts on the three considered outcomes.

Jointly the two tests suggest that neither by their observable nor by their unobservable determinants are "don't know" answers correlated with - and therefore likely substitutes of - valid answers or complete item nonresponses. Therefore "don't know" answers must be viewed as independent outcomes in their own right. Missing values due to "don't know" replies cannot simply be mixed with item nonresponses. The test results show that the two processes are determined by different observable and unobservable mechanisms.[21] Therefore our results yield additional support to the conclusion of Juster and Smith (1997) at in their analysis of responses to follow-up bracket questions in surveys (p.1272): "This marked contrast in the behavior of DK and REF responses suggests that the two need to be handled separately when imputations are being done", where DK represents don't know and REF refusal to respond.

[21] This confirms Juster and Smith (1997), who showed that "don't know" respondents and non-respondents differ in their willingness to provide responses to "bracket" questions (i.e. follow-up questions asked when initially no valid response is received): Whereas almost 80 percent of initial "don't know" respondents provided complete bracket data, the share among non-respondents reached only 40 percent.

Table 10: Summary of Hill-Cramer-Ridder Test Results

Dependent Variable	No. of Obs.	H_0: $\beta^{\text{Don't Know}} = \beta^{\text{Response}}$ LR[1] (DF[2]; p-value)	H_0: $\beta^{\text{Don't Know}} = \beta^{\text{Item-Nonresponse}}$ LR[1] (DF[2]; p-value)
Pooled wealth measures	12,613	233.02 (21; 0.000)	158.22 (21; 0.000)
Stocks and Bonds	636	39.51 (20; 0.006)	42.18 (20; 0.003)
Home Loan Savings	1,001	60.81 (20; 0.000)	100.98 (20; 0.000)
Savings Account	2,064	65.75 (20; 0.000)	45.87 (20; 0.001)
Owned home: Market value	1,065	65.31 (17; 0.000)	21.44 (17; 0.207)
Total household wealth	2,427	87.74 (21; 0.000)	31.24 (21; 0.070)

Notes:
[1] LR represents the value of the likelihood ratio test statistic.
[2] The degrees of freedom differ accross wealth measures, since due to collinearity and small number of cases the full model (see Table 2) could not be estimated. The full model was estimated for the pooled wealth measures and for total household wealth. For testing Savings, Home Loan Savings and Stocks and Bonds, the indicator of missing respondent schooling was omitted. In the case of ownership of occupied flat or home, the indicators of self administered survey and higher respondent schooling were also dropped from the econometric model.

Source: Own calculations based on GSOEP.

6 Conclusions

Even though item nonresponse affects any analysis using survey data it has found little attention as a behavioral phenomenon in its own right.[22] In this study we present a number of results that are new to the literature. We first summarize existing theoretical frameworks for item nonresponse behavior, i.e. the cognitive model and the rational choice approach.

The empirical literature on item nonresponse is limited and generally focuses on measures of labor income. We address this limitation by investigating the frequency and determinants of item nonresponse for a variety of financial outcomes. We find significant heterogeneity in nonresponse intensities across outcomes. This conclusion from descriptive statistics is confirmed in regressions of nonresponse behavior where much explanatory power derives from the

[22] Certainly a vast statistical literature has developed following Rubin's influential work on missing data imputation (Rubin 1987). However the issue there is to find the best possible correction given that the data is missing. Our interest is to explain at least in part why it is missing in order to improve data collection efforts.

consideration of outcome specific fixed effects. We confirm several results of the literature regarding correlates of item nonresponse. An investigation of the homogeneity of nonresponse determinants across outcomes yields new insights: Estimating a fully interacted model shows clearly that a number of the established correlates of item nonresponse behavior vary depending on the specific item under consideration.

We investigate whether the match of interviewer and respondent characteristics affects the quality of survey outcomes. Robust findings on this matter would be valuable to reduce the cost and to increase the quality of information gathered from social surveys. The analysis yields that nonresponse rates tend to be higher if the interviewer is female in particular when the respondent is female as well. Having a respondent and an interviewer with the same employment status or the same educational level does not significantly affect nonresponse outcomes. However, our measures of employment and educational attainment may be too rough to reflect the impact of potential matching effects on nonresponse behavior. With respect to age differences there is some evidence that matching a younger interviewer to an older respondent may increase response propensities particularly with respect to wealth outcomes. Interestingly, the personal acquaintance of the respondent with the interviewer is beneficial for wealth but not for income outcomes.

Our third research question concerns "don't know" answers in questionnaires. A Hausman test of the independence of irrelevant alternatives assumption and the (Hill-) Cramer-Ridder test suggest strongly that "don't know" responses cannot be viewed as a subcategory of valid answers nor as comparable to item nonresponse. Therefore simple statements as to how offering "don't know" answer options affects the set of valid answers are not possible.

In the end researchers have to acknowledge that the group of respondents who refuse to answer a survey question is not a random draw from the population, that the group varies depending on the question looked at, that those answering "don't know" differ from non-respondents, and that simply omitting these individuals from the analysis may well bias results. Much attention has been devoted to developing appropriate imputation mechanisms when data is missing. Our results suggest that imputation procedures should differentiate between missing values due to "don't know" answers and due to outright nonresponse.

Bibliography

Biewen, Martin, 2001, Item nonresponse and inequality measurement: Evidence from the German earnings distribution, *Allgemeines Statistisches Archiv* 85, 409-425.

Cramer, J.S. and G. Ridder, 1991, Pooling states in the multinomial logit model, *Journal of Econometrics* 47, 267-272.

De Maio, Theresa J., 1980, Refusals: Who, where and why, *Public Opinion Quarterly* 44, 223-233.

Dillman, Donald A., 1978, *Mail and Telephone Surveys: The Total Design Method*, John Wiley and Sons, New York.

Esser, Hartmut, 1984, Determinanten des Interviewer- und Befragtenverhaltens: Probleme der theoretischen Erklärung und empirischen Untersuchung von Interviewereffekten (Determinants of interviewer- and respondent behavior: Problems of theoretical explanations and empirical analyses of interviewer effects), in: K.Mayer, P. Schmidt (eds.) *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*, 26-71, Frankfurt.

Ferber, Robert, 1966, Item Nonresponse in a Consumer Survey, *Public Opinion Quarterly* 30, 399-415.

Flinn, Christopher J. and James J. Heckman, 1983, Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?, *Journal of Labor Economics* 1(1), 28-42.

Hanefeld, Ute, 1987, *Das sozio-oekonomische Panel: Grundlagen und Konzeption*, Campus Verlag, Frankfurt / Main.

Hausman, Jerry A. and Daniel McFadden, 1984, Specification Tests for the Multinomial Logit Model, *Econometrica* 52(5), 1219-1240.

Hill, M. Anne, 1983, Female Labor Force Participation in Developing and Developed Countries - Consideration of the Informal Sector, *Review of Economics and Statistics* 6(3), 459-468.

Hill, Daniel and Robert J. Willis, 2001, Reducing Panel Attrition: A Search for Effective Policy Instruments, *Journal of Human Resources* 36(3), 416-438.

Horowitz, Joel L. and Charles F. Manski, 1998, Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations, *Journal of Econometrics* 84, 37-58.

Journal of Human Resources, 1998, Special Issue: Attrition in Longitudinal Surveys, *Journal of Human Resources* 33(2).

Journal of Human Resources, 2001, Special Issue on Data Quality, *Journal of Human Resources* 36(3).

Juster, F. Thomas and James P. Smith, 1997, Improving the Quality of Economic Data: Lessons from the HRS and AHEAD, *Journal of the American Statistical Association* 92(440), 1268-1278.

Lachman, R., Lachman J.T., and E.C. Butterfield, 1979, *Cognitive Psychology and Information Processing*, Hillsdale, NJ: Erlbaum.

Lillard, Lee, James P. Smith, and Finis Welch, 1986, What do we really know about wages? The Importance of Nonreporting and Census Imputation, *Journal of Political Economy* 94(3), 489-506.

Loosveldt, Geert, Jan Pickery, and Jaak Billiet, 1999, Item nonresponse as a predictor of unit nonresponse in a panel survey, Paper presented at the International Conference on Survey Nonresponse, Portland Oregon, USA.

Reinecke, Jost, 1991, *Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte*, Westdeutscher Verlag, Opladen.

Rendtel, Ulrich, 1995, *Panelausfälle und Panelrepräsentativität*, Campus Verlag, Frankfurt, New York.

Riphahn, Regina T., 1997, Disability Retirement and Unemployment - Substitute Pathways for Labor Force Exit?, 1997, *Applied Economics* 29(5), 551-561.

Rubin, Donald B., 1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley Publishers, New York.

Schräpler, Jörg-Peter and Gert G. Wagner, 2001, Das Verhalten von Interviewern - Darstellung und ausgewählte Analysen am Beispiel des "Interviewerpanels" des Sozio-ökonomischen Panels, *Allgemeines Statistisches Archiv* 85, 45-66.

Schräpler, Jörg-Peter, 2001, Respondent Behavior in Panel Studies. A Case Study of the German Socio-Economic Panel (GSOEP), *DIW Discussion Papers* No. 244, DIW - Berlin.

SOEP Group, 2001, The German Socio-Economic Panel (GSOEP) after more than 15 years - Overview. In: Elke Holst, Dean R. Lillard and Thomas A. DiPrete (eds.): Proceedings of the 2000 Fourth International Conference of German Socio-Economic Panel Study Users (GSOEP2000), *Vierteljahrshefte zur Wirtschaftsforschung (Quarterly Journal of Economic Research)* 70(1), 7-14.

Sousa-Poza, Alfonso and Fred Henneberger, 2000, Wage data collected by telephone interviews: an empirical analysis of the item nonresponse problem and its implications for the estimation of wage functions, *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 136(1), 79-98.

STATA Manual, 2000, Reference Volume 2, Stata Press: College Station (TX).

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz, 1996, *Thinking About Answers. The Application of Cognitive Processes to Survey Methodology*, Jossey Bass Publishers, San Francisco.

Trometer, Reiner, 1996, *Warum sind Befragte "meinungslos"? Kognitive und kommunikative Prozesse im Interview*, (Why do respondents not have an opinion? Cognitive and communicative processes in interviews) Ph.D. Dissertation University of Mannheim, *mimeo*.

Zweimüller, Josef, 1992, Survey nonresponse and biases in wage regressions, *Economics Letters* 39, 105-109.

Appendix

Table A1: Multinomial logit estimation results for the pooled wealth measure

Dep. Variable	=	Pooled wealth outcomes, valid answer is the reference category.
LR chi2(42)	=	347.90
Prob > chi2	=	0.0000
Pseudo R2	=	0.0287

	Don't Know		Nonresponse	
Independent Variable	Coeff.	t	Coeff.	t
respondent female interviewer male	0.642	6.410	-0.056	-0.430
respondent male interviewer female	0.105	1.010	-0.018	-0.140
respondent female intervwr. female	0.476	4.300	0.456	3.550
respondent part time employed	0.185	1.520	0.272	1.640
respondent not employed	-0.028	-0.290	0.079	0.660
interviewer part time employed	0.490	4.260	-0.359	-2.190
interviewer not employed	0.583	6.110	0.029	0.250
same employment status	0.000	-0.010	0.022	0.250
respondent medium level schooling	0.160	1.920	-0.131	-1.200
respondent high schooling	-0.057	-0.570	-0.406	-3.020
interviewer medium level schooling	0.074	0.900	-0.136	-1.270
interviewer high schooling	0.024	0.230	-0.003	-0.020
same schooling	-0.050	-0.630	0.055	0.540
respondent age	-0.005	-1.180	0.006	1.100
age difference: intervw. - respond.	0.007	1.810	-0.006	-1.220
change of interviewer	0.228	2.440	-0.192	-1.390
public sector employee	-0.251	-2.520	-0.051	-0.420
self administered survey	0.114	0.960	0.151	1.000
lives in small town	0.561	7.530	-0.063	-0.730
household size	0.043	1.550	-0.199	-5.140
respondent schooling missing	0.081	0.250	-1.734	-1.720
constant	-3.282	-12.210	-2.612	-7.81
Log Likelihood	-5887.54			
Number of observations	12,613			

Source: Own calculation based on GSOEP.

III. The Interaction between Item, Questionnaire and Unit Nonresponse in the German SOEP

Oliver Serfling

Abstract

This study investigates respondents' behavior on financial items with respect to item nonresponse, questionnaire nonresponse, and panel attrition. We define questionnaire nonresponse as a new category of respondents' behavior. Using financial items from the household questionnaires of the German Socio-Economic Panel (SOEP), we test whether item nonresponse is positively correlated with questionnaire and unit nonresponse, and if questionnaire nonresponse is a predictor for subsequent panel attrition. Second, we test whether both nonresponse mechanisms may affect studies on item nonresponse due to endogenous sample selection.

Acknowledgements
This paper is part of a research project granted by WWZ-Forum, University of Basel. The author thanks participants of the "Workshop on Item-Nonresponse and Data Quality in Large Social Surveys", Basel, October 9-12, 2003, colleagues at the Department of Economics and Business Administration, University of Basel, and especially Regina Riphahn for very helpful comments.

1 Introduction: The Nonresponse Problem

The phenomena of unit nonresponse or panel attrition (UNR) and item nonresponse (INR) have been widely studied by the survey literature. Nevertheless, the literature on the interaction of both phenomena is still scarce. This short study attempts to provide some preliminary evidence on the interrelation of various nonresponse types. We first examine, whether *panel attrition*, *questionnaire nonresponse*, i.e. respondents' selective response to single questionnaires of a multi-questionnaire survey, *and item nonresponse* are positively correlated and/or driven by a similar decision process. If so, panel attrition may cause endogenous sample selection with respect to item nonresponse. Hence, studies on determinants of item nonresponse using panel data are likely to be biased. Detecting such bias is our second research aim.

This article makes a variety of contributions to the literature: It examines a broad set of financial, i.e. income and wealth, items from the German Socio-Economic Panel (SOEP). Besides panel attrition, we examine for the first time respondents' behavior with respect to a separate wealth questionnaire in a multi-questionnaire survey, which we name "questionnaire nonresponse" (QNR). In addition, we provide some evidence that sample selection may lead to biased results in item nonresponse analyses.

The paper is organized as follows: First, we define several types of nonresponse in a survey and discuss how they might be interacted in a panel survey, in Section 2. An extensive discussion of the literature is skipped and saved for the following paper of this dissertation. The research hypotheses, the empirical strategy and our data are described in Section 3. Section 4 presents and discusses our empirical findings, while the last section summarizes and concludes.

2 Issues of Nonresponse

2.1 Definition of Nonresponse

Unit nonresponse (UNR) or *panel attrition* describes the drop-out of a household or person from the respondents group. *Questionnaire nonresponse (QNR)* may only occur in surveys which consist of several separate questionnaires. The respondent or household takes part in the interview, but completely refuses to fill in a whole special-topic questionnaire. This type of nonresponse has - to our knowledge - not yet been analyzed in the nonresponse literature. *Item nonresponse (INR)* is an interview participants' refusal to answer a specific question of an interview participant.

2.2 Interrelation of Item and Unit Nonresponse

The literature provides scanty evidence on the relationship between item and unit nonresponse. Mostly, it is hypothesized that both types of nonresponse result from the same decision process, which is driven by interest, motivation, and ability of the respondent (cf. Loosveldt et al. 2002: pp. 546). In that case, item and subsequent unit nonresponse should be positively correlated. Some panel studies observe the joint decline of item and unit nonresponse rates over time (see e.g. Van den Eeden 2002). This may be explained by self-selection of respondents and supports the aforementioned hypothesis. Schräpler (2003a) finds a small but significant negative correlation between refusing the gross income statement and participation in the next wave of the SOEP over the first twelve years, and Frick & Grabka (2005) find a positive correlation between item nonresponse in an aggregated measure of "total income" and subsequent attrition from SOEP.

However, there also exists empirical evidence that does not support the findings above: Dolton et al. (1998) found that item nonresponse rate and interview duration do not have explanatory power for panel attrition. Van den Eeden (2002) concedes that item nonresponse as a proxy for motivation has only extremely low explanatory power in a regression of unit nonresponse.

This study examines the correlation between item-, questionnaire-, and unit nonresponse. We consider various possible determinants of the nonresponse decision of the respondent which may be influenced by the characteristics of the respondent and the interviewer as well as by the interview situation. These mechanisms are extensively discussed in the nonresponse literature[23].

3 Empirical Approach

Our first research hypothesis is that item and unit nonresponse are positively correlated. Hence, item nonresponse should be a precursor of panel attrition, and in the year before interview refusal drop-outs should have higher INR-propensities than stayers. Therefore, we test whether the INR rate for drop-outs is significantly higher than the INR rate for stayers, using a t-test. Our item nonresponse rate is calculated as the share of item nonresponses on 12 income-related items from the household questionnaire. These items typically suffer from INR[24]. Since not all of those income items are applicable to each respondent, the number of relevant questions varies across respondents[25]. Second, therefore, we apply the same test procedure at the question level, hoping to identify those items for which nonresponse is positively correlated with subsequent attrition. In a third step, we regress the unit nonresponse indicator on last year's item nonresponse rate, using a logit approach, and test for sign and significance of its marginal effect. We use a broad set of potential determinants of panel attrition

[23] See e.g. Schräpler (2003b) and the sources cited there for a careful discussion of nonresponse mechanisms.

[24] The items are presented along with the results in Table 1.

[25] Nevertheless, the person specific INR rate is standardized to the number of applicable questions.

behavior, in order to reduce the heterogeneity of our data with respect to the nonresponse decision.

Our second hypothesis is that QNR is an intermediate category between INR and UNR, indicating a cooperation level lower than INR but higher than UNR. Therefore, INR should be a precursor of QNR, and QNR itself should be a precursor of subsequent UNR. We test this hypothesis using the test procedure described above, i.e. t-tests and logit regression.

The conclusion of the first and second hypothesis leads to our third hypothesis: unit- and questionnaire Nonresponse may lead to endogenously selected samples, which may cause biased estimates in INR-regressions. We apply a Heckman-type bivariate probit selection model (see Van den Veen and Van Praag (1981)), which consists of two estimation equations: (1) the INR specification equation[26]:

$$INR_{i,t} = \begin{cases} 1 \text{ if } y_{i,t}{*} < 0 \\ 0 \text{ otherwise} \end{cases} \qquad (1)$$

$$\text{with:} \qquad y_{i,t}{*} = \alpha + X_{i,t}\,\beta + \mu_{i,t}$$

and (2) a selection equation[27]:

$$UR_{i,t} = (\gamma + Z_{i,t-1}\,\delta + \eta_{i,t} \geq 0) \qquad (2)$$

which determines whether the individual is observed at time t (unit responded: $UR_{i,t}$=1). The regressors Z=(X,W)' of the selection equation consist of the regressors of the specification equation X and additional regressors W which have explanatory power for unit nonresponse without affecting item nonresponse and thus being instruments for panel attrition. Furthermore, it is assumed that the error terms μ and η are bivariately standard normally distributed with correlation ρ. A self-selection bias exists if the error terms are correlated ($\rho \neq 0$). The significance of ρ is tested using a likelihood-ratio test.

4 Data and Sampling

In this study we are mainly interested in income and wealth items, since these are relevant for many economic research questions and typically affected by nonresponse. We use data from the 1988 wave as well as from the previous and following survey waves of the SOEP household questionnaires. In the 1988 panel wave, the special topical module covered wealth, and was designed as a separate household questionnaire. To circumvent language problems, we restrict our sample to German households from the representative subsample of the native German population (SOEP sample "A") who participated in the 1987 survey. This sample includes 3,394 households in 1987 and reduces to 3,308 participating households in 1989 due to panel attrition, including losses due to

[26] With α being the constant, $X_{i,t}$ are the explanatory variables for interview i in period t, β is the vector of regression coefficients and $\mu_{i,t}$ the error term of the specification equation.

[27] γ is the constant, δ the coefficient vector and $\eta_{i,t}$ is the error term of the selection model

death, emigration, and household dissolution. Second, we restrict our sample to face-to-face interviews since this mode is used in the majority of cases (68 to 77 per cent) and permits us to control for interviewer effects while omitting mode effects[28]. Finally, the sample had to be restricted to observations where the same person answered the household questionnaire in two subsequent waves[29]. Due to all these restrictions the number of analyzable households declines by about one third. Additionally, we use data from the supplemental interviewer dataset, to measure interviewer and interaction effects. The unit nonresponse indicator (UNR) is coded 1 if the participating household dropped out after the considered wave. The questionnaire nonresponse indicator (QNR) is coded 1 for households that completed the 1988 household questionnaire but refused to fill in the wealth questionnaire in that same interview. Item nonresponse (INR) is coded 1 if an answer to an applicable item was denied[30]. For the analysis of item nonresponse in the wealth questionnaire, we constructed three wealth categories "property", "savings" and "total household wealth" which consist of up to 4 items each in the wealth questionnaire[31].

5 Empirical Analysis of Nonresponse Interaction

5.1 INR as a Precursor of UNR

Row 1 of Table 1 gives the difference between the item nonresponse rates of subsequent drop-outs and stayers in 1987 (column 1) and 1988 (column 2), respectively. The t-test results show that the null hypothesis (the difference in the mean INR rates for stayers and drop-outs is zero) cannot be rejected. Looking at the question-specific INR rates, only for two items in 1987 are differences significantly different from zero and thus confirm our hypothesis: e.g. the nonresponse rate on the special welfare benefits item was 50 percentage points higher for subsequent drop-outs than for stayers. The same holds for maintenance expenditures on property, even though with a lower difference and at a lower significance level. In 1988, none of the item-specific INR rates were significantly different for drop-outs and stayers (column 2). In the wealth questionnaire the item nonresponse among subsequent drop-outs was 40.9 percentage points higher for the item "stocks and bonds" and 1.4 percentage points for "total household wealth". So far, we have found no clear evidence supporting the first hypothesis of a positive correlation between UNR and INR.

[28] Even if face-to-face is the standard interview mode in SOEP, we have to concede that respondents with lower willingness to cooperate may have opted for paper and pencil interview.

[29] It is assumed that continuity of the head of household is uncorrelated with response behavior.

[30] In the wealth questionnaire of 1988 the option to answer "don't know" was provided. We treat this category as a valid response.

[31] For a more careful discussion of the problems of using wealth items instead, see Serfling (2004).

Table 1: Differences in current Item Nonresponse Rates for subsequent Drop-outs and Stayers

Item	(1) INR 1987, UNR 1988 Mean diff.	t		# of cases	(2) INR 1988, UNR 1989 Mean diff.	t		(3) INR 1988, QNR 1988 Mean diff.	t		# of cases
person specific INR rate in Household Questionnaire	-0.008	-0.5		2459	0.012	0.8		0.030	2.8	***	2353
Item specific: Household Questionnaire											
Welfare benefits [1]	0.002	0.3		2252	0.002	0.3		0.006	0.3		2126
General welfare benefits [2]	0.016	0.2		61	0.096	0.5		0.094	-		54
Special welfare benefits [2]	-0.500	-7.6	***	61	0.096	0.5		0.094	-		54
Child benefits [1]	0.006	0.5		1439	0.003	0.3		0.003	0.3		1391
Child benefits [2]	0.004	0.1		813	0.003	0.1		0.003	0.3		735
Rental or lease incomes [1]	0.007	0.5		2252	0.003	0.3		0.003	0.4		2126
Rental or lease incomes [2]	0.012	0.2		255	0.022	-		0.023	0.4		273
Maintenance exp. on property [2][3]	-0.425	-2.2	**	255	0.136	-		0.140	1.1		273
Annuity & interest payments [2][3]	-0.188	-0.6		255	0.331	-		-0.047	-0.3		273
Interest payments [2][3]	-0.148	-0.4		255	0.368	-		-0.009	-0.1		273
Interest and dividend income [3]	-0.074	-1.0		1674	-.013	-0.2		0.066	1.3		1615
Monthly household net income [2]	0.033	1.1		2252	-.004	-0.1		-0.142	-6.7	***	2126
Item specific: Wealth Questionnaire											
Ownership of occupied flat or home: rateable value					0.012	0.4					902
Ownership of occupied flat or home: market value					0.005	0.2					902
Property					-	-					268
Farm					0.184	-					50
Equity in a business					-.265	-0.9					134
Savings account					0.034	1.0					1770
Home loan savings certificates (*Bausparvertrag*)					-.027	0.2					817
Stocks and bonds					-.409	-1.7	*				562
Life Insurance: Originally insured amount					0.011	0.4					1124
Life Insurance: Current monthly payment					0.029	0.6					1124
Household debt					0.013	0.3					640
Total household wealth					-.014	0.7	*				2072
Inheritances since 1960					0.064	0.4					331

Notes:
answer possibilities: [1] yes / no; [2] amount; [3] last year (retrospective question)
Significance levels: * 10 %, ** 5 %, *** 1 %
Number of observations for columns 2 and 3 are equal since both samples are conditioned on participation in 1988.

Source: Own calculations based on SOEP waves 1987 and 1988

Table 2: Determinants of Unit and Questionnaire Nonresponse (marginal effects of logit regression)

Explanatory variables:	(1) UNR after 1987 ME	t		(2) UNR after 1988 ME	t		(3) QNR in wealth questionnaire in wave 1988 ME	t	
Item nonresponse rate	-0.067	-		-0.674	-		-1.284	-	***
QNR in 1988				0.112	0.64				
Sex									
R female I male	0.060	0.98		-0.172	-	**	0.084	1.77	*
R male I female	0.107	1.51		0.002	0.03		0.096	2.00	**
R female I female	0.048	0.71		-0.099	-		0.099	1.88	*
Age									
R age	0.000	0.17		0.002	0.50		0.003	0.99	
age difference: R - I	0.001	0.84		0.001	0.39		0.000	0.00	
Employment status									
R part time employed	-0.002	-		0.067	0.35		-0.025	-	
R not employed	0.021	0.34		0.170	1.37		-0.090	-	**
I part time employed	-0.016	-		0.040	0.39		-0.055	-	
I not employed	-0.049	-		-0.054	-		-0.098	-	*
same employment status	0.026	0.57		0.121	1.59		-0.026	-	
Schooling									
R medium level schooling	-0.056	-		-0.069	-0.8		-0.087	-	
R high schooling	-0.065	-		-0.198	-	*	0.063	1.44	
I medium level schooling	0.018	0.35		-0.035	-		0.059	1.08	
I high schooling	-0.029	-		-0.087	-		-0.017	-	
same schooling	-0.026	-		-0.108	-		0.097	2.06	**
Situation Effects									
Change of I	0.007	0.12		-0.103	-		-0.062	-	
R public sector employee	-0.052	-		0.221	2.18	**	-0.139	-	*
Self administered survey	-0.027	-		-0.192	-		0.011	0.23	
HH in small town	-0.004	-		-0.061	-		0.024	0.55	
R's household size	-0.080	-	***	-0.119	-	***	0.037	2.31	**
Number of I contacts	-0.021	-		0.038	1.86	*	0.023	1.80	*
R living in high-rise buildings	-0.013	-		0.127	1.79	*	0.056	1.22	
R living in residential area	-0.054	-		0.111	1.32		0.010	0.24	
interview duration (min.)	0.002	0.89		0.002	0.52		-0.002	-	
Constant (coefficient)	-2.591	-		-4.932	-	**	-5.871	-	***
No. of obs.	2172			2107			2107		
Pseudo R^2	0.17			0.16			0.10		
Log Likelihood	-130.6			-138.9			-228.04		
LR – Test (df)	55.30 (25)			53.6 (26)			53.5 (25)		
$p > \chi^2$	0.00			0.00			0.00		

Notes:
Significance levels: * 10 %, ** 5 %, *** 1 %
I: interviewer; R: respondent; HH: household; ME: marginal effects
It should be noted that explanatory variables for the model presented in column 3 are taken from the same year 1988, while those for columns 1 and 2 are taken from the base year.

Source: Own calculations based on SOEP waves 1987 and 1988

With respect to the correlation between INR and QNR it is obvious that it points in the opposite direction than hypothesized: the person-specific item nonresponse rate for questionnaire nonrespondents is 3 percentage points lower than for questionnaire respondents (Table 1, column 3, row 1). Only those refusing the net household income question seem to be more likely to opt for QNR.

Against the presented results above it may be argued that unit nonresponse is also affected by other determinants than the INR propensity. We therefore reduce the heterogeneity in attrition behavior by controlling for respondent, interviewer and situation characteristics, as well as their interactions, and additionally for the duration of the conducted interview. The marginal effects of the logit regressions are presented in Table 2.

To address our first hypothesis, we have to check whether the INR rate is significantly positively correlated with UNR, when controlling for the above-mentioned covariates. In columns 1 and 2, the INR rate of the interviews in 1987 and 1988 is negatively correlated with subsequent UNR, which is contradictory to our hypothesis. We concede that the coefficient, and thus the marginal effect, is not precisely estimated, such that the null-hypothesis of INR rate and UNR being uncorrelated cannot be rejected. However, this does not support our hypothesis either, since it predicts a positive correlation. In the UNR model specification of column 2, we have also used the QNR indicator as explanatory variable to test our second hypothesis. Even if the effect of questionnaire nonresponse is estimated to be positive, we cannot reject the hypothesis that QNR in 1988 and subsequent UNR are uncorrelated. Robustness checks[32] showed that the effects of the INR rate and QNR indicator that were identified remained unaffected with respect to magnitude and significance if only one of them was used in the model specification.

In column 3 of Table 2, the results of a regression of the QNR indicator on our set of possible determinants are presented. Here, the INR rate, derived from the household questionnaire in the same interview, has a highly significant negative effect on questionnaire nonresponse. This is unambiguous evidence against our first hypothesis and affirms the results from our sample t-tests provided in Table 1 above.

With regard to the effects of control variables, we find only household size having a negative significant effect on unit nonresponse in both years, but having a positive significant effect on questionnaire nonresponse in 1988. The effects of other controls suffer mostly from imprecise estimates (i.e. insignificance). Nonetheless, our models have significant explanatory power for UNR and QNR as indicated by McFadden's pseudo R^2 statistic and likelihood ratio test given at the bottom of Table 2.

[32] These are not provided here, but available from the author upon request.

5.2 Attrition Bias

Adressing our third hypothesis, we estimate Heckman-type bivariate probit models for the occurrence of item nonresponse on several financial items and test for correlation in the error terms of the selection and specification equation as described in Section 3. The specification equation describes the potential determinants of item nonresponse. As regressors we use gender and age of interviewer and respondent as well as interactions thereof, situation effects such as self-administered survey and household size, the employment status and schooling degree of the respondent.

The variables: "number of interviewer contacts before first successful interview", "household living in a residential area" (in contrast to living on the country or in an industrial area) and "type of building the household lives in" (high-rise building or not) are used as instruments for the selection equation, since they have some explanatory power for UNR without affecting INR-results.

Table 3: Tests of sample selection bias in INR models due to panel attrition. Results of bivariate probit.

Questionnaire Items	panel wave	No. obs. (thereof UNR)	ρ	Std.err	H_0: $\rho=0$ p-value
Individual Questionnaire					
Gross earnings last month	1988	2459 (24)	-0.6912	0.437	0.309
Gross earnings last month	1989	2317 (23)	0.0008	1.774	0.999
Net earnings last month	1988	2459 (24)	-0.8654	0.596	0.580
Net earnings last month	1989	2317 (23)	-0.7926	0.335	0.232
All applicable income questions	1989	11942 (142)	-0.9351	0.073	0.004 ***
Household Questionnaire					
Net income of household	1988	2219 (35)	0.9319	0.313	0.609
All applicable income questions	1988	11779 (179)	-0.3623	0.349	0.446
All applicable income questions	1989	15050 (157)	0.0159	0.918	0.986
Wealth Questionnaire					
Total wealth of household	1988	2160 (94) 1)	0.8500	0.177	0.015 **
Property and total wealth, pooled	1988	2411 (108) 1)	0.8749	0.149	0.003 ***

Notes:
1) Number of cases in brackets consist of UNR + QNR.
Significance levels: * 10 %, ** 5 %, *** 1 %

The results – presented in Table 3 - indicate a sample selection bias for the items in the wealth questionnaire, but not for the repeating part of the household questionnaire. When all applicable financial items were pooled, we derived a correlation coefficient ρ of -0.36 (standard error: 0.35) in 1988's and 0.02 (standard error: 0.92) in the 1989 household questionnaire. For both coefficients the null hypothesis of being zero could not be rejected on any level of significance. When it comes to the item "total wealth of household" in the wealth questionnaire, we derived a positive correlation of 0.85 (standard error: 0.18), which is significantly different from zero at the 95 percent level of confidence. This indicates endogenous sample selection and therefore biased estimates in INR-regressions if panel attrition and questionnaire nonresponse is neglected.

6 Summary and Conclusion

The literature focusing on the interactions among nonresponse types is scarce and partly ambiguous. We introduce the nonresponse category QNR, i.e. the refusal of a mono-thematic questionnaire in a multi-questionnaire survey, and assume this to be an intermediate category between INR and UNR. We contribute to the literature by providing empirical evidence for: (1) the correlation of item and unit nonresponse, (2) the correlation of item and questionnaire nonresponse, (3) the correlation of questionnaire and unit nonresponse and (4) sample selection with respect to item nonresponse due to panel attrition (UNR) and questionnaire nonresponse. First, we tested whether INR, QNR, and UNR are positively correlated in the considered SOEP waves 1987 and 1988.

In summary, we do not find evidence for positive correlations of INR and UNR, INR and QNR, nor QNR and UNR. Instead we find slightly negative correlations of the INR rate with subsequent UNR. When it comes to questionnaire nonresponse, we find a significant negative correlation with item nonresponse. This leads to the conclusion that people may be willing to fill in the special topics questionnaire, because they know they are not going to provide certain answers. These results are derived from univariate statistics as well as from multivariate regressions.

Second, we tried to identify sample selection bias due to panel attrition in the results of INR regressions. We find that the items in the repeating household questionnaire are unaffected by panel attrition. The wealth questionnaire is subject to two possible biasing sample selection processes: panel attrition and questionnaire nonresponse. Hence, we could identify a bias in the estimates of item nonresponse on the total household wealth question.

This study contributes to the scarce literature on the dynamic effects of several nonresponse types. It has provided several interesting results on the interaction of unit-, questionnaire- and item nonresponse. Since the literature is ambiguous concerning this research area, further research would be beneficial for the understanding of respondents' behavior with respect to nonresponse over time. For a check on the generalizability of these results, it may be desirable to test whether the negative correlation between several response types may also be found in other waves of the SOEP, or in other survey studies.

Bibliography

Dolton, P., Lindeboom, M., and G. J. van den Berg, 1998, Survey Attrition: A taxonomy and the search for valid instruments to correct for biases, Tinbergen Institute Discussion Paper, TI 98-109/3.

Frick, J. and M. Grabka, 2005, Item-nonresponse on income questions in panel surveys: incidence, imputation, and the impact on inequality and mobility, Allgemeines Statistisches Archiv 89(1), 49-61.

Loosveldt, G., Pickery, J., and J. Billiet, 2002, Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey, Journal of Official Statistics, 18(4), 545-557.

Schräpler, J.-P., 2003a, Gross-Income-Nonresponse in the German Socio-Economic Panel (GSOEP) - Refusal or Don't Know? Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 123(1).

Schräpler, J.-P., 2003b, Respondent behavior in Panel Studies - A Case Study for Income-Nonresponse by means of the British Household Panel Study (BHPS). Working Papers of the Institute for Social and Economic Research, paper 2003-8. Colchester: University of Essex;
URL: http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2003-08.pdf

Serfling, O., 2004, The interaction between unit and item nonresponse in view of the reverse cooperation continuum, WWZ-Discussion Paper 04/01, Center of Economics and Business Management, University of Basel;
URL: http://www.wwz.unibas.ch/forum/discussion/volltext/0401.pdf .

Van den Eeden, P., 2002, Unit- and Item-Nonresponse Bias in a longitudinal case, presented at ICIS (International Conference on Improving Surveys), 25-28 August 2002, Copenhagen;
URL: http://www.icis.dk/ICIS_papers/F3_5_3.pdf.

Van den Veen, W. and B. Van den Praag, 1981, The Demand for Deductibles in Private Health Insurance – A Probit Model with Sample Selection, Journal of Econometrics 7, 229-252.

IV. The Interaction between Unit and Item Nonresponse in View of the Reverse Cooperation Continuum

- Evidence from the German Socioeconomic Panel (GSOEP)

Oliver Serfling

Abstract

This study examines the interplay of unit and item nonresponse in surveys by addressing two research questions: First, is item nonresponse (INR) a precursor of unit nonresponse (UNR), as predicted by the theory of a latent cooperation continuum, or is the interrelation of another type? Second, are the results of models for item nonresponse behaviour affected by a selectivity bias due to panel attrition?

For this purpose, we investigate the response behaviour of the original first-wave participants of the German Socio-Economic Panel (GSOEP) over the first 19 panel years. We find that item nonresponse on the income question is mostly positively correlated with subsequent interview refusal. Nonetheless, we find selective attrition with respect to income INR only for a minor number of panel waves. We introduce and test the hypothesis that two types of respondents simultaneously coexist in the panel group and drop-out from the panel: one type with high INR propensity, and another type with low INR propensity. We find that the correlation between item and unit nonresponse is inversely U-shaped, which supports the hypothesis of the coexistence of both types of cooperation. This coexistence may reduce the selectivity of panel attrition with respect to INR. In contrast to our expectations, we find similar INR correlation patterns with regard to panel mortality, which allows for the conclusion that a bigger part of panel mortality is non-natural, but hidden interview refusal.

Acknowledgements

This paper is part of a research project granted by WWZ-Forum, University of Basel. The author thanks three anonymous referees, participants of the "Workshop on Item-Nonresponse and Data Quality in Large Social Surveys", Basel, October 9-12, 2003, colleagues at the Department of Economics and Business Administration, University of Basel, and especially Regina Riphahn for very helpful comments.

1 Introduction: The Nonresponse Problem

Survey data from household panels form the basis of many empirical studies on household or personal income. Such data is mostly surveyed by personal or telephone interviews. Respondents’ behaviour during the interview process may affect directly the outcomes and therefore quality and representativeness of the acquired data. Since the quality of a dataset determines the reliability of results of empirical studies, it is worth investigating the data collection process with respect to respondents' behaviour. Considerable research has been conducted on the phenomenon of unit nonresponse (UNR), and the determinants of item nonresponse (INR) are also being investigated by a growing number of studies. However, the interaction of unit and item nonresponse has been widely neglected. This study therefore pursues as first research aim the question whether panel attrition and item nonresponse are correlated or driven by a similar decision process. If so, panel attrition may cause endogenous sample selection with respect to item nonresponse and this selective attrition may cumulate over subsequent panel waves: If the attrition rate of interviewees with high INR-propensity is 10% higher than for other participants, nearly 30% more of those respondents with high INR-propensity would have quit the panel by wave 4 (see Rendtel (1989) for a similar example with respect to income). It is evident that such a selection process is non-ignorable and may possibly lead to biased regression coefficients in studies on the determinants of item nonresponse. Detecting such a bias is our second research aim.

Understanding the relationship between both types of nonresponse may, on the one hand, permit the development of techniques that jointly reduce item and unit nonresponse. At the same time, our results may improve the researchers' ability to deal with the nonresponse problem in their own analyses and to judge the need of adjustment by choosing the appropriate adjustment techniques. And lastly, it furthers the understanding of motivation and cooperation processes of respondents and the interaction with their environmental situation. Since the interview situation compares to a lot of social interaction processes, this knowledge may also be applicable to other fields of research (e.g. consumer behaviour).

This paper adds to the literature in various ways: It discusses different measures of item nonresponse in household panels. By evaluating previous studies on the interrelation of unit and item nonresponse it can be shown that – contrary to intuition - INR is not necessarily a good predictor of subsequent panel attrition. By reverting the hypothesis of a cooperation continuum, we sketch a rationale for a possibly negative association between both types of nonresponse. Using data from the participants of the first panel survey in 1984, i.e. households and their members, of the German Socio-Economic Panel (GSOEP) with repeated observations for the first 19 panel waves until 2003, we test shape and sign of the correlation of INR and UNR. The richness of our data permits us to control for effects of socio-demographic characteristics of respondents and interviewers, their interactions, the interview situation and the (re)contact process. In addition, we provide evidence that sample selection may lead to biased results in item nonresponse analyses.

The paper begins in section two with an explanation of the nonresponse phenomena drawing on a brief review from cognitive-psychology and rational choice theory. Previous hypotheses and findings of studies on the nonresponse-interrelation are summarised. We discuss the hypothesis of a latent cooperation continuum and motivate the hypothesis of a reverse cooperation continuum. The empirical strategy underlying this paper is presented in section three. The fourth section describes data and sampling criteria. The study proceeds by empirically addressing the two research questions in section five. Univariate descriptive statistics as well as multivariate regressions are employed to provide evidence for the effect of INR on UNR. To identify selective attrition bias, Heckman-type selection models are used, instrumented with characteristics from the last observed interview and the (re)contact attempt. The last section summarizes the key findings and points to the areas in need of further investigation.

2 Towards a Theory of Nonresponse

Many studies on unit and item nonresponse are of a descriptive nature and are criticised for their lack of theoretical reflection. Although a large and interdisciplinary theoretical effort has been made in psychology and sociology, a unique theory of "survey questioning" still does not exist (cf. Schnell, Hill, Esser 1995). In the following, we define the types of nonresponse, review the two prominent theoretical frameworks for the explanation of respondent behaviour and the literature addressing the interrelation of unit and item nonresponse. We conclude this section by proposing our research hypotheses.

2.1 Types of Nonresponse in Panel Data

Nonresponse may occur at several stages of the interview process: Unit Nonresponse (UNR) in general describes the drop-out of a household or person from the respondents group. When it comes to panel surveys, where the same individual is repeatedly interviewed, one can distinguish between initial unit nonresponse and panel attrition. In contrast to initial unit nonresponse, it is a prerequisite for panel attrition that the respondent has participated in at least one previous interview. This panel attrition can be divided up into (natural) panel mortality and interview refusal. Panel mortality describes all drop-outs that occur due to problems in the re-contact or re-interview process, like e.g. death, bad health or non-reachability of the interviewee. Even if this type of panel attrition may have a large stochastic component, it may still be partly correlated with observable attributes (e.g. age or health-status), but should not be related to item nonresponse. Nevertheless, interview refusal may also be communicated to the interviewer "by a policy of closed doors" (cf. Rendtel (1988), p. 38), and would falsely be classified as "non-reachability" (panel mortality) by the survey administration. Such refusal may be related to item nonresponse.

However, for the purposes of this study, which focuses on respondents behaviour, the refusal of the request to participate in the subsequent interview is more interesting, since it may provide information on the determinants of the interviewee´s propensity to cooperate. Therefore, we define UNR as being an actively communicated refusal of interview participation of previous panel study respondents.

Item Nonresponse (INR) describes the fact that a respondent is taking part in the interview, but refuses to answer a specific question. This does not comprise the choice of "don't know" or "no opinion" - options offered by the interviewer or the questionnaire. INR may also include a stochastic component comparing to natural panel mortality, since interviewers and respondents may erroneously skip items. Erroneous INR may be related to socio-demographic characteristics (e.g. older interviewees and stressed interviewers may more often fail to process the item) but should be uncorrelated with the attitude of the respondent towards the survey or the item.

2.2 Determinants of Nonresponse

The cognitive model of respondent behaviour extends prior psychological models of thought processes (cf. Lachman et al. 1979) and structures the question-answering process, by defining the tasks a respondent has to accomplish before providing an answer: After hearing or reading a question the respondent must interpret it, recognise the issue addressed, contextualise the meaning of the question within the interview, gather the information needed to give an appropriate answer, take into account subjective motives like self-representation or social desirability and modify the "true" answer to then communicate the so derived information to the interviewer (cf. Sudman et al. 1996). Rational choice theory assumes that a respondent evaluates his response alternatives and accounts for the expected costs and benefits of his possible actions. Then he opts for the alternative with the highest subjective expected utility (cf. Esser 1986).

By broadening the definitions of costs and benefits, the cognitive model can be integrated into the rational choice framework, if we assume that people with cognitive difficulties have to put more effort in answering a given question(naire), or that items / questionnaires requiring a high cognitive effort will involve higher costs for respondents with given cognitive ability. This mechanism may be relevant for item and unit nonresponse.

Even if most of the hypothesised determinants are unobservable, the nonresponse literature has shown that nonresponse can be partly explained by observable proxies such as characteristics of the interview situation, the personal characteristics of respondents and interviewers, as well as their interactions (see e.g. Groves 1989).

2.3 Relationship of Item and Unit Nonresponse

2.3.1 Positive Correlation of INR and UNR

In the literature, evidence on the relationship between unit and item nonresponse is scarce. It is often suggested that both types of nonresponse result from the same decision process, which is driven by interest, motivation and ability of the respondent (cf. Loosveldt et al. 2002). Some panel studies observe the joint decline of item and unit nonresponse rates over time (see e.g. Van den Eeden 2002). This finding may be explained by self-selection of respondents: over time only motivated respondents stay in the group of panel participants, and they have low item nonresponse propensities.

The cooperation continuum hypothesis, introduced by Burton et al. (1999), can be regarded in a similar vein: According to this hypothesis, potential survey respondents can be placed on and move along a cooperation continuum of item and unit nonresponse probability correlations. This continuum spans the categories from "will always take part and answer any question" over "hard to persuade and will refuse a lot" to "will never take part". The authors use empirical evidence from the British Household Panel Study (BHPS) to support the conclusion that persons with a high motivation to participate are also likely to respond and vice versa. Respondents with no missing data on key variables were also very likely to complete a full interview in the next wave. Panel drop-outs had a higher item nonresponse in the prewave. Intermittent respondents, i.e. respondents who suspended at least one previous interview, had higher item nonresponse than regular panel participants. Additionally, the conversion of initial drop-outs led to higher item nonresponse rates of these persons.

Summing up, the hypothesis of the cooperation continuum suggests a positive correlation of unobservable *a priori* probabilities of INR and UNR, which is depicted in Figure 1. Empirical evidence for this is provided by several studies: Loosveldt et al. (2002) find that item nonresponse on difficult questions in the first panel wave significantly raises the refusal probability in the second wave of the Belgian General Election Study. Schräpler (2003) finds a small but significant negative correlation between refusing the gross income statement and participation in the next wave of German Socio-Economic Panel (GSOEP) over the first twelve years. This finding is also confirmed by Frick and Grabka (2005).

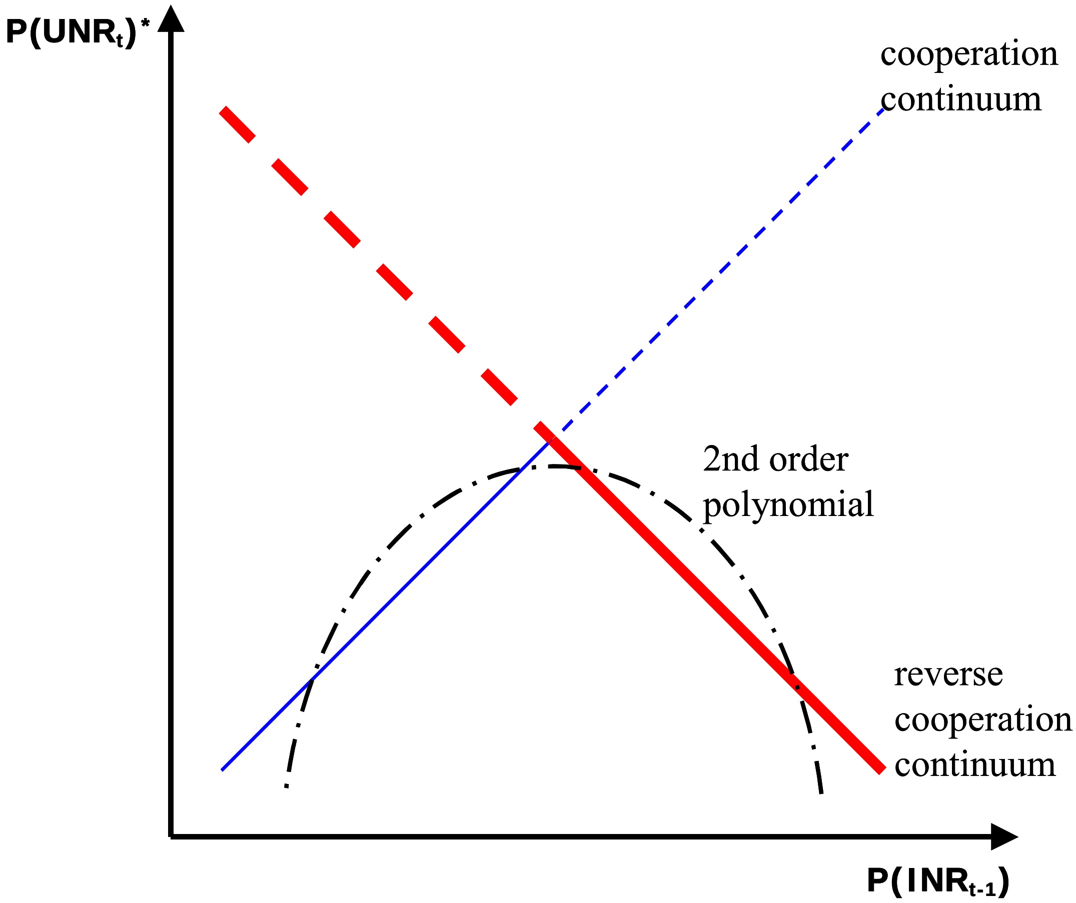
2.3.2 Negative and Nonlinear Interaction of INR and UNR

On the other hand, empirical evidence also exists which does not support the cooperation continuum hypothesis. Dolton et al. (1998) found that item nonresponse rate and interview duration do not explain panel attrition, not even in the first wave of the panel. Van den Eeden (2002) hypothesized that both INR and UNR are caused by the lack of respondents' motivation. Using data from the first seven waves of the Longitudinal Aging Study of Amsterdam (LASA), he regressed unit nonresponse on the INR rate, which he claims to be a proxy for motivation. Nonetheless, his results showed only an extremely low explanatory power of its model, so he concludes that INR does not express the lack of motivation.

Based on these findings and contrary to the theory of the cooperation continuum we argue that the correlation of INR and UNR probabilities can also reverse (see Figure 1). In other words, interviewees may show appetite to participate in an interview but are unwilling to provide informative answers and vice versa. We label the phenomenon describing that there is a negative correlation of the unobserved *a priori* probabilities of unit and item nonresponse the "reverse cooperation continuum". From this perspective, respondents are only willing to take part in the interview because they know that they are not willing to provide informative answers. This behaviour may be rational in the sense of a cost-benefit calculus: The respondent participates in the interview to receive the appreciation of the interviewer or the survey organisation and uses item nonresponse as a strategy to minimize subjective expected costs of answering questions (low P(UNR) with high P(INR), as illustrated by the bold faced line in Figure 1). On the upper part of this line, people are extremely conscientious and willing to answer every question posed. However, since they do not know if they are able to provide an exact answer to every question or know exactly the total (time-) costs of answering every question, they are likely to refuse participation (UNR concurrently with low P(INR)). Clearly, this low INR-probability stays unobserved if the respondent decides to refuse the interview. Therefore, only the lower part of the bold line of figure 1 may be observable in the panel data. In the light of this reverse cooperation continuum, refusals should have lower item nonresponse rates in the year before dropping out than seen from the cooperation continuum perspective.

If both types of respondents, "cooperators" and "reverse cooperators", appear in a panel sample, the INR-propensity will reflect two opposing mechanisms as a predictor of UNR. The effects of both mechanisms on UNR may balance. Hence, in a linear model, the coefficient of INR may be close to zero, or insignificant as observed in some of the studies quoted above. As said, it is unlikely to observe respondents with high a priori unit nonresponse probability in a panel study. Therefore, the observable part of the interactions between unit and item nonresponse has an inverse U-shaped pattern and can be empirically approximated by a second order polynomial of the INR propensity in a model of UNR (see Figure 1).

Figure 1: Possible correlations of latent unit nonresponse and pre-wave´s item nonresponse probabilities



Notes:
Dashed lines are less likely to be observed in a panel study because of high unit nonresponse probability.
Second order polynomial describes occurrence of both respondent types for lower unit nonresponse probability.

2.4 Hypotheses

To investigate the relationship between item nonresponse and unit nonresponse probabilities, we formulate five hypotheses which we test below:

H1: INR is a precursor of panel attrition, i.e. in the year before drop-out, refusals show a higher INR-propensity than subsequent respondents, as predicted by the cooperation continuum-hypothesis.

H2: The drop-out decision of respondents who move along a cooperation continuum of positive INR and UNR correlation, is taken within the first few waves. Therefore, the relative importance of INR on subsequent UNR decreases and vanishes in the course of panel duration.

H3: A portion of respondents behave as predicted by our reverse cooperation continuum hypothesis. Accordingly some refusals have lower INR-propensity than subsequent respondents in the year before UNR. Both types of respondents coexist in the panel: one "cooperation continuum type" and one "reverse cooperation continuum type".

H4: In contrast to interview refusal (UNR), panel mortality is uncorrelated with previous INR and mainly explicable by problems of the survey organisation to re-interview the respondent.

H5: UNR biases the results of INR-analyses due to self-selection of panel participants. This bias grows in the course of time, since selective drop-outs cumulate over subsequent waves.

3 Empirical Approach

To address the first and second hypotheses, we first look at differences in the item nonresponse propensities in wave t-1 for respondents and drop-outs of subsequent wave t. As proxies for the latent INR-propensity we use:

a) the dichotomous outcome of item nonresponse on the income question (IncINR)
b) the share of item nonresponses on all applicable questions in the survey (INR rate)

The income question is one of the most sensitive items in a comprehensive survey and suffers from item nonresponse. Therefore it is commonly used as a measure for the INR behaviour of the respondent. The shortcoming of this discrete binary INR measure is that it mixes a latent cooperation level with the question-specific effect of income questions, e.g. the negative attitude towards income questions or privacy concerns. Furthermore, the income item is very often solely applicable to employed respondents. Therefore, we also use the second measure, the INR rate, which is computed as the share of item nonresponses on the number of questions applicable to that person, for all survey items where at least one INR occurred. The offering of a "don't know" or "no opinion" - option may influence the frequency of item nonresponse. In a recent study, Riphahn and Serfling (2005) have shown that INR and "don’t know"-responses differ with respect to their observable and unobservable determinants. Therefore, and because of the very small (<5) number of items with a "don't know" option in the GSOEP questionnaires, we treat "don't know" statements as valid responses.

With respect to the first hypothesis (H1), we use a t-test for the equality of the share of income INRs of subsequent participants and refusals. Additionally, we test the equality of the mean INR-rate for both groups. Addressing the second hypothesis, we reiterate both tests for each panel wave.

To seize the effect of possible determinants on the nonresponse decision, we formulate a model of the response decision process using a random utility hypothesis relying on the rational choice theory. Accordingly, an individual does not respond if the expected costs of participation exceed the expected utility:

$$P(\mathit{Nonresponse}) = P(\mathit{Resp.\ Costs} > \mathit{Resp.\ Benefits}) = P(y^* > 0) \qquad (3.1)$$

(Since we use this framework to examine both UNR and INR, we sketch it in the broader term of "nonresponse".)

To evaluate the covariates of this nonresponse decision we formulate a binary choice model where the respondent decides for (item or unit) nonresponse if a latent response index y^* exceeds zero:

(Nonresponse = 1 if $y^* > 0$ and Nonresponse = 0 if $y^* \leq 0$).

This latent response propensity y^* should reflect the utility surplus of nonresponse, which is given by the cost-benefit considerations of the interviewee. The cost and benefit elements of this latent nonresponse are represented by: (1) socio-demographic characteristics of the respondent i at time t (R_{it}) like age, sex, education and family status, (2) characteristics of the re-contact process and the interview situation (S_{it}), and in cases where an interviewer j was present we use (3) the characteristics of the interviewer (I_{jt}) and (4) interactions ($R_{it}*I_{jt}$) between interviewer j's and respondent i's characteristics and (5) a baseline cost-surplus of response at time t (α_t) (panel effect):

$$y_{it}* = \alpha_t + R_{it}\beta_1 + S_{it}\beta_2 + I_{jt}\beta_3 + (R_{it} * I_{jt})\beta_4 + \varepsilon_{it} \qquad (3.2)$$

We assume a Gaussian standard normal distribution for the error term ε_{it} in the econometric model, which transforms (3.2) into a probit model. Within this general framework, our investigation focuses on the next research hypotheses: To address the second hypothesis, we include the Income-INR dummy in 3.2 and separately estimate this equation (model 1) for each wave. To address the third hypothesis, we alternatively include the INR-Rate and its square (model 2), test for sign and significance of their parameters, as well as for joint significance. Hypothesizing an inverse U-shaped relationship between INR and UNR, we expect the coefficient for the linear effect to be positive and the coefficient for the quadratic effect to be negative. With respect to the regressions on the household level, we had to consider that the attrition decision is taken at the household level. Therefore, the occurrence of UNR across individuals of the same household may not be treated as independent, since the household mostly drops out completely. Nonetheless, the households themselves are assumed to act independently. We therefore correct the variance estimate for this within household correlation of the UNR decision (i.e. intra-cluster correlation, see Rogers (1993) and Williams (2000)) in the regressions on the individual level.

The results for wave-specific regressions may not be reliable with respect to the results for Income-INR and INR-Rate since wave-specific effects with influence on the nonresponse decision may exist. Hence, we regress the UNR indicator on Income-INR and INR-Rate for all respondents who dropped out until 2002 or are still participating in model 3. To determine the influence of panel effects, the set of explanatory variables is augmented by panel fixed-effects in a second specification, and their joint significance is tested.

To check our fourth hypothesis - whether panel mortality is uncorrelated with INR-behaviour, that is - we extend our binary choice model presented above to a multinomial choice model. The set of choice-options is extended by the category "mortality" which was ignored in models 1-3. In model 4, we regress the participation status-measure (with the discrete outcomes 1: Response, 2: UNR, 3: Mortality) on the set of explanatory variables from model 3, employing a multinomial logit estimation. We check the separate and joint significance of the INR-proxies.

To address the second research question of this study, we put the fifth hypothesis to the test: the influence of sample selection on the results of an INR-analysis. Therefore, we regress the dichotomous income INR indicator on possible determinants of INR by taking into account sample selection. Hence, we use a Heckman-type bivariate probit selection model which is illustrated in the following:

Applying standard estimation methods to non-randomly selected samples leads to biased coefficients if the expectation of the error term is non-zero and dependent on a selectivity process that is correlated with the regressors of interest. Heckman (1979) introduces a two-stage estimator that allows to consistently estimate behavioural functions based on non-randomly drawn samples with a least squares method, imposing distributional assumptions on the error term structure. As an extension of the Heckman approach with respect to dichotomous dependent variables, Van den Veen and Van Praag (1981) introduced a bivariate probit estimator to obtain unbiased coefficient estimates for selective samples. This bivariate probit model consists of two estimation equations: first, a specification equation with the function of interest, here: the probability of item nonresponse, and second, a selection equation, which determines the probability of observing an observation's outcome in the specification equation. The structural threshold model for the dichotomous outcome compares to 3.1 and 3.2:

$$INR_{i,t} = \begin{cases} 1 \ if \ y_{i,t}^* < 0 \\ 0 \ otherwise \end{cases} \qquad (3.3)$$

$$with: \ y_{i,t}^* = \alpha + X_{i,t}\,\beta + \mu_{i,t}$$

where α is the constant, $X_{i,t}$ are the explanatory variables for individual i in period t, β is the vector of regression coefficients and $\mu_{i,t}$ the error term of the specification equation.

Item nonresponse ($INR_{i,t}=1$) of respondent i at time t occurs if the unobserved subjective expected utility of answering a question ($y_{i,t}^*$) is negative. The opposite applies to the selection equation, which can shortly be written as:

$$UR_{i,t} = (\gamma + Z_{i,t-1}\,\delta + \eta_{i,t} \geq 0) \qquad (3.4)$$

where γ is the constant, δ the coefficient vector and $\eta_{i,t}$ is the error term of the selection model. This equation determines whether the individual is observed at time t (unit responded: $UR_{i,t}=1$) or dropped out of the sample (unit nonresponse: $UR_{i,t}=0$). Following the rational choice framework, the respondent will remain in the sample if the unobserved subjective utility is positive or zero and drop out if it is negative.

The regressors of the selection equation Z=(X,W)' consist of the regressors of the specification equation X and additional regressors W which have explanatory power for unit nonresponse without affecting item nonresponse and thus being instruments for panel attrition. Furthermore, it is assumed that the error terms μ and η are bivariately standard normally distributed with correlation ρ.

$$\begin{pmatrix} \mu_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \tag{3.2.3}$$

As Heckman (1979) has shown, a self-selection bias exists if the error terms are correlated ($\rho \neq 0$). The significance of ρ is tested by means of a likelihood-ratio test, after the maximum-likelihood estimation of the bivariate probit model.

To illustrate the magnitude of the attrition bias, we finally compare our estimation results for the determinants of item nonresponse derived from the bivariate probit approach with those obtained when ignoring panel attrition.

4 Data and Sampling

The data analysed in this study are taken from the first 19 waves of the German Socioeconomic Panel (GSOEP) between 1984 and 2003. The GSOEP data are collected annually and contain information on household and individual characteristics, which are surveyed in two separate questionnaires. The household questionnaire is answered by the head of the household, while the individual questionnaires are filled in by every member of the household who reached age 16. In 1984, the GSOEP started with two subsamples A (German natives) and B (immigrants, guest-workers), surveying 5,921 households (among those 1,393 immigrant households) in West-Germany with 12,290 individuals (thereof 3,175 living in immigrant households). The GSOEP administration uses a method-mix to gather the desired information, including paper and pencil interviews, face-to-face interviews, and telephone interviews. Within face-to-face interviews, there is the possibility that the respondent administered the questionnaire on his own, while the interviewer was present (for further description of the GSOEP see Hanefeld (1987)). This allows us to determine mode-effects on the nonresponse behaviour. The survey is supplemented by an interviewer dataset (see Schräpler and Wagner 2001), so measures for interviewers and interactions between interviewers and respondents are available for face-to-face interviews.

For the purpose of this study, we use only the original participants of the 1984 interview and observe their response behaviour until 2003. Using the gross information from the panel interviews, we define three types of respondents: (R) respondents who were participating in the subsequent panel interview, (UNR) respondents who communicated that they were not willing to participate, and (MORT) drop-outs due to natural or non-natural panel mortality (non-reachable or not interviewable). Since the survey administration follows participants when their former household splits up (e.g. children leaving parents' home, divorce, etc.) "new" panel households are generated. We expect the response behaviour of these households to be different from their parental household, and dropped the "new" households with their "old" members from the sample. Since the response behaviour is attributed to the interviewee, even in the household questionnaire, we restricted our household sample to observations where the same person answered the household questionnaire all subsequent waves. This involved the loss of about half of the household observations. Lastly, we dropped some respondents and households who intermitted the survey for one or more years. The distribution of response types over all 19 panel waves is presented in Table 1a.

The INR measures are constructed as follows: Income INR in the household questionnaire was coded 1 if the question on the total income of all household members was not answered and 0 otherwise. In the personal questionnaire, IncINR was coded 1 if neither the gross wage nor the net wage question were answered. Both questions were only applicable to those who were employed, which was determined by filter questions. The INR-Rate was computed for both, the household and the personal questionnaire, as the number of non-responded but applicable items divided by the whole number of applicable items. Here, only those items where at least one item nonresponse occurred in the survey were taken into account. As stated earlier, "don't know" responses were not seen as INR. In the GSOEP survey a "don't know" option was not offered with the income questions in the personal and household questionnaires.

5 Empirical Analysis of Nonresponse Interaction

This analysis addresses two main research questions: First, is there evidence, that item nonresponse is a precursor of unit nonresponse, and does it have explanatory power for subsequent panel attrition behaviour, or is there another type of interaction? Second, does attrition cause endogenous sample selection which biases the coefficients of INR-regressions?

5.1 Item Nonresponse as a Precursor of Panel Attrition

According to the theory of the cooperation continuum, panel attrition should be preceded by a higher item nonresponse propensity. In Table 1b, we present the t-test results for the test of equality in mean income item nonresponse for drop-outs and stayers over all 19 panel years, separately for the household and personal questionnaires. It is obvious that the mean number of income item nonresponses of subsequent drop-outs exceeds that of subsequent respondents. The overall household's income nonresponse probability is with 10.8% more than twice as large for subsequent drop-outs than for respondents. The same can be seen in the individual questionnaires, albeit at lower nonresponse frequency levels. In the three waves, where the difference in the Income INR frequency was negative, it was found to be not significantly different from zero. The same holds for the difference in the INR-Rates of drop-outs and stayers, which are presented in table 1c: The INR-Rate of drop-outs is significantly higher than that of stayers. For the household sample, we see a period between 1988 and 1995 where the differences are mostly insignificant.

Due to sample selection, we expect that the difference in INR behaviour of participants and refusals vanishes within the first waves of the panel since those who are not cooperative should drop out. Evidence for this hypothesis is not obvious. The difference in the frequency of Income INR and the INR-Rate does not decline over the panel waves. With respect to the household questionnaires, we find large significant differences in the first years of the panel (1984-1987) and the last observed waves (1997-2002). With respect to the individual questionnaires, we see a slight decline of the differences in both measures, frequency of income INR and INR-Rate. Nonetheless, these differences are significant with respect to the INR-Rate, and mostly significant with respect to income INR. So far, we found evidence for H1 but no evidence for H2. Against this descriptive evidence, it may be argued that unit nonresponse is also affected by other determinants than the INR-propensity. Therefore, we regressed the UNR indicator on an extensive model of possible determinants of interview refusal in the probit model framework. In order to reduce the mentioned heterogeneity with respect to the UNR-decision and to avoid an omitted variable bias, we included a set of 29 explanatory variables, describing the respondent characteristics, interview situation characteristics, the re-interview attempt, interviewers characteristics, and interactions between interviewers and interviewees. (The detailed list of regressors is depicted in Table 3).

Table 2 briefly summarises only the marginal effects of the INR-proxies (Income INR dummy in model 1 and INR-Rate and its square in model 2) on the unit nonresponse probability. The results of model 1 show clear evidence for H1: The effect of income INR is positive if significantly different from zero, and has an obvious decreasing trend over the panel years, with exception of a large positive significant effect in 2002. The results until 2001 support our second hypothesis.

Concerning the third hypothesis, model 2 is informative. We proposed the coexistence of two types of respondent cooperation: one generating higher INR before drop-out, the other generating low INR before drop-out. As argued in section 3, we see evidence for this hypothesis if the linear effect of the INR-Rate is positive and the effect of the squared INR-Rate is negative. Additionally, we expect that both coefficients should be jointly significant and the indicated maximum should lie in a reasonable range for the INR-Rate (between 0 and 100) to support our coexistence hypothesis. The last columns of Table 2 show the calculated maximum and the χ^2 distributed test statistic of a Wald test on joint significance. As it can be seen, we find evidence consistent with H3 in 7 out of 19 panel waves with the household questionnaires and 11 out of 19 waves with the individual questionnaires. These findings are consistent with our third hypothesis. Counter-evidence is scarce: in the personal questionnaires we find in 1992 and 1995 a maximum out of the domain of INR-Rates (>100), which we interpret as evidence against the reverse cooperation continuum. The same holds for the 1997 and 2001 personal interview and the 1992 household interview, where we find significant continuous positive effects of the INR-Rate without a maximum. These results may be seen as evidence in favour of the cooperation continuum hypothesis. Additionally, in the household interview of the year 2000 we find a significant U-shaped correlation pattern of INR and UNR with a minimum. The reported significant effects are robust with respect to models with smaller subsets of explanatory variables (not presented here).

These findings indicate that the reverse cooperation continuum hypothesis, the coexistence and simultaneous drop-out of cooperators and reverse cooperators may have explanatory power in their own right. Nonetheless, 13 out of 38 regression results provided no evidence neither for the first nor for the third hypothesis.

Table 1: Differences in income INR frequency and INR rate of subsequent refusals and participants

	(a) Participation Status				(b) Income INR					(c) INR-Rate				
		... with subsequent												
Wave	Total # of participants ...	Response	UNR	Mortality	E(INR|UNR=0)	E(INR|UNR=1)	diff.	t-value		E(INR|UNR=0)	E(INR|UNR=1)	diff	t-value	
Household Questionnaire:														
1984	2'635	1'990	549	96	5.7	10.0	4.3	3.7	***	1.8	2.2	0.5	2.4	**
1985	1'990	1'707	248	35	4.9	14.9	10.0	6.2	***	2.5	4.7	2.2	5.4	***
1986	1'707	1'574	99	34	5.2	9.1	3.9	1.7	*	2.2	4.1	1.9	3.7	***
1987	1'574	1'439	100	35	4.3	13.0	8.7	3.9	***	2.1	4.7	2.6	4.5	***
1988	1'439	1'321	96	22	4.1	5.2	1.1	0.5		2.6	3.8	1.2	1.7	*
1989	1'321	1'248	55	18	4.0	5.5	1.5	0.5		2.2	3.4	1.2	1.4	
1990	1'248	1'182	41	25	3.5	9.8	6.3	2.1	**	2.3	3.1	0.9	0.9	
1991	1'182	1'122	38	22	4.4	7.9	3.5	1.0		2.7	3.7	1.1	1.2	
1992	1'122	1'081	23	18	4.2	17.4	13.1	3.0	***	3.0	8.3	5.4	3.8	***
1993	1'081	1'014	50	17	4.2	4.0	-0.2	-0.1		2.2	1.9	-0.2	-0.3	
1994	1'014	957	38	19	4.3	5.3	1.0	0.3		2.8	4.3	1.4	1.4	
1995	957	912	25	20	5.1	8.0	2.9	0.6		3.1	3.8	0.7	0.6	
1996	912	875	30	7	5.6	26.7	21.1	4.7	***	2.4	5.0	2.6	2.9	***
1997	875	829	32	14	4.3	6.3	1.9	0.5		4.1	8.4	4.4	3.1	***
1998	829	785	33	11	3.4	12.1	8.7	2.6	**	2.8	8.0	5.3	5.3	***
1999	785	745	27	13	4.0	11.1	7.1	1.8	*	2.5	4.8	2.3	2.3	**
2000	745	709	24	12	4.4	12.5	8.1	1.9	*	2.5	4.0	1.5	1.2	
2001	709	673	26	10	4.0	23.1	19.1	4.6	***	2.1	6.9	4.8	4.8	***
2002	673	643	25	5	3.6	16.0	12.4	3.1	***	1.6	4.5	2.9	3.9	***
all	22’798	20’806	1559	433	4.5	10.8	6.3	11.2	***	2.4	3.7	1.3	8.7	***
Individual Questionnaire:														
1984	9'883	8'421	1084	378	3.5	5.5	2.0	3.3	***	7.3	8.5	1.3	5.51	***
1985	8'421	7'527	706	188	3.7	7.8	4.1	5.3	***	2.2	3.3	1.1	6.76	***
1986	7'527	6'997	375	155	3.3	6.7	3.3	3.4	***	2.4	4.0	1.6	7.79	***
1987	6'997	6'419	407	171	2.8	5.9	3.1	3.6	***	1.9	3.3	1.4	7.32	***
1988	6'419	5'937	356	126	2.8	6.7	3.9	4.2	***	1.9	2.7	0.8	4.49	***
1989	5'937	5'605	230	102	2.8	6.5	3.7	3.3	***	2.4	4.9	2.5	8.51	***
1990	5'605	5'327	180	98	2.4	8.9	6.5	5.5	***	2.7	4.7	2.0	5.43	***
1991	5'327	5'075	159	93	2.7	4.4	1.7	1.3		2.3	3.7	1.4	4.44	***
1992	5'075	4'832	149	94	2.6	7.4	4.8	3.6	***	2.0	4.0	2.1	5.8	***
1993	4'832	4'553	188	91	2.2	4.3	2.1	1.9	*	2.4	4.1	1.7	4.9	***
1994	4'553	4'287	170	96	1.9	4.7	2.8	2.6	**	1.6	4.0	2.4	7.94	***
1995	4'287	4'071	135	81	4.0	3.0	-1.0	-0.6		2.8	5.6	2.8	7.91	***
1996	4'071	3'870	145	56	2.1	2.8	0.7	0.6		1.8	3.4	1.6	6.13	***
1997	3'870	3'658	151	61	2.5	5.3	2.8	2.2	**	2.8	4.0	1.3	4.25	***
1998	3'658	3'437	151	70	2.1	4.0	1.9	1.6		4.5	6.0	1.5	5.13	***
1999	3'437	3'279	99	59	2.1	2.0	-0.1	-0.1		1.3	2.6	1.3	4.4	***
2000	3'279	3'123	98	58	2.2	6.1	3.9	2.6	**	1.7	2.9	1.2	3.41	***
2001	3'123	2'964	117	42	2.5	4.3	1.7	1.2		1.9	2.9	1.0	2.71	***
2002	2'964	2'796	124	44	2.4	9.7	7.3	4.9	***	1.9	3.1	1.2	4.01	***
all	99'265	92'178	5024	2063	2.8	6.0	3.2	13.0	***	2.7	4.8	2.1	30.3	***

Notes: Significance levels: * 10 %, ** 5 %, *** 1 %; Income INR frequency, INR-Rates in %

Table 2: Marginal effects of income INR (Model 1) on UNR and coefficients of INR Rate (Model 2) of UNR probits

		Model 1			Model 2								
		Income INR			INRRate			$INRRate^2$			Max.	Wald-Test	
Wave	# obs.	ME	t		coef.	t		coef.	t			χ^2	
Household Questionnaire													
1984	2619	0.117	3.16	***	0.022	1.81	**	0.000	-0.62		35.8	5.34	*
1985	1958	0.148	3.62	***	0.034	3.54	***	0.000	-1.37	*	49.3	18.33	***
1986	1676	0.016	0.65		0.076	3.53	***	-0.002	-2.23	**	17.6	15.47	***
1987	1542	0.093	2.22	**	0.053	3.03	***	-0.001	-1.66	*	26.0	13.00	***
1988	1420	-0.002	-0.08		0.004	0.31		-0.000	-0.04		182.6	0.25	
1989	1306	-0.005	-0.34		-0.012	0.63		0.000	0.60			0.40	
1990	1070	0.035	0.93		-0.004	0.21		0.000	0.10			0.07	
1991	1142	0.009	0.54		0.009	0.30		0.000	0.00			0.42	
1992	1107	0.028	1.18		0.032	1.25		0.000	0.10		+ +	9.16	**
1993	1067	-0.011	-0.65		0.004	0.10		-0.001	-0.53		1.7	0.75	
1994	992	-0.006	-0.44		-0.023	0.84		0.000	0.23			1.46	
1995	930	0.019	0.76		0.015	0.53		-0.000	-0.06		198.0	0.72	
1996	908	0.053	1.58	*	0.074	1.82	**	-0.002	-1.06		18.3	4.53	
1997	864	-0.002	-0.14		0.030	1.30		0.000	-0.70		38.7	2.51	
1998	821	0.067	1.25		0.060	2.56	***	0.000	-0.74		80.6	14.84	***
1999	775	0.031	0.84		0.021	0.59		0.000	-0.05		181.6	1.31	
2000	736	0.008	0.83		-0.029	0.56		0.002	1.27		Min	5.03	*
2001	702	0.070	1.44	*	0.115	2.91	***	-0.002	-1.91	**	23.7	10.88	***
2002	668	0.047	1.08		0.086	1.71	**	-0.001	-0.55		34.9	7.42	**
All	22'496	0.067	6.88	***	0.023	5.99	***	0.000	-2.76	***	41.5	51.53	***
Individual Questionnaire													
1984	10747	0.056	2.70	***	0.015	2.85	***	-0.000	-1.25		99.5	15.21	***
1985	9312	0.051	2.54	***	0.040	3.86	***	-0.001	-2.50	***	27.5	19.65	***
1986	8269	0.027	1.89	**	0.024	2.27	**	-0.000	-0.19		80.2	13.81	***
1987	7582	0.042	2.13	**	0.041	3.60	***	-0.001	-1.81	**	31.2	19.24	***
1988	6930	0.025	1.51	*	0.013	0.98		-0.000	-0.58		21.9	1.11	
1989	6351	0.008	0.72		0.041	3.23	***	-0.001	-2.39	**	26.0	11.89	***
1990	5936	0.045	2.25	**	0.017	1.32	*	-0.000	-0.05		97.0	8.34	**
1991	5578	0.012	0.80		0.033	2.13	**	-0.001	-1.34	*	22.7	5.01	*
1992	5264	0.036	1.98	**	0.019	1.63	*	-0.000	-0.13		340.2	10.76	***
1993	4967	0.026	1.26		0.023	1.82	**	-0.000	-1.55	*	23.7	3.31	
1994	4639	0.027	1.34	*	0.046	3.15	***	-0.001	-1.50	*	41.7	16.63	***
1995	4358	-0.009	-1.02		0.030	1.89	**	-0.000	-0.13		232.2	15.33	***
1996	4135	0.008	0.48		0.050	2.59	***	-0.001	-1.14		22.9	9.77	***
1997	3914	0.027	1.30		0.003	0.16		0.000	0.41		+ +	1.42	
1998	3670	0.032	1.05		0.051	2.33	**	-0.001	-1.30		32.0	7.94	**
1999	3444	-0.002	-0.16		0.051	1.88	**	-0.001	-1.48	*	23.8	3.65	
2000	3263	0.020	1.06		0.060	2.03	**	-0.003	-1.83	**	10.6	4.17	
2001	3113	0.005	0.30		0.016	1.00		0.000	0.37		+ +	3.46	
2002	2920	0.081	2.19	**	0.102	3.17	***	-0.004	-2.24	**	13.2	12.28	***
All	104'392	0.037	6.27	***	0.025	10.6	***	-0.000	-5.47	***	55.8	137.24	***

Note: Significance levels: * 10 %, ** 5 %, *** 1 %; ME: marginal effect, evaluated at mean.
+ +: strictly increasing effect; Min: convex effect with minimum.

The evidence from the results of these wave-specific regressions has the shortcoming that the identification of wave-specific fixed effects is not possible. Therefore, we restricted the coefficients to be wave-invariant and performed a pooled regression of the UNR indicator for all drop-outs until 2002 and the respondents of 2002 in model 3. (The outcomes of the RHS-variables stem from the last observed interview, i.e. 2002 for all respondents (UNR=0) and the year before the drop-out for all UNR=1). Additionally, we included both INR-proxies, Income INR-dummy and INR-Rate, to determine which effect is stronger. (Since the correlation between Income INR and INR-Rate is low, collinearity should not be an issue for that regression. Otherwise, the estimated standard errors would be too large, which does not lead to spurious evidence for hypothesis H3.) The models 3b and 3d include panel wave fixed-effects to check the robustness of our results.

Table 3 presents the marginal effects on the UNR probability evaluated at the mean of our data. The results for model 3a show that there is an inverse U-shaped correlation pattern of INR-Rate and UNR-probability on the household level. This result is robust with respect to panel effects, which are additionally controlled in model 3b. In both models, the effect of the INR-Rate on UNR is jointly significant and has a maximum at an INR-Rate of 33.3%, and 24.2% respectively (see rows at the bottom of table 3). Apparently, income INR has no strong additional explanatory power for subsequent interview refusal. This changes when it comes to the personal questionnaires: the UNR-probability raises by 6.6% if the income question in the preceding interview remained unanswered, and this effect increases to 12.4% in the model including panel effects (compare models 3c and 3d). In contrast, the INR-Rate and its square loses explanatory power as soon as panel fixed-effects are controlled. With respect to other determinants of UNR, we can see that unemployed and older respondents are more likely to stay in the panel group: The UNR-probability decreases by 1.5 percentage points per year. In contrast, married respondents, unemployed interviewers, same school level of respondent and interviewer, survey modes without interviewer, number of interview attempts before contact, and a change of the interviewer induce higher UNR-probability. We find significant interview situation effects and respondent-interviewer interaction effects except for model 3b. These findings are in line with the survey literature.

Table 3: Marginal effects of UNR probit on income INR and INR rate (Model 3), sample: drop-outs until 2002 and respondents in 2002.

LHS: P(UNR=1 | X)	Household Questionnaire (Model 3a)			(Model 3b)			Individual Questionnaire (Model 3c)			(Model 3d)		
	ME	t		ME	t		ME	t		ME	t	
INR measures												
INR Rate	0.02	4.9	***	0.01	2.2	**	0.02	8.1	***	0.00	0.9	
INR Rate^2	-0.00	-3.3	***	-0.00	-2.1	**	-0.00	-5.9	***	-0.00	-0.4	
Income INR	0.08	1.8	*	0.06	1.1		0.07	1.9	*	0.12	3.0	***
Respondent characteristics												
sex: female	0.07	2.7	***	0.07	2.1	**	0.01	0.7		0.08	5.2	***
age	-0.02	-13	***	-0.00	-0.8		-0.01	17.2	***	-0.00	-2.3	**
medium schooling	-0.05	-1.7	*	-0.00	-0.0		0.02	0.9		0.06	3.1	***
higher schooling	-0.12	-3.3	***	0.03	0.6		0.01	0.5		0.08	2.8	***
university graduate	0.03	0.6		-0.04	-0.5		-0.09	-3.1	**	-0.07	-1.9	*
part time employed	-0.02	-0.5		0.04	0.6		-0.10	-4.8	***	-0.03	-1.4	
not employed	-0.04	-1.1		-0.14	-3.5	***	-0.09	-5.4	***	-0.16	-8.9	***
public employee	-0.02	-0.5		0.07	1.4		0.08	4.1	***	0.05	1.9	*
Married	0.16	5.0	***	0.34	8.3	***	0.04	2.3	**	0.11	6.4	***
head of household	-	-		-	-		0.02	0.9		-	-	
Interviewer characteristics												
sex: female	-0.06	-2.2	**	0.03	0.9		-0.03	-1.6		0.03	1.6	
medium schooling	0.00	0.1		0.05	1.4		0.05	2.3	**	0.04	2.0	**
higher schooling	-0.06	-1.7		0.03	0.7		-0.03	-1.1		0.02	0.7	
part time employed	-0.04	-1.0		0.05	1.1		-0.02	-0.8		0.03	1.2	
not employed	0.15	5.0	***	0.07	1.9	*	0.09	4.1	***	0.02	0.8	
married	-0.06	-2.3	**	-0.01	-0.2		-0.01	-0.6		-0.01	-0.3	
interv. experience	-0.00	-1.8	*	0.00	1.6		0.00	0.5		0.00	4.2	***
Respondent-Interviewer Interactions												
same sex	0.02	0.9		-0.01	-0.2		-0.00	-0.2		-0.00	-0.4	
R-I age difference	-0.01	-9.9	***	0.00	0.2		-0.01	-7.8	***	0.00	2.9	***
same schooling	0.04	1.4		0.10	2.9	**	0.03	1.8	*	0.07	3.9	***
same employment	-0.00	-0.1		-0.02	-0.8		0.01	0.5		0.01	0.9	
same family status	-0.01	-0.6		0.01	0.4		-0.04	-2.7	***	-0.02	-1.4	
Interview situation												
w/o interviewer	0.24	4.6	***	0.24	3.9	***	0.16	5.0	***	0.12	3.5	***
self administered	-0.02	-0.8		0.01	0.2		0.01	0.6		0.07	3.7	***
# of intvw. contacts	0.05	7.4	***	0.07	8.7	***	0.01	3.0	***	0.02	3.9	***
interviewer change	0.19	8.0	***	-0.01	-0.4		0.24	14.1	***	0.15	7.8	***
HH size	-0.01	-0.5		0.00	0.2		0.01	1.9	*	0.01	1.1	
in high-rise build.	0.01	0.5		-0.08	-2.4	**	0.02	1.1		-0.03	-1.5	
in residential area	-0.02	-1.0		-0.04	-1.3		-0.01	-0.6		-0.00	-0.1	
wave fixed effects	-	-		yes		***	-	-		yes		***
number of obs		2,635			2,635			9,883			9,883	
LR χ^2 (d.f.), $p > \chi^2$	629	(31)	0.0	1770	(49)	0.0	1091	(32)	0.0	1776	(50)	.0
McFadden's R^2		0.18			0.50			0.16			0.39	
Wald-Tests on joint significance	χ^2	$p>\chi^2$		χ^2	$p>\chi^2$		χ^2	$p>\chi^2$		χ^2	$p>\chi^2$	
R-I Interactions	105.7	0.00	***	8.40	0.14		108.6	0.00	***	25.7	0.00	***
Interview Situation	98.9	0.00	***	86.2	0.00	***	245.6	0.00	***	83.5	0.00	***
INRR, INRR^2	25.4	0.00	***	5.1	0.08	*	68.2	0.00	***	1.1	0.59	
Max INR-Rate(%)		33.3			24.2			47.5			76.6	

Notes for Table 3:
Significance levels: * 10 %, ** 5 %, *** 1 %; I: interviewer; R: respondent; HH: household; ME: Marginal effect, evaluated at mean; The estimated variances for the individual sample are corrected for intra-household correlation.

Addressing our fourth hypothesis, we regressed the status of the interview attempt (which can have the outcomes: (1) response, (2) UNR and (3) panel mortality) on a reduced set of explanatory variables for the cumulated drop-outs until 2002 and respondents of 2002 in a multinomial logit framework. According to our theoretical considerations, we expect the INR-indicators to be insignificant with respect to the odds of category 3 (panel-mortality).

Table 4: Results of multinomial logit of participation status on model 3 (significant effects only)

	Household Questionnaire				**Individual Questionnaire**			
Participation status:	**UNR**		**MORT**		**UNR**		**MORT**	
	coef.	**t**	**coef.**	**t**	**coef.**	**t**	**coef.**	**t**
INR Rate	0.11	1.1	0.08	0.8	0.17	2.8 ***	0.16	2.7 ***
INR Rate^2	-0.00	-0.5	-0.00	-0.3	-0.01	-2.2 **	-0.01	-2.2 **
Income INR	1.40	1.6 *	1.22	1.4	1.13	2.8 ***	0.65	1.5
R female	-0.92	-1.5	-1.38	-2.3 **	0.20	0.7	-0.29	-1.0
I female	0.90	1.3	0.82	1.2	-0.35	-1.3	-0.49	-1.9 *
R's age	0.03	0.9	0.03	1.0	0.03	1.6 *	0.03	2.0 **
R medium schooling	-1.57	-1.7 *	-1.69	-1.8 *	-0.04	-0.2	-0.43	-1.6
R married	-1.22	-1.8 *	-3.21	-4.5 ***	0.45	1.4	-0.04	-0.1
I married	-1.52	-2.4 **	-1.71	-2.6 ***	-0.09	-0.3	-0.08	-0.3
HH in hirise building	-1.34	-1.8 *	-0.93	-1.2	0.17	0.5	0.34	1.0
# interview contacts	-0.70	-2.2 **	-1.24	-3.9 ***	-0.29	-2.1 **	-0.39	-2.8
interviewer change	1.64	2.7 ***	1.64	2.6 ***	2.74	11.3 ***	2.23	9.1 ***
HH size	0.37	1.4	0.38	1.4	0.19	1.9 *	0.17	1.7 *
I's experience (years)	-0.00	-1.8 *	-0.00	-2.2 **	0.00	0.5	-0.00	-0.5
w/o interviewer	3.58	2.3 **	2.22	1.4	0.10	0.1	-0.50	-0.5
Wave fixed effects	yes	***	yes	***	yes	***	yes	***
Maximum INR-Rate		29.7		32.4		12.6		12.2
Wald-Test on joint significance of:		χ^2		χ^2		χ^2		χ^2
INRR, INRR^2		2.1		1.2		8.6 **		7.9 **
INRR, INRR^2, IncINR		11.0 **		6.9 *		21.0 ***		12.3 ***
# obs.		2,635	(LogL:	-849.2)		9,883	(LogL:	-4263.9)
Mc Faddens R^2		0.66	(p>χ^2	0.000)		0.58	(p>χ^2:	0.000)

Note: UNR: interview refusal, MORT: panel mortality; Significance levels: * 10 %, ** 5 %, *** 1 %;
The estimated variances for the individual sample are corrected for intra-household correlation.

As the results in Table 4 show, this hypothesis holds only for the household questionnaires. With regard to the personal questionnaires we find strong counter-evidence for this hypothesis: We observe the same pattern of INR-effects on panel mortality as for unit nonresponse. The effect of the INR-Rate on panel mortality is maximum at an INR-Rate of 12.9% (32.4% for the household sample, respectively) and Income INR has also a positive effect on panel-mortality. Even if the coefficients are not significant in the household sample, all three INR parameters are jointly significant as can be seen from the Wald tests provided at the bottom of table 4. It seems that a better part of panel-mortality is non-natural, e.g. the respondent communicates his decision to refuse participation by a policy of closed doors as mentioned in Rendtel (1988), or pretends to be in bad shape and not interviewable.

5.2 Sample Selection due to Attrition

Correlation of item nonresponse with panel attrition leads to endogenously selected samples with biased regression coefficients in item nonresponse regressions. Since we presented some evidence for the correlation between different INR-proxies and UNR, as well as panel-mortality, we now intend to find out whether panel attrition biases the results of a regression of INR using the bivariate probit approach described in section 3. On the right hand-side the specification equation consists of the (potential) determinants of income item nonresponse, where we use mostly the same as for unit nonresponse. The identifiability of such selection models depends on the explanatory power of the involved exclusion restrictions. Auxilliary regressions - not presented here - have shown that the family status of the respondent (married or not), the number of contact attempts of the interviewer, household living in a residential area (as opposed to living on the countryside or in an industrial area), type of the building the household lives in (high-rise building or not), the size of the household, experience of the interviewer, and change of the interviewer have explanatory power for the UNR-behaviour, and are neither separately nor jointly significant in INR-regressions. We therefore use these variables as instruments to identify the selection equation. Table 5 presents the correlation coefficient ρ and its significance. If ρ is significantly different from zero, the panel attrition between 1984 and the considered wave was selective with regard to income item nonresponse of the respective year. This means that ignoring panel attrition would lead to biased parameter estimates of the INR-model. Astoundingly, only in a few waves is ρ significantly different from zero, which indicates a selection bias. A significant negative correlation between the selection- and INR-model, such as in the 1994 and 1998 household questionnaire, indicates that the nonresponse mechanisms point in the same direction: those participants who are likely to participate in subsequent interviews are unlikely to generate item nonresponses. Astoundingly, the correlation for those households who dropped out until 1986 and for those individuals who dropped out until 1985 and 1987 is positive. This implies that the INR-mechanism was contrary to the UNR-mechanism for the cumulated drop-outs and respondents. These results are in contrast to our hypothesis H5: if panel drop-outs were selective with respect to income INR, we would have expected a significant negative correlation ρ over all subsequent panel waves. These results may be evidence in favour of our theory of the coexistence of a reverse cooperation continuum: not only low cooperative respondents drop-out from the panel group, but also highly motivated interviewees leave the panel. The effect of drop-outs on the bias of the regression coefficients for the INR model is therefore not linear. The biasing effects of the drop-outs mostly cancel. The number of low motivated respondents (with high INR propensity) is balanced by the number of highly motivated respondents (with low INR propensity). This reduces or cancels the biasing effect of UNR on INR analysis.

Table 5: Rho-parameters of bivariate probit selection models for item nonresponse on the income question (drop-outs due to UNR and panel mortality)

Wave	Household Questionnaire				Individual Questionnaire			
	# obs.	thereof refusals	ρ	χ^2	# obs.	thereof refusals	ρ	χ^2
1985	2635	645	0.39	1.643	9883	1462	0.74	19.788 ***
1986	2635	928	0.38	3.121 *	9883	2356	0.05	0.062
1987	2635	1061	-0.11	0.274	9883	2886	0.39	4.101 **
1988	2635	1196	-0.28	1.654	9883	3464	0.10	0.256
1989	2635	1314	-0.13	0.324	9883	3946	0.06	0.076
1990	2635	1387	0.02	0.006	9883	4278	-0.16	0.690
1991	2635	1453	-0.05	0.046	9883	4556	-0.08	0.188
1992	2635	1513	-0.06	0.065	9883	4808	0.10	0.402
1993	2635	1554	-0.05	0.054	9883	5051	-0.19	0.976
1994	2635	1621	-0.41	3.239 **	9883	5330	0.15	0.535
1995	2635	1678	-0.10	0.267	9883	5596	-0.15	0.603
1996	2635	1723	0.25	1.606	9883	5812	0.09	0.170
1997	2635	1760	-0.24	0.887	9883	6013	0.24	1.160
1998	2635	1806	-0.47	3.876 **	9883	6225	-0.44	2.321
1999	2635	1850	0.00	0.000	9883	6446	-0.28	1.487

Note:
Significance levels: * 10 %, ** 5 %, *** 1 %
The estimated variances for the individual sample are corrected for intra-household correlation.

In Table 6 we present the coefficients of uncorrected and corrected probit estimates of INR on the income question for the year 1998 in the household questionnaire. The uncorrected estimates are calculated with standard probit, ignoring panel attrition. The corrected coefficients result from the bivariate probit with selection correction. The results of the selection model can be extracted from column 3. The correlation between specification and selection equation is with -0.47 the strongest plausible correlation found across all samples and waves.

However, the bias in the coefficients of the INR-regression is mostly negligible and mostly differs by less than 1/10. Only the coefficient for full-time employed respondents is smaller by one third if selective attrition is ignored.

Table 6: Results from uncorrected and attrition corrected probit INR regression and selection model, household questionnaire 1998

	Uncorrected Probit			Corrected Probit			Selection Model		
	coef.	T		coef.	t		coef.	t	
INR-Model:									
R female	0.608	2.97	***	0.654	3.36	***	-0.114	-1.46	
R med. level schooling	0.035	0.16		-0.143	-0.65		0.308	3.87	***
R higher schooling	0.033	0.12		-0.050	-0.19		0.314	3.23	***
R university graduate	-0.320	-0.77		-0.346	-0.92		0.083	0.61	
R fully employed	0.426	1.86	*	0.622	2.78	***	0.045	0.54	
I female	0.056	0.23		-0.041	-0.18		0.273	3.78	***
I med. level schooling	-0.187	-0.82		-0.130	-0.62		0.025	0.3	
I higher shooling	-0.999	-2.38	**	-0.900	-2.35	**	0.200	2.1	**
I part time employed	-0.664	-1.38		-0.634	-1.44		0.166	1.57	*
I unemployed	0.091	0.34		0.225	0.89		-0.409	-4.76	***
I married	-0.557	-2.48	**	-0.641	-3.07	***	0.401	5.04	***
R-I same sex	0.249	1.28		0.258	1.43		-0.057	-0.91	
R-I age difference	-0.011	-1.54		-0.016	-2.33	**	0.037	12.69	***
R-I same schooling	0.059	0.25		0.087	0.41		0.060	0.79	
R-I same employment	-0.012	-0.06		-0.011	-0.06		0.037	0.57	
R-I same family status	0.143	0.74		0.104	0.58		0.083	1.31	
mode w/o Interviewer				0.392	1.08		-0.192	-1.15	
self administered surv.	0.056	0.23		0.008	0.04		0.224	3.01	***
Constant	-1.978	-6.44	***	-1.427	-3.74	***	-3.666	-13.48	***
Selection Model:									
R age							0.050	14.44	***
R public employee							0.227	2.41	**
R married							0.184	2.02	**
HH size							0.043	1.26	
in high-rise building							-0.069	-0.87	
HH in residential area							-0.051	-0.8	
# of interview contacts							-0.001	-0.04	
I's experience (years)							0.000	2.71	***
change of Interviewer							-1.327	-15.86	***
Correlation		ρ (s.e.):			-0.472			(0.198)	
Test on H_0: ρ =0		χ^2 (p> χ^2):			3.88			0.049	**
# obs. (censored obs.)		829			2635			(1806)	
LogL		-112.9						-1324.7	
Wald Test model sig: χ^2, p	38.8	0.003	***	38.8	0.003	***			

Note: Significance levels: * 10 %, ** 5 %, *** 1%

Summing up, our results show that an analysis of the determinants of item nonresponse based on data from only one panel wave and neglecting self-selection of respondents may suffer from attrition bias. Due to the coexistence of two types of refusals, the existence and strength of this bias is lower than expected. For the analysis of income INR in the GSOEP, endogenous sample selection is ignorable in most cases.

6 Summary and Conclusion

The nonresponse literature which pays attention to the interaction of unit- and item nonresponse and to the problem of endogenous sample selection with respect to item nonresponse is rare. This paper contributes to this discussion by answering two questions: first, how are item and unit nonresponse interrelated and second, does panel attrition cause a selection bias in the results of INR studies?

The hypothesis of a latent cooperation continuum predicts a positive correlation between the propensities of unit and item nonresponse. The data used in this study provides some evidence in favour of this hypothesis: the income nonresponse frequency and the INR-Rates of drop-outs are in the majority of cases larger than for subsequent respondents. Income nonresponse has explanatory power in UNR regressions, which in general tends to be decreasing over time. This may point to endogenous sample selection. Nonetheless, the findings are discontinuous and not undividedly supported by the literature. We argue that the cooperation continuum could in principle also be inverse, which we name "reverse cooperation continuum". And we hypothesise that both cooperation hypotheses may apply to different types of respondents in the same panel. Therefore, the interaction of INR and UNR propensities may be described by an inversely U-shaped pattern. Evidence for this is provided by applying probit regressions to unit nonresponse behaviour for the original households and participants for the first 19 panel waves of the GSOEP. Astoundingly, we find the same patterns of INR-interrelation with respect to panel mortality. We conclude that a better part of this panel mortality is non-natural, namely hidden interview refusal. With respect to other determinants of UNR, our findings are in line with the survey literature. Regarding the interview-mode effects, we find that the presence of an interviewer is beneficial for both, item and unit response.

Addressing the second question, we identified a sample selection bias due to panel attrition in the results of INR-regressions only for a minor number of panel waves. Selective panel attrition of interview refusal and panel mortality does not cumulate over subsequent waves, since the bias does not increase, but vanishes. Even in waves where the correlation in the unit response and item nonresponse models was found to be significantly negative, the bias in the coefficients of INR regressions was mostly not substantial.

However, this study has shown that the correlation between item nonresponse and panel attrition propensities is mostly not linear, but quadratic. This indicates that two types of respondents may coexist in a panel study: one type behaving as predicted by cooperation continuum theory with attrition following high item nonresponse. The other type behaving in a reverse manner with refusal following no or low INR. The coexistence and drop-out of both respondent types alleviates the biasing effect of panel attrition on the regression coefficients of INR-analyses.

The researcher as well as the survey conductor may wish to find out which nonresponse type is prevalent in their panel group to derive a judgment about the magnitude of the attrition bias and may impose correction methods. The survey organisation should take care of the residual-category of panel-mortality: there are some hints that people use non-reachability as a strategy to communicate their interview refusal. A short interview with the (reachable) refusals asking for their attitudes towards surveys in general, their satisfaction with respective survey, and their reasons for interview refusal may provide some insight on their attrition decision and might reduce the stochastic component in INR and UNR analysis. Accordingly, this will lead to more reliable results for the determinants of item nonresponse. Finally, this knowledge enables survey conductors to improve the data quality in their surveys and the researchers to improve their analyses.

Bibliography

Burton, Jonathan, Heather Laurie and Nick Moon, 1999, "Don't ask me nothin' about nothin', I just might tell you the truth" - The Interaction between unit nonresponse and item nonresponse, Paper presented on the *International Conference on Survey Nonresponse*, Portland, Oregon.

Dolton, Peter, Maarten Lindeboom, and Gerard J. van den Berg, 1998, Survey Attrition: A taxonomy and the search for valid instruments to correct for biases, *Tinbergen Institute Discussion Paper*, TI 98-109/3.

Eeden, Pieter van den, 2002, Unit- and Item-Nonresponse Bias in a longitudinal case, *presented at ICIS (International Conference on Improving Surveys)*, 25-28 August 2002, Copenhagen;
URL: http://www.icis.dk/ICIS_papers/F3_5_3.pdf.

Esser, Hartmut, 1986, Können Befragte lügen? Zum Konzept des wahren Wertes im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung (May respondents lie? To the concept of the true value in a framework of explaining situation effects in an interview), *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, No. 38, 314-336.

Frick, Joachim R. and Markus M. Grabka, 2005, Item-nonresponse on income questions in panel surveys: incidence, imputation, and the impact on inequality and mobility, *Allgemeines Statistisches Archiv*, forthcoming.

Groves, Robert M., 1989, Survey errors and survey costs, New York: Wiley.

Hanefeld, Ute, 1987, Das sozio-oekonomische Panel: Grundlagen und Konzeption (The Socio-economic panel: Fundamentals and Concept), Frankfurt /M.: Campus-Verlag.

Heckman, James J., 1979, Sample Selection as a specification error, *Econometrica*, Vol. 47 (1), 153-162.

Lachman, R., J. T. Lachman and E.C. Butterfield, 1979, Cognitive Psychology and Information Processing, Hillsdale, NJ: Erlbaum.

Loosveldt, Geert, Jan Pickery and Jaak Billiet, 2002, Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey, *Journal of Official Statistics*, 18(4), 545-557.

Rendtel, Ulrich, 1988, Panelmortalität - Eine Analyse der Antwortausfälle beim Sozio-ökonomischen Panel in der 2. und 3. Befragungswelle (Panel mortality - an analysis of response drop-outs at the German Socio-Economic Panel in the second and third wave), *Vierteljahresheft zur Wirtschaftsforschung*, Heft 1/2, 37-59.

Rendtel, Ulrich, 1989, Vertrauensbildung und Teilnahmeentscheidung in Panelstudien – Über die Entwicklung der Antwortbereitschaft im Sozio-Ökonomischen Panel (trust building and decision to participate in panel studies), Arbeitspapier Nr. 313, Sonderforschungsbereich 3: Mikroanalytische Grundlagen der Gesellschaftspolitik der Universitäten Frankfurt und Mannheim.

Riphahn, Regina T. and Oliver Serfling, 2005, Item Nonresponse on Income and Wealth Questions, *Empirical Economics 30(2)*, 521-538.

Rogers, W. H., 1993, Regression standard errors in clustered samples. Stata Technical Bulletin 13: 19–23. Reprinted in Stata Technical Bulletin Reprints, Vol. 3, 88–94.

Schnell, Rainer, Paul B. Hill, and Elke Esser, 1995, Methoden der empirischen Sozialforschung (Methods of emprical social sciences), 5.Ed., München; Wien: Oldenbourg-Verlag.

Schräpler, Jörg-Peter, 2003, Gross Income Nonresponse in the German Socio-Economic Panel - Refusal or don't know, in: Schmollers Jahrbuch 123, pp. 109-124, Berlin: Duncker & Humblot.

Schräpler, Jörg-Peter and Gert G. Wagner, 2001, Das Verhalten von Interviewern - Darstellung und ausgewählte Analysen am Beispiel des "Interviewerpanels" des Sozio-ökonomischen Panels (Behaviour of interviewers - Report and selected analyses using the interviewer dataset of GSOEP), *Allgemeines Statistisches Archiv 85*, 45-66.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz, 1996, Thinking about Answers – The Application of Cognitive Processes to Survey Methodology, San Francisco: Jossey-Bass Publishers.

Veen, Wynand van den and Bernard M.S. Van den Praag, 1981, The Demand for Deductibles in private health insurance – A Probit Model with Sample Selection, *Journal of Econometrics 7*, 229-252.

Williams, R. L., 2000, A note on robust variance estimation for cluster-correlated data. Biometrics 56, 645–646.

V. Rounding Behavior of Respondents in Household Surveys - Evidence from Swiss Data

Oliver Serfling

University of Basel, Switzerland

Abstract

Rounding is a frequent phenomenon in self-reported income data. In the first five waves of the Swiss Household Panel (SHP), in 87 of 100 income statements at least half of the provided digits are rounded off (zeros). In empirical studies the rounding error is typically ignored. This can have damaging effects on the consistency of estimates if rounding does not occur at random and is correlated with the underlying true value.

In this study, we discuss several rounding intensity measures and develop a framework for testing discrete outcomes with respect to ordinality and distinguishability for a given model. Using raw data from the SHP, we identify socio-demographic patterns of rounding intensity. We show that interviewer effects as well as interaction effects between respondents and interviewers are statistically significant, but negligible with respect to their magnitude. With respect to the ascertained positive correlation of the provided income figure and its rounding intensity, we attempt to disentangle the amount effect from a number-of-digits effect. We find a nonlinear relationship between the income statement and rounding intensity, while a number-of-digits effect remains significant. Additionally, household wealth position has a negative influence on the precision of the statement. Lastly, we find a positive panel duration effect on precision, being robust with respect to selective panel attrition.

1 The Rounding Phenomenon

Rounding is an omnipresent phenomenon when measurements of continuous variables are involved. The informational value of data decreases, the more coarsely the data are rounded. Even though this phenomenon has been addressed in the statistical literature for more than a century, it has not become a central issue in the measurement error literature. This is mostly because occurrence of rounding was seen as a random event and its interfering impact on statistical inference therefore was seen as negligible.

This study provides evidence that rounding in survey data does not occur randomly at all. Using data from the Swiss Household Panel (SHP), we show that rounding behavior is correlated with the figure provided and follows consistent patterns according to the interview situation characteristics and the persons involved. This raises doubt on questions as to whether standard assumptions on measurement errors are likely to be violated, which may have damaging effects on mean and variance of regression coefficients. Moreover, rounding may reflect respondents' motivation towards survey participation and may be an indicator of subsequent item- or survey-nonresponse

With our analysis we further the understanding of the interview process and the interactions between respondent and interviewer occurring during the course of a survey interview. Our results may, on the one hand, help to improve the data collection process with respect to data quality. If rounding is a phenomenon resulting from lack of respondents´ motivation, our results may show what factors determine the differences in rounding behavior and how interviewers may motivate the respondents to provide a greater extent of precision. On the other hand, our results may sensitize researchers who work with survey data to take such measurement errors into account in their own analyses.

This study contributes to the literature in various ways: It provides descriptive evidence for Swiss data - the Swiss Household panel (SHP) - and shows that the rounding error is correlated with the provided amount, correlated with respondent characteristics, serially correlated over time, and weakly positively correlated with subsequent nonresponse. We discuss several rounding intensity measures. A framework for testing the ordinality and distinguishability of spuriously ordered discrete outcomes of a variable is developed to determine the reliability of rounding intensity measures for different number of digits. We evaluate the effects of socio-demographic characteristics of the respondent on rounding behavior. Concerning the role of the interviewer in the interview process, we investigate interviewer fixed effects, influences of the socio-demographic characteristics of the interviewers as well as interactions between interviewers and interviewees on rounding behavior, by using additional data from an interviewer survey. Previous findings suggest positive correlations between the provided income figure and the intensity of its rounding. As will be laid out in the following, this observation may be composed of a net effect of the possibility to make a rounded statement (having a lot of digits that can be rounded) and an income effect, which affects rounding behavior. Therefore, we attempt to disentangle the income effect of the provided figure from the number-of-digits effect. Finally, we examine the panel duration or respondents' experience effect and check for its robustness with respect to selective panel attrition.

The remainder of this paper is structured as follows. This introduction proceeds by presenting the literature that gives evidence of the damaging effects of rounding for statistical inference. Subsequently, prior studies on rounding patterns in survey data are introduced. In Section 2, we define the term "rounding" and discuss several approaches to measure it. Additionally, we present the research questions and hypotheses of this analysis. In the third section, we describe the data underlying this study. First descriptive evidence on rounding in the SHP is presented. Section 4, introduces two test strategies for ordinality of discrete outcomes, and lays out the empirical strategy as regards the research questions. The results are presented in Section 5. The last section discusses and summarizes the results, and points to the limitations of this study as well as to further research needs.

1.1 Rounding as a Measurement Error

While the rounding (or grouping) effects on the moments of univariate distributions were discussed already in the late 19th century, the more recent studies concentrate on variance inflation and possible biasing effects of rounding in multivariate regressions. Sheppard (1898) first analyzed the effect of grouping on the moments of the normal distribution and found that the effect on the mean was in most circumstances negligible, but the variance estimation increased by a factor of $1/12w^2$ in proportion to the rounding interval w, i.e. the range of numbers which will be expressed by the rounded figure according to a rounding rule. He introduced a correction method for rounded data (Sheppard's Correction). Student (1908) provides a prominent example of grouping the measurements of the length of the left middle finger of 3,000 criminals in groups of 2-inches length. He finds that the means of samples drawn from this population occur as multiples of 0.25. The standard error increases by nearly 8 per-cent. Kendall (1938) discussed the conditions under which Sheppard's correction is valid while Eisenhart (1947) stressed that the rounding lattice, i.e. the unit to which the continuous measure is rounded, needs to be imposed at random, otherwise the Sheppard correction may be wrong and will even adjust in the wrong direction. Tricker (1984) analyzed the effects on distributions other than the Gaussian normal and found the skewness of the respective distribution to have influence on the rounding bias. For right or left skewed distributions like the log-normal, gamma or exponential distribution, the first moment estimator is biased. Polasek (1987) examined the effect of systematical rounding over all observations versus uniform rounding, i.e. rounding in just one observation, using local Bayesian sensitivity analysis. He found that uniform rounding errors affect only the intercept in a linear regression, while a unique rounding error affects all coefficients. Other studies dealt with the improvements in the empirical treatment of rounded data. Dempster and Rubin (1983) tested the appropriateness of Sheppard's corrections for eliminating rounding error in least squares regressions. Heitjan and Rubin (1990) analyzed the effects of age heaping from coarse data by employing multiple imputation techniques. In more recent studies, the measurement errors in labor force survey data are evaluated by comparing reported data in surveys with recorded data (see e.g. Duncan and Hill 1985, or Bound and Krueger 1991). Rodgers et al. (1993) tested classical assumptions on measurement errors in surveys: normal distribution with mean zero, uncorrelatedness with either the true value of the variable or other explanatory variables, no serial correlation, and homoscedasticity. They found that most of these assumptions are violated and that respondents with lower-than-average earnings are prone to over-report their income, while opposite behavior was found at the upper end of the earnings scale. This is also confirmed by the findings of Hanisch and Rendtel (2002), who compared Finnish survey data with register data. Nonzero covariances of the measurement error with other explanatory variables leads to the conclusion that regression coefficients may either be upward or downward biased, a result in opposition to the conventional wisdom that measurement errors in variables will only bias coefficients toward zero. In a similar vein, Rowe and Gribble (1994) test the effects of rounding on the evaluation of trends in income inequality in the Canadian Survey of Consumer Finances. They demonstrate that wage inflation together with fixed rounding points can lead to spurious trends or discontinuities in the rounding bias when time series data are collected. This effect may occur, even though rounding patterns or the underlying earnings distribution remains stable. They found such rounding bias and rounding variance making contributions to the mean squared error in a comparable magnitude as sampling variance, and conclude that rounding should be taken into account with the measurement of trends in income statistics and income inequality.

To put structure on the discussion of measurement errors, Henderson and Jarrett (2003) have classified survey errors in three categories: (a) *measurement error* which occurs when a continuous true value X is reported with error as continuous Z, (b) *misreporting error* when continuous X is reported as a discrete interval with midpoint Z and (c) *misclassification error* which occurs when the discrete true value X is reported as a wrong discrete value Z. Following this scheme, the rounding error belongs to the class of misreporting errors. In contrast to common measurement errors, Henderson and Jarrett emphasize that misreporting errors are not independent of X, which underlines the findings quoted above.

1.2 Rounding as Respondent Behavior

The literature on measurement errors in surveys has mostly concentrated on the (visual) design of questionnaires, response alternatives, recall bias, interviewer effects, interview mode effects, and both unit and item nonresponse, while the problem of rounding in surveys has been widely neglected (see e.g. Biemer et al. 1991, also as a hint of the neglect of this problem). One reason for this wide neglect of the rounding phenomenon as measurement error may be that the influence on quality of data and its inference was expected to be small. The psychological literature provided only few concepts which can be applied for the understanding of rounding behavior: Max Wertheimer's classical essay on "Numbers and numerical concepts in primitive people" (1912) points out that the understanding of numbers relies heavily on the cultural understanding of quantities, and that there exist ideal types that act as anchoring points for perception. Rosch (1975) claims that the theory of cognitive reference points may apply to rounding behavior: The decimal system is by definition constructed as multiples of 10, which are therefore presumed to be reference numbers. Individuals tend to perceive other numbers in relation to those reference points.

Within the survey literature the problem of rounding has gained little attention: Schweitzer and Severance-Lossin (1996) investigated the occurrence of rounding in gross earnings data. In data from the Current Populations Survey (CPS) they find strong evidence that rounding behavior is positively correlated with the observed earnings level and varies systematically with observable characteristics of the respondents. Schräpler (1999) investigates the rounding behavior of respondents in the first 12 waves of the German Socio-Economic Panel (GSOEP) with respect to the gross income question. The pattern of rounding behavior seems to be stable over subsequent waves and has significant differences by nationality of the respondents (Germans vs. guest workers). It also differs by sex (men are less precise), age (the elderly are more precise), the occupational position, interview mode and length, and income of the respondent. He also finds significant experience effects of the respondents and ambiguous but significant panel duration effects. Rietveld (2001) examines the rounding of arrival and departure times in a Dutch travel survey. Interestingly, departure times seem to be rounded much more frequently than arrival times. By analyzing the distribution of reported trip duration he provides some evidence that the probability of upward rounding is considerably lower than downward rounding, which implies that the mean duration time is biased. Hanisch (2004) reports that approximately 80 per cent of gross wage and earnings and approx. 95% of net disposable household income are rounded after one or two significant digits in the Finish sample of the ECHP (European Community Household Panel). The relative precision does not vary considerably with average money amount provided by the respondent. In contrast to Schräpler (1999), he finds that males tend to provide more exact values than women. He also finds nationality, interview mode, and job effects on rounding intensity. Against his expectation, the respondents' experience with panel participation did not have a monotone effect on rounding behavior. In the most recent study, Kroh (2004) analyses influences of the interview on the reliability of self-reported body weight. He finds significant socio-demographic effects, like sex, education and marital status. Additionally, he found overweight people more prone to provide rounded weights.

2 Definition, Theoretical Framework, Measurement, and Research Goals

2.1 Formalization of the Rounding Phenomenon

The objective of the survey is to obtain a true numerical value z for an item of interest. The respondent communicates a value z^* which differs from the true value z. We assume that the true value z lies within a rounding interval with width d around z^* with full confidence:

$$P\left(z^* - \tfrac{d}{2} \leq z < z^* + \tfrac{d}{2}\right) = 1 \qquad (2.1.1)$$

Several rounding rules, i.e. the rules governing the transformation of a continuous figure into a rounded discrete figure, are possible and discussed with greater detail in Eisenhart (1947). Within this study we restrict rounding to the provision of zeros at the rear end of an integer figure, which strictly implies that the respondent rounds his statement to multiples of 10^n, where n is a positive integer value including 0. It is also possible that respondents round to multiples of $5*10^n$. Preliminary evidence has shown that a "5" as first significant digit (seen from the right to the left) does not occur with a significantly higher frequency than the digits "1"…"9", which encourages our assumption above.

The width d of the rounding interval is identified by the number of rounded digits (NRD, i.e. the number of zero-digits at the rear of the provided figure):

$$d = \begin{cases} 0 & \text{if } NRD = 0 \\ 10^{(NRD)} & \text{otherwise} \end{cases} \qquad (2.1.2)$$

where the number of rounded digits may be determined by modulo operation, dividing $z*$ by $10^{(NRD)}$ with significant part q and remainder r, which satisfies:

$$z^* \stackrel{!}{=} 10^{(NRD)} q + r \quad \text{with } r=0 \qquad (2.1.3a)$$

and

$$z^* \stackrel{!}{=} 10^{(NRD+1)} q + r \quad \text{with } r \neq 0$$

$$\text{with: } NRD, q, r \in \mathbf{Z} \qquad (2.1.3b)$$

Z defines the set of natural numbers, i.e. positive, integer values; q and r are any positive integer value satisfying 2.1.3a or b, respectively.

The number of significant digits (NSD) is the number of non-zero digits at the front of the provided figure. Hence, the total number of digits (ND) is the sum of rounded and significant digits: ND = NSD + NRD.

The resulting rounding error ε_r is defined as being the difference of z^*-z, therefore:

$$z^* = z + \varepsilon_r \qquad (2.1.4)$$

For this rounding error we assume a uniform distribution around zero, implying $E(\varepsilon_r)=0$ and a variance larger than zero: $\sigma^2_{\varepsilon r}>0$.

These assumptions have the following strong implications:

(1) The respondent does not deliberately misreport the figure of interest, but is able to provide a rounded value. So systematically upward and downward rounding, which would imply a mean rounding error being significantly different from zero, is restricted by assumption. This assumption may be violated in practice, as it is indicated by the results of Duncan/Hill (1985), Bound/Krueger (1991), and Rodgers et al. (1993). However, in this study we are concerned with the evaluation of rounding errors and not misreporting errors in general.

(2) A zero at the rear end of the provided figure z^* indicates rounding. If it is assumed that the digits - with exception of the first (leading) digit - are distributed uniformly between 0 and 9, i.e. they appear with a probability of 1/10, then this assumption holds only in 9 out of 10 cases. This *rounding assumption error* inflates the variance of the rounding error, since real rounding occurs less often than hypothesized. The probability of the occurrence of zeros at the rear of the true figure is unknown, since we have no information about the "true" value. We assume that this probability should be very low for net amounts and larger for gross amounts. For the purpose of this study, we control for the gross/net type of the income statement. Beyond that we neglect this rounding assumption error.

2.2 Theoretical Framework

According to the literature from cognitive and social psychology, the answering process in a survey interview can be split up into four tasks that the respondent has to process to yield an informative answer. These tasks involve comprehension of the question, gathering the required information, assessing the correspondence between desired and retrieved information, editing and communicating the answer (see Sudman, Bradburn and Schwarz (1996) or Tourangeau (1984)). While cognitive abilities are a prerequisite to give an answer, the rational choice theory (applied to survey research by Esser 1984) suggests that even if cognitive resources are available to perform these steps, the respondent’s answer may deviate from the desired true value. Comprehension of the question, information retrieval, formatting the answer, and the disclosure of privacy (which comes along with sensitive questions) may be seen as costs of answering that have to be balanced by positive sanctions of the interviewer (e.g. a smile) or the survey institute (e.g. a small gift). Hence, the problem of determining the deviation of the reported from the true value (misreporting error) is reduced to a simple cost-benefit analysis. It has been shown that the rational-choice framework has explanatory power for item-nonresponse behavior of respondents (see e.g. Riphahn and Serfling (2005)). When it comes to the analysis of rounding in reported data, we find a situation where the respondent has already understood the question, assessed the desired information and also judged to be willing to give an answer, since he communicated one. Hence, rounding may stem only from errors in the information retrieval task (recall error, time-costs for information retrieval are too high) or from the editing and communicating stage, where the respondent may decide that he wishes to maintain privacy by rounding the figure. Assuming that the respondent is free to make a decision about his rounding intensity (RI) this behavior may be formalized by the random utility maximization hypothesis, where the respondent aims to find the optimal rounding intensity that maximizes his expected utility. The utility function can be seen as a continuous function of costs and benefits, which leads to a maximum, when the following optimization criterion is satisfied:

$$\frac{\partial Benefit(RI)}{\partial RI} - \frac{\partial Costs(RI)}{\partial RI} \overset{!}{=} 0 \qquad (2.2.1)$$

If the cost and benefit elements are attributed to characteristics of the interviewer, respondent and interactions, this calculus leads to a latent optimal rounding intensity:

$$RI_{it}^{*} = \alpha_t + R_{it}\beta_1 + I_{jt}\beta_2 + (R_{it} * I_{jt})\beta_3 + \varepsilon_{it} \qquad (2.2.2)$$

where α_t is the baseline cost-surplus in answering the question at time t, R_{it} are the characteristics of the respondent i, I_j are the characteristics of the interviewer j, (R*I) are the interaction of both and ε_{it} is white noise.

2.3 Appropriateness of Rounding Measures

As seen above, we discuss rounding behavior in terms of an unobserved latent rounding intensity. When it comes to the communication of an income statement, the respondent has to map this latent rounding intensity into a discrete measure of the number of digits he rounds off, or the number of significant digits he provides. Assuming that the respondent does not deliberately misreport the figure in question and is free to decide on his rounding intensity, the number of significant digits (NSD) is censored by the total number of digits (ND) of the (true) figure. For NRD the maximum is ND-1, since at least the first digit has to be different from zero by definition. We define a rounding measure as being appropriate, if it is positively related to the latent rounding intensity, which requires that a higher outcome of the rounding measure compares to a higher level of rounding intensity and vice versa. This condition will be discussed in Section 4 in the terms of “ordinality” and “distinguishability” of categories.

Several measures of rounding intensity are used in the literature, though a careful discussion of the appropriateness of such measures is missing. In the following we introduce the most frequently used measures and discuss their properties:

Number of significant digits (used by Hanisch, 2004) is a discrete absolute measure of the precision of the income statement. In contrast, the **Number of rounded digits** (used by Schräpler 1999) is an absolute discrete measure for the rounding intensity. Rietveld (2001) uses a dichotomized outcome of more than two rounded digits (=1, 0 otherwise).

The shortcoming of these measures is that they are truncated by the total number of digits (ND) of the statement: the maximum of NRD is ND-1 (since at least one significant digit has to be provided) and NSD has a maximum of ND. The ignorance of ND in the above measures leads to the (unpleasant) phenomenon that two significant digits lead to the same measure of precision, regardless of whether the provided figure has 2 or 7 digits. The same holds for the number of rounded digits (2 rounded digits are equal, whether occurring at a 3- or 7-digit number) as a measure of rounding intensity. Another problem arises when there is low variation in NSD. If e.g., the respondents follow the strategy always providing 2 significant digits, then NRD is perfectly collinear with ND. Therefore, a regression of NRD on a model of explanatory variables would not yield the effects of the explanatory variables on the abstract concept of rounding intensity, but explain the amount of the provided figure in terms of the number of digits. The same holds for NSD if the rounding strategy of the respondents is to always round a certain number of digits.

We argue that one should consider relative measures when it comes to the determination of a latent rounding propensity of the interviewee.

One possible relative continuous measure is the **percentage of rounding error**, defined as the width of the rounding interval divided by the provided figure. It has the property that the percentage decreases if the provided figure is large. Given 3 rounded digits (rounded on 1,000s) of a 7-digit number would lead to 0.1% if "1,000,000" is provided and 0.01% if "9,999,000" is provided. This could be appropriate if the aim is to detect the relative damage of rounding to the data, but does not necessarily capture properly the willingness to round.

For the aims of this study, one can also think of a discrete relative measure like the share of the number of rounded digits in the total number of digits minus 1: RQ=NRD/(ND-1). This **rounding quotient (RQ)** takes account of the number of digits, i.e., the possibility for the respondent to round-off digits, and is limited to the interval [0,1]. One should bear in mind that even if it appears to be a continuous measure, the rounding quotient for a reasonable range of income statements can only take on 7 different values (fractions). Therefore, it shares the properties of discrete measures. Unfortunately, it also shares the shortcomings of the other discrete measures NRD and NSD: since the numerator of the fraction is truncated by its denominator not every outcome of RQ can be chosen by the respondent, e.g. a 5-digit figure can never be rounded by 2/3. Therefore, it remains highly questionable as to whether providing 4 of 6 digits reflects a lower rounding intensity than providing 5 of 7 digits, even if the number of zero digits is equal. (The same holds for the values 4/6 and 4/7 with respect to the number of significant digits.) Second, providing an exact figure (NRD=0) yields the same outcome (RQ=0), regardless of the number of digits. One may want a measure which reflects lower rounding intensity with 0/2 than 0/6 digits. Moreover, a lot of outcomes of the RQ are uniquely identified by a unique NRD/ND-combination, which may cause problems when estimating ordered regression models.

Taking these shortcomings into account, we define a third relative discrete measure, which we name **Rounding Strain Measure (RSM)**, as the difference of the rounded and the significant digits minus 1: RSM=NRD-(NSD-1). If the observed maximum number of digits is 7, the domain of RSM spans from -6 to 6. Negative values indicate NRD<NSD-1, i.e. a low rounding intensity, positive values indicating higher rounding. Since the odd outcomes of this measure can only be realized by an odd number of digits (the same holds for even numbers, respectively) and because the more extreme categories are unlikely to be observed, we aggregate this measure by pooling the outcomes according to the rule: 1 = {-4, -5, -6}; 2 = {-2, -3}; 3 = {-1, 0, 1}; 4 = {2, 3}; 5 = {4, 5, 6}. This gives us a discrete measure of five categories, which should reflect the latent rounding intensity and should be ordered. This measure has two disadvantages: first, figures with 2 or 3 digits are mapped to the middle-category "3". Second, its interpretation is not very illustrative.

2.4 Research Questions and Hypotheses

Before addressing the four main research questions on the determinants of rounding behavior, we take a look at descriptive statistics for the occurrence of rounding in the SHP data, the correlation of rounding error and rounding intensity with the provided value, respondent characteristics, and autocorrelation of the rounding error. This may yield evidence on how standard assumptions on misreporting errors in empirical analyses may be violated, and therefore if rounding may have damaging effects on statistical inference with such data. Moreover, we hypothesize that rounding indicates a lack of motivation for making a precise statement. In the course of time, i.e. over subsequent panel interviews, the motivation towards the survey participation and/or provision of informative answers may decrease, which may lead to unit nonresponse, item nonresponse, or the statement of "don't know". Therefore, such response behavior should be positively correlated. Descriptive evidence on these issues is provided in section 3 and tables 2.

The first main research goal of this study is to find an appropriate rounding intensity measure that meets the requirements outlined in the preceding section: it should reflect the latent rounding intensity (RI*) as well as possible and should therefore take account of the ordering of the observable discrete rounding outcomes. From a methodological point of view, ordinality and distinguishability of discrete – apparently ordered - outcomes is investigated.

Our second research question is, whether rounding in income statements occurs at random, or if socio-demographic patterns of rounding behavior exist. As hypothesized above, the respondent makes his decision in the light of cost-benefit considerations. Some of the cost-benefit determinants may be attributed to socio-demographic characteristics of the respondent; for instance, over the course of life the variability of personal income may change, e.g. pension payments may be more smoothed than work incomes, which induce differing (time-) costs of information retrieval, when asked about income. Moreover, the opportunity costs of leisure may be higher for mid-aged persons than for young or old persons. Therefore, we hypothesize that the respondents' age may have a non-linear effect on rounding intensity, with lower rounding intensities (higher precision) for respondents belonging to the tails of the age distribution. Additionally, we hypothesize that higher education of the respondent should decrease information retrieval costs and rounding intensity. The opportunity costs of leisure time may be influenced by the health status. For sick people, the time costs of the interview may be lower, since the utility of leisure (assuming that participating in interviews is not seen as a leisure activity) is lower. Therefore, RI should be lower for respondents who are in a poorer physical condition

Another possible source of influences on the interview process is the person of the interviewer. Other studies on data quality aspects, like item nonresponse, identify the interviewer as a possible source of differences in response behavior (see e.g. Riphahn and Serfling, 2005). Therefore, we hypothesize that the interviewer characteristics, as well as the interaction of interviewer and respondent characteristics may impact the rounding intensity. Since it is the job of the interviewer to collect valuable data, the job experience of the interviewer may have a (non-linear) impact on data quality. It is known from survey literature that notions, views, and values can be conferred upon the respondent (see e.g. Rice 1929 for a classical example). Hence, we hypothesize that interviewers who declared in the interviewer survey that they would not participate in an SHP interview or respond to the income question should attract higher rounding intensities. The workload of the interviewer may also impact the results of the interview. Nicoletti and Buck (2004) have shown that excessive workload for the interviewer negatively affects the cooperation probability of respondents. We hypothesize that interviewers with high workload, in terms of number of conducted interviews in the respective field period, will increase rounding. Moreover, we hypothesize that there is a nonlinear field-period effect with respect to the number of interviews conducted. Since the field period, where interviews take place, is less than half a year, we expect that the interviewers lose a part of their routine, which is recovered within the first interviews of the new field period. After a while, a fatigue effect will take place and the quality of the interviews will decrease. With respect to the rounding intensity we expect a decrease to a minimum and an increase afterwards.

The social distance between interviewers and respondents may have an influence on the costs and benefits of answering a question (see e.g. Dohrenwend et al. 1968). In light of the RC-framework one can hypothesize that low social distance should build a more trustful situation, thereby reducing the costs of disclosure of privacy, resulting in lower rounding intensity.

Our third research aim concerns the number-of-digits effect on rounding probability. Some of the studies presented in Section 1 report positive correlations between the income amount and the intensity of rounding. From the point of view of the interview process, this finding may result from two possible sources: first, respondents think it is satisfactory to provide one or two significant digits (see theory of satisficing, Krosnick et al. 1996), and therefore tend to round larger figures more strongly. Second, wealthier people may have a general tendency to be less precise, since the cost of disclosure of privacy and the opportunity costs of time are higher. The provided income statement consists of both mechanisms. Hence, we aim to disentangle the income effect of the provided figure from the number-of-digits effect and evaluate its sign and significance. We hypothesize that there exists both an income effect and a number-of-digits effect. We additionally expect the rounding intensity to be higher if the respondent has indicated that he has provided an estimated amount or a gross figure.

Our fourth and final research aim concerns the panel duration effect. This can be seen as a bundle of experience effects of the respondents, the interviewers, and confidence building between interviewer and interviewee, which may increase the willingness-to-cooperate (see Rendtel (1988)). Moreover, at least with relatively young panel studies, one can imagine growing experience in the survey organization, e.g., optimization of the operational procedures and learning progress by the panel administration. Rendtel (1988, 1989) reports such positive panel duration effects with respect to (unit) response probabilities. According to the cooperation continuum hypothesis (see Burton et al. 1999 or Serfling 2005) this result may be driven by selective attrition. Those respondents who negatively influence data quality are the less cooperative ones and drop out of the panel early. In the course of time, only the highly-cooperative respondents remain in the panel group and produce “better” data. Therefore, we hypothesize that rounding probability is positively related to the willingness to cooperate, which is negatively related to the panel attrition probability. Hence, a negative panel duration effect on respondents’ rounding intensity should become insignificant (or positive) only if the highly motivated respondents are observed.

3 Data

3.1 The Swiss Household Panel

For the analysis of measurement error with respect to rounding, we use data from the Swiss Household Panel Study (SHP). The SHP is an annually collected comprehensive survey and provides data on the living conditions in Switzerland. It comprises subjective and objective information on the housing situation, the living standard, income and its components, socio-demographics, education, employment, politics, values, and leisure (see Budowski et al. 2001). As with most surveys the questions are divided into three separate questionnaires: grid, household, and personal questionnaire. For this study, we use several income questions from the personal questionnaire, which had to be answered by each household member who had reached the age of 14. As a result of cost vs. data quality considerations, the SHP is completely surveyed by CATI (Computer Assisted Telephone Interviews) (see Scherpenzeel et al. 1999). Therefore, we will not be able to identify interview mode effects on rounding behavior. The sample size of the SHP started with 7,799 persons in 1999 and decreased continuously to 5,220 persons in 2003, since a refreshment sample was not drawn until that date. Additionally, in the second wave (2000) the interviewers were also surveyed. A questionnaire consisting of 24 questions regarding socio-demographics, interviewer experience, occupation, and opinions towards the survey had to be answered by the 53 interviewers who worked for the SHP. Forty-five of them returned the questionnaire, among them 41 who filled it in completely. Since the interviewers were surveyed only in wave 2, some information on the interviewers who worked only for wave 1, 3, and 4 is missing. Since we match interviewers with their interviewees, we lose information for another portion of personal interviews due to questionnaire nonresponse by 7 interviewers. Our sample is confronted with missing interviewer information for 1,211 out of 7,799 cases in wave 1999, which reduces to approx. 700 missing cases in the waves 2001 and 2002.

Another specialty of a Swiss survey is that in the Swiss Federation four official languages coexist: German, French, and Italian, as well as Rhaeto-Romanic (a Romance language consisting of Friulian, Tyrolese, Ladin, and the Romansh dialects). Interviewers have one of the first three languages as their mother-tongue, but must be able to converse in a second official language.

This study aims to examine the determinants of rounding behavior of respondents with income statements. For such an analysis the provided public use file of the SHP is useless, since most of the income statements therein are constructed (net from gross income, yearly income from monthly statements, etc.), imputed or otherwise changed due to reliability checks (an extensive description of the methods applied can be found in Gabadinho/Budowski 2002 and Budowski/Gabadinho/Tillmann 2002). These procedures lead to artificial rounding (and as well to artificial precision) in the income variables. Therefore we use the raw interview data, which was graciously provided by the SHP-team, for which the author is especially grateful. Within the SHP-interview, the income-section starts with a filter question regarding whether or not the respondent has received income from dependent employment in the past year. Then she is asked to provide her monthly personal total income or, if it is easier, she is encouraged to provide her annual income. In the next question she is asked for how many months she has received this income.

Table 1: Descriptive Statistics of explanatory variables

Variable	mean	s.d.	min	max	Variable	Mean	s.d.	min	max
Income Statement					**Interviewer Characteristics**				
Number of digits	4.03	0.69	2	6	Age	48.47	15.62	16	76
Significant digits (NSD)	1.74	0.69	1	6	Male	0.31	0.46	0	1
Rounded digits (NRD)	2.29	0.91	0	5	Workload	207.8	148.0	1	872
Rounding quotient (RQ)	0.75	0.24	0	1	Experience (years)	0.09	0.56	0	5
Strain measure (RSM)	3.46	0.65	1	5	1st lang. german (ref.)	0.66	0.47	0	1
Estimated amount (0/1)	0.01	0.11	0	1	1st language: French	0.26	0.44	0	1
Gross amount (0/1)	0.49	0.50	0	1	1st language: Italian	0.04	0.20	0	1
Ln(amount)	4.03	0.69	2	6	1st lang.: non-swiss	0.04	0.19	0	1
Respondent Characteristics					bilingual speaker	0.46	0.50	0	1
Age	43.4	17.12	13	93	self: no SHP partic.[1)]	0.01	0.07	0	1
Male (0/1)	0.48	0.50	0	1	self: income INR[2)]	0.07	0.26	0	1
Primary education	0.22	0.42	0	1	No. of interview	103.2	103.1	1	872
Secondary education	0.42	0.49	0	1	work in mornings	0.45	0.50	0	1
Tertiary education (ref.)	0.36	0.48	0	1	work around noon	0.56	0.50	0	1
1st language: French	0.27	0.45	0	1	work at evenings	0.62	0.49	0	1
1st language: Italian	0.05	0.22	0	1	**Respondent-Interviewer Interaction**				
1st lang.: German (ref.)	0.68	0.47	0	1	R-I age difference	-3.91	21.67	-59	67
Health: very good	0.29	0.46	0	1	same sex (0/1)	0.44	0.50	0	1
Health: good	0.56	0.50	0	1	same education (0/1)	0.37	0.48	0	1
Health: fair, so so (ref.)	0.13	0.34	0	1	same language (0/1)	0.87	0.34	0	1
Health: not very good	0.02	0.12	0	1	**Panel Participation in**				
naturalized immigrant	0.10	0.30	0	1	wave 1999 (ref.)	0.30	0.46	0	1
Immigrant	0.01	0.08	0	1	wave No. 2	0.22	0.41	0	1
Swiss native (ref.)	0.90	0.30	0	1	wave No. 3	0.21	0.40	0	1
Log. HH net equiv. inc.	9.74	3.28	0	14	wave No. 4	0.18	0.39	0	1
HH net eq. inc. missing	0.01	0.30	0	1	wave No. 5	0.17	0.38	0	1

Notes: ref.: reference category; s.d.: standard deviation.

1) The interviewer declared that he himself would not participate in an SHP-interview if requested

2) The interviewer declared that he himself would not provide his income in an SHP-interview (item nonresponse, INR)

Source: Own calculations based on the Swiss Household Panel.

In a third step it is asked whether the income statement relates to the gross or net income. If the respondent is unwilling or unable to provide her income, she is asked: "Could you estimate your yearly professional income as an employee?". This procedure is repeated for income acquired through self-employment, payments from OASI/DI (old-age, survivors and disability insurance schemes) and other sources of income.

To gather as much information as possible for the income statement variable of interest, we use the first income statement made by the respondent in the interview, starting with income from employment, social transfers etc., and ending with other sources of income. If none of these items were answered, we use the estimated income amounts of these categories in the same order. Two dummies are coded, one for gross income-statements and another one for estimated income statements.

The sample consists of the pooled observations of all 32,393 respondents of the first five waves of the Swiss Household Panel. Of these, 1,747 observations had to be dropped due to missings in the income statement. In addition, 3,542 cases in which a 1 digit-figure was provided had to be deleted, since such statements should not occur for any of the considered income variables. Lastly, 30 observations with a 7-digit number were dropped in order to avoid small cell problems with the regression methods used. The analysis sample therefore consists of 27,075 observations, thereof 25'119 statements from the earnings question. Table 1 shows the mean and standard deviations of the variables used in the following analyses.

3.2 Rounding and Rounding Error Correlations at First Glance

Table 2a shows the distribution of the number of digits, number of significant digits, number of rounded digits, and the rounding strain measure.

Table 2a: Distribution of the rounding measures

	Number of digits		Significant digits		Rounded digits		Strain measure	
	freq	**%**	**freq**	**%**	**freq**	**%**	**freq**	**%**
0					854	3.2		
1			10'198	37.7	2'966	11.0	10	0.04
2	737	2.7	14'322	52.9	13'363	49.4	704	2.60
3	2'889	10.7	1'955	7.2	7'661	28.3	14'807	54.69
4	19'419	71.7	589	2.2	1'919	7.1	9'895	36.55
5	2'967	11.0	10	0.0	312	1.2	1'659	6.13
6	1'063	3.9	1	0.0	0	0.0		
Total	27'075	100.0	27'075	100.0	27'075	100.0	27'075	100.00
Mean	4.03		1.74		2.29		3.46	
Median	4		2		2		3	
St. Dev	0.694		0.687		0.908		0.652	

Table 2b shows the mean and standard deviation of these rounding measures (including the rounding quotient) separately by the number of provided digits. Interestingly, the number of significant digits is lower with 5 digits than with 4 or 6 digits figure. On average, 56% of provided digits are rounded. The highest precision is attributed to 2-digit figures, where less than the half (=1) digits are rounded.

Table 2b: Mean and standard deviation of rounding measures by number of digits

NoD	NRD		NSD		RQ		RSM	
	mean	**s.d.**	**mean**	**s.d.**	**mean**	**s.d.**	**Mean**	**s.d.**
2	0.78	0.413	1.22	0.413	0.78	0.413	3.00	0.000
3	1.59	0.568	1.41	0.568	0.79	0.284	3.59	0.568
4	2.18	0.687	1.82	0.687	0.73	0.229	3.29	0.513
5	3.35	0.678	1.65	0.678	0.84	0.169	4.35	0.678
6	4.12	0.694	1.88	0.694	0.82	0.140	4.12	0.677
Total	2.29	0.908	1.74	0.687	0.75	0.238	3.46	0.652

Notes: s.d. = standard deviation.

Similar to Rodgers et al. (1993) we examine if and to what extent the rounding errors that we observe in the SHP data conform to the assumptions "routinely made" (Rodgers et al., p. 1212) about misreporting errors, namely:

a) $\sigma_{\varepsilon z} = 0$: the rounding error of a variable is independent of its true value z

b) $\sigma_{\varepsilon x} = 0$: the rounding error is (linearly) independent of other (explanatory) variables

c) $\sigma_{\varepsilon_t \varepsilon_{t-n}} = 0$: there exists no serial correlation in the rounding errors between two different times, t and t-n.

To assess whether the rounding error is independent of the provided figure and characteristics of the respondent, Table 2c is informative: it presents the correlations between the rounding interval width d (see 2.1.1), the stated income, the four rounding intensity measures (RQ, RS, NRD, NSD), households´ net equivalent income, and the respondent characteristics: age, sex, and education. The correlation coefficients are presented for the 25'119 statements for the total personal income question if they were significantly different from zero at the 5%-level of significance, subdivided by net/gross and yearly/monthly statements. In each of these 4 income categories, we find a high correlation (0.33-0.50) between the provided income figure and the interval width. This can be seen as first a evidence against assumption a). With respect to the characteristics of the respondent, we find positive correlations with age and male gender, and negative correlations with education, across several rounding intensity measures. This indicates that also assumption b) is likely to be violated. As expected, the correlations are stronger with net than gross income statements, since gross income normally comes with more zero digits at the rear. Therefore, the gross/net type of the statement should be considered in the analysis of respondent behavior.

Another interesting fact comes with the negative correlation of -0.12 between the provided income figure and the indicator that this figure was estimated (0/1). If choosing the estimation category is similar to rounding, this may indicate that another common assumption of the absence of systematic rounding bias, i.e. $E(\varepsilon)=0$, may be violated. This unexpected negative correlation can be seen as a hint that rounded statements may be downwardly biased

Table 2c: Correlation between rounding measures and respondent characteristics of the total income item (correlations significant at 5%-level, only)

		Net Total Income						Gross Total Income					
		Inc.	d	NSD	NRD	RQ	RSM	Inc.	d	NSD	NRD	RQ	RSM
Yearly	Income	1	**0.48**	0.15	0.43	0.09	0.10	1	**0.33**	0.12	0.39		
	EST [1)]	**-0.12**	-0.05	-0.10		0.09							
	HH net eq. inc	0.52	0.36		0.24			0.42	0.14		0.16		
	Age	0.25	0.12		0.27		0.16	0.15	0.09		0.13		
	Male	0.39	0.16		0.33	0.06	0.19	0.35	0.17	0.09	0.24		
	Low school.	-0.22	-0.09		-0.26	-0.05	-0.16	-0.17	-0.08	-0.06	-0.10		
	Med.school.	-0.09	-0.07					-0.19	-0.05	-0.09	-0.10		0.08
Monthly	Income	1	**0.45**		0.28	0.07	0.15	1	**0.50**		0.29	0.12	0.18
	HH net eq. inc	0.32	0.14	-0.05	0.18	0.08	0.13	0.22	0.14		0.17	0.05	0.12
	Age	0.1	0.05	0.23	0.06	-0.15	-0.09	0.1	0.06	0.03	0.15		0.07
	Male	0.21	0.10	-0.06	0.13	0.07	0.09	0.19	0.09	-0.05	0.16	0.08	0.12
	Low school.	-0.2	-0.08	0.04	-0.28	-0.11	-0.11	-0.14	-0.06	0.04	-0.22	-0.09	-0.08
	Med. school.			0.04	0.05		-0.03	-0.05	-0.04	0.02			-0.07

Notes: Number of income observations: 25'119, thereof 11'676 monthly gross, 10'278 monthly net, 1'761 yearly gross, 1'404 yearly net. Bold figures are mentioned in the text. d = Rounding interval width. RSM= rounding strain measure, RQ = rounding quotient, NRD = number of rounded digits, NSD = number of significant digits, Inc.= provided income amount.
[1)] The net total income was estimated by the respondent (0/1 dummy).

Table 2d provides the Pearson correlation coefficients for the measures of unit nonresponse (UNR) in wave t, item nonresponse (INR) and "don't know"-statement (DK) on the personal income question, and the rounding intensity measures: estimated value (EST), rounding interval width (d), rounding strain (RSM), number of rounded digits (NRD), and the number of significant digits (NSD) for the contemporaneous wave t and two lagged waves $t-1$ and $t-2$. Again, only those coefficients are presented, which are significantly different from zero at a significance level of 5%.

We find large autocorrelations of the rounding measures, between 0.21 for the rounding interval and 0.43 for the number of rounded digits. This shows that this common assumption on the misreporting error is likely to be violated. This may have an impact especially on time series analysis with such data (see Rodgers et al., 1993).

As hypothesized above, rounding may reflect the motivation and could therefore be related to subsequent nonresponse. We provide descriptive evidence on that issue with regard to the correlation coefficients between the nonresponse, don't know and rounding indicators in table 2d. The evidence provided by the correlations is not very supportive for our hypothesis. The correlations between the rounding and the nonresponse measures are very close to zero, but - when they are significant - they have the expected sign. Unit nonresponse is negatively correlated with previous NSD, i.e. the more precise/higher motivated respondents are less inclined to drop out. The correlation for UNR is highest with INR on the income question (0.06). The rounding measures d, RSM and NRD are positively correlated with INR and DK and with the subsequent declaration of an estimated amount (EST). Analogously, the NSD measure is negatively correlated. Looking at the column-wise highest correlations, one may speculate if there might be a temporal ordering in the several measures: rounding → estimated statements → "don't know" → INR → UNR. However, more sophisticated methods need to be employed to analyze the complex structure of interactions of response behavior in surveys, but are out of the scope of this study.

Table 2d: Correlations and Autocorrelations of Nonresponse- and Rounding-measures (correlations significant at 5%-level, only)

	Nonresponse			Rounding Measures				
	UNR_t	INR_t	DK_t	EST_t	d_t	RSM_t	NRD_t	NSD_t
INR_{t-1}	0.06	0.43	0.05	0.03	0.02	0.03	0.03	
DK_{t-1}		0.07	0.08					
EST_{t-1}		0.04	0.05	0.07		0.04	0.03	-0.04
d_{t-1}		0.04			0.21	0.16	0.19	-0.05
RSM_{t-1}		0.04	0.02	0.05	0.13	0.31	0.32	-0.21
NRD_{t-1}		0.04		0.03	0.16	0.30	0.43	-0.16
NSD_{t-1}	-0.01		-0.02	-0.03	-0.04	-0.22	-0.17	0.29
INR_{t-2}	0.02	0.42	0.07		0.02	0.02		
DK_{t-2}		0.02	0.12					
EST_{t-2}		0.06		0.04		0.03		-0.03
d_{t-2}		0.03	0.03	0.02	0.14	0.13	0.16	-0.04
RSM_{t-2}	-0.03	0.04	0.04	0.04	0.11	0.27	0.28	-0.16
NRD_{t-2}	-0.03	0.04	0.03	0.03	0.15	0.26	0.36	-0.12
NSD_{t-2}			-0.02	-0.03	-0.03	-0.18	-0.14	0.23

Notes: Abbreviations: UNR_t = Unit nonresponse in wave t, INR = Item nonresponse on the personal total income question, DK = "don't know" statement, d = rounding interval width, RSM = rounding strain measure, NRD = number of rounded digits, NSD = number of significant digits, EST = estimated amount provided.
Grey shadowed cells: autocorrelations.
Number of observations: For the calculation of the Pearson-correlation coefficients all 32'393 observations (all respondents who participated in the SHP study between 1999 and 2003, pooled) were used.

4 Empirical Strategy

4.1 An Informal Test on Ordinality and Distinguishability

The first research goal is to test whether the above discussed rounding measures are appropriate ordinal measures reflecting the latent rounding intensity of a respondent. Key conditions for the appropriateness of these measures are the distinguishability and ordinality of their outcomes. Distinguishability is defined by Anderson (1984): Two outcome categories are distinguishable with respect to a set of explanatory variables X, if X is predictive between the two categories. Additionally, we define ordinality as a restricted form of distinguishability: two outcomes of a categorical variable are ordered with respect to a set of explanatory variables X, if a positive prediction of X leads to a higher probability of observing the "higher" category, and vice versa. If the ordinality assumption does not hold, ordinal regression models are mis-specified, which may lead to biased parameters and test statistics. A test of ordinality of discrete outcomes of a variable was - to the best knowledge of the author - not yet developed in the literature, since in most fields of application the ordering of categories is obvious or seems to be obvious. We contribute to the literature of categorical dependent variables by illustrating in the following two approaches for determining whether the dependent variable is ordered. The first approach is based on the stereotype model, labeled STOT (=stereotype ordinality test), the second uses the linear predictions of the unobserved latent variable in the ordered regression model framework, labeled OPOT (=ordered probit ordinality test).

As explanatory variables X for the latent rounding intensity we use a rich set of the respondents´ characteristics, interviewer characteristics, and interactions of both, in order to reduce the possibility of an omitted variable bias. These explanatory variables are discussed in the results section five in greater detail.

The **stereotype (ordered) regression model (SORM)**, developed by Anderson (1984), can be thought of as a constrained multinomial logit model. The multinomial logit (MNL) model of the form:

$$P(y=m\mid X)=\frac{\exp(\beta_m{}'X)}{1+\sum_{j=2}^{J}\exp(\beta_j{}'X)} \qquad (4.1.1)$$

(with j=1…J categories, the parameter vector for the base category (here j=1) being constrained to zero ($\beta_1 = 0$)) estimates J-1 parameter vectors. The stereotype model now imposes a proportional regression assumption:

$$\beta_j = -\phi_j\beta \qquad (4.1.2)$$

where ϕ can be seen as an outcome specific scale factor of the regression coefficient vector. Therefore, instead of J-1 β-vectors as in the multinomial logit case, only one vector of regression parameters, but J additional ϕ_j parameters, have to be estimated or need to be constrained. This allows for differing relative importance of the regressors between the categories. The resulting stereotype model has the form:

$$P(y=m \mid X)=\frac{\phi_m \exp(\beta' X)}{\sum_{j=1}^{J} \phi_j \exp(\beta' X)} \qquad (4.1.3)$$

In order to identify the model, two additional constraints on the φ-parameters have to be imposed, commonly $\phi_1=1$ and $\phi_J = 0$. The resulting model can be transformed to an ordered regression model if the ordering constraint ORD: $1 = \phi_1 > \phi_2 > \ldots > \phi_J = 0$ is imposed. For testing purposes, the stereotype model may be fitted without the ordering constraint, and the ordering of the categories can be tested. Anderson (1984) proposes a Likelihood-Ratio Test to compare the likelihood of an ordering constrained stereotype model (L_{ORD}) with the likelihood of an unconstrained stereotype model (L_1) using the test statistic $2(L_1-L_{ORD})$. For the purposes of this paper we propose to use a t-test of the ordering hypotheses of two adjacent categories: H_0: $\phi_j - \phi_{j+1} \leq 0$ against H_1: $\phi_j - \phi_{j+1} > 0$. H_0 should be rejected at a desired level of significance if the categories are ordered. Additionally the distinguishability of both categories may be tested by H_2: $\phi_j - \phi_{j+1} = 0$ against H_3: $\phi_j - \phi_{j+1} \neq 0$.

It may also be desirable to test if the "proportional regression" constraint (see 4.1.2) of the stereotype model holds. Test procedures have not been developed yet, but since the stereotype model is a constrained multinomial logit model (both models are nested), a Likelihood-Ratio test over the Log-Likelihood of both models (MNL and stereotype model) may yield insight into the appropriateness of the parallel regression constraint. If the constraint is satisfied, the log-likelihoods of both models should be close. The stereotype regression model was programmed for the use with STATA by Lunt (2001), which is gratefully acknowledged by the author.

Our second approach uses the **ordered regression model (ORM, i.e. ordered logit, ordered probit)**, where the probability of chosing a category m is:

$$P(y=m \mid X)=F(\mu_m-\beta' X)-F(\mu_{m-1}-\beta' X) \qquad (4.1.4)$$

(with $\mu_J = +\infty$ and F(.) being the CDF of the standard normal distribution for the ordered probit and logistic distribution in the ordered logit model). Here, only one parameter set β and J-1 additional threshold parameters μ_j are estimated. The parallel regression assumption (also known as the proportional odds assumption in the ordered logit case) is imposed, which implies that the effect of a regressor on the probability of observing an outcome category j is independent of the category involved (see 4.1.4). The difference in the probabilities of two categories depends only on the differences in the covariates (see McCullagh, 1980). Commonly, the regression constant is omitted ($\beta_0 = 0$) and μ_1 is estimated;

however, constraining $\mu_1 = 0$ and estimating β_0 is also possible and has no consequences for the coefficients. The thresholds μ_j link the underlying, unobservable values y* to the observed categories of y:

$$y^* = \beta' X + \varepsilon \quad with \quad y = j \quad if \quad \mu_{j-1} < y^* \le \mu_j \qquad (4.1.5)$$

Therefore, the predicted latent variable $\hat{y}^* = \hat{\beta}' X$ should have higher values for higher observable categories and vice versa. After fitting the data with a given model X, the latent variable is predicted, and for each observed category of y_j the mean of y* is calculated: $\overline{\hat{y}_j^*} = E(\hat{y}^* \mid y = j)$. According to Long (1997) the variance of y* can be estimated by the quadratic form:

$$Var(y^*) = \hat{\beta}' \Sigma_X \hat{\beta} + \sigma_\varepsilon^2 \qquad (4.1.6)$$

(where Σ_X denotes the covariance matrix for the X's in the observed data, and σ_ε^2 denotes the variance of the error term, which is 1 for the ordered probit and $\pi^2/3$ for the ordered logit model.)

The standard error of the predicted latent variable is σ_{y*} for each category, which allows us to use a t-test for each pair of two adjacent categories i and j. We test the following two hypotheses at a level of significance of $\alpha=0.05$ for all categories 1…J:

Ordinality:

H_0: $\overline{\hat{y}_i^*} \le \overline{\hat{y}_j^*}$ against H_1: $\overline{\hat{y}_i^*} > \overline{\hat{y}_j^*}$ for $j = i + 1$

Distinguishability:

H_2: $\overline{\hat{y}_i^*} = \overline{\hat{y}_j^*}$ against H_3: $\overline{\hat{y}_i^*} \ne \overline{\hat{y}_j^*}$ for $j = i + 1$ (if H_0 is not rejected)

Our aim is to find the correct ordering of the categories within the ordered probit framework. Since the predicted latent variable is valid only for the given data and the given set of covariates, and the magnitude of the possible bias of spuriously ordered categories remains unclear, we do not solely rely on the results of the above tests for each category, based on a singular estimation of the model. We chose a more conservative strategy, akin to stepwise regression, where the test procedure takes place in several iterations. In any iteration, all J consecutive categories are tested for correct ordering and distinguishability (H0, H2). Then, the dependent variable is recoded according to the rule: if H_0 is rejected, the categories i and j are interchanged. If neither H_0 nor H_2 is rejected, the categories i and j are pooled. Only if H_0 is not rejected but H_2 is, do we take this as evidence that the implied ordering is correct and leave the data unaltered. This data recoding (interchanging or pooling) is only imposed for the category with the clearest decision in terms of the test statistic (largest p-value) of the respective test. Afterwards, the model is re-estimated and the algorithm iterates test and recode until H_0 cannot be rejected and H_2 is rejected for each category. This results in re-ordered categories of the dependent variable whose categories are ordinal and distinguishable for the given ordered probit model.

A first insight on whether an ordered regression model may be misspecified may be given by a test on whether the parallel regression assumption of the ordered regression model holds. Brant (1990) developed a Wald-Test on the equality of the β-vectors across the j outcomes of the dependent variable y_j. The intuition of this test is that the regression parameters of consecutive binary cumulated logits should not vary a lot. Hence, the null hypothesis of this test is that regression parameters are independent of the categories. This test procedure was implemented for the use within the ordered logit framework in STATA by Long et al. (2003). Unfortunately, the Brant test can be computed only if all of the independent variables in the ordered model are retained for each category. This is unlikely to be the case with models that have few observations in the extreme categories and many independent variables. Therefore, we limited the set of explanatory variables to respondent characteristics for the Brant-test procedure. (This should not affect the generality of the results, since respondent characteristics turn out to have the strongest explanatory power.)

To evaluate direction and magnitude of the possible bias due to spurious ordering, we compare the regression coefficients of two models for the rounding quotient, the first with the originally ordered categories, the second with re-ordered categories according to the ordering-proposition of our ordinality test, using the ordered probit approach.

4.2 Determinants of Rounding Intensity

If our test of ordinality fails for any of the discrete rounding intensity measures (i.e. NRD, NSD, RQ, RSM) one should consider the rounding measure as a nominal measure and fit a multinomial logit model (MNL). One disadvantage of the MNL is that the problem of small cells arises for the extreme categories, even with 27,075 observations. (Another point is that the reader may find the bundle of parameter vectors awkward to interpret.)

We argue that the non-ordinality of the count-measures NRD and NSD in our behavioral response model derives from neglect of the number of digits (the rounding possibility): the counted outcomes “number of significant digits” and “number of rounded digits” are ordered naturally (3 is always more than 2!). In our response model we model the latent rounding intensity of the respondent, as explained by characteristics of the respondent, the interviewer, interactions thereof, and the interview situation. Failing to ascertain ordinality of the outcomes means that the modeled latent (unobservable) rounding intensity y* is not monotonically connected to the observed categories y. This is expected in situations where the outcome of n rounded digits is evaluated for differing numbers of digits. Clearly, the rounding intensity for 2 rounded out of 3 digits is higher than for 2 rounded out of 7 digits. Therefore, the model fit should strongly increase if we allow the regression coefficients for the determinants of rounding intensity to vary for the number of digits. In this way, the latent unobservable variable should be correctly associated with the observed outcomes. Hence, we estimate an ordered probit model of the number of rounded digits (NRD) with a full set of dummy-interactions of all explanatory variables for 2, 3, 5, and 6 digits (a 4 digit number as reference group). Additionally, we control for both base effects for the number of digits and for the respective explanatory variable. To avoid the problem of small cells, observations with 7 digits are dropped from the sample, which induces a loss of 30 observations. After the estimation, the set of ND-interactions (omitting the base effect) are tested on joint significance using an LR-Test to determine whether the effect of the respective explanatory variable varies with the length of the stated income. The baseline specification of the used model consists of the number-of-digit dummies and a dummy for "estimated" values, where the respondent was asked to provide a best guess of the value. Since the provided income statement may also stem from a question, for which the respondent indicated that the figure provided was a gross-amount (these naturally having more zeros at the end), a full set of ND-interactions of a "gross-value" dummy is also added to the baseline specification:

$$y_i^* = \beta_1 EST_i + \sum_{j=1}^{NRD\max} \beta_{1+j}(ND_{ij} \cdot GROSS_i) + \sum_{j=1}^{NRD\max} \beta_{1+NRD\max+j} ND_{ij} + \ldots + \varepsilon_i \qquad (4.2.1)$$

where EST=1 if the respondent i declared that the provided amount was a guess, 0 otherwise; ND_{ij} is the number-of-digits dummy, which is 1 if the number of digits (for the statement of respondent i) = j and 0 otherwise; GROSS = 1 if the provided value was a gross amount, and 0 otherwise; ε_i denotes the error term of the econometric model.

Concerning the second research question, the existence of socio-demographic patterns of rounding, we expand the set of explanatory variables by the respondent characteristics age, sex, education, mother tongue, health status, and nationality (model 1). The joint significance of effects is tested by a likelihood-ratio (LR) test. To put the non-linearity hypothesis of the age effect to the test, we add the squared age of the respondent to the model and calculate its maximum.

To specify the influence of the interviewer on rounding, the model above is augmented with interviewer fixed effects (model 2) and their joint significance is tested. In an alternate specification we integrate the interviewer characteristics: sex, age, language, interviewer-experience, workload, and opinions towards the SHP panel study. As a measure of social distance, interviewer-respondent interaction characteristics are added (model 3). Tests of joint significance of the interviewer coefficients are presented.

Addressing our third research question, the disentangling of the amount-effect from the number-of-digits effect, we augment the model with logarithmic income splines for each figure length of 2, 3, 5, and 6 digits (model 4), estimate a model with a higher order polynomial of the provided amount (model 6), add household wealth to both models (giving models 5 and 7) and check whether the amount and number of digit effects remain significant. As a measure for households' wealth status, we use the log of the household's net equivalence income, here defined as: "the yearly household net income, equivalized by the modified OECD-scale, after deduction of social security contributions, taxes not deducted" (SHP 2005, p. 61). Additionally, we control for the observations where this equivalence income is missing due to item nonresponse on the income question in the household questionnaire. Again, we present parameter tests on joint significance of the ND-dummies, the log-amount splines, the amount polynomial, and the equivalent income.

Addressing the question of panel duration effects, dummies for the waves 2 to 5 are included (first wave as reference) (model 8) and their coefficients are tested for joint significance. To check for robustness with respect to selective attrition, the estimation is repeated for only those respondents who participated in all 5 waves (model 9). A change in the results for the latter specification would indicate panel-selection effects with respect to rounding.

Since the coefficients of the ordered probit can be interpreted only with respect to sign and significance, but predicted probabilities and marginal effects have to be calculated and presented for each of the 5 outcome categories (number of rounded digits), we reduce the information by providing tables and figures for the impact on the mean number of rounded digits. Since the NRD is a counted outcome, the arithmetic mean is interpretable:

$$\overline{NRD} = \sum_{m=0}^{NRD\max} P(y = m \mid X) * m \qquad (4.2.2)$$

Hence, the marginal change in the mean NRD can be calculated as the weighted sum of the marginal effects for each outcome:

$$\frac{\partial \overline{NRD}}{\partial x_i} = \sum_{m=0}^{NRD\max} \frac{\partial P(y = m \mid X)}{\partial x_i} * m \qquad (4.2.3)$$

This gives a single measure on the direction and strength of the impact of a marginal change in x_i, evaluated at the mean of the data, on the mean number of roundings.

5 Results

5.1 Ordinality and Distinguishability

The first research aim was to find an appropriate measure that reflects the latent rounding intensity of a respondent. Two strategies to test ordinality of the outcomes of a categorical variable have been developed in Section 4.1. In Table 3a, we present goodness of fit statistics for fitting an ordered probit model of the originally ordered categories (row A), the results of the Brant test on the parallel regression assumption (row B), the order of the categories implied by the OPOT results (row C), the results of fitting a stereotype regression model (row D), and the implied ordering of the categories implied by STOT (row E).

With respect to the model fit of our rounding model consisting of 42 determinants (see row A), the number of rounded digits (NRD) model seems to be the most advantageous: the null hypothesis of the likelihood ratio test on the joint significance of all 42 slope coefficients is strongest rejected, McFaddens R^2 reaches its maximum with 0.106, and the Bayesian Information Criterion (BIC') reaches its lowest value. In contrast, the Akaike Information Criterion (AIC, being the lowest) and the percentage of correctly predicted outcomes (Count R^2, being highest) indicate a better fit for a model with the rounding strain measure as dependent variable. (For a general discussion of goodness of fit measures, see e.g. Long (1997), for information-based measures, see Judge et al. (1985), and for the Schwartz Criterion (BIC and BIC'), see Raftery (1996).)

Nonetheless, the Brant test (see row B) for the parallel regression assumption in the ordered regression model fails for each of the provided measures: the null hypothesis, i.e. regression coefficients are equal in the cumulated logits over all categories, is rejected at any level of significance. Concerning our tests on ordinality, based on the developed OPOT approach (row C) and the stereotype model (STOT, row E), we find that the results of both approaches correspond, with exception for those from the RQ-model. The tests for the NSD-model imply that the outcomes 1-4 are ordered, but the extreme categories of 5 and 6 significant digits yield lower latency levels. Therefore, the ordered probit approach implies the ordering: 6<1<2<5<3<4, while the stereotype model could only be fitted if observations with outcomes 5 and 6 were dropped. The results for the NRD-model suggest that the number of rounded digits are mostly ordered along with the latent rounding intensity. Only the extreme category of zero rounded digits cannot be identified: the oprobit approach finds that the predicted latent variable for the "0" outcome is not significantly different from that of "1" and implies to pool those outcomes. In the stereotype regression model, the phi-parameter for "1" (ϕ_2) with 1.016 slightly exceeds ϕ_1, which is constrained to 1 in the model. Therefore, the stereotype model implies the ordering: 1<0<2<3<4<5. As proposed above, this wrong ordering of the NRD-outcomes may derive from the neglect of the number of digits of the provided figure. As a test of this proposition, we have interacted the whole set of explanatory variables with number of digits dummies and imposed the two tests (STOT and OPOT). Unfortunately, due to small cells, those tests could only be computed if the "0"-category was dropped and then yielded correct ordering of the categories 1 to 5.

With respect to the rounding quotient (RQ), we find the outcomes to be completely unordered. Moreover, both ordinality test procedures (OPOT and STOT) imply a different re-ordering of the categories. (Robustness checks have shown that the results of the OPOT are confirmed if the test procedure is alternatively carried out within the ordered logit framework. Hence, our results are robust with respect to the underlying distribution of the error term.)

Table 3a: Tests on Ordinality in Response Models for several Rounding Measures

Dependent Variables:	NSD	NRD	RQ	RSM
A) Ordered Probit				
Log Likelihood	-25'930.28	-31'376.307	-37'286.494	-24'820.932
LR Test χ^2 (d.f.)	1'259.52 (42)	7'427.895 (42)	1'424.497 (42)	2'712.610 (42)
McFaddens R^2	0.024	0.106	0.019	0.052
Count R^2	0.531	0.511	0.423	0.564
AIC	1.919	2.321	2.758	1.837
BIC'	-830.857	-6'999.228	-995.829	-2'283.943
B) Brant Test				
$p>\chi^2$	0.000 *** [1)]	0.000 ***	0.000 ***	0.000 *** [2)]
χ^2 (d.f.)	735.53 (24)	949.92 (48)	3904.42 (112)	699.04 (24)
C) Ordering based on OPOT	6, 1, 2, 5, 3, 4	(0=1), 2, 3, 4, 5	0, 1/3, 1/2, 2/3, 1, 1/4, 1/5, 3/4, 2/5, 3/5, 4/5	2, 3, 4, 1, 5
D) Stereotype Model	[1)]			
LogL SO	-25'800.00	-31'800.00	-37'600.000	-25'100.00
LRT null χ^2 (d.f.)	1'294.56 (14)	6'701.56 (38)	755.08 (43)	2'088.47 (37)
LRT MNL χ^2 (d.f.)	185.53 (22)	675.78 (132)	5455.50 (297)	571.54 (99)
$\phi1$ (s.e.)	1 .	1 .	1 .	1 .
$\phi2$	0.892 (0.012)	1.016 (0.011)	1.813 (0.230)	3.463 (2.482)
$\phi3$	0.518 (0.025)	0.728 (0.014)	1.180 (0.077)	2.221 (1.590)
$\phi4$	0	0.558 (0.019)	0.358 (0.019)	1.579 (1.130)
$\phi5$		0.221 (0.033)	1.504 (0.129)	0 .
$\phi6$		0 .	0.399 (0.021)	
$\phi7$			0.737 (0.040)	
$\phi8$			-0.022 (0.002)	
$\phi9$			0.352 (0.018)	
$\phi10$			0.513 (0.027)	
$\phi11$			0 .	
E) Ordering implied by ϕ-coefficients, STOT	1, 2, 3, 4	1, 0, 2, 3, 4, 5	1/5, 2/5, 1/4, 0, 3/5, 4/5, 1/2, 1/3, 3/4, 1, 2/3	2, 3, 4, 1, 5

Notes: NSD: number of significant digits, NRD: number of rounded digits, RQ: rounding quotient = NRD/(NRD+NSD-1), RSM: rounding strain measure= NRD-(NSD-1) (recoded to 4 categories).

1) Categories 5 and 6 were omitted since Brant-test statistic and stereotype regression model could not be computed due to small cells.

2) Brant test without category 1, due to small cells.

We interpret this finding such that the ordered models with RQ as dependent variable are not correctly identified. This is because some outcomes (e.g. 1/6) are completely identified by the number of digits and not by respondent behavior. Accordingly, the RQ-model also has the worst model fit (row A) in terms of the pseudo R^2 measures, AIC, BIC', and the clearest rejection of the parallel regression assumption in the Brant test (row B).

The results for the rounding strain measure indicate a fair ordering of the outcomes. Only the "1" is arranged between 4 and 5. But, the "1" outcome is only observed for 10 observations. Therefore, we conclude that this re-ordering artifact stems from a small cells problem. As a measure of the appropriateness of the stereotype model, the Likelihood-Ratio Tests provided in rows D of Table 3a are informative. If the imposed proportional regression assumption of the SRM holds, its log-likelihood should not be significantly lower than in the multinomial logit case. Therefore, the log-likelihood is tested (1) against a null model consisting of a regression constant only and (2) the MNL model. In all models, both hypotheses are rejected and the log-likelihoods indicate that the stereotype model has more explanatory power than a constant-only model, but not as much explanatory power than a MNL.

To derive an impression of the effects of spuriously ordered categorical variables with ordered regression models, we present the estimation coefficients of the originally ordered and re-ordered outcomes of the rounding quotient in Table 3b. Both models were estimated using an ordered probit model and the reordering of the categories was done according to the results of the OPOT. One may find the coefficients not being reliable, since the STOT approach implied a different ordering and the parallel regression assumption is rejected by the Brant-test for both models. Nonetheless, the change in the coefficients between the two regressions is significant. Mostly the coefficients have the same sign, but are more distinctive in the re-ordered model, i.e. partly twice as large. Other than the estimated mean of the coefficients, their estimated standard errors remain fixed. This leads to an increase in the significance of the coefficients. A substantial change can be found for the gross-dummy which changes from -0.049, significant at the 1%-level, to +0.017 not significant. Furthermore, the coefficient for the quadratic age of the respondent and the number of the interview alternates its sign from positive to negative, which has an impact on the monotonicity or convexity of respective effect.

In conclusion, the several tests provided here have shown that there are two potentially correct ordinal measures for rounding intensity: the number of rounded digits (NRD) and the rounding strain measure (RSM). Therefore, we use both measures for the assessment of the rounding behavior of respondents. The RSM should directly reflect the rounding intensity, while NRD should reflect the rounding intensity if the number of digits are taken into account. For the following analysis we use the RSM while omitting the irregular category 1, which involves a loss of 10 observations, and the NRD in a model where the explanatory variables are interacted with the number of digits, to overcome ordering irregularities.

Table 3b: Comparison of the ordered probit regression results for the originally ordered and reordered categories of the rounding quotient (RQ) (only coefficients that are significant at the 10%-level in either of both models)

	Originally ordered RQ			Re-ordered RQ		
	coef.	**s.e.**		**coef.**	**s.e.**	
Statement Characteristics						
Estimated amount (0/1)	0.680	0.065	***	1.062	0.059	***
Gross amount (0/1)	-0.049	0.015	***	0.017	0.015	
Respondent Characteristics						
Age	0.026	0.002	***	0.047	0.002	***
Age^2	0.000	0.000	***	-0.001	0.000	***
Male (0/1)	0.157	0.014	***	0.318	0.014	***
Primary education	-0.153	0.021	***	-0.212	0.020	***
Secondary education	-0.104	0.022	***	-0.249	0.022	***
1^{st} language: French	-0.092	0.027	***	-0.104	0.026	***
1^{st} language: Italian	-0.061	0.052		-0.089	0.050	*
Health: very good	0.132	0.023	***	0.191	0.022	***
Health: good	0.108	0.021	***	0.140	0.020	***
Foreigner	-0.007	0.023		-0.041	0.023	*
Log(HH's net equiv inc.)	0.135	0.015	***	0.307	0.015	***
HH net eq. inc. missing	1.589	0.164	***	3.417	0.160	***
Interviewer Characteristics						
Age	0.002	0.001	*	0.002	0.001	
Male	0.033	0.017	*	0.067	0.016	***
Experience (years)	-0.083	0.058		-0.049	0.057	
$Experience^2$	0.019	0.012	*	0.010	0.012	
Bilingual speaker	-0.026	0.017		-0.027	0.016	*
No. of Interview	0.000	0.000		0.001	0.000	***
No. of $Interview^2$	0.000	0.000		-0.001	0.000	
Respondent-Interviewer Interaction						
R-I age difference	0.002	0.001	*	0.002	0.001	**
Same sex	-0.013	0.015		-0.032	0.014	**
Same education	-0.003	0.023		0.038	0.022	*
Panel Duration						
Wave No. 2	-0.050	0.022	**	-0.022	0.021	
Wave No. 3	-0.058	0.023	**	-0.025	0.022	
Wave No. 4	-0.022	0.024		0.017	0.023	
Wave No. 5	-0.233	0.025	***	-0.134	0.024	***
Threshold Parameters: $\mu1$	-0.060	0.179		2.386	0.175	
$\mu2$	-0.059	0.179		2.916	0.175	
$\mu3$	-0.049	0.179		3.175	0.175	
$\mu4$	0.462	0.179		4.500	0.175	
$\mu5$	0.463	0.179		5.903	0.176	
$\mu6$	0.715	0.179		5.909	0.176	
$\mu7$	0.747	0.179		5.909	0.176	
$\mu8$	2.029	0.180		6.474	0.178	
$\mu9$	2.161	0.180		6.477	0.178	
$\mu10$	2.217	0.180		6.605	0.178	
# obs.	27'075			27'075		
Log-Likelihood	-37'297.54			-36'308.349		
LR-Test χ^2 (39 d.f.); $p > \chi^2$	1413.39 ; 0.000			3391.60 ; 0.000		
McFaddens R^2	0.0186			0.0446		

Notes: significance levels: 10% *; 5% **; 1% ***; s.e. = standard error.

5.2 Determinants of Rounding Intensity

The estimation results of the stepwise augmented ordered probit models are summarized in table 4a with respect to the set of explanatory variables used, number of observations and measures of model fit, separately for the NRD and RSM outcome. The related LR-tests of joint significance are presented together with table 4b. The most complete model 10 is presented in full detail, with coefficients and marginal effects evaluated at the mean of the data for each outcome, for the RSM measure in table 5. Table 6a presents coefficients, z-values for the base effect, sign and significance indicators for the ND-interactions, and tests for joint significance of the ND-interaction for each variable for the ordered probit model with NRD. The marginal effect of a change in an explanatory variable on the mean number of rounded digits is presented in Table 6b. In the following, we interpret with these tools the results in the light of our research questions and hypotheses. Overall, we can state that the results of the two models (NRD and RSM) are very similar. Nonetheless, the results of the NRD-model have the higher interpretative value since marginal effects on the mean number of rounded digits can directly be calculated. All results presented are robust with respect to the specification of the set of explanatory variables.

Table 4a: Explanatory Power of Different Models for NRD and RSM, oprobit

	Set of explanatory variables							Model fit				
Model No.	**Respondent**	**WaveFE**	**Interviewer**	**R-I Interaction**	**Amount**	**HH income**	**I's work time**	**# observations**	**LogL**	**LR-Test χ^2 (df)**	**P > χ^2**	**McFadden's R^2**
NRD:												
1	Nx							27,075	-26,071	18,037 (70)	0.000	0.257
2	Nx		FE					26,769	-25,739	17,957 (120)	0.000	0.259
3	Nx		Nx	Nx				27,075	-26,105	18,151 (155)	0.000	0.259
4	Nx		Nx	Nx	SNx			27,075	-25,777	18,625 (170)	0.000	0.265
5	Nx		Nx	Nx	SNx	Nx		27,075	-25,733	18,715 (180)	0.000	0.267
6	Nx		Nx	Nx	P			27.075	-25.774	18,632 (170)	0.000	0.266
7	Nx		Nx	Nx	P	Nx		27,075	-25,733	18,715 (180)	0.000	0.267
8	Nx	Nx	Nx		SNx			27,075	-25,690	18,800 (160)	0.000	0.268
9	B	Nx	B		SNx			16,407 [1)]	-15,768	10,410 (60)	0.000	0.248
10	Nx	Nx	Nx	Nx	SNx	Nx	Nx	27,075	-25,674	18,832 (195)	0.000	0.268
RSM:												
1	B							27,065	-24,919	2,336 (14)	0.000	0.045
2	B		FE					27,065	-24,880	2,416 (67)	0.000	0.046
3	B		B	B				27,065	-24,894	2,388 (33)	0.000	0.046
4	B		B	B	SNx			27,065	-20,887	10,402 (42)	0.000	0.199
5	B		B	B	SNx	B		27,065	-20,867	10,443 (44)	0.000	0.200
6	B		B	B	P			27.065	-20,872	10,430 (42)	0.000	0.200
7	B		B	B	P	B		27,065	-20,853	10,470 (44)	0.000	0.201
8	B	B	B		SNx			27,065	-20,815	10,547 (44)	0.000	0.202
9	B	B	B		SNx			16,399 [1)]	-12,817	6,421 (43)	0.000	0.200
10	B	B	B	B	SNx	B	B	27,065	-20,812	10,553 (51)	0.000	0.202

Notes: Besides the set of explanatory variables, each regression model consists of the baseline specification: "number of digit"-dummies, "gross"-statement dummy interacted with ND and "estimated"-statement dummy. Model 10 is presented in Table 5 with greater detail.

Legend: B = base effect only, Nx = base effect with full set of ND-interactions, FE = interviewer fixed effects instead of characteristics, SNx = log(amount)-splines fully interacted, P = 5th order amount polynomial, (df) = degrees of freedom.

[1)] Only Respondents, who participated in all 5 wave interviews

Table 4b: Likelihood-Ratio Tests on joint significance of coefficients

Test # relating to Model	Test on joint significance of:	NRD, interacted		RSM	
		LR χ^2 (df)	p> χ^2	LR χ^2 (df)	p> χ^2
1a	Respondents' age, age^2	476.7 (2)	0.000 ***	511.8 (2)	0.000 ***
1b	Respondents' education	135.9 (2)	0.000 ***	269.9 (2)	0.000 ***
1c	Respondents' health	29.6 (3)	0.000 ***	69.9 (3)	0.000 ***
1d	Respondents' nationality	0.9 (2)	0.644	7.5 (2)	0.024 **
1e	Respondents' interview language	56.6 (2)	0.000 ***	89.5 (2)	0.000 ***
2a	Interviewer fixed effects	74.4 (53)	0.014 **	79.7 (53)	0.010 **
3a	Interviewer effects	36.9 (13)	0.000 ***	24.6 (15)	0.056 *
3b	R-I interaction effects	3.7 (4)	0.442	4.3 (4)	0.367
3c	Number of digit (ND) effects	90.8 (4)	0.000 ***	6232 (4)	0.000 ***
4a	Number of digit (ND) effects	112.4 (4)	0.000 ***	165.8 (4)	0.000 ***
4b	Ln(amount)-splines	462.3 (5)	0.000 ***	347.5 (5)	0.000 ***
5a	Ln(amount)-splines	398.2 (5)	0.000 ***	334.5 (5)	0.000 ***
5b	Ln(HH net equiv. inc)	88.8 (10)	0.000 ***	40.5 (2)	0.000 ***
6a	Amount polynomial (5th order)	468.2 (5)	0.000 ***	376.3 (5)	0.000 ***
6b	Number of digit (ND) effects	41.8 (4)	0.000 ***	2612 (4)	0.000 ***
7a	Amount polynomial (5th order)	397.9 (5)	0.000 ***	361.7 (5)	0.000 ***
7b	Ln(HH net equiv. inc)	82.0 (10)	0.000 ***	39.1 (2)	0.000 ***
7c	Number of digit (ND) effects	17.8 (4)	0.001 ***	2583 (4)	0.000 ***
8a	Panel fixed effects	111.8 (4)	0.000 ***	108.4 (4)	0.000 ***
8b	Wave 5	17.5 (4)	0.002 ***	-8.6 (1) [1)]	0.000 ***
9a	Panel fixed effects	78.9 (4)	0.000 ***	71.8 (4)	0.000 ***
9b	Wave 5	13.9 (4)	0.007 ***	-6.4 (1) [1)]	0.000 ***
10a	Respondents' age, age^2	194.4 (2)	0.000 ***	132.4 (2)	0.000 ***
10b	Respondents' education	44.2 (2)	0.000 ***	48.9 (2)	0.000 ***
10c	Respondents' health	21.2 (3)	0.000 ***	12.4 (2)	0.006 ***
10d	Respondents' nationality	1.9 (2)	0.382	0.9 (2)	0.642
10e	Respondents' interview language	19.4 (2)	0.000 ***	10.0 (2)	0.006 ***
10f	Interviewers' experience, exp.^{2}	4.2 (2)	0.122	3.9 (2)	0.142
10g	Interviewers' mother tongue	4.6 (4)	0.331	4.0 (4)	0.401
10h	Interviewers' opinion towards SHP	1.0 (2)	0.603	1.1 (2)	0.566
10i	Interviewers' workload, # interviews	6.5 (3)	0.092 *	1.4 (2)	0.501
10j	R-I Interaction effects	1.4 (4)	0.840	4.2 (4)	0.385
10k	Number of digit (ND) effects	108.2 (4)	0.000 ***	157.0 (4)	0.000 ***
10l	Ln(amount)-splines	101.9 (4)	0.000 ***	339.4 (5)	0.000 ***
10m	Panel fixed effects	109.1 (4)	0.000 ***	106.7 (4)	0.000 ***

Notes: [1)] t-value instead of LR χ^2

Table 5: Regression coefficients and marginal effects of ordered probit regression on rounding strain (RSM) (model 10)

Dep. Var.: RSM	Regression results			Marginal effects, evaluated at mean							
				RS=2		RS=3		RS=4		RS=5	
	coef.	t		ME		ME		ME		ME	
2 digits (ND2)	1.814	4.45	***	-0.013	***	-0.507	***	0.207	**	0.312	**
3 digits (ND3)	2.900	11.70	***	-0.024	***	-0.638	***	0.024		0.638	***
5 digits (ND5)	1.629	5.06	***	-0.017	***	-0.510	***	0.315	***	0.212	**
6 digits (ND6)	4.923	4.26	***	-0.018	***	-0.608	***	-0.360	***	0.986	***
Estimated value (0/1)	0.471	6.83	***	-0.008	***	-0.177	***	0.159	***	0.027	***
Gross value (0/1)	-0.035	-2.06	**	0.001	**	0.013	**	-0.013	**	-0.001	**
Ln(amount) (base)	0.268	15.00	***	-0.008	***	-0.098	***	0.097	***	0.009	***
Ln(amount) * ND2	-0.289	-2.77	***	0.008	***	0.106	***	-0.105	***	-0.010	***
Ln(amount) * ND3	-0.263	-6.93	***	0.008	***	0.096	***	-0.095	***	-0.009	***
Ln(amount) * ND5	0.004	0.13		0.000		-0.002		0.002		0.000	
Ln(amount) * ND6	-0.356	-3.64	***	0.010	***	0.131	***	-0.129	***	-0.012	***
Respondent Characteristics											
Age	0.028	9.65	***	-0.001	***	-0.010	***	0.010	***	0.001	***
Age^2	0.000	-11.46	***	0.000	***	0.000	***	0.000	***	0.000	***
Male (0/1)	0.033	1.94	*	-0.001	*	-0.012	*	0.012	*	0.001	*
Primary education	-0.164	-6.98	***	0.005	***	0.059	***	-0.059	***	-0.005	***
Secondary education	-0.033	-1.34		0.001		0.012		-0.012		-0.001	
1st language: French	-0.096	-3.13	***	0.003	***	0.035	***	-0.035	***	-0.003	***
1st language: Italian	-0.033	-0.57		0.001		0.012		-0.012		-0.001	
Health: very good	0.079	3.15	***	-0.002	***	-0.029	***	0.029	***	0.003	***
Health: good	0.067	2.93	***	-0.002	***	-0.025	***	0.024	***	0.002	***
Health: not very good	-0.024	-0.37		0.001		0.009		-0.009		-0.001	
Naturalized immigrant	-0.034	-0.36		0.001		0.012		-0.012		-0.001	
Immigrant	0.022	0.85		-0.001		-0.008		0.008		0.001	
log(HH's net eq. inc.)	0.036	2.04	**	-0.001	**	-0.013	**	0.013	**	0.001	**
HH net eq. inc. missing	0.524	2.76	***	-0.010	***	-0.196	***	0.178	***	0.029	*
Interviewer Characteristics											
Experience (years)	-0.111	-1.66	*	0.003	*	0.041	*	-0.040	*	-0.004	*
$Experience^2$	0.025	1.85	*	-0.001	*	-0.009	*	0.009	*	0.001	*
SHP Participation?	-0.006	-0.05		0.000		0.002		-0.002		0.000	
Response on income?	-0.044	-1.07		0.001		0.016		-0.016		-0.001	
Respondent-Interviewer Interaction											
Same education	-0.045	-1.80	*	0.001	*	0.016	*	-0.016	*	-0.001	*
Panel Duration											
Wave No. 2	-0.041	-1.71	*	0.001	*	0.015	*	-0.015	*	-0.001	*
Wave No. 3	-0.031	-1.28		0.001		0.011		-0.011		-0.001	
Wave No. 4	-0.006	-0.23		0.000		0.002		-0.002		0.000	
Wave No. 5	-0.229	-8.42	***	0.008	***	0.081	***	-0.083	***	-0.007	***
Threshold Parameters				# obs.:	27'065						
μ1	1.049			Log-Likelihood:	-20'811.736						
μ2	3.464			LR-Test χ^2 (39 d.f.) $p > \chi^2$	10'552.93	0.000	***				
μ3	5.558			McFaddens R^2:	0.2023						

Table 6a: Ordered probit on number of rounded digits (NRD), fully interacted (model 10)

Dep. Var.: NRD	ND2	ND3	Base Effect (ND4) coeff.	z		ND5	ND6	Test on joint sig. of ND-Interactions χ^2	p	
Statement Characteristics										
Number of digits (base)	-	+++	(base	cat.)		+++	+++	108.12	0.000	***
"Estimated value" (0/1)			0.411	(6.08)	***					
"Gross value" (0/1)	+++	---	-0.014	(-0.74)		0	---	4.33	0.364	
Ln(amount)	---	---	0.344	(20.13)	***	---	---	100.27	0.000	***
Respondent Characteristics										
Age	+++	0	0.029	(9.25)	***	-	---	3.37	0.497	
Age^2	---	+++	-0.000	(-12.6)	***	+++	+++	9.27	0.055	*
Male (0/1)	+++	---	0.038	(2.02)	**	+ +	---	4.90	0.297	
Primary education	+++	---	-0.163	(-6.27)	***	+++	+++	6.83	0.145	
Secondary education	+++	0	-0.100	(-3.74)	***	+++	+++	22.53	0.000	***
1st language: French	+++	+++	-0.140	(-4.18)	***	0	+++	8.90	0.064	*
1st language: Italian	+ +	-	-0.010	(-0.17)		---	+++	2.70	0.609	
Health: very good	0	0	0.120	(4.43)	***	---	+++	2.69	0.610	
Health: good	-	0	0.093	(3.86)	***	- -	+++	1.65	0.800	
Health: not very good	+++	+++	0.031	(0.49)		0	0	1.10	0.895	
Naturalized immigrant	0	---	0.140	(1.34)	*	---	---	3.94	0.414	
Immigrant	+++	+++	-0.008	(-0.28)		+ +	+++	3.07	0.546	
log(HH's net equiv inc.)	---	-	0.069	(3.34)	***	---	+++	21.25	0.000	***
HH net eq. inc. missing	---	---	0.927	(4.17)	***	---	+++	22.05	0.000	***
Interviewer Characteristics										
Age	---	---	0.001	(0.96)		---	+++	3.81	0.432	
Male	---	0	0.056	(2.61)	***	---	---	21.78	0.000	***
Experience (years)	0	+++	-0.135	(-1.90)	**	0	+++	3.48	0.482	
$Experience^2$	0	---	0.029	(2.02)	**	0	---	2.82	0.588	
"SHP Participation?"	+++	0	-0.045	(-0.32)		-	+++	2.07	0.723	
"Response on income?"	0	0	-0.043	(-0.97)		0	---	2.29	0.683	
Respondent-Interviewer Interaction										
Same education	+++	+++	-0.018	(-0.64)		-	---	3.22	0.522	
Panel Duration										
Wave No. 2	+++	---	-0.013	(-0.46)		+ +	---	4.75	0.314	
Wave No. 3	+++	0	-0.032	(-1.15)		+++	---	3.98	0.409	
Wave No. 4	+++	0	-0.012	(-0.41)		+ +	- -	1.72	0.787	
Wave No. 5	+++	+++	-0.249	(-8.29)	***	---	---	17.54	0.002	***
5 Threshold parameters:	0.876	1.896	3.781	5.827	7.800					
Log L intercept only:	-35,090.254					Log L full model:		-25,674.223		
Likelihood ratio (195):	18,832.063					p > LR:		0.000		
McFaddens R^2:	0.268					Count R^2:		0.573		

Legend: significance levels: 10% *; 5% **; 1% ***; positive coefficient, significant at 10% + ; pos. sig. 5%: ++; negative significant at 1%: - - -, etc.

Table 6b: Marginal Effects on Mean NRD, significant ND-interactions only

Δ mean(NRD) / Δ x	ND2	ND3	Base Effect (ND4) ME	z		ND5	ND6
Statement effects:							
Number of digits (base)	0.1518	0.9495	(base	cat)		2.377**	2.853***
"Estimated value" (0/1)			0.1917	2.38	***		
"Gross value" (0/1)			-0.0123	-0.64			
Ln(value)	0.1610*	-0.0048	0.1789	8.49	***	0.1070***	-0.1807*
Respondent Characteristics							
Age	0.0209	0.0183**	0.0158	4.23	***	0.0108	-0.0025
Age^2	-0.0003	-0.002**	-0.0002	-5.54	***	-0.0001	0.0001
Male (0/1)			0.0239	1.23			
Primary education			-0.0840	-3.20	***		
Secondary education	-0.0311	-0.0205	-0.0443	-1.53	*	0.0529	0.1041
1st language: French			-0.0582	-1.73	**		
1st language: Italian			-0.0120	-0.19			
Health: very good			0.0537	1.87	**		
Health: good			0.0462	1.80	**		
Health: not very good			0.0194	0.27			
Immigrant			0.0220	0.21			
Foreigner			0.0108	0.37			
log(HH's net equiv inc.)	0.0003	0.0242	0.0358	1.47	*	-0.0546	0.1740**
HH net eq. inc. missing	-0.0297	0.2367	0.4704	2.02	**	-0.5371**	2.0193
Interviewer Characteristics							
Age			0.0001	0.08			
Male	0.0047	0.0329	0.0258	1.04		-0.0536	-0.0919
Workload			0.0000	0.06			
Experience (years)			-0.0469	-0.63			
$Experience^2$			0.0114	0.75			
1st language: French			-0.0027	-0.08			
1st language: Italian			-0.0381	-0.55			
1st language: non-Swiss			-0.0061	-0.08			
Bilingual speaker			-0.0078	-0.36			
"SHP participation?"			0.0047	0.04			
"Response on income?"			-0.0320	-0.70			
No. of interview			0.0001	0.21			
No. of $interview^2$			0.0000	.			
Working on mornings			0.0222	0.75			
Working around noon			-0.0113	-0.46			
Working at evenings			0.0117	0.33			
Respondent-Interviewer Interaction							
R-We age difference			0.0003	0.21			
Same sex			-0.0131	-0.70			
Same education			-0.0150	-0.53			
Same language			-0.0013	-0.02			
Panel Duration							
Wave No. 2			-0.0147	-0.55			
Wave No. 3			-0.0173	-0.63			
Wave No. 4			-0.0040	-0.14			
Wave No. 5	-0.0474	-0.0263	-0.1377	-4.49	***	-0.255***	-0.23***

Legend: significance levels: 10% *; 5% **; 1% ***

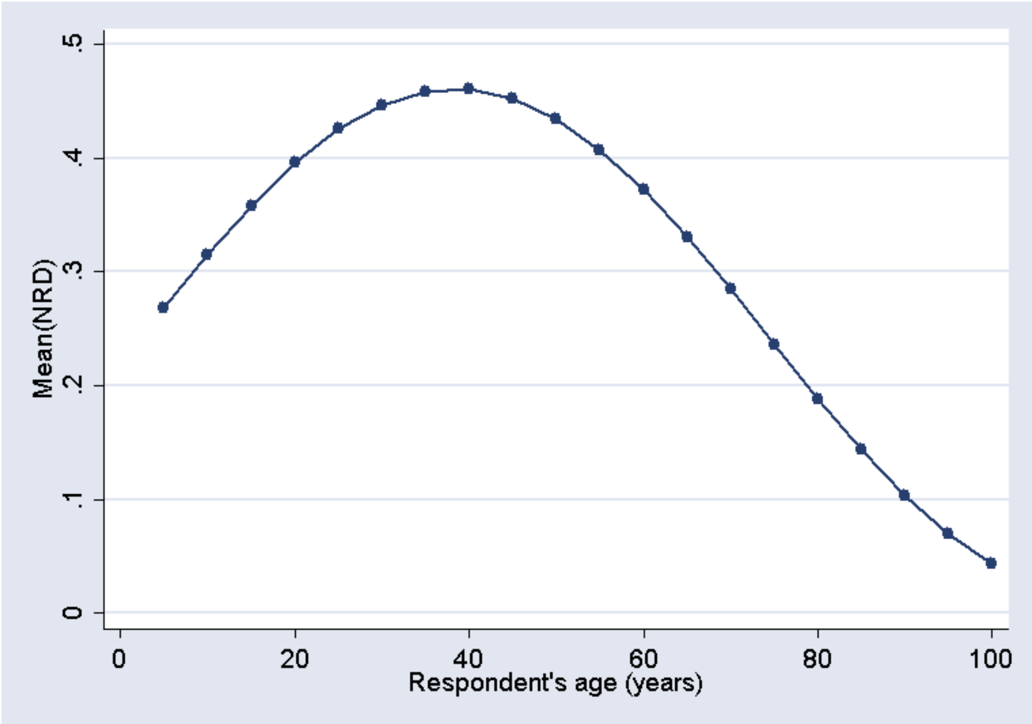
Notes for Table 5: Model 10 consists also of the explanatory variables: Age and gender of the interviewer, workload, number of interview and number of interview squared, Interviewer language (4 dummies: French, Italian, non-swiss, and bilingual speaker), working time of the interviewer (3 dummies: on mornings, around noon, at evenings), 2 missing indicators for interviewer and interviewer characteristics, and the respondent-interviewer interaction effects: age difference, same sex, and same language, which were found to be insignificant and omitted in this table.

Notes for Table 6a: Model 10 consists also of the explanatory variables: workload, number of the interview and number of interview squared, Interviewer language (4 dummies: French, Italian, non-swiss, and bilingual speaker), working time of the interviewer (3 dummies: on mornings, around noon, at evenings), 2 missing indicators for interviewer and interviewer characteristics, and the respondent-interviewer interaction effects: age difference, same sex, and same language, which were found to be insignificant and omitted in this table. Extrema of the second order polynomials: Age has a maximum at 38 years (base category of 4 digits), 41 (2 digits), 45 (3 digits), 81 (5 digits), and a minimum at 44 years concerning the results of 6 digits. Interviewers experience has a minimum at 2.3 years (base), 1.3 (2 digits), 1.6 (5 digits) and maxima at 0.2 (3 digits) and 3.4 (6 digits).

5.2.1 Respondent-, Interviewer- and Interaction-Effects

The estimation results presented in Table 4a show the significant explanatory power of a model consisting of the baseline covariates (NoD-, gross-dummy, estimated amount-dummies) and respondents' socio-demographic characteristics (model 1). The effects of age, education, health and language are jointly significantly different from zero (see tests 1a-1c and 1e in Table 4b). The rounding behavior of naturalized immigrants and immigrants does weakly differ from that of Swiss natives in the RSM model, being jointly significant at the 5%-level of significance (test 1d in Table 4b), while their separate regression coefficients are insignificant. Insignificance is also found in the NRD model, with respect to the coefficients as well as for the LR test on joint significance. As predicted, the age of the respondent has a concave effect with a maximum at 44 years on the rounding strain and 38 years for the base specification of a 4 digit number in the NRD model (see bottom of Table 6a). Since the age effect significantly differs by the number of digits, the age with the largest impact on the NRD varies between 41 and 81 years. This is unambiguous evidence for age as an influential factor on the cost-benefit situation in the RC-framework. The shape of the age impact on the mean NRD is shown in Figure 1.

Figure 1: Influence of Age on number of rounded digits

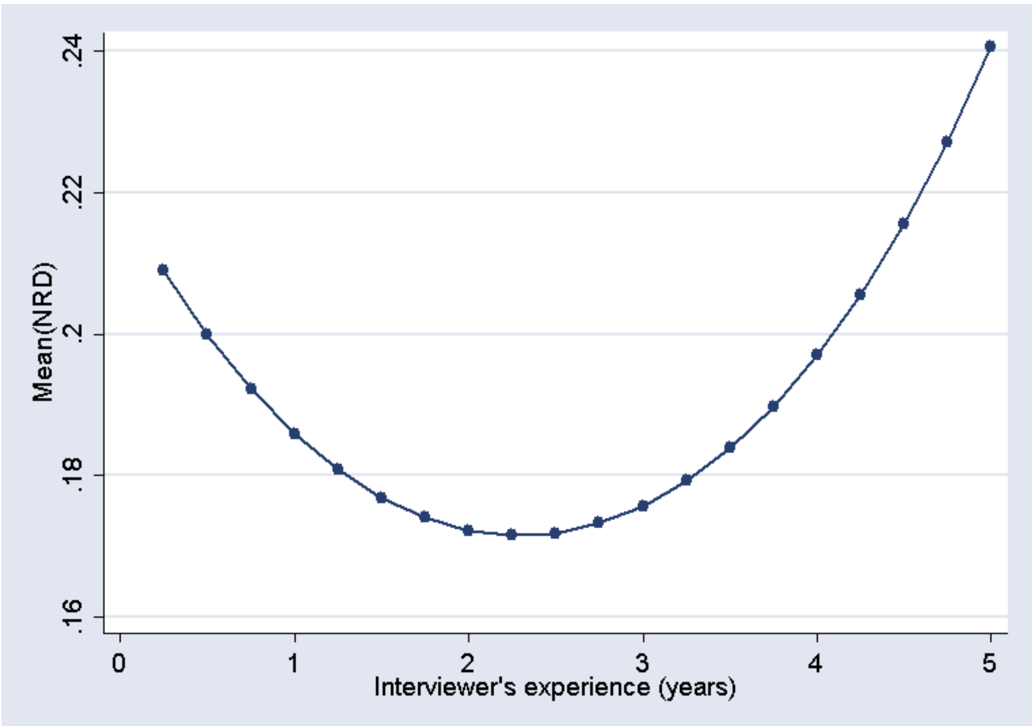


The statements of male respondents are more probable to be found in the upper RSM categories 4 and 5, as shown by the marginal effects in Table 5. Males have a weakly positive significant influence on the number of rounded digits in the base category, while respective ND-interactions are not significant: being male raises the mean number of rounded digits by 0.02, evaluated at the average of the data (see Table 6b). Our second hypothesis that higher education should reduce information costs and rounding intensity is rejected: in comparison to tertiary education, respondents with primary or secondary education have a lower rounding intensity. French speaking respondents have a lower rounding intensity than German-speaking (reference group) or Italian speaking respondents (who do not differ significantly from the German speaking group). With respect to the health status of respondents, there is some evidence in favor of our hypothesis: respondents with a good or very good health status tend to round more intensively, with an effect of 0.04-0.05 rounded digits at the mean (see Table 6b). Respondents who declare fair (reference group) or bad health status are more precise. When it comes to the interviewer effects, we find significant patterns of interviewer characteristics and interviewer fixed effects, but none for respondent-interviewer interactions (see tests 2a-3b in Table 4b). Even if the interviewer fixed effects in model 2 are jointly significant, we cannot determine a better model fit in terms of Mc Fadden's R^2, compared to models 3. There is a small quadratic interviewer experience effect, only significant at a 15%-level of significance (see test 10f). Minimum rounding occurs at 2.2 years of interviewer experience in the RSM model (2.3 years in the NRD-model), as can be seen at the bottom of tables 5 and 6, and in figure 2.

Interestingly, the interviewers who declared that they themselves would not participate in an SHP-interview or provide information about their income do not get less precise income statements from their interviewees. From the point of view of theory we would have expected that such interviewers are able to confer their motivation towards the survey upon the respondent, and that this should result in higher rounding, if item nonresponse and rounding are positively related. Still, Scherpenzeel (2002) reports that these interviewers have invoked a higher item nonresponse and a higher frequency of "don't know" statements on income questions. Thus, we are not able to observe the rounding propensity of these respondents.

Age and mother tongue of the interviewer do not impact on the rounding probability. The workload and number of the interview jointly have a weak impact on rounding intensity, different from zero at the 10%-level of significance in the NRD model, and insignificant in the RSM model. Our hypothesis that an added nonlinear experience effect with a minimum goes along with the number of conducted interviews in the field-period is not supported.

Figure 2: Marginal effect of Interviewer's Experience on number of rounded digits



Our last hypothesis, the social distance of interviewer and respondent having an influence on rounding, is rejected. The estimated coefficients have the expected sign: age difference between respondent and interviewer increases rounding intensity, same sex, education and mother tongue decreases NRD and RSM. But they are not significantly different from zero, neither separately nor jointly.

5.2.2 Number-of-Digits and Income Effects

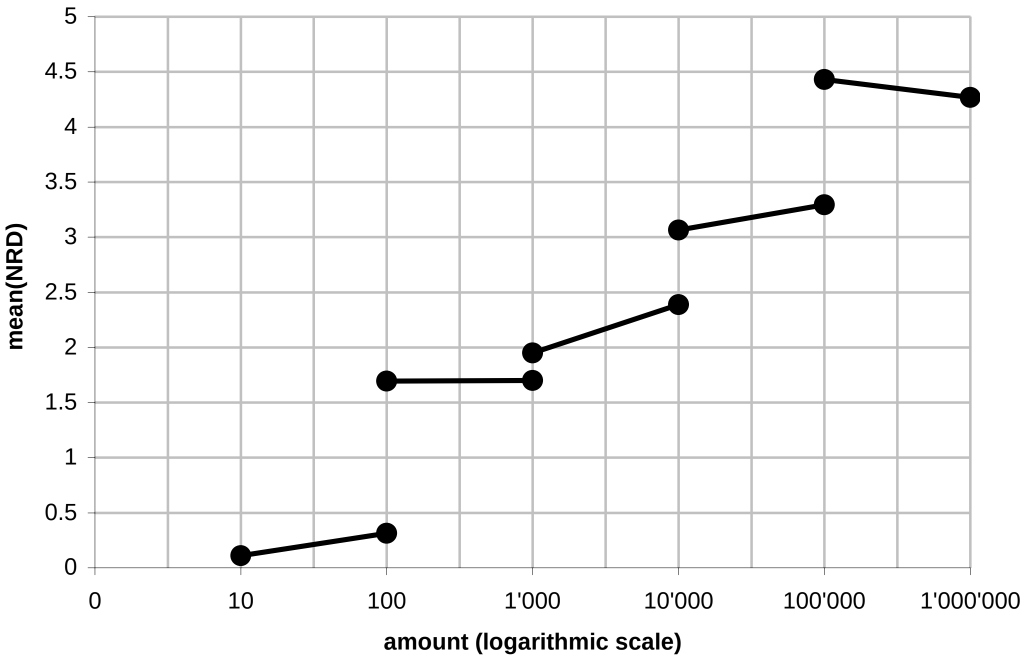
Focusing on our third research aim, we attempt to disentangle the income effect from the number-of-digits effect. Concerning the latter, we find strong significant effects in all models presented in Table 4a. This result was expected for the NRD model since the dependent variable NRD is constrained by the number of digits (see discussion in section 2.3). We therefore expect negative effects of the dummies for 2 and 3 digits and positive effects of the dummies for 5 and 6 digit-statements, with 4 digits being the reference group. Astoundingly, the ND3-dummy has a positive effect, significant at the 1%-level. Looking at the means of the rounding quotient in Table 2b, we see a possible explanation for that result: the unconditioned mean RQ for 4 digits is only 0.02 higher than for 3 digits. With respect to the rounding strain measure we also find positive, highly significant coefficients for all of the ND-dummies. This result was not expected prima vista, since RSM is a relative rounding intensity measure. In model 3, the dummy for 2 digits is negative and significant (not presented). The sign of this coefficient changes when the amount is additionally controlled for (models 4-10). These results imply that the relative rounding intensity is higher for 3, 5, and 6 digit-statements than for 4 digit statements, which is the reference category. Two digit statements have a negative impact on rounding intensity (RSM) which turns positive if the amount is controlled for in the following.

To determine the income amount effect on rounding intensity, we first augment the baseline model with logarithmic income splines, i.e. ND-dummy-interactions with the logarithmic provided amount. We find a strong positive base effect (evaluated for 4 digits) which is significantly smaller for 2, 3, 5, and 6 number of digits for the NRD model (see table 6a). These results are confirmed by the RSM model (table 5), with the exception that the log-amount interaction with 5 digits is not significantly different from zero. Table 7 and Figure 3 show the predicted income effect on the mean NRD, evaluated at the mean of the sample. It is obvious that there is an income effect in addition to the number-of-digits effect. This income effect has different slopes for each income interval. Interestingly, the slope for 6-digit numbers is negative, i.e. the precision of the statement rises when the provided amount grows from 100,000 to a million Swiss francs. The second surprise comes along with the predicted impact of 3-digit figures on the NRD: it is nearly as high as for 4-digit numbers and does not vary within the interval of 3-digit numbers (slope nearly zero). The substantial effect of the number of digits is also shown in this figure: if there were no ND-effect but a log-linear income effect, the income splines should be connected at their endpoints.

Table 7: Income splines

Amount	Log-amount	ND	y*	Mean NRD	P(NRD=m | Xb)					
					m=0	m=1	m=2	m=3	M=4	m=5
10	2.30	2	-0.406	0.111	0.900	0.089	0.011	0.000	0.000	0.000
99	4.60	2	0.244	0.313	0.736	0.214	0.049	0.000	0.000	0.000
100	4.61	3	2.371	1.695	0.068	0.250	0.603	0.079	0.000	0.000
999	6.91	3	2.378	1.699	0.067	0.248	0.605	0.080	0.000	0.000
1'000	6.91	4	2.794	1.951	0.028	0.157	0.654	0.161	0.001	0.000
9'999	9.21	4	3.587	2.387	0.003	0.042	0.531	0.411	0.013	0.000
10'000	9.21	5	4.939	3.065	0.000	0.001	0.122	0.689	0.185	0.002
99'999	11.51	5	5.407	3.294	0.000	0.000	0.052	0.611	0.329	0.008
100'000	11.51	6	7.702	4.431	0.000	0.000	0.000	0.030	0.508	0.461
999'999	13.82	6	7.365	4.270	0.000	0.000	0.000	0.062	0.606	0.332

Figure 3: Income splines



To check for the robustness of this result, we estimate a fifth-order polynomial of the income amount in model 6. In both models (NRD and RSM), this has substantial explanatory power, which is depicted by the Likelihood-ratios presented with tests 6a and 7a in Table 4b. The coefficients of the amount-polynomial have alternating signs, with a positive linear coefficient, and are separately significant at any level of significance (not presented). The number-of-digits effect also remains highly significant. This means that the income polynomial has jump discontinuities at the points 10^{ND} for any number of digits (ND). In addition to the results from the log-income splines, this finding supports our educated guess that there exists an additional number-of-digits effect besides a non-linear income effect. Admittedly, we are not able to disentangle both effects, since they are naturally connected and we do not know whether the income effect has a log-linear or 5^{th} order polynomial shape. We solely can state that the influence of income, depicted by the rounded figure, has a non-linear effect on rounding.

In addition, we test whether the wealth position of the household the respondent lives in has an influence on rounding. Therefore, we augment models 4 and 6 with a measure of the households' wealth position that is the logarithmic households' net equivalence income, according to the OECD-scale. Since the households' total income item suffers from item nonresponse in the household questionnaire of the SHP, we control for such nonresponses. The sample correlation between household's net equivalence income and the amount of the income statement is 0.44, so collinearity issues should not arise. As can be seen from the tests belonging to model 5 and 7 in table 4b, the household income has additional explanatory power for the rounding behavior, while the explanatory power of the income amount spline and polynomial slightly decreases, but remain substantial. An increase in the household's wealth by 10% would increase the number of rounded digits by 0.36 for a 4-digit number and by 1.75 digits for a 6-digit statement. The effect of item nonresponse on the households' income item is in the same direction: respondents whose head of household has not responded to the income question in the household questionnaire tend to stronger rounding of their income statement if they provide a 4-digit statement (+ 0.47 rounded digits) and to lower rounding if they provide a 5-digit statement (-0.54 rounded digits).

These results can be seen supportive for our hypothesis that rounding behavior of respondents is driven by the possibility to round off digits (number-of-digits effect) and higher opportunity costs of the interview as predicted by the rational-choice theory. This leads to higher rounding intensities with higher incomes and with a higher wealth level of the household (income effect). Besides, the provided income amount has a nonlinear impact on rounding behavior, which may also explain the highly significant ND-dummies. We find that rounding behavior is also correlated with the income-nonresponse behavior of the household.

5.2.3 Panel Duration Effects

As hypothesized above, panel duration can positively affect respondents' willingness to cooperate and thus influence the quality of the reported data. This may result from a variety of learning and experience effects of the persons involved in the survey interview (i.e. interviewer, respondent, survey administration). If data on such experiences in the several stages of the interview exists and if there is variability across these measures, it is possible to disentangle these different experience effects. With our data from the SHP we face the problem that panel duration equals respondent experience, since there has been no refreshment of the panel until 2003, and so every respondent started in the first wave 1999. The same holds for the experience of the survey organization and administration, who also started their learning process with the panel in 1999, or with a pre-test in 1998, respectively. Since no change of the interviewer occurred in the data, a measure of the number of jointly completed interviews between interviewer and respondent as a measure of the level of confidence-building would also be perfectly multicollinear with panel duration. Therefore, the possible experience effects with several causes are lumped together within our data, and we are not able to disentangle them, except for the professional experience of the interviewer, which is additionally controlled.

Concerning the resulting panel duration effect, we find no continuous negative influence of panel duration on rounding intensity, as can be seen in Tables 5 and 6a. The wave dummies for waves 2-4 have a negative sign, but are not significantly different from the reference wave 1. Only the last wave 5 has a significant negative base effect on the rounding intensity. This negative effect is stronger for 5 and 6 digit statements and lower for 2 and 3 digits, and drives the joint significance of all wave dummies (see tests 5a and 7m in Table 4b). One can now speculate whether this effect is due to selective attrition between wave 4 and 5, such that the high-rounding respondents have quitted the panel and only the highly motivated stayed. Following this hypothesis, the systematic effect should vanish if only the highly motivated respondents, who participated in all 5 waves, are taken into account. Model 9 re-estimates model 8, restricting the sample to the 16,407 respondents of all 5 waves (see table 4a). Due to the small cells problem, the ND-interactions for the respondent and interviewer characteristics had to be dropped in the NRD model, but the base effects remained in the model. A comparison of the test results 8a with 9a (table 4b) shows that the panel effect remains significant at any level of confidence. Also, the variation of the wave 5-effect remains stable over the number of digits and the rounding strain measure, as the results of the joint significance tests 5b and 6b suggest. These results are also confirmed by the RSM model.

In conclusion, the positive panel effect of wave 5 in terms of income statement precision does not arise due to self selective panel attrition. Other reasons have to be found for the respondents' willingness to be precise in the interviews of the year 2003.

6 Conclusion and Further Research Needs

Self-reported income data is mostly rounded by the respondent. This can have damaging effects on statistical inference with such data, such as variance inflation. Moreover, it may even lead to upward or downward biased estimators, particularly when rounding does not occur randomly and is correlated with the true value of the provided figure. The aim of this study was to investigate if rounding occurs at random or if it follows consistent patterns and how such patterns may be explained.

We provided evidence on rounding behavior of respondents for Swiss data for the first time and showed that the rounding intensity and rounding error are correlated with the provided value and with observational characteristics of the respondent. We also found the rounding intensity to be autocorrelated over subsequent waves. Thus, standard assumptions on misreporting errors are violated, which in turn has an impact on the validity of mean and variance of estimates. We also found weak positive correlations between rounding intensity and subsequent unit and item nonresponse. This may indicate that rounding is connected to the motivation of the respondent and may be a weak predictor for future item or interview refusal.

From a methodological point of view, we discussed the appropriateness of measures for rounding intensity. All such measures are discrete and seemingly ordered. We argue that discrete rounding measures may be spuriously ordered with respect to the underlying rounding intensity, defined by a behavioral model of respondent behavior. Since this issue was - to our knowledge - not discussed in the hitherto existing literature, we introduced two test strategies for the distinguishability and ordinality of the discrete outcomes of a variable for a given behavioral model. It is ascertained that the outcomes of the fraction of rounded digits on the total number of digits, i.e. the rounding quotient (RQ), are not ordered in the sense of a latent rounding intensity for a given set of explanatory variables. In contrast, we find that our suggested rounding strain measure, i.e. the aggregated difference between the number of rounded digits and the number of significant digits, is ordered and associated with rounding intensity.

The analysis of the occurrence of rounding with respect to socio-demographic characteristics of the respondent reveals strong rounding patterns, which are explicable in the light of cost-benefit considerations of the respondent and are mostly in line with results from previous studies. The age of the respondent has an inverse U-shaped nonlinear effect on the rounding intensity with a maximum at around 40 years. Also, the results for the health status of the respondent are in line with theoretical assumptions: people in good health have higher benefits from leisure time, i.e. the costs of precise income statements are higher in the interview, and therefore the rounding intensity is higher than that of respondents in fair or bad health. In contrast to the theory, we find the effect of education on rounding to be in the opposite direction: the higher educated respondents tend to round more intensively. Additionally, significant interviewer characteristic effects as well as interviewer fixed effects on rounding behavior are ascertained. We find a negative non-linear interviewer experience effect on rounding, which has a minimum at 2.2 years of interviewer experience. Other interviewer effects, which are found to have some explanatory power in other studies on respondent behavior (e.g., with respect to item nonresponse), are not significantly different from zero. This may be due to the fact that the used SHP data are collected using CATI (computer-assisted telephone interviews) and the possibility of influences of the interviewer may be systematically lower than in face-to-face interviews. Accordingly, the social distance of interviewer and respondent has only negligible and insignificant effects on the rounding probability.

The third large contribution of this study to the literature of respondent behavior was to disentangle the income effect from the number-of-digits effect. We first discussed that the positive effect of the provided figure on the probability of its rounding may consist of two mechanisms: First, respondents who provide higher income statements have more possibilities to round off digits. Second, the income itself has an impact on the cost-benefit situation of the respondent and positively affects its rounding propensity. Moreover, we showed that the slopes of the income effect differ by the income interval and become negative for the highest interval of 6-digit numbers. This nonlinearity of this effect hints at the coexistence of an income and a number-of-digits effect. We also identified a negative effect of the last observed panel wave 5 on the rounding propensity. This finding is inexplicable, but robust with respect to selective panel attrition.

Finally, we found no strong correlations between rounding and subsequent nonresponse. Nevertheless, the signs point in the right direction which lets us conclude that rounding, declaration of estimated amounts, "don't know" statements, item nonresponse and unit nonresponse may be positively correlated with the respondents' motivation.

However, the findings of this study are limited by two assumptions: first, a zero digit at the rear of the figure indicates rounding, but the figure may truly be a multiple of 10, 100, etc.. We assume that the occurrence of this rounding assumption error is distributed uniformly over the observations and should not constrain the generalizability of our results. Second, we assume that the rounding error is distributed uniformly within an interval around the provided figure. This strongly implies that systematic upward and downward rounding cannot occur. An evaluation of such a systematic rounding bias cannot be done with our data since information on the "true" value, e.g., by register data, is needed. Indeed, Hanisch and Rendtel (2002) provide some evidence with Finish register data that respondents tend to round their income statement upwards in the lower percentiles of the distribution, while those respondents at upper percentiles tend to downward biased rounding. This confirms prior studies of Rodgers et al. (1993) on misreporting error.

We have illustrated that rounded digits in self-reported income statements are not a stochastic event, but are correlated with the provided number itself, with the individuals that interact in the interview, and with the interview situation itself. Therefore, treating rounding errors as noise in the data may have damaging effects on statistical inference. The researcher may prefer to use the estimation methods developed for interval data (see e.g. Manski, Tamer 2002) rather than standard estimation methods. Evaluation studies on the impact of roundings in income data on the statistical inference would give the researcher guidance in choosing his methods.

From the view of the survey research literature another question may deserve attention: does rounding behavior reflect the motivation of the respondent, and how can motivation be increased to obtain more precise data and to avoid subsequent nonresponse?

Bibliography

Anderson, J. A., 1984, Regression and Ordered Categorical Variables, *Journal of the Royal Statistical Society - Series B*, 46(1), pp.1-30.

Biemer, Paul P. et al., 1991, Measurement Errors in Surveys, New York: Wiley.

Brant, R., 1990, Assessing proportionality in the proportional odds model for ordinal logistic regression, *Biometrics*, 46, pp. 1171-1178.

Bound, J. and A. B. Krueger, 1991, The Extent of Measurement Error in Longitudinal Earnings Data: Do two Wrongs make a Right?, Journal of Labor Economics, 9, pp. 1-24.

Budowski, Monica et al., 2001, The Swiss Household Panel 1999-2003: Data for research on micro social change, ZUMA-Nachrichten, Vol. 50 (49), pp. 100-125.

Budowski, Monica, Alexis Gabadinho and Robin Tillmann, 2002, Construction of Income Variables for Wave 1 (Survey 1999), *Working Paper No. 2/02*, Swiss Houeshold Panel, Neuchâtel.
URL: http://www.swisspanel.ch/file/working_papers/WP2_02.pdf

Burton, Jonathan, Heather Laurie, and Nick Moon, 1999, "Don't ask me nothin' about nothin', I just might tell you the truth" - The Interaction between unit nonresponse and item nonresponse, *Paper presented on the International Conference on Survey Nonresponse*, Portland, Oregon.

Dempster, Arthur P. and Donald B. Rubin, 1983, Rounding Error in Regression: The Appropriateness of Sheppard's Correction, *Journal of the Royal Statistical Society - Series B*, 45(1), pp.51-59.

Dohrenwend, Barbara Snell et al., 1968, Social Distance and Interviewer Effects, Public Opinion Quarterly, 32, pp. 410-422.

Duncan, G.J. and D.H. Hill, 1985, An Investigation of the Extent and Consequences of Measurement Error in Labor Economic Survey Data, Journal of Labor Economics, 3, pp. 508-522.

Eisenhart, Churchill, 1947, Effects of Rounding or Grouping Data, in: Eisenhart, C., M. Hostay and W. Wallis: Selected Techniques of Statistical Analysis, Chapter 4, McGraw-Hill, New York.

Esser, Hartmut, 1984, Determinanten des Interviewer- und Befragtenverhaltens: Probleme der theoretischen Erklärung und empirischen Untersuchung von Interviewereffekten (Determinants of interviewer- and respondent behavior:

Problems of theoretical explanations and empirical analyses of interviewer effects), in: K. Mayer, P. Schmidt (eds.) *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*, pp. 26-71, Frankfurt /M.

Gabadinho, Alexis and Monica Budowski, 2002, Construction of Income Variables for Wave 2 (Survey 2000), *Working Paper No. 3/02*, Swiss Household Panel, Neuchâtel. URL: http://www.swisspanel.ch/file/working_papers/WP3_02.pdf

Hanisch, Jens U., 2004, Rounded Responses to income questions, *Statistisches Archiv – Journal f the German Statistical Association*, forthcoming.

Hanisch, Jens U. and Ulrich Rendtel, 2002, Quality of Income Data from Panel Surveys with Respect to Rounding, *Chintex working paper* No. 6, URL: http://www.destatis.de/chintex/download/paper6.pdf

Heitjan, Daniel F. and Donald B. Rubin, 1990, Inference form Coarse Data via Multiple Imputation with application to age Heaping, *Journal of the American Statistical Association*, Vol. 85(410), pp. 304-314.

Henderson, Brent and Richard Jarrett, 2003, Models with errors due to misreported Measurements, *Australian & New Zealand Journal of Statistics 45(4)*, pp. 431-444.

Judge, George G. et al., 1985, The theory and practice of econometrics, (2nd ed.), New York: Wiley.

Kendall, M.G., 1938, The Conditions under which Sheppard's Corrections are Valid, *Journal of the Royal Statistical Society*, Vol. 101, pp. 592-605.

Kroh, Martin, 2004, Intervieweffekte bei der Erhebung der Körpergewichts - Die Qualität von umfragebasierten Gewichtsangaben, *DIW discussion paper No. 439*, German Institute for Economic Research (DIW),
URL: http://www.diw.de/deutsch/produkte/publikationen/diskussionspapiere/docs/papers/dp439.pdf

Krosnick, Jon, Narayan, S., & Smith, W., 1996, Satisficing in Surveys: Initial Evidence. *Advances in Survey Research: New Directions in Evaluation*, 70, pp. 29-44.

Long, J. Scott, 1997, Regression Models for Categorical and Limited Dependent Variables, SAGE Publications, Thousand Oaks (CA).

Long, J. Scott and Freese, Jeremy, 2003, Regression Models for Categorical and Limited Dependent Variables using STATA - Revised Edition, Stata Press, College Station (TX).

Lunt, Mark, 2001, Stereotype Ordinal Regression, STATA Technical Bulletin No. 61, Stata-Press, College Station (TX).

Manski, Charles F. and Elie Tamer, 2002, Inference on Regressions with interval data on a regressor or outcome, *Econometrica*, Vol. 70(2), pp. 519-546.

McCullagh, Peter, 1980, Regression Models for Ordinal Data, *Journal of the Royal Statistical Society - Series B*, 42(2), pp. 109-142.

Nicoletti, Cheti and Nicholas H. Buck, 2004, Explaining Interviewee Contact and Co-operation in the British and German Household Panels, in: Ehling, Manfred and Ulrich Rendtel (eds.): Harmonisation of Panel Surveys and Data Quality, Statistisches Bundesamt, Wiesbaden.

Polasek, Wolfgang, 1987, Bounds on Rounding Errors in Linear Regression Models, *The Statistician 36(2/3)*, pp. 221-227.

Raftery, A. E., 1996, Bayesian Model selection in social research, in: P.V. Mardsen (Ed.): Sociological methodology, Vol. 25, pp.111-163.

Rendtel, Ulrich, 1988, Panelmortalität - Eine Analyse der Antwortausfälle beim Sozio-ökonomischen Panel in der 2. und 3. Befragungswelle (Panel mortality - an analysis of response drop-outs at the German Socio-Economic Panel in the second and third wave), *Vierteljahresheft zur Wirtschaftsforschung*, Heft 1/2, 37-59.

Rendtel, Ulrich, 1989, Vertrauensbildung und Teilnahmeentscheidung in Panelstudien – Über die Entwicklung der Antwortbereitschaft im Sozio-Ökonomischen Panel (trust building and decision to participate in panel studies), *Arbeitspapier Nr. 313*, Sonderforschungsbereich 3: Mikroanalytische Grundlagen der Gesellschaftspolitik der Universitäten Frankfurt und Mannheim.

Rice, Stuart A., 1929, Contagious Bias in the Interview - A Methodological Note, American Journal of Sociology, pp. 420-423.

Rietveld, Piet, 2001, Rounding of Arrival and Departure Times in Travel Surveys – an interpretation in terms of scheduled activities, discussion paper, URL: http://www.tinbergen.nl/discussionpapers/0110.pdf

Riphahn, Regina T. and Oliver Serfling, 2005, Item Nonresponse on Income and Wealth Questions, *Empirical Economics*, forthcoming.

Rodgers, Willard L., Charles Brown and Greg J. Duncan, 1993, Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages, *Journal of the American Statistical Association*, Vol. 88, No. 424.

Rosch, Eleanor, 1975, Cognitive Reference Points, *Cognitive Psychology*, Vol. 7(4), pp. 532-547.

Rowe, Geoff and Steve Gribble, 1994, Income Statistics From Survey Data: Effects of Respondent Rounding, Proceedings *of the American Statistical Association, Section on Business and Economic Statistics*, pp. 77-82 .

Schräpler, Jörg-Peter, 1999, Thesis paper, Ruhr-University Bochum.

Scherpenzeel, Annette, B. Wernli and P. Eichenberger, 1999, An experimental study of data collection strategies: report of the first stage, *Working Paper 3/99*, Swiss Household Panel, Neuchâtel, Switzerland.
URL: http://www.swisspanel.ch/file/working_papers/WP3_99.pdf

Scherpenzeel, Annette, 2002, Interviewer Effects in data from the Swiss Household Panel, *Working Paper 8/02*, Swiss Household Panel, Neuchâtel, Switzerland.
URL: http://www.swisspanel.ch/file/working_papers/WP8_02.pdf

Schweitzer, Mark E. and Eric K. Severance-Lossin, 1996, Roundings in Earnings Data, *working paper 96/12*, Federal Reserve Bank of Cleveland.

Serfling, Oliver, 2005, The Interaction between Item, Questionnaire and Unit Nonresponse in the German SOEP, *Schmollers Jahrbuch - Journal of Applied Social Science Studies* , 125. Jg. Vol. (1) 2005, Berlin.

Sheppard, W.F., 1898, On the Calculation of the most Probable Values of Frequency-Constants, for Data arranged according to equidistant Divisions of a Scale, Proceedings of the London Mathematical Society, Vol. 29, pp. 353-380.

SHP - Swiss Household Panel, 2005, Household Questionnaire, Wave 6, Swiss Household Panel, Neuchâtel, Switzerland.
URL: http://www.swisspanel.ch/file/doc/q_pdf/QuestionML-H-W6.pdf

Student, 1908, The Probable Error of a Mean, Biometrika, Vol. 6 (1), pp. 1-25.

Sudman, Seymour, Norman M. Bradburn und Norbert Schwarz, 1996, Thinking about Answers – The Application of Cognitive Processes to Survey Methodology, San Francisco: Jossey-Bass Publishers.

Tourangeau, R., 1984, Cognitive Sciences and Survey Methods, in: T. Jabine et al.: Cognitive Aspects of Survey Methodology - Building a Bridge between Disciplines, Waschington D.C., pp. 73-100.

Tricker, A.R., 1984, Effects of Rounding on the moments of a probability distribution, *The Statistician*, Vol. 33(4), pp. 381-390.

Wertheimer, Max, 1912, Über das Denken der Naturvölker, Zahlen und Zahlengebilde, *Zeitschrift für Psychologie*, 60, pp.321-378; reprinted as "Numbers and Numerical Concepts in Primitive Peoples", in: Ellis, W.D. (1938): A source book of Gestalt psychology, New York: Harcourt, Brace & Co.

VI. Conclusion

The four essays of this dissertation provide a number of new and unique insights on diverse issues of respondent behavior and aspects of data quality in surveys.

The first paper examines the determinants of item nonresponse on several questions of households' wealth, households' income, and respondents' income. It firstly provides empirical evidence that the mechanisms behind item nonresponses and "don't know" statements differ. The item nonresponse intensity is found to be item specific. The interactions between the respondent and the interviewer and the interview situation are evaluated and it is found that the gender of both, respondent and interviewer, and the age difference between interviewer and respondent have an influence on the occurrence of nonresponse.

The second and third paper show that the correlation of item nonresponse with subsequent unit nonresponse is not necessarily positive and linear. The analysis shows a negative correlation of item nonresponse with the newly introduced category of (wealth-) questionnaire nonresponse. With respect to subsequent unit nonresponse it is shown that the correlation pattern with the INR rate is nonlinear. It obeys an inverse-U-shaped pattern, which is explained by simultaneous drop-out of two types of respondents: those with low INR propensity and those with high INR propensity.

Finally, the fourth study examines the quality of income data provided by the respondents with respect to rounding. It finds that rounding does not occur at random, but is explicable by cost/benefit considerations of the respondent. The magnitude of rounding is also correlated with the income figure and autocorrelated. This provides evidence that the rounding error is likely to harm estimates of empirical studies with rounded data. From a methodological point of view, this study contributes to the check of ordinality of discrete outcomes of a variable and the adequacy of ordered regression models.

All four papers of this dissertation contribute to the understanding of the social interaction processes which occur during a survey interview. The benefit of the insights provided in this dissertation is threefold: first, the findings enable survey institutions to advance the data collection process, in order to reduce data deficiencies and increase the informational value of the survey: We have shown that pairing interviewers based on gender and age may reduce income INR, Face-to-face interviews are beneficial for reducing INR and UNR, and experienced interviewers improve the quality of the collected data.

Second, the results could support the sophistication of imputation procedures for missing or misreported data. Since we have shown that the mechanisms behind item nonresponses and "don't know" statements originate from different response processes, the origin of missing statements should be considered by imputation methods.

Third, the methods employed in the studies may improve researchers' ability to rigorously deal with misreports in his or her own empirical analyses, e.g. to use selection models with pre-interview data as instruments.

Nonetheless, further research is needed to derive more concrete advise on how to design a survey study to increase the quality of the data collected. Since the approach underlying the studies herein is of empirical nature, the results of these studies are restricted to observable and surveyed characteristics of respondents, interviewers and the interview situation. It is likely that this reflects only part of the story, since a lot of possible determinants may be unobservable, non-measurable or not surveyed. This opens avenues of qualitative research in the disciplines of e.g. sociology and psychology.

Curriculum Vitae

Oliver Serfling wurde am 27.5.1975 in Mannheim-Neckarau als zweiter Sohn von Versicherungskauffrau Christa Serfling (geb. Bös) und Fernmeldeamtsrat Hans-Werner Serfling geboren. Am 30.06.1994 erlangte er die allgemeine Hochschulreife (Abitur) an der Otto-Hahn-Schule in Hanau. Von 1995 bis 2001 studierte Oliver Serfling Volkswirtschaftslehre mit den Vertiefungsrichtungen Finanzwissenschaft, Konjunktur, Wachstum und Verteilung, sowie Sozialpolitik an der Johann-Wolfgang Goethe Universität in Frankfurt am Main. Während seines Studiums arbeitete er u.a. an einer Studie zum ersten Armuts- und Reichtumsbericht der Bundesregierung mit und sammelte Erfahrungen in der Telekommunikations- und Finanzanalystenbranche sowie in der Politikberatung.

Im Frühjahr 2006 schloss er sein Studium mit Prädikat ab und graduierte zum Diplom-Volkswirt. Anschliessend begann er seine Promotion unter der Leitung von Prof. Regina T. Riphahn, Ph.D. an der Johannes-Gutenberg Universität in Mainz. Ab 1.1.2002 setzte er sein Promotionsstudium am Wirtschaftswissenschaftlichen Zentrum der Universität Basel fort, wo er im August 2005 die vorliegende Dissertation fertig stellte. Zum Zeitpunkt der Drucklegung des Werkes war Oliver Serfling Parlamentarischer Referent beim Hessischen Landtag in Wiesbaden.