SONDERFORSCHUNGSBEREICH 504
Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 02-35

Bracketing effects in categorized survey questions and the measurement of economic quantities

Winter, Joachim*

Financial support was provided by the European Union (TMR project

*Mannheim Research Institute for the Economics of Aging (MEA), email: winter@uni-mannheim.de

Universität Mannheim
L 13,15
68131 Mannheim

Bracketing effects in categorized survey questions and the measurement of economic quantities*

Joachim K. Winter**
University of Mannheim

First draft: August 2001
This version: May 2002

Abstract: In households surveys, quantities of interest are frequently elicited using categorized (range-card) formats rather than open-ended questions. One advantage of this format is that it typically reduces item non-response. Unfortunately, results from research in social psychology suggest that the choice of bracket values in range-card questions is likely to influence responses. Not much is known about the influence of bracketing effects on the measurement of economic quantities and regression analysis. Based on data from controlled survey experiments, this paper shows that bracketing effects exist in interval data obtained from household surveys and that they are of significant size. I also discuss strategies for limiting bracketing effects in range-card questions and for developing valid econometric models for biased interval data.

Keywords: survey methodology, bracketing, measurement error, interval data

JEL classification: C81

---

* I am grateful to Rob Alessie, Marcel Das, Michael Hurd, Guido Imbens, Arie Kapteyn, Daniel McFadden, Paul Ruud, Matt Schonlau, Norbert Schwarz, Arthur van Soest and to seminar participants at UC Berkeley, RAND (Santa Monica), and Ente Luigi Einaudi (Rome) for helpful comments and discussions. Household expenditure data from the Dutch Budget Survey 1998 have been made available through the Netherlands Scientific Statistical Agency. Financial support was provided by the European Union (TMR project "Savings, pensions, and portfolio choice") and by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 504 at the University of Mannheim). I appreciate the hospitality of the CentER for Economic Research at Tilburg University where much of this research was conducted.

** *Address*: Mannheim Research Institute for the Economics of Aging (MEA)
University of Mannheim
D-68131 Mannheim, Germany
E-mail: winter@uni-mannheim.de

1 Introduction

In household surveys, quantities such as income, consumption or wealth, or components of these quantities, are frequently elicited not by using open-end question formats, but by giving respondents categorized reponse options such as range cards. One reason for using range cards instead of open-ended formats is that response rates are typically higher. Since range cards are difficult to implement in telephone interviews, an alternative format has been developed, so-called "unfolding brackets" questions that require only a yes-no answer at each of several stages. Unfolding brackets questions can be designed such that the same information as in a range-card question is obtained (provided that the entire sequence of questions is completed). Like range-card questions, they yield interval data on continuous quantities.

Range cards and unfolding brackets are also used in "follow-up" or "default" questions. In this case, survey respondents are first asked to report some quantity using an open-ended question format. Only those respondents who give "don't know" answers are presented with a follow-up bracketed question format such as a range card. Even though responses to range-card or unfolding brackets questions generate only interval data and are therefore less informative than the continuous response to an open-ended question, there is still some information revealed about initial non-respondents that can be used in the statistical and econometric analysis of survey data. It has been argued that follow-up brackets can reduce the problem of non-response and improve measurement of economic quantities considerably (e. g., Juster and Smith (1997)).

Mainly because of their good properties with respect to non-response, range-card and unfolding brackets design are now used quite frequently in household surveys, and researchers are confronted with interval data on continuous quantities. A unified framework for the parametric and nonparametric analysis of interval data has been introduced by Vazquez Alvarez *et al.* (1999) and Manski and Tamer (2002).[1]

Unfortunately, results from research in social psychology suggest that the choice of starting values in unfolding brackets sequences and of bracket values in range-card questions are likely to influence responses. The former phenomenon is known as anchoring bias (even though it is also related to acquiesence bias), the latter is often referred to as bracketing effects. In the context of survey questions on economic quantities, bias in the responses to unfolding brackets questions induced by the cognitive processes governing survey response behavior has received some attention; see Hurd *et al.* (1998) and Hurd (1999). However, response bias that may arise in range-card questions due to bracketing effects has been neglected so far in the economics

[1] There is of course an older literature on parametric approaches for interval data, as discussed by Hsiao (1983). Parametric approaches are routinely used in applied work, and they are also covered in standard textbooks, such as Verbeek (2000).

literature. The present paper closes this gap. Based on data from two controlled survey experiments, I characterize response bias in interval data obtained from range-card questions on economics quantities. Furthermore, I illustrate the effects of such bias on econometric analysis, and I discuss implications for survey design and applied research using interval data. Both survey experiments were conducted within the Dutch CentERpanel in February and June 2001. Appendix 6.3 provides a detailed description of the CentERpanel.

The practical implications of the results presented in this paper are strong. Unless it can be ruled out that response bias is present in interval data, none of the nonparametric and semi-parametric estimators for interval data discussed by Manski and Tamer (2002) and, *a fortiori*, no traditional parametric regression model provide consistent point estimates or bounds or identification regions for the parameters of interest in regressions that involve interval data (either as dependent or explanatory variables). I discuss how the absence of bracketing effects can be tested in practice in section 6. Furthermore, I dicuss how the framework of Manski and Tamer can be extended to the case of interval data that are subject to response bias, which might open the possibility to construct both nonparametric and parametric estimators for biased interval data.

The remainder of this paper is structured as follows. In section 2, I review the econometric literature on interval data and the psychology literature on bracketing effects in range-card questions. The two survey experiments that document the existence of bracketing effects in economic variables elicited in household surveys are presented in sections 3 and 4. Effects of response bias on subsequent empirical research that uses interval data are illustrated in section 5. The results of this paper are summarized in section 6, together with recommendations for survey practice and directions for future research.

2 Prior literature on interval data and bracketing effects

In this section, I first present an abstract framework for the econometric analysis of interval data (section 2.1). It will become apparent immediately that identification and consistency results fail in the presence of response bias. In section 2.2, I discuss the psychological literature on the cognitive processes that induce response bias in range-card questions.

2.1 Econometric analysis of interval data

To structure the subsequent discussion, it is helpful to have an abstract framework for the econometric analysis of interval data at hand. I therefore present the structure of the Manski and Tamer (2002) model briefly. I focus on the assumptions about the data generating process and their role in obtaining identification results for parametric and nonparametric regressions.

Consider a data generation process in which some continuous variable of interest, v, is either fully observed or measured on a interval scale. Let a population be characterized by a probability distribution $P(y, x, v, v_l, v_u)$, with $y \in I\!R$, $x \in I\!R^k$, and $v, v_l, v_u \in I\!R$. Let a random sample be drawn from P. The variables of interest are y , x, and v. While y and x are observed directly, v is unobserved. However, two interval bounds $v_l < v_u$ for v are observed. Note that this framework also allows for the case in which $v_l = -\infty$ and $v_u = \infty$ which corresponds to a selection model, as discussed by Horowitz and Manski (1998) and Horowitz and Manski (2000).

There are two abstract regression problems that contain an interval-measured variable v. The first problem is concerned with estimating $E(y|x, v)$ and has v on the right-hand side. The second is concerned with estimating $E(v|x)$, i. e., v is the dependent variable.

In order to make inferences about $E(y|x, v)$, a researcher can impose parametric assumptions and use a maximum likelihood approach for estimation, as in Hsiao (1983). Such assumptions would imply that $P(y, v|x, v_l, v_u)$ is contained in a finite-dimensional family of distributions. A more robust alternative is the nonparametric approach proposed by Horowitz and Manski (1998), Vazquez Alvarez *et al.* (1999), Horowitz and Manski (2000) and Manski and Tamer (2002). The latter framework is the most general and also covers the case of estimating $E(v|x)$.

In order to derive general identification results for interval data, Manski and Tamer (2002) formulate the following set of assumptions.

> **Assumption I (correct interval reporting):** $P(v_l \leq v \leq v_u) = 1$. This assumption implies that the unobserved true values of v are contained in the observed intervals $[v_l, v_u]$.
>
> **Assumption M (monotonicity):** $E(y|x, v)$ exists and is weakly increasing in v. This is a weak assumption on the shape of $E(y|x, v)$ which is typically imposed in order to obtain non-parametric estimators of regressions.
>
> **Assumption MI (mean independence):** $E(y|x, v, v_l, v_u) = E(y|x, v)$. This assumption implies that observation of $[v_l, v_u]$ would be superfluous for predicting y if v were observed directly. Manski and Tamer (2002) point out that while this assumption might seem innocuous, it is not warranted when observations on v are missing non-randomly, i. e., in the case of endogenous censoring.

Based on this set of assumptions, Manski and Tamer (2002) derive a series of identification results which I summarize only very briefly. In the absence of any other information, assumption I alone implies nonparametric bounds on $E(v|x)$, and assumptions I, M, and MI together imply nonparametric bounds on $E(y|x, v)$. When y is binary, identification results for $E(y|x, v)$ can be obtained in the semiparametric regression model of Manski (1975, 1985). With interval-measured data on v, identification regions can be obtained using a modified version of the maximum score estimator. Similarly, identification regions for the parameters of interest exist in parametric models of $E(y|x, v)$ and $E(v|x)$, and these can be obtained using modified minimum distance or modified method of moments estimators.

Results from survey research by psychologists, as reviewed in detail in section 2.2, suggest that assumption I need not hold in practice – that is, respondents do not always report correctly the bracket in which the true value of the quantity of interest falls. Such a phenomenon will be called "response bias" in the sequel. Psychologists call this phenomenon "bracketing effects" since the source of the bias is related to the location of the bracket bounds (as will become clear below). Formally, this type of response bias implies that $P(v_l \leq v \leq v_u) < 1$, i. e., that assumption I is violated. Consequently, all identification results obtained by Manski and Tamer – and results from traditional regression models that can be obtained as specializations of their framework – do not hold.[2]

2.2 Psychological research on survey response behavior and bracketing effects

Why should bracket bounds in range-card questions influence responses? If respondents are perfectly certain about the quantity in question, they should be able to indicate the correct bracket. However, respondents are rarely certain about quantities they are asked to report in household surveys. Therefore, the formation of answers to survey questions is a complicated process. As a starting point for thinking about ways to avoid bracketing effects *ex ante*, or to correct for resulting bias in survey data *ex post*, it is useful to review existing research by psychologists and survey methodologists in some detail.[3]

An important insight from survey research is that the process of forming the response to a survey question consists of several steps, each of which might contribute to the fact that answers often do not provide reliable measures of the quantity in question. Survey respondents first have to understand the question and determine which quantity they are to report on. To do so, they draw on a wide range of contextual information in ways that researchers are often unaware of. Second, respondents have to recall information on the quantity from memory. In many instances, respondents will have imperfect recall and need to apply various inference and estimation strategies to arrive at an answer; this is the third step of the response process. Fourth, once respondents have arrived at an answer, they need to map it onto the

[2] An obvious strategy to rescue identification results and to construct estimators for regressions with biased interval data would be to exploit knowledge of $P(v_l \leq v \leq v_u)$, the probability of misreporting. I return to this possibility in section 6.

[3] Comprehensive reviews of this literature can be found in Sudman *et al.* (1996), Tourangeau *et al.* (2000), and Schwarz and Oyserman (2001). The following discussion is based on the latter.

response alternatives provided by the researcher (unless the question format is open-ended). Finally, respondents may edit their answer because of social desirability and self-representation concerns (i. e., even though they might be aware of the "true" value of some quantity, they on purpose or unconsciously report a higher or lower value).

It is well established in survey research that at different stages of this response process, respondents extract information from the questionnaire and use this information to construct their response. Therefore, survey responses typically reflect not only the respondent's (possibly imperfect) knowledge about quantities that a researcher is interested in, but also information contained in the questionnaire.[4] Framing effects – which are related to the first step of the response process – are a well-known example: Depending on how a question is framed, respondents might arrive at different conclusions about what quantity they are to report.

For the response biases considered in this paper, the third step of the response process is the critical one: At the inference and estimation stage, respondents often pick up numerical clues contained in the survey questionnaire and use this information in forming their responses. For instance, in unfolding brackets or double referendum questions that consist of two or more yes-no questions, responses to the second and later questions are biased towards the first "bid" presented. In this case, the cognitive processes are well understood. Respondents apply the anchoring heuristic (Tversky and Kahneman, 1974) to determine an answer about which they are uncertain. The first bid – a numerical clue contained in the survey – serves as an "anchor", and respondents conclude that their answer should be closer to that value than they would have said if they had not seen the anchor.[5] In line with the multi-step model of survey response behavior in which uncertainty plays a key role, anchoring effects have been shown to be less severe or even disappear when subjects are certain about a quantity in question; see Mussweiler and Strack (2000).

In the case of range-card questions, numerical clues are provided by the bracket bounds. The cognitive processes that might result in response bias are summarized by Sudman *et al.* (1996), p. 218: "Every contribution to the survey interview, including formal features of the questionnaire, is relevant to their task [...] Respondents assume that the range of response alternatives reflects the researcher's knowledge of or expectations about the distribution of behavior in the 'real world'. Thus, response alternatives in the middle range of the scale would reflect the 'average' or 'typical' behavioral frequency whereas the extremes of the scale would

[4] As noted above, variations in survey response that can be attributed to features of the survey design are called "response bias" in the present paper.

[5] This is a very robust result – even if subjects are explicitly made aware of the fact that an anchor does not contain any information relevant to their task, responses are still biased towards that anchor; see Jacowitz and Kahneman (1995) for a review of the psychlogical literature on anchoring. For detailed discussions of anchoring bias in unfolding brackets questions and similar designs such as double referendum formats in contingent valuation studies, see Green *et al.* (1998), Hurd *et al.* (1998), and Hurd (1999).

correspond to the extremes of the distribution." For example, the location of bounds in a bracketed response format might by taken to reveal some features of the distribution of the underlying quantity in the population, extreme (bottom and top) brackets being associated with "extreme" types of the population (heavy drinkers in the case of alcohol consumption, very rich households in the case of income or wealth).

The first experiment that demonstrated the effects of bracket bounds on survey responses was conducted by Schwarz *et al*. (1985). They asked a sample of adults to report how many hours a day they spent watching television. Half the sample received a scale ranging from 'up to a half hour' to 'more than two and a half hours', the other half received a scale ranging from 'up to two and a half hours' to 'more than four and a half hours'. The range of response alternatives contained in the survey design had a significant impact on the reports. These "bracketing effects" have been replicated in numerous studies, including the experiments reported in section 3.

Bracketing effects not only have the potential to bias responses to some bracketed question itself; they can also have spill-over effects to subsequent questions. In an experimental study, Menon *et al*. (1997) first asked subjects about past expenses on some consumer product (e. g., shampoo) or some social activity (e. g., going to the movies), using range-card questions with different bracket bounds. They then asked the same subjects for estimates on future expenses for the same products or activities, using an open-ended format. Responses to the bracketed retrospective question revealed the expected bracketing effects. More surprisingly, the location of the bracket bounds in the first question on past expenditure had a significant influence on subjects' estimates of future expenditure on the same consumption item in the second question. Menon *et al.* argue that response alternatives provided in a survey question carry information about the population distribution that is used by the respondent not only to answer that question (this finding is in line with results from earlier research), but also to formulate responses to subsequent, related questions.

From the perspective of survey design, the fact that numerical clues such as bracket bounds can influence response behavior is an important insight. A constructive perspective is opened up by identifying conditions under which bracketing effects can be reduced or even disappear. As in the case of anchoring bias, the role of uncertainty is crucial: When respondents are uncertain, they use heuristics such as the anchoring or infer the population distribution from bracket locations in range card questions when they formulate their answers. In contrast, respondents who understand a question fully, who have ample time to think about their answer, and who feel certain about the quantity in question are less likely to use such estimation strategies. Consequently, they are less likely to be influenced by the location of bracket bounds in range-card questions; see Menon *et al*. (1995).

In addition to these cognitive factors, motivational factors and constraints on cognitive resources such as time pressure also determine whether respondents use information contained in the questionnaire when they produce their responses. If respondents are highly motivated and have ample time, they are more likely to use a more elaborate response process and thus less likely to be subject to anchoring bias or bracketing effects, as demonstrated in the case of range-card questions by Stocké (2001). These findings on the role of motivational factors are in line with work by Philipson (1997), Philipson and Malani (1999), and Philipson (2001) who stress – from a purely economic perspective – that survey respondents cannot be trusted to produce reliable response unless they are sufficiently motivated. Philipson (2001) presents experimental evidence that data quality is improved considerably by using incentive payments or telling survey respondents that their answers might be checked for accuracy.

Studies in psychology and survey research have traditionally focused on frequency reports. The question on the number of hours spent watching TV, mentioned above, is one example. Other applications include asking for the frequency with which patients have experienced certain symptoms in a given period, or asking for the number of sexual partners people have had. In all these cases, response scales contained in the survey questions have been found to influence frequency reports significantly. Surprisingly little is known, however, about the relevance of bracketing effects on the measurement of economic quantities. To the best of my knowledge, there exist no experimental studies other than Menon *et al.* (1997) that investigate the effect of bracket bounds in range-card questions on quantities such as income, saving, wealth, or consumption expenditure. The survey experiments presented in the following sections fill this gap.

3 Experiment 1

The first experiment reported in this paper was conducted using the Dutch CentERpanel in February 2001. It uses questions on six different quantities, three on the number of hours spent for certain activities (similar to earlier experiments in the psychology literature), and three on household expenditure categories.

3.1 Design

The experiment is based on earlier work by Schwarz *et al.* (1985) who asked a sample of adults to report how many hours a day they spent watching television, using range-card question formats with low and high scales. They found that these ranges had a significant impact on the reports.

In the present experiment, I used three questions on time use. Specifically, I asked for the number of hours spent watching TV last week, reading books last month, and filling in tax forms last year (Questions 1 through 3, respectively). In addition, I used three questions on quantities that are of interest in applied research in economics, namely expenditure on food during the last month, and expenditure on travel and clothing, both during the last year (Questions 4 through 6). The exact wordings of these questions are summarized in table 1.

Before answering each of these questions, subjects were asked whether they had spent time on the activity or spend money on the expenditure category, to avoid a dominating influence of zero responses on results. For those subjects who reported that they had used time or spend money on the respective item, a range-card question was presented. All questions used a forced design; i. e., respondents were not given a "Don't know" option.

Each of the six questions used three experimental treatments with different levels of the bracket values; these six treatments were assigned randomly and independently across questions. The values of the bracket points are listed in table 2. The order of questions was the same for all subjects.

The number of observations for each question and treatment is reported in table 2. The numbers are smaller for the expenditure questions since I used additional treatments in these questions. I do not report on these treatments in the present paper, but since treatments were assigned randomly, the results reported below are unaffected.[6]

3.2 Continuous responses obtained from control groups

To assess the direction of response bias and to determine which brackets are affected, I use information from separate control groups. These subjects received the same six questions in an unfolding brackets design, i. e., they first received an open-ended question asking for the quantity in question, and if they responded "Don't know", they entered a series of unfolding brackets. While these experiments are not the subject of the present paper, the responses to the corresponding open-ended questions can be used to obtain external information on the distribution of the quantities in question. Descriptive statistics on the open-ended responses obtained from this control group are reported in table 3.

In the sequel, I used only the responses to the open-ended questions from the control group, ignoring responses to follow-up unfolding brackets questions (since the latter are subject to large anchoring effects given the nature of the experimental design). Using responses to the open-ended question in the control group as a "gold standard" in the analysis of the range-card responses requires that two conditions are met.

First, non-response to the open-ended question should be random. To test for selection effects based on observable characteristics, I ran probit regressions with a binary indicator of non-

[6] These additional treatments used bracket values that were conditioned on prior information on the households income. An analysis of bracketing effects in these treatments is the subject of ongoing research.

response to the open-ended question as the dependent variable. Explanatory variables were the available personal and household characteristics of the subject.[7] In the probit non-response regressions for five of the six variables, the observable characteristics were jointly insignificant.[8] The observables were jointly significant only for the question on time spent filling in last year's tax form. In that regression, the only individually significant variables were the education dummies. Even though in general, a more detailed analysis of non-response behavior would be warranted, these findings suggest that in the present experiments, random non-response is a valid assumption.

Second, responses to the open-ended question should ideally be measured without error. The high proportion of focal point responses suggests that they are not. The use of focal values might be an indication of respondents' uncertainty about the quantity in question, as discussed by Hurd *et al*. (1998) and Battistin *et al*. (2001). From a pure measurement perspective, the effects of focal values could be purged using the method proposed by Heitjan and Rubin (1990, 1991). Such a rather involved approach was not attempted in the present paper, for the following reason. If the coarse data structure implied by heaping at focal values turned out to be non-ignorable (in the sense of Heitjan and Rubin, 1991), the location of the "true" distribution recovered by purging focal responses would be shifted relative to that of the open-ended responses. While this might change inference obtained from comparing bracketed responses with open-ended control data, the substantive results reported below would not be affected. As reported below, the distributions of the responses to the low and high bracket configurations are typically located to the right and to the left of the open-ended distribution, respectively. Hence, if the latter distribution is shifted because of measurement error, in particular because of heaping at focal points, the difference to the distribution of bracketed response gets smaller in one direction, but larger in the other. Similarly, the substantive results would also not be affected if measurement error in the open-ended question implied that the observed distribution is compressed relative to the true distribution.

This being said, a formal analysis of response bias, defined as the difference between the responses to range-card questions and the true distribution, would require to correct responses to the open-ended question for measurement error such as heaping due to focal points. Moreover, non-response would have to be addressed formally, for example by obtaining bounds for the true distribution following Horowitz and Manski (1995). This is left to future research. The following analysis uses the distribution of responses to the open-ended question as a standard for comparison without any modification.

[7] Specifically, these were age (and age squared), gender, a four-level indicator variable of educational attainment transformed into three dummy variables, a dummy for homemaker, a dummy for household head, a dummy for retired subjects, root household size, and the log of net household income in the previous month.

[8] Detailed regression results are available on request.

3.3 Results

A first check for the existence of bracketing effects is given by the location of the bracket associated to the median response in each treatment. These median brackets are reported in table 4.

For the number of hours respondents watched TV during the last week, the median responses in the three treatments fall in disjoint brackets, indicating that the distributions are different. This is a clear replication of earlier results on bracketing effects. In contrast, for the number of hours spent reading books, the median brackets are mutually consistent across treatments. The same holds true for the number of hours spent filling tax returns. The reason for this consistency is that a large enough number of households appear to belong to extreme brackets (top or bottom), and were willing to state this truthfully. These results already indicate that bracketing effects are specific to the exact location of the bracket bounds relative to the underlying distribution. I return to this issue in Section 6.

Turning to the expenditure categories, the results for food expenditure indicate strong bracketing effects. The brackets associated with median response in the three treatments are disjoint. With respect to travel expenditure, the fact that the median response in the low treatment falls in the (open) top bracket is consistent with the other two treatments, but median responses in the medium and high treatments are inconsistent with each other. Finally, responses to the bracketed questions on clothing expenditure are mutually consistent if one looks only at the medium bracket.

It should be noted that consistency of median brackets need not imply that the entire distribution is elicited without bias, while the fact that median brackets are disjoint typically implies that the entire distributions are signifcantly different, and that bracketing effects exist. A closer analysis confirms most of the results obtained by just looking at median response brackets, but also offers some additional insights.

Table 5 lists the distribution of responses to the range-card questions (expressed by the number of the bracket, not the underlying quantity), together with information on the control group. For the latter, the open-ended responses are discretized in three different ways so that they correspond to the three treatment groups. Specifically, the table reports the brackets associated with the 10, 25, 50, 75 and 90 percentiles of the responses. The bottom panel of Table 5 also contains χ^2-statistics for the null hypothesis of identical distribution of the responses in a bracketed design and in the open-ended control group. The most striking findings arise in the cases of TV watching, food expenditure and clothing expenditure where responses to bracketed questions are significantly different from open-ended responses in all three treatments.

Similar information is contained, in graphical form, in the odd-numbered figures 1 through 11. These figures depict histograms that correspond to table 5. In these figures, the top row contains the distribution of responses to the range-card questions for the three treatments (low, medium, and high levels of brackets). The bottom row the corresponding distribution of discretized open-ended responses, i. e., the bottom distributions are all based on the same observations but look differently because the brackets are different.

Comparing distributions of bracketed and the corresponding discretized open-ended responses pairwise reveals that bracketing effects arise mostly at either the bottom and/or top end of the distribution. For example, in the low treatment in the TV question, people appear to be unwilling to place themselves in the top bracket (see Figure 1a). Interpretations that are consistent with this finding include that (i) they are unwilling to admit that they are among the most intensive TV watchers, or (ii) they know that they are not among the most intensive TV watchers, and therefore reason that they should not place themselves into the top bracket. These interpretations are in line with existing research on survey response behavior reported in section 2.2. Similarly, in the case of food expenditure, there appears to be a strong tendency for respondents not to place themselves into extreme brackets (see Figure 7). Significant differences between the responses to the bracketed designs and the open-ended responses are also present in all other four questions, but in some cases not in all three treatments.

As discussed above, responses to the open-ended question might be biased because of selection effects associated with non-response or measurement error. This might remove some of the significant differences between bracketed and open-ended responses, and hence change some of the interpretations provided in the preceeding paragraphs. However, the striking differences between the treatment groups do not depend on a valid comparison with control group data.

The measurement effects of response bias in range-card questions can be seen better in the cumulative distribution functions reported in the even-numbered figures 2 through 12. In all cases, the distributions of the range-card responses are more compressed than those of the open-ended responses. Despite the fact that the open-ended distributions are only an approximation to that of the true quantities, the effects are striking.

Equally important, the magnitude of the response bias apparently depends on the location of the brackets in range-card questions relative to the population distribution. This effect is perferctly in line with psychological models of survey response behavior that predict that respondents extract information on the population distribution from brackets and form their response using that information. For instance, in the question of time used to fill in tax forms in the previous year, the brackets were, in all three treatments much higher than the population distribution. In addition, it might have been relatively easy to answer this question. Therefore, differences between the three treatment groups and to the control group are small – most subjects correctly placed themselves in the bottom bracket in all three groups. In the questions on expenditure items, the exact location of the bracket bounds relative to the continuous response distribution also affects response bias.

An extreme case is the low treatment in the clothing expenditure question, see figures 11 and 12. Here, the bracket distribution was much too low – the median response in the open-ended question was 1750 guilders, while the top bracket started at 350 guilders. While more than half of the subjects placed themselves correctly in the top bracket, the distribution obtained from bracketed responses is still shifted to the left by a large amount. In addition, not much can be learned from a distribution that has most mass in a half-open interval that starts below the median of the continuous response. Similarly, in the high treatment brackets were deliberately set much too high. As a result, responses at the low end of the distribution were largerly correct. However, not much can be learned from a interval data that has more than half of the responses in a large bottom interval.

4 Experiment 2

The second experiment reported in this paper was conducted using the Dutch CentER panel in July 2001. It focused exclusively on one quantity of specific economic interest: monthly household consumption. The experiment uses an extreme option in designing a survey on household consumption – a so-called "one-shot" question, adopted from an experimental module for the U.S. Health and Retirement Study (HRS). The text of this question is as follows:

> *Think about how much you and your household spent on everything in the past month. Please think about all bills such as rent, mortgage loan payments, utility, insurance and other bills, as well as all expenses such as food, clothing, transportation, entertainment and any other expenses you and your household may have. Roughly, how much would that amount to?*

While such a question is easy to administer, with low cost and a low rate of item non-response, the danger that respondents miss important components of their expenditure is quite high, and underreporting appears to be very likely. The alternative is to use more questions that refer to expenditure on more disaggregated categories. In an independent experiment administered using the CentERpanel, aggregation bias in a one-shot question has been documented; see Winter (2002). Those results indicate that a one-shot question is not very well suited to elicit household expenditure on non-durables. However, for the present paper, the facts that the one-shot question is relatively vague and that respondents are uncertain about their answer, are advantageous since the objective is to evaluate the effects of the level of bracket bounds on respondents' reports.

The results of Experiment 1, as discussed in Section 3, indicate that the location of bracket values relative to the distribution of the quantity of interest in the population has a crucial effect on the existence and strength of bracketing biases, in line with findings from the socio-psychological literature on conversational aspects of survey response behavior. The experiment presented in this section sharpens the findings of Experiment 1 by using an objective procedure to create the bracket values for alternative treatments. This procedure is based on the (expected) distribution of monthly consumption in the population.

4.1 Design

The experiment has three treatments, one with low levels of the bracket bounds, one with medium and one with high levels. The number of subjects in the low, medium and high treatment groups was 96, 97, and 87, respectively.

While in experiment 1, the bracket bounds were evenly spaced, bracket bounds in this experiment are based on percentiles of the expected distribution of monthly food expenditure in the population. These percentiles have a distance of 10 points each, ranging from the 5th to the 65th and from the 35th to the 95th percentile in the low and high treatments, respectively. In the medium treatment, bounds are located in 10-percent steps symetrically around the expected median, starting at the 20-percentile and ending at the 80-percentile. With percentiles that are evenly spaced, the shape of the expected distribution implies that the bracket bounds are not evenly spaced and that the range covered by each bracket increases away from the expected median. Table 6 reports the percentiles used and the bracket bounds. An important advantage of this design is that four brackets overlap exactly in both treatments, which facilitates comparing the response distributions.

The expected distribution of monthly total household non-durables consumption was constructed using data on household expenditure from the Dutch expenditure survey. The latest data from that survey that were available when this paper was written are from 1998, and all values are converted to 2001 prices using the consumer price index. Also, these data had to be converted from annual to monthly values since the experiments focused on monthly expenditures. Since these data are obtained from budget surveys, they provide very reliable measures of household consumption. However, even ignoring any errors induced by predicting monthly consumption in May 2001 from annual data for 1998, one expects that the responses to the one-shot question differ from the expectation due to aggregation bias, see Winter (2002). Such deviations do not affect the substantive results reported below: All tests for, and interpretations of, response bias induced by range-card questions are based on the 2001 experimental responses. The 1998 data are only used to determine the bracket bounds for the range-card treatments.

Similar to experiment 1, there was also a control group that received an open-ended question. Due to the overall design of the experiments run as part of the CentER panel in June 2001, the control group was significantly larger than the treatment groups, with a total of 753 open-ended responses. Table 7 summarizes these responses. As in experiment 1, the control group data originate from a separate experiment on default bracketing that was administered on the same weekend. Therefore, there was an option to reply "Don't know" to the open-ended question, in contrast to the bracketed design where such an option was not given.

The same caveats as in experiment 1 apply to the interpretation of the distribution of the control group data. The non-response rate to the open-ended question, 32.6%, is rather high, indicating a fair degree of uncertainty about the quantity in question.[9] A non-response probit model on the same set of observable household characteristics as those used in 3 was jointly insignificant, so the following analysis uses open-ended responses under the assumption of random non-response. Focal point responses are rather frequent, as in the responses to the open-ended questions in experiment 1. No attempt to correct for heaping effects was undertaken, with the same rationale as above. The results on differences between distributions from alternative range-card configurations are quite large. While measurement error in the control group might distort the comparisons between treatment and control groups, there is an offsetting effect: As the difference between one treatment group response distribution and the control distribution gets smaller, the difference for the other treatment group gets larger.

4.2 Results

A first indication that the distributions of bracketed responses in the low and high treatments are different is given by the median responses. In the low treatment, the median response falls into the third bracket, corresponding to $(2590, 3080)$ guilders, while in the high treatment, the median response falls in the second bracket, which is $(3490, 3920)$ guilders. These brackets are disjoint. The bracket that contains the median response in the medium treatment is $(2830, 3290)$.

Comparing the two bracketed treatments with the continuous responses to the open-ended question, the χ^2-tests reported in table 8 indicate that responses in the low treatment are significantly different from the continuous responses ($p = 0.022$), while responses in the high treatment are not. The brackets corresponding to the 10, 25, 50, 75 and 90 percentiles of the responses (also table 8) indicate that differences in bracketed responses from open-ended responses are primarily to be found in the upper end of the distribution, see also figure 13.

Figure 14 contains cumulative distribution functions constructed from the three bracketed treatments and the open-ended control group. Again, it is apparent that the responses in the low and high treatments are different. In contrast to the distribution functions obtained from bracketed responses in some of the extreme treatments in experiment 1, the shapes of the

[9] However, a non-response rate of slightly more than 30% is similar in magnitude to the non-response rate of 35.8% reported by Hurd *et al.* . (1998) for a very similar one-shot question on consumption in an experimental module of the AHEAD survey.

distributions obtained here are similar to each other. This corresponds to how the bracket treatments were constructed, using percentiles of an expected distribution. Note also that the responses to the medium treatment, which was designed to be on top of the expected distribution, is not shifted by a large amount (for instance, the median bracket contains the median of the open-ended distribution), so reporting appears to be not distorted in the center of the distribution. However, the distribution of bracketed responses in the medium treatment is compressed relative to the open-ended distribution which confirms that response bias occurs mostly in the top and bottom brackets. This implies that response bias affects the tails of the distribution if brackets are placed on top of the expected distribution.

From Figure 13, it appears that bracketing bias arises mainly from underreporting in the low treatment while open-ended responses in the high treatment are in line with the control group. As noted before, the responses to the open-ended question might be biased because of selection effects associated with non-response or measurement error. To the extent that they are biased, the last statement might have to be modified. However, this would not remove the significant difference between responses in the experimental treatments, and it would not change the conclusion that the choice of bracket values affects the measurement of the underlying quantity.

5 Estimation of econometric models with biased interval data

Given the evidence on bracketing effects in range-card questions, and the result that response bias invalidates inference from regression with interval data described in section 2.1, it is interesting to check whether response bias is relevant in practice. This section contains an illustrative example. In order to make this example representative for current practice in applied work, I stick to the parametric framework. The regression analysis with interval data using ordered probit models is standard practice, as reviewed for example by Verbeek (2000), pp. 192-4.

Suppose three researchers want to explore factors that determine monthly household consumption. These researchers use interval data on consumption obtained from range-card questions in three household surveys administered at the same time using random samples of the same population. Researcher A has interval data on monthly consumption obtained from a survey that used a bracket configuration corresponding to the low treatment in experiment 2, and researchers B and C use data corresponding to the medium and high treatments, respectively. The covariates they use – age, household size and net household income – stem from identical question designs and are assumed to be measured without error. This set-up corresponds exactly to the data obtained in experiment 2.

Table 8 reports, in the top panel, the results of ordered probit regressions with known thresholds obtained by the three researchers.[10] These results are strikingly different, both with respect to the significance of coefficients and their magnitudes. For instance, researcher C would conclude from high treatment data that there is no significant age effect on household consumption, controlling for income while reseacher A and B would find significant effects.

These illustrative regressions have important implications. First, bracketing effects induced by the location of brackets in range-card questions can distort econometric analysis in practice. Second, if a researcher has only data from one bracket configuration, he cannot infer from his regression results alone whether they are affected by reponse bias. Only a comparison of regressions with data from different designs can raise doubts about the validity of interval data.

The bottom panel provides estimates of ordered probit regressions based on the open-ended control data which have been discretized using the thresholds given by the low, medium and high experimental designs, respectively (i. e., the same data are used in all three regressions). Note two results: First, there are differences between the regression results with range-card data and the discretized open-ended data in all three treatments. Second, the thresholds used for discretizing the open-ended data also affect inference with respect to age effects. The latter observation indicates that the age effect is not correctly specified in the underlying model (such misspecifications are not unlikely in practice). This by itself does not invalidate the implication of these illustrative regressions – rather to the contrary. While with continuous data, some simple analysis would reveal the misspecification of the age effect rather quickly, researchers A, B and C would not be able to detect the misspecification of the age effect from their samples alone.

Moreover, response bias in range-card questions on the continuous variable depends on its level (as shown in the previous sections). Therefore, classical measurement error frameworks in which response errors occur randomly provide no help for the present problems. The next section contains a more general discussion.

6 Conclusions for survey practice and future research

The experimental evidence on bracketing effects in range-card questions reported in this paper has strong implications for empirical research with interval data obtained from household surveys. These results call for strategies to limit bracketing effects *ex ante* and to account for response bias in interval data *ex post*. In this concluding section, I first discuss some

[10] This model is not the most appropriate one for the analysis of household consumption. This is not the point of the present analysis, though.

implications for applied research. Then I present strategies for avoiding or reducing bracketing effects in the first place. Finally, I discuss how estimators for regressions with biased interval data might be constructed.

6.1 Implications for applied research

The evidence from controlled survey experiments presented in this paper should have convinced the reader that in any field survey, there is a good chance that responses to range-card questions are biased, i. e., that some respondents do not report the bracket into which their true response falls. There are conditions under which such bracketing effects are less likely to bias responses (these are discussed below in detail). However, from a practical perspective, it is important to realize that if only one bracket configuration is used in a survey and no unbiased validation data are available, one cannot test for the absence of bracketing effects.

Testing for the absence of bracketing effects requires either experimental strategies (i. e., at least two versions of a range-card question with different bracket configurations that are assigned randomly) or unbiased, continuous validation data. The latter option is rarely available in practice. Usually, either open-ended or range-card questions are used. If range-card questions are used as follow-up questions in case of non-response to an open-ended question, the distribution of the open-ended responses might serve as a standard against which the bracketed responses can be tested. This requires some technical conditions (discussed below in section 6.3) which might fail in practice.

The main recommendation for survey practice is, therefore, that there should always be experimental variation with respect to bracket bounds in range-card questions. Such experimental variation does not generally reduce the information contained in interval data obtained from range-card questions, but it allows to test for the absence of bracketing effects. This advice is, with respect to range-card questions, similar to Danny Kahneman's statement on unfolding brackets questions, as reported by Hurd *et al.* . (1998): "Collecting bracket responses without varying the anchors is criminally negligent."

6.2 Strategies for limiting bracketing effects in range-card questions

The results of experiment 1 indicate that the exact location of bracket bounds is crucial for the nature of bracket effects. Consider two polar extremes. If brackets are far off the distribution of a quantity in the population, responses are driven into either the bottom or the top bracket, and misreporting is reduced. However, since these brackets are either wide or half-open, and cover a large part of the distribution's mass, resulting interval data do not contain much information, so they are not useful for substantive analysis. In contrast, if bracket bounds are slightly off the distribution of the underlying quantity, as in the questions on time spent watching TV and the food expenditure question, bracket effects that bias measurement are very likely to show up, and they affect mainly the tails of the distribution. This case of a more concentrated response distribution is similar to the effects of anchoring of the responses to unfolding brackets questions, as documented by Hurd *et al.* (1998). In some applications, confining response bias to the tails of the distribution might be acceptable, in others, such as the analysis of wealth holdings, the tails are of particular interest.

In summary, a design that has the midpoint of the bracket contribution at the median of the expected distribution has the advantage of restricting bracketing effects mostly to the tails. If substantive analysis focuses on logs rather than levels, it might be preferable to construct brackets using the expected log distribution, although there exists no experimental evidence on response bias in this case. Another design tool that might affect bias induced by bracketing effects is optimal spacing of brackets, following work by Cox (1957). In Cox' design, brackets cover smaller quantiles of a distribution the further away from the median they are.[11] There is no experimental evidence on this design either, and it is not apparent how it might affect the magnitude of response bias arising from bracketing effects.

Even if it could be established that certain bracket configurations reduce the magnitude of response bias in interval data, they will be difficult to implement in many practical applications. Prior knowledge about the expected true response distribution is rare, as illustrated in experiment 2 in which the expected distribution was shifted relative to the distribution obtained from the open-ended question.

Another possibility to reduce bracketing effects that does not require a priori knowledge about the expected distribution is to increase the number of brackets. Results from psychological research and also the experiments reported in this study indicate that response bias is mostly associated with bottom and top brackets, so it stands to reason that increasing the number brackets might be helpful. Preliminary results from the SAVE study, a household survey on savings and asset holdings conducted in Germany in the Spring of 2001, indicate that this might indeed be the case.[12] The SAVE survey included experimental modules that assigned open-ended and range-card questions on wealth components (the latter with 14 progressively spaced brackets) randomly to participants. Data from these experimental modules show that responses from the range-card questions are consistent with those from the open-ended questions, suggesting that bracketing effects disappear as the number of brackets increase. However, non-response rates on both formats are similar, so using range-cards might loose its

[11] For instance, in the case of a normally distributed quantity and six brackets, Cox suggests having the bounds so that they cover 7.4%, 18.1%, 24.5%, 24.5%, 18.1%, and 7.4% of the mass (i. e., the brackets are symmetric about the median).

[12] See Börsch-Supan and Essig (2002) for first results of the SAVE survey. The results on bracketing effects, obtained from the experimental modules by the author of this paper, are not yet documented.

main advantages of improving response rates or, in a follow-up design, of eliciting information from respondents who do not give an answer to an open-ended question. More data from experimental surveys are required to understand the apparent trade-off between non-response and response bias in range card questions, and how this trade-off is influenced by the number of brackets.

Further research is also warranted on the psychological mechanisms that generate bracketing effects. In ongoing research, Schwarz and Winter (2001) study the effects of a person's perceived typicality on response behavior. The main idea is that respondents extract information about the distribution of the quantity of interest in the population or some other reference group from the brackets and use it as a frame of reference. While this model would predict the results obtained in the present study, other interesting implications arise. For example, bracketing biases should be attenuated for respondents who think of themselves as "atypical" because they know their income is out of range. An additional implication of this model is that the location of brackets in a range-card question might also influence responses in subsequent survey questions, see also the study by Menon *et al*. (1997).

Despite these open research questions, a few general suggestions from research on survey response behavior can be applied to the administration of range-card questions. As pointed out in section 2, there is a lot of support for a model of the survey response process in which bracketing effects arise because respondents who are uncertain about the quantity in question resort to information contained in the questionnaire, in particular numerical clues. In the case of range-card questions, respondents infer the location of a quantity's population distribution from bracket bounds. Accordingly, bracketing effects are reduced or disappear if survey respondents are certain about the quantity in question, if they have ample time to answer, and if they are highly motivated to provide correct answers. If at all possible, vague questions such as the "total consumption" one-shot question used in experimental HRS and AHEAD modules and in the experiments reported in this paper should therefore be avoided in practice.[13]

Finally, with the increasing use of internet surveys and their potential for using graphical interfaces (e. g., Couper *et al.* , 2001), range-card questions are likely to become more important in applied work. The alternative, unfolding brackets techniques, was developed primarily because range cards are difficult to administer in telephone interviews. This restriction does not apply in computer-assisted or internet surveys with graphical interfaces. As this paper has shown, responses to range-card questions are subject to potential bias that correspondes to anchoring bias in unfolding brackets questions. Important questions for survey design concern the relative magnitude of these biases and the trade-off with non-response. Such a comparison

---

[13] Browning *et al*. (2002) discuss how a question on total consumption might be phrased to improve respondents' understanding. See also Winter (2002).

is conceptually difficult, however. In the survey experiments reported in the existing literature and also in this paper, brackets in range-card questions are set either such that the presence of bracketing effects can be demonstrated (this is the research strategy in psychology where the cognitive processes that generate response bias are of interest) or using heuristic strategies that replicate designs typically found in field surveys (this was the strategy in the present paper). To make a horse-race between two survey designs such as unfolding brackets and range cards meaningful, both would have to be designed such that response bias is minimal *ex ante*. As the discussion in this section has shown, there is not much systematic knowledge of how range-card questions should be designed in order to limit bracketing effects and response bias.

6.3 Strategies for the econometric analysis of biased interval data

Even though the magnitude of bracketing effects might be reduced by careful survey design and administration, it would be helpful to have econometric tools that allow valid inference with potentially biased interval data. The framework of Manski and Tamer (2002) can serve as a natural starting point for developing such methods. As pointed out before, response bias induced by bracketing effects implies that $P(v_l \leq v \leq v_u) < 1$, i. e., that assumption I of Manski and Tamer is violated. Consequently, all identification results in their abstract framework as well as consistency results for estimators in more traditional parametric regression models fail to hold.

One might argue that the fact that some respondents report intervals that do not contain the true value of arises from random errors. Classical measurement error (or more correctly, classical misclassification) in the case of discrete dependent and independent variables has been explored by several authors and can be dealt with in a straightforward fashion.[14] These approaches typically assume that classification error is random or, more generally, that the probability of misclassification is a function of the latent continuous variable. However, they are not readily applicable in the present context: Results from the social psychology literature on survey response bias, as well as the experimental evidence in this paper, suggest that misreporting is the outcome of a complicated cognitive process, which implies that the probability of misclassification is a potentially discountinous, non-monotonic function of the

[14] Contributions to the literatur on misclassification in discrete variables include Aigner (1973), Copas (1988), Poterba and Summers (1995), and Bollinger (1996), and Hausman *et al*. (1998) on binary dependent variables, Bollinger (1996) on binary independent variables, Abrevaya and Hausman (1999) on duration data, and Ekholm and Palmgren (1987), and Ramalho (2002) on multinomial dependent variables.

underlying continuous variable, covariates such as demographic variables, and unobservables such as cognitive ability.[15]

An obvious strategy to rescue identification results and to construct estimators for regressions with biased interval data would be to exploit knowledge of $P(v \notin [v_l, v_u]) = 1-P(v_l \leq v \leq v_u)$, the probability of misreporting. If point estimates of, or upper bounds on, this probability were available, it is conceivable that identification regions and bounds for estimates of regressions like $E(y|x,v)$ and $E(v|x)$ could be established. Whether the resulting regions and bounds would be helpful in practice depends of course on the magnitude of $P(v \notin [v_l, v_u])$. In addition to extending non-parametric estimators to the case of biased interval data, knowledge of $P(v \notin [v_l, v_u])$ should also allow to construct parametric estimators (say, generalizations of ordered probit).

A more fundamental problem is estimation of $P(v \notin [v_l, v_u])$. If only interval data obtained from a single bracket configuration are available, the probability of misreporting is not identified (unless one is willing to impose strong distributional assumptions). If external continuously measured validation data are available, the probability of misreporting could potentially be estimated using a "double sampling" approach, see Ekholm and Palmgren (1987) among others. However, this is rarely the case in practice – survey experiments such as those reported in this paper are an example where subjects are assigned randomly to open-ended and range-card questions. In most field surveys, however, bracketed questions – unfolding brackets or range cards – are presented only to those respondents who do not answer an open-ended question. Using data from open-ended responses to infer misreporting probabilities in the sub-sample of bracketed responses is possible but requires additional ignorability assumptions on the selection process that might fail in practice. Specifically, initial non-respondents presented with follow-up range-card questions might be different from respondents who anser the open-ended question with respect to unobservable characteristics that are correlated with the level of the quantity in question (think of cognitive ability and income).

A practical obstacle to approaches that use estimates of $P(v \notin [v_l, v_u])$ is that initial non-respondents are likely to be more uncertain about the quantity in question, and therefore they are also more likely to incorporate numerical clues in their response (see section 2). Consequently, their responses are also more likely to be subject to bracketing effects. The probability of misreporting – even if somehow identified – should therefore be higher in follow-up brackets than in forced brackets. This might imply that identification regions and parameter bounds are not informative.

[15] Note that anchoring bias in unfolding brackets questions does not follow classical assumptions since anchoring to the first threshold biases subsequent responses in the bracketing sequence. In this case, assumption MI of the Manski-Tamer framework is violated as well.

An attractive approach to obtain estimates of misreporting probabilities is available when surveys contain different bracket configurations that are randomized. In regions in which the brackets in different configurations overlap, misreporting probabilities can be inferred from comparing responses across bracket configurations. In the extreme, when responses are consistent across the entire range of brackets, one might set $P(v \notin [v_l, v_u]) = 0$ so that the original Manski-Tamer framework is restored.

In the framework sketched in the previous paragraphs, a promising strategy to narrow identification regions and parameter bounds in regressions with interval data is to impose assumptions on the data generating process that are informed by knowledge of the cognitive processes that generate survey responses. For instance, if one could assume that bracketing effects shift responses at most by one bracket, probabilities of misreporting could be bounded using observed interval data alone. This approach has been used in the literature on classification errors in discrete dependent variables; see Hausman *et al*. (1998). Whether such a strong assumption is justified requires more research, as do the other extensions of the Manski-Tamer regression framework to the case of biased interval data that I proposed in the previous paragraphs.

Appendix: The CentERpanel

The CentERpanel is an internet-based telepanel. It consists of some 2000 households in the Netherlands. Every week, the members of those households fill in a questionnaire at their home computers or set-top boxes attached to a TV set. Households may use their own computers, or they are provided with PCs or set-top boxes by CentERdata (the agency running the panel, a unit of Tilburg University). In this way, each year about fifty questionnaires of up to 30 minutes each are answered by the respondents. The advantage of such a survey method is that computer-assisted interviewing is combined with panel research: quick results, consistency checks, reliable ways to measure changes, and relatively low attrition rates. CentERdata tries to make sure that respondents remain members of the panel for longer periods, and that they are motivated to answer the questionnaires carefully, thus providing valid data.[16]

It is known which household member has answered the questionnaire on a particular weekend. In most cases, this will be the person responsible for the household's finances (the "financial officer"), but it could be other members as well. It is also possible to request that a questionnaire be answered only by some specific household member (say, the financial officer).

The CentERpanel was established in 1991 and since then has been used in many research projects. Large, complex research projects (like the CentER Savings Project) profit from the possibility of large-scale data collection. Small projects – such as those reported below – profit from the fact that telepanel surveys are quick and efficient. In addition, data obtained in small-scale projects such as experimental surveys can be matched with existing data from the CentER Savings Survey. In experimental surveys, questions can be conditioned on existing information about household characteristics, including variables that might be difficult to obtain in other methodological studies on survey design, such as household income.

The CentERpanel is representative of the Dutch population. Detailed tabulations of the distributions of key demographic variables (such as age, sex, education, region) in the CentERpanel and in population data provided by Statistics Netherlands can be found on CentERdata's website at `http://cdata4.kub.nl/eng/representative`.

[16] This is a situation in which some effort is invested in alleviating the principal-agent situation of survey data production discussed by Philipson (2001).

References

Abrevaya, J. and J. A. Hausman (1999): Semiparametric estimation with mismeasured dependent variables: An application to duration models for unemployment spells. *Annales d'Economie et de Statistique*, 55-56, 243–275.

Aigner, D. J. (1973): Regression with a binary independent variable subject to errors of classification. *Journal of Econometrics*, 1, 49–59.

Battistin, E., R. Miniaci, and G. Weber (2001): What do we learn from recall consumption data? Unpublished manuscript, Padua University.

Bollinger, C. R. (1996): Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics*, 73, 387–399.

Börsch-Supan, A. and L. Essig (2002): *Sparen in Deutschland: Ergebnisse der ersten SAVE-Studie*. Köln: Deutsches Institut für Altersvorsorge.

Browning, M., T. F. Crossley, and G. Weber (2002): Asking consumption questions in general purpose surveys. Unpublished manuscript.

Copas, J. B. (1988): Binary regression models for contaminated data. *Journal of the Royal Statistical Society B*, 50(2), 225–265.

Couper, M. P., M. W. Traugott, and M. J. Lamias (2001): Web survey design and administration. *Public Opinion Quarterly*, 65, 230–253.

Cox, D. R. (1957): Note on grouping. *Journal of the American Statistical Association*, 52, 543–547.

Ekholm, A. and J. Palmgren (1987): Correction for misclassification using doubly sampled data. *Journal of Official Statistics*, 3(4), 419–429.

Green, D., K. E. Jacowitz, D. Kahneman, and D. McFadden (1998): Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, 20, 85–116.

Hausman, J., J. Abrevaya, and F. M. Scott-Morton (1998): Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, 87(2), 239–269.

Heitjan, D. F. and D. B. Rubin (1990): Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, 85(410), 304–314.

Heitjan, D. F. and D. B. Rubin (1991): Ignorability and coarse data. *Annals of Statistics*, 19(4), 2244–2253.

Horowitz, J. L. and C. F. Manski (1995): Identification and robustness with contaminated and corrupted data. *Econometrica*, 63(2), 281–302.

Horowitz, J. L. and C. 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.

Horowitz, J. L. and C. F. Manski (2000): Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449), 77–84.

Hsiao, C. (1983): Regression analysis with a categorized explanatory variable. In S. Karlin, T. Amemiya, and L. Goodman (Eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press.

Hurd, M. D. (1999): Anchoring and acquiescence bias in measuring assets in household surveys. *Journal of Risk and Uncertainty*, 19(1/3), 111–136.

Hurd, M. D., D. McFadden, H. Chand, L. Gan, A. Merrill, and M. Roberts (1998): Consumption and savings balances of the elderly: Experimental evidence on survey response bias. In D. Wise (Ed.), *Frontiers in the Economics of Aging*, 353–387. Chicago, IL: University of Chicago Press.

Jacowitz, K. E. and D. Kahneman (1995): Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166.

Juster, F. T. and J. 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.

Manski, C. F. (1975): Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3, 205–228.

Manski, C. F. (1985): Semiparametric analysis of discrete response: Asymptotic properties of the Maximum Score Estimator. *Journal of Econometrics*, 27, 313–333.

Manski, C. F. and E. Tamer (2002): Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2), 519–546.

Menon, G., O. Raghubir, and N. Schwarz (1995): Behavioral frequency judgements: An accessibility–diagnostic framework. *Journal of Consumer Research*, 22, 212–228.

Menon, G., O. Raghubir, and N. Schwarz (1997): How much will I spend? Factors affecting consumers' estimates of future expense. *Journal of Consumer Psychology*, 6(2), 141–164.

Mussweiler, T. and F. Strack (2000): Numeric judgements under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology*, 36, 495–518.

Philipson, T. (1997): Data markets and the production of surveys. *Review of Economic Studies*, 64(1), 47–72.

Philipson, T. (2001): Data markets, missing data, and incentive pay. *Econometrica*, 69(4), 1099–1111.

Philipson, T. and A. Malani (1999): Measurement errors: A principal investigator-agent approach. *Journal of Econometrics*, 91, 273–298.

Poterba, J. M. and L. H. Summers (1995): Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. *Review of Economics and Statistics*, 77(2), 207–216.

Ramalho, E. A. (2002): Regression models for choice-based samples with misclassification in the response variable. *Journal of Econometrics*, 106, 171–201.

Schwarz, N., H. J. Hippler, B. Deutsch, and F. Strack (1985): Response categories: Effects on behavioral reports and comparative judgements. *Public Opinion Quarterly*, 49, 388–395.

Schwarz, N. and D. Oyserman (2001): Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, forthcoming.

Schwarz, N. and J. K. Winter (2001): Own typicality and bracketing effects in survey questions. Unpublished manuscript, University of Michigan and University of Mannheim.

Stocké, V. (2001): The influence of frequency scales on the behavior: A theoretical model and its empirical examination. Discussion Paper No. 01-24, Sonderforschungsbereich 504, Universität Mannheim.

Sudman, S., N. M. Bradburn, and N. Schwarz (1996): *Thinking About Answers*. San Francisco, CA: Jossey-Bass.

Tourangeau, R., L. J. Rips, and K. Rasinski (2000): *The Psychology of Survey Response*. New York, NY and Cambridge, UK: Cambridge University Press.

Vazquez Alvarez, R., B. Melenberg, and A. von Soest (1999): Bounds on quantiles in the presence of full and partial item non-response. Discussion Paper No. 9938, CentER, Tilburg University.

Verbeek, M. (2000): *A Guide to Modern Econometrics*. Chichester, UK: Wiley.

Winter, J. K. (2002): Design effects in survey-based measures of household consumption. Discussion Paper No. 02-34, Sonderforschungsbereich 504, University of Mannheim.

Table 1: Experiment 1, wording of questions

1. TV – time use
   How many hours did you watch TV last week?
2. Books – time use
   How many hours did you read books last month?
3. Tax-form – time use
   How many hours did you spend last year filling in your tax form?
4. Food – expenditures
   Which amount in guilders did your household spend on food last month?
5. Travel – expenditures
   Which amount in guilders did your household spend last year on holidays, camping equipment (also trailer) and weekend trips?
6. Clothing – expenditures
   Which amount in guilders did your household spend last year on clothing and footwear?

Source: Experiments conducted using the CentER Panel, February 2001

Table 2: Experiment 1, design

Bracket number	Time use questions (hours)			Expenditure questions (guilders)		
	TV	Books	Tax form	Food	Travel	Clothing
	Upper bracket bound (low treatment)					
1	2	2	1	275	825	50
2	3	3	2	325	1025	100
3	4	4	3	375	1225	150
4	5	5	4	425	1425	200
5	6	6	5	475	1625	250
6	7	7	6	525	1825	300
7	8	8	7	575	2025	350
8	∞	∞	∞	∞	∞	∞
	Upper bracket bound (medium treatment)					
1	6	6	3	450	2050	1025
2	9	9	4	500	2250	1225
3	12	12	5	550	2450	1425
4	15	15	6	600	2650	1625
5	18	18	7	650	2850	1825
6	21	21	8	700	3050	2025
7	24	24	9	750	3250	2225
8	∞	∞	∞	∞	∞	∞
	Upper bracket bound (high treatment)					
1	15	15	4	675	3650	3000
2	20	20	6	725	3850	3200
3	25	25	8	775	4050	3400
4	30	30	10	825	4250	3600
5	35	35	12	875	4450	3800
6	40	40	14	925	4650	4000
7	45	45	16	975	4850	4200
8	∞	∞	∞	∞	∞	∞
Number of observations						
Low treatment	99	63	75	37	35	44
Medium treatment	76	62	48	43	25	40
High treatment	102	58	69	33	30	61
Total	277	183	192	113	90	145

Source: Experiments conducted using the CentER Panel, February 2001

Table 3: Experiment 1, descriptive statistics for control group with open-ended questions

	Time use questions						Expenditure questions					
	TV		Books		Tax form		Food		Travel		Clothing	
	N	%	N	%	N	%	N	%	N	%	N	%
Total responses	808		559		576		413		322		407	
Don't know	29	3.6	32	5.7	104	18.0	88	21.3	41	12.73	134	32.92
Continuous responses	779	96.4	527	94.3	472	81.9	325	78.7	281	87.27	273	
Focal responses												
Multiples of 5	453	58.2	357	67.7	52	11.0	324	99.7	279	99.3	271	99.3
Multiples of 10	291	37.4	242	45.9	24	5.0	317	97.5	277	98.6	269	98.5
Multiples of 50	12	1.5	31	5.9	0		306	94.2	276	98.2	267	97.8
Multiples of 100	5	0.6	9	1.7	0		248	76.3	266	94.7	244	89.4
Multiples of 500	0		0		0		88	27.1	238	84.7	187	68.5
Multiples of 1000	0		0		0		46	14.2	174	61.9	113	41.4
	Hours						guilders					
Minimum		1		1		1		100		100		1
Maximum		100		200		30		2500		35000		10000
Median		15		12		2		600		3500		1750
Mean		17.23		19.82		3.69		706.65		4658.53		2165.69
St. dev.		12.38		21.11		3.63		383.64		4654.31		1888.44

Source: Experiments conducted using the CentER Panel, February 2001

Table 4: Experiment 1, brackets containing median response

	Time use questions			Expenditure questions		
	TV	Books	Tax form	Food	Travel	Clothing
Low treatment	(7, 8)	(8, ∞)	(1, 2)	(475, 525)	(2025, ∞)	(350, ∞)
Medium treatment	(9, 12)	(9, 12)	(0, 3)	(600, 650)	(2850, 3050)	(1025, 1225)
High treatment	(15, 20)	(0, 15)	(0, 4)	(675, 725)	(3650, 3850)	(0, 3000)
Open-ended question	15	12	2	600	3500	1750

Source: Experiments conducted using the CentER Panel, February 2001

Table 5: Experiment 1, distributions of responses to bracketed and open-ended questions

	Time use questions						Expenditure questions					
	TV		Books		Tax form		Food		Travel		Clothing	
	number of bracket in which percentiles fall											
	brack	open	brack	open	brack	open	brack	open	brack	open	brack	open
Low treatment												
10-percentile	2	4	2	3	1	1	2	1	3	3	4	8
25-percentile	4	8	4	5	1	2	3	4	5	7	6.5	8
median	7	8	8	8	2	2	6	8	8	8	8	8
75-percentile	8	8	8	8	4	4	8	8	8	8	8	8
90-percentile	8	8	8	8	5	8	8	8	8	8	8	8
Medium treatment												
10-percentile	1	1	1	1	1	1	1	1	1	1	1	1
25-percentile	2	3	1	1	1	1	2	1	2	1	1	1
median	3	4	3	3	1	1	5	4	6	8	2	5
75-percentile	5	6	4	8	2	2	7	8	8	8	4	8
90-percentile	6	8	7	8	4	6	8	8	8	8	7.5	8
High treatment												
10-percentile	1	1	1	1	1	1	1	1	1	1	1	1
25-percentile	1	1	1	1	1	1	1	1	1	1	1	1
median	2	1	2	1	1	1	2	1	2	1	1	1
75-percentile	3	3	4	3	1	1	4	8	7	8	2	1
90-percentile	4	4	6	7	2	3	6	8	8	8	4	8
	χ^2-tests for identical distributions (d.f. = 7)											
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Low treatment	78.2	0.000	10.7	0.149	18.4	0.010	23.7	0.001	12.1	0.098	36.6	0.000
Medium treatment	39.4	0.000	46.8	0.000	24.6	0.001	47.7	0.000	45.0	0.000	45.3	0.000
High treatment	14.5	0.021	17.5	0.014	9.6	0.210	25.7	0.001	42.0	0.000	461	0.000

Source: Experiments conducted using the CentER Panel, February 2001
Notes: For the control group (continuous responses to open-ended questions), the distribution of responses is based on interval data constructed using the bracket values of the treatment groups.

Table 6: Experiment 2, design

	Total household non-durables expenditure, last month (guilders)					
	Low treatment		Medium treatment		High treatment	
Bracket	percentile	upper bound	percentile	upper bound	percentile	upper bound
1	5	1890	20	2830	35	3490
2	15	2590	30	3290	45	3920
3	25	3080	40	3690	55	4380
4	35	3490	50	4090	65	4800
5	45	3920	60	4590	75	5340
6	55	4380	70	5110	85	6130
7	65	4800	80	5700	95	7710
8	100	∞	100	∞	100	∞
Number of observations	96		97		87	

Source: Experiments conducted using the CentER Panel, June 2001.
Notes: The upper bound of each bracket obtained from a percentile of the expected distribution of food expenditure in the population (after rounding to the next multiple of 10 Guilders).

Table 7: Experiment 2, descriptive statistics for control group with open-ended question

	Total household non-durables expenditure, last month	
	N	%
Total responses	1117	
Don’t know	364	32.6
Continuous responses	753	77.4
Focal responses		
Multiples of 100	714	94.8
Multiples of 500	533	70.8
Multiples of 1000	352	46.8
	guilders	
Minimum		0
Maximum		40000
Median		3000
Mean		3891
St. dev.		3741.9

Source: Experiments conducted using the CentER Panel, June 2001.

Table 8: Experiment 2, distributions of responses to bracketed and open-ended questions

	Total household non-durables expenditures, last month	
	brackets	open-ended
	number of bracket in which percentiles fall	
Low treatment		
10-percentile	1	1
25-percentile	2	2
median	3	3
75-percentile	6	7
90-percentile	8	8
High treatment		
10-percentile	1	1
25-percentile	1	1
median	2	1
75-percentile	4	4
90-percentile	6	6
	χ^2-tests for identical distributions (d.f. = 7)	
	χ^2	p
Low treatment	16.35	0.022
High treatment	5.54	0.594

Source: Experiments conducted using the CentER Panel, June 2001.

Table 9: Experiment 2, some illustrative interval regressions

	low treatment			medium treatment			high treatment		
	estimate	p-value		estimate	p-value		estimate	p-value	
experimental groups (range-card questions)									
age	201.79	0.003	***	191.62	0.036	**	174.39	0.104	
age squared	-2.01	0.003	***	-2.12	0.027	**	-1.76	0.109	
root HH size	652.10	0.015	**	1210.10	0.003	***	920.78	0.081	*
net income	0.20	0.002	***	0.00	0.826		0.05	0.062	*
constant	-3356.7	0.023	**	-2519.4	0.205		-1993.9	0.413	
N	96			97			87		
control group (open-ended question, discretized)									
age	49.25	0.037	**	63.54	0.022	**	61.69	0.091	*
age squared	-0.38	0.092	*	-0.48	0.070	*	-0.45	0.199	
root HH size	1518.73	0.000	***	1395.9	0.000	***	1697.31	0.000	***
net income	0.02	0.000	***	0.02	0.000	***	0.024	0.001	***
constant	-401.3	0.496		-767.19	0.279		-1257.7	0.180	
N	744			744			744		

Source: Experiments conducted using the CentER Panel, June 2001.

Figure 1: Experiment 1, response distributions (time use – TV, last week)

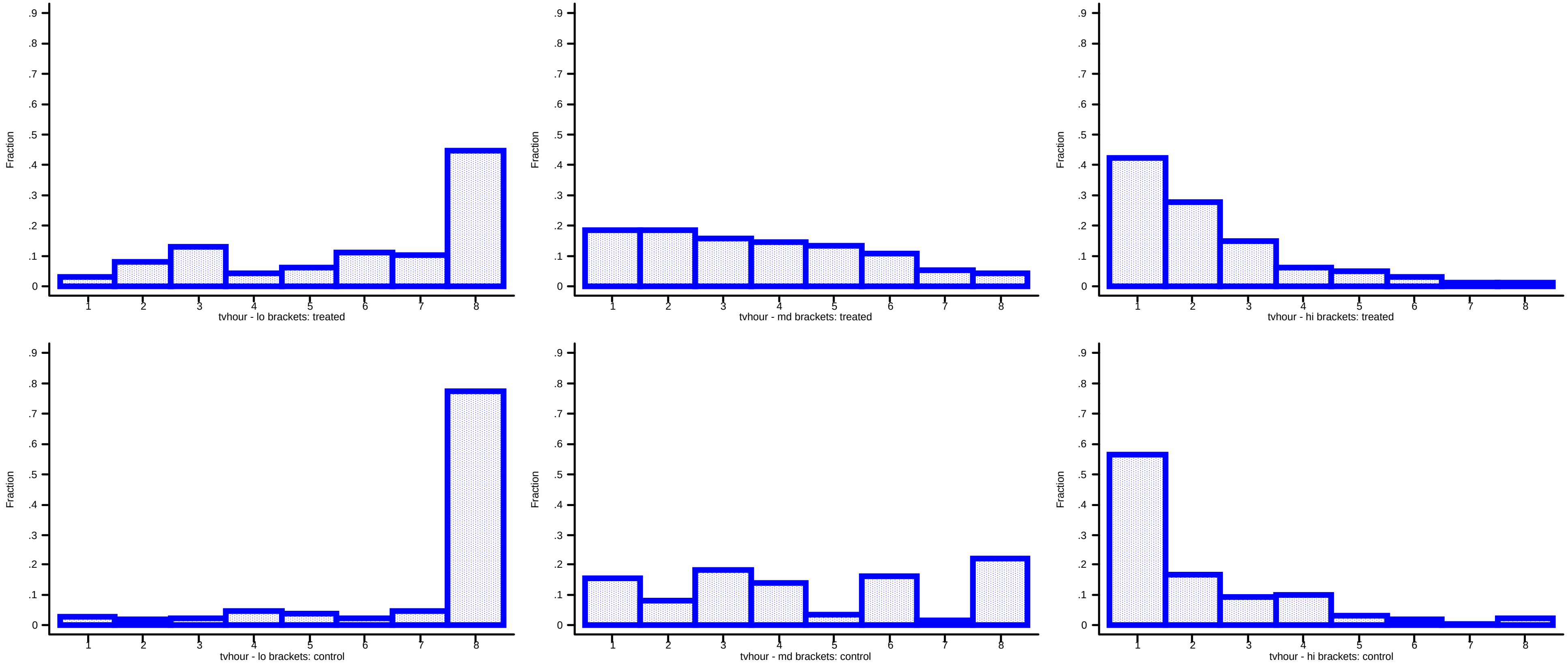


Figure 2: Experiment 1, cumulative distribution functions (time use – TV, last week)

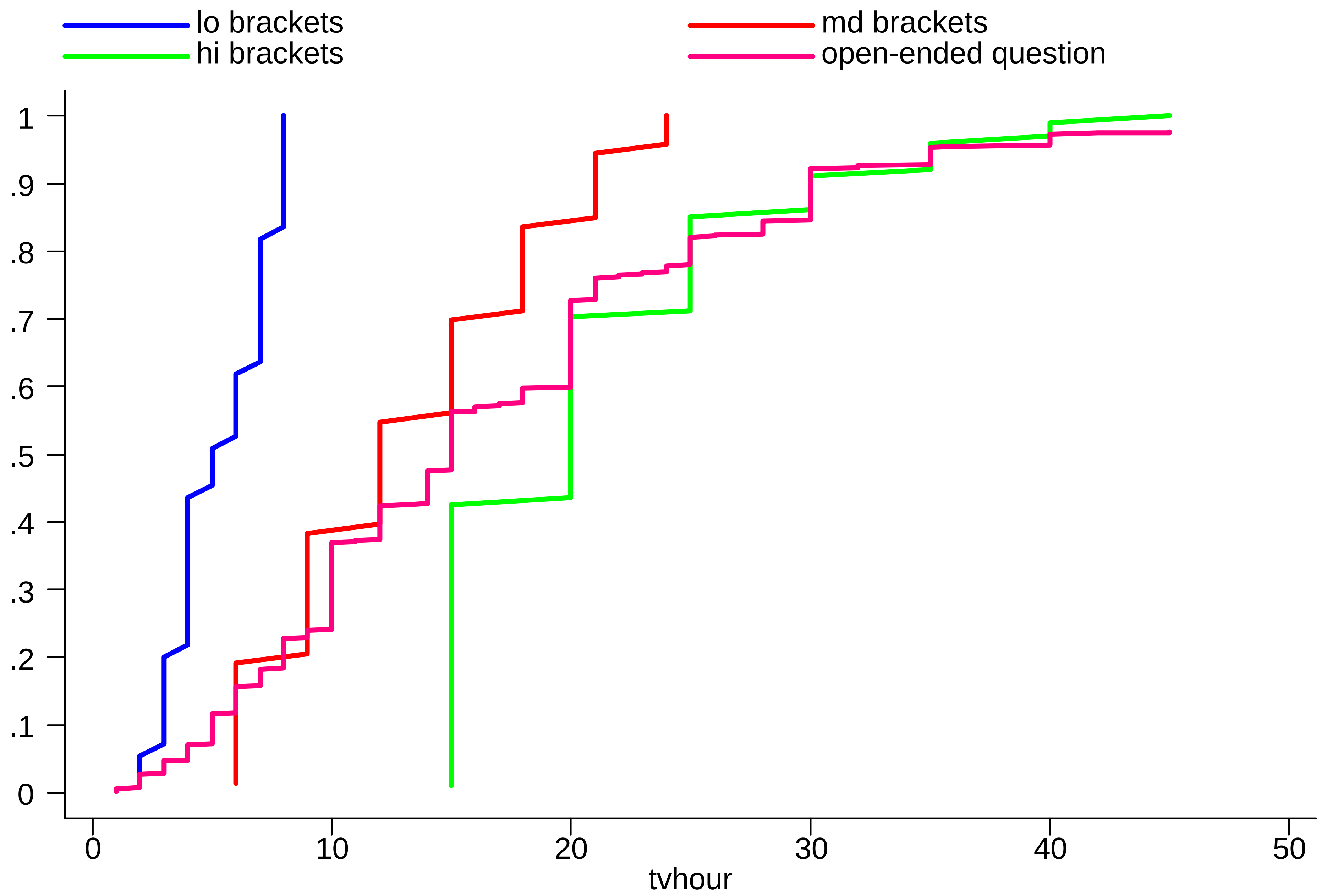


Figure 3: Experiment 1, response distributions (time use – books, last month)

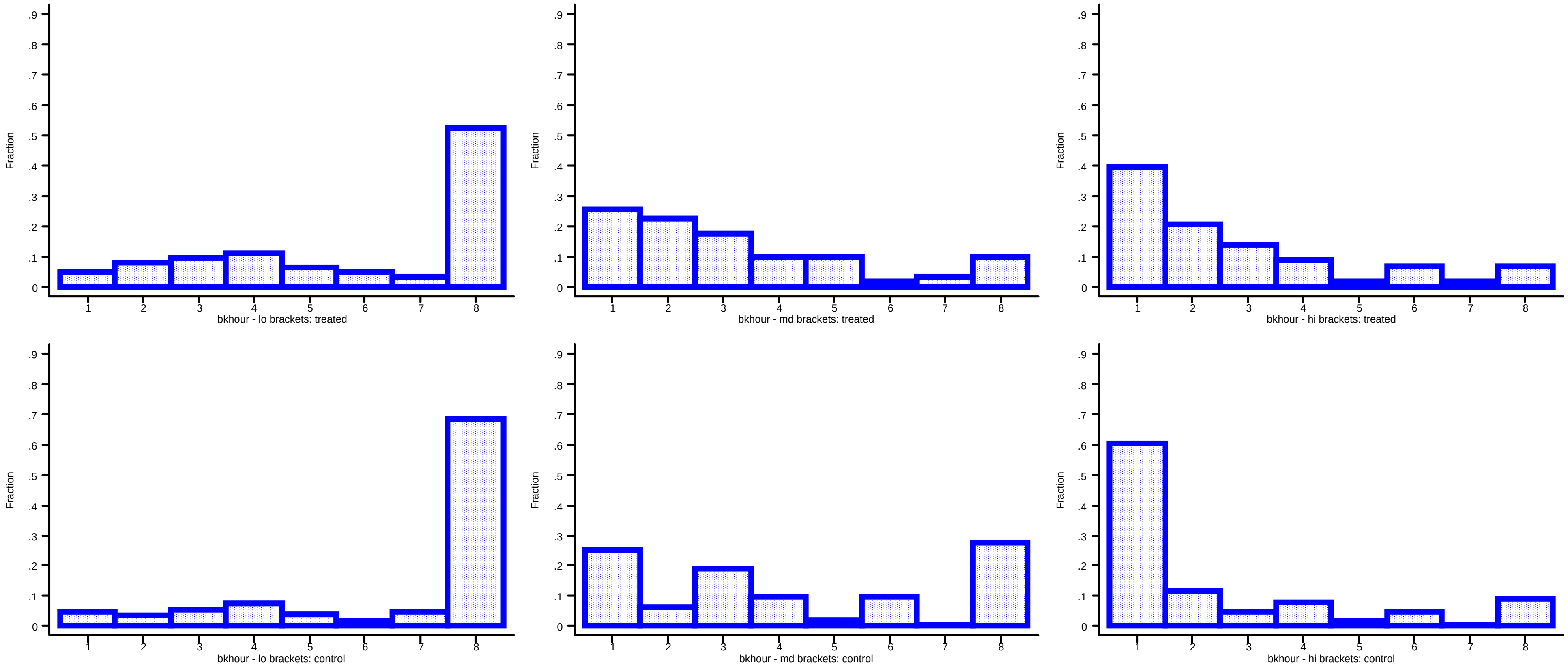


Figure 4: Experiment 1, cumulative distribution functions (time use – books, last month)

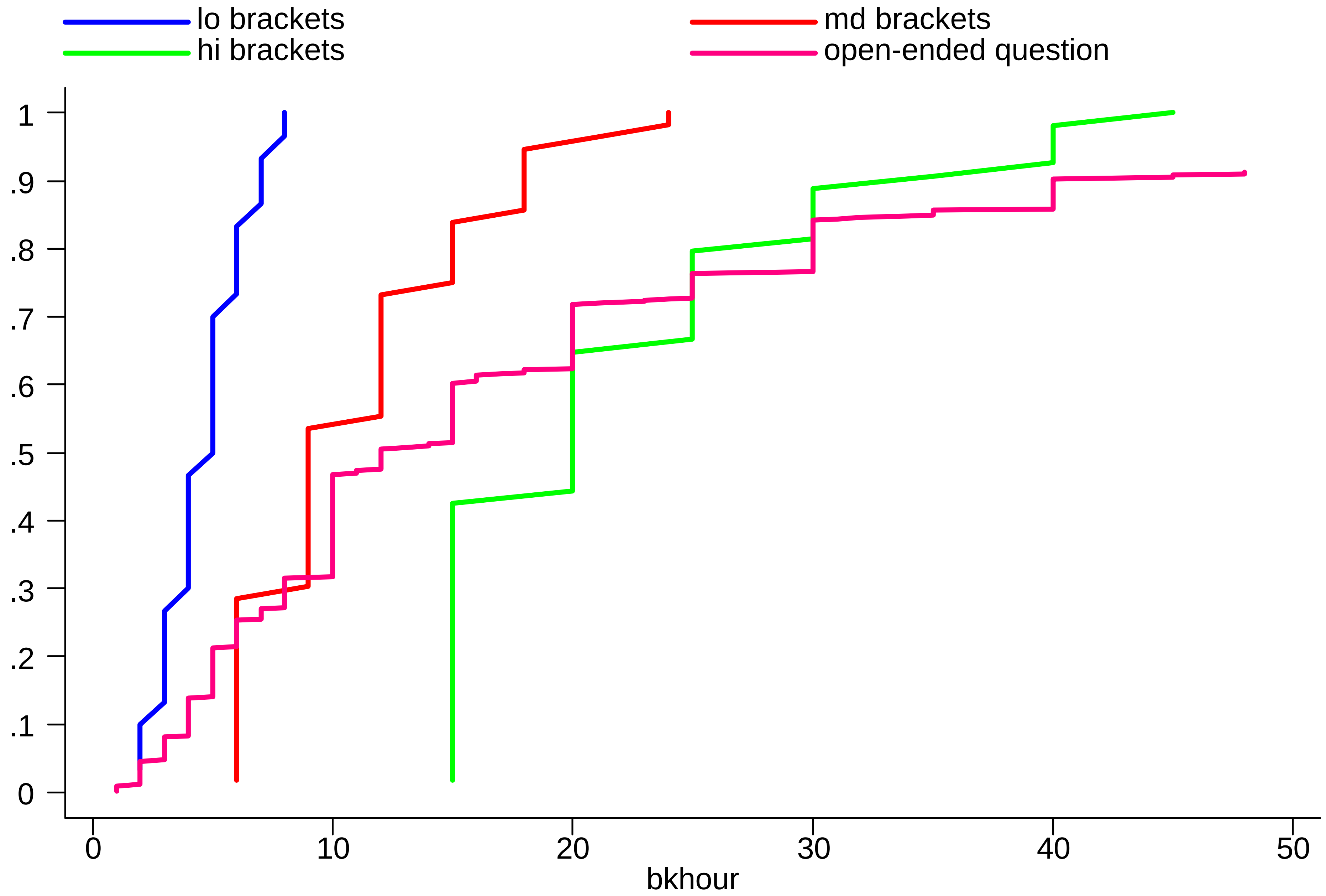


Figure 5: Experiment 1, response distributions (time use – tax form, last year)

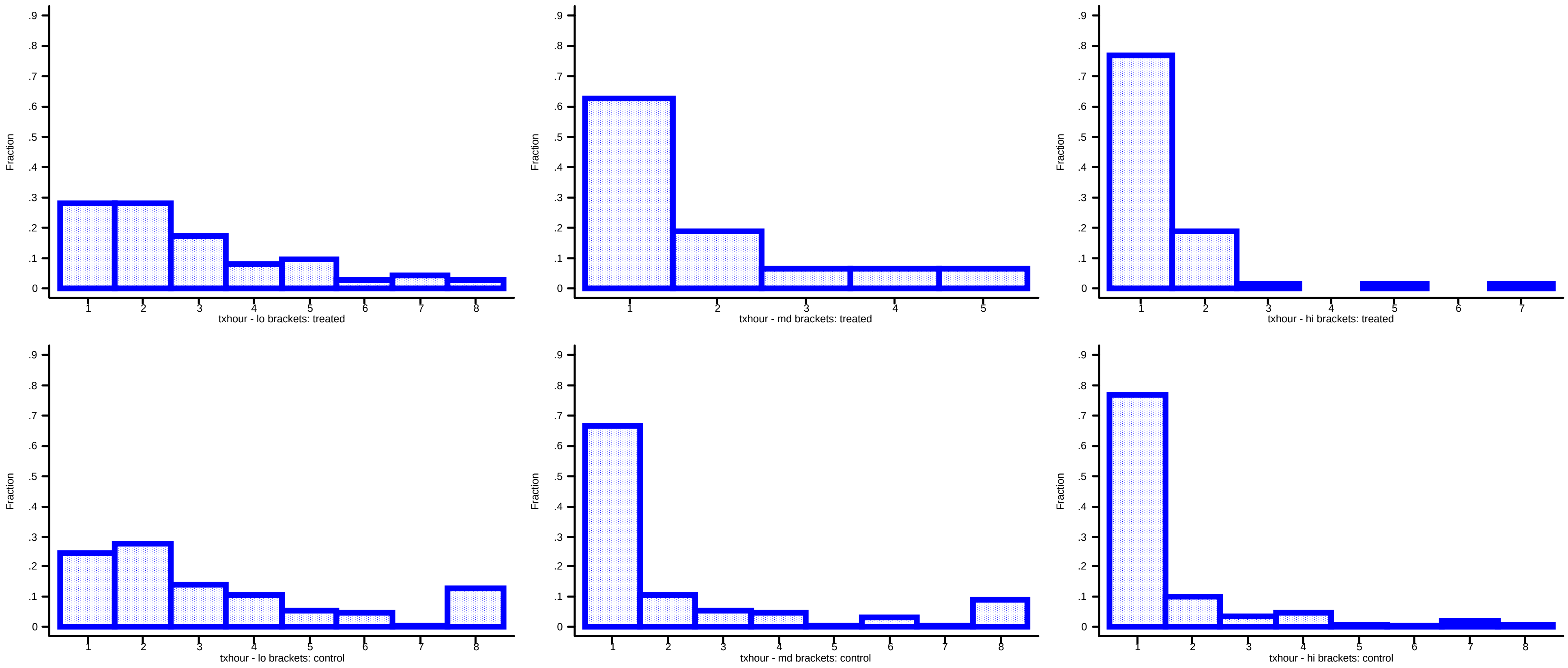


Figure 6: Experiment 1, cumulative distribution functions (time use – tax form, last year)

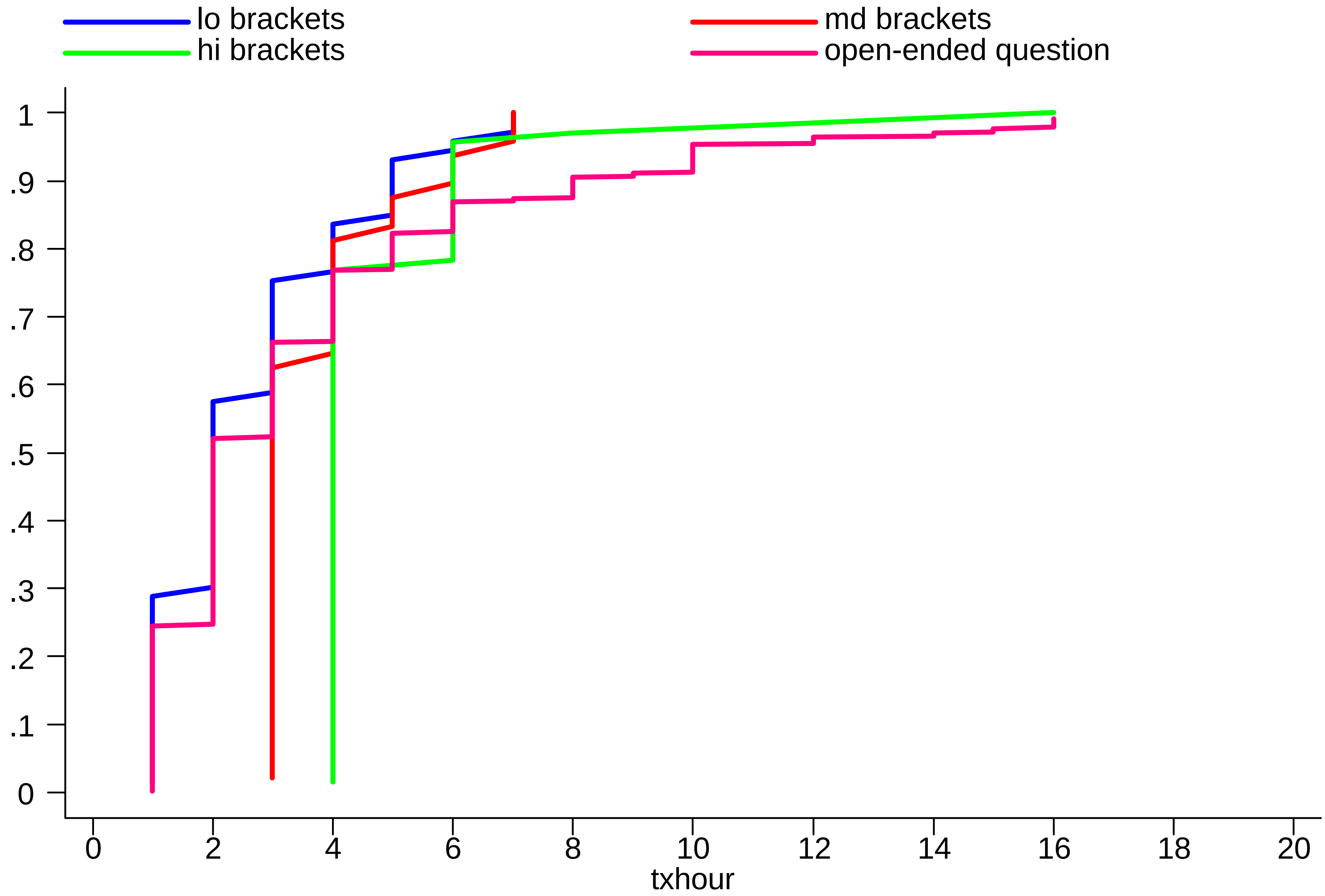


Figure 7: Experiment 1, response distributions (food expenditure, last month)

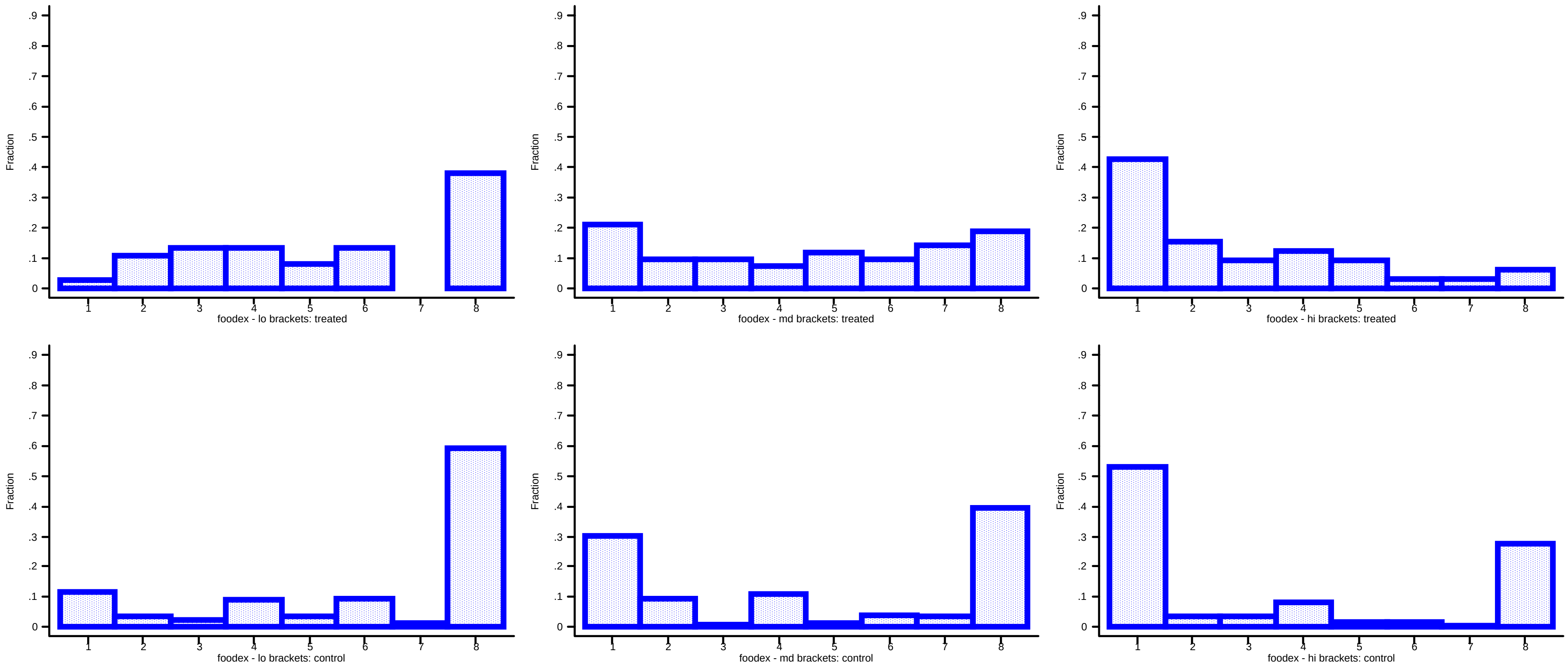


Figure 8: Experiment 1, cumulative distribution functions (food expenditure, last month)

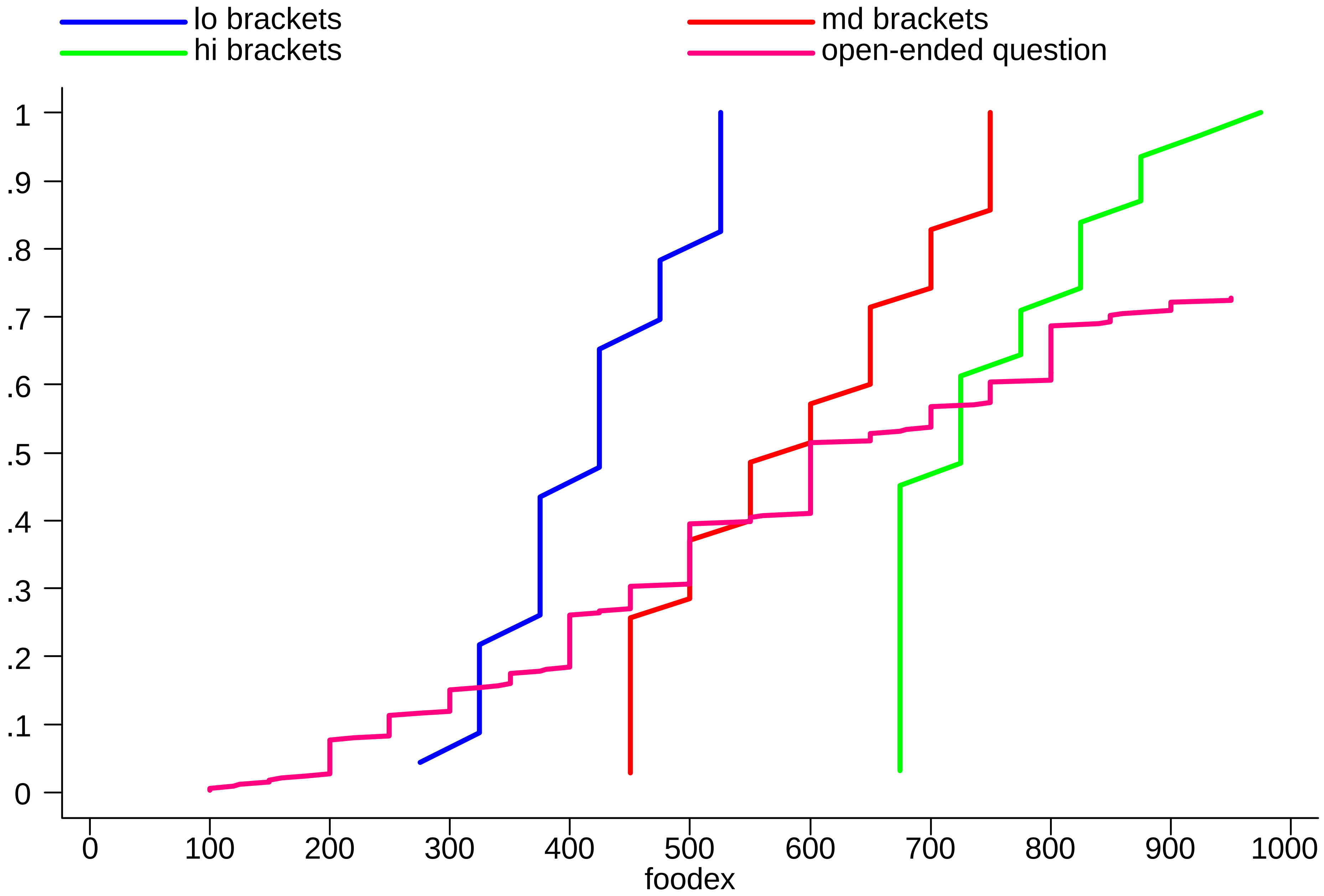


Figure 9: Experiment 1, response distributions (travel expenditure, last year)

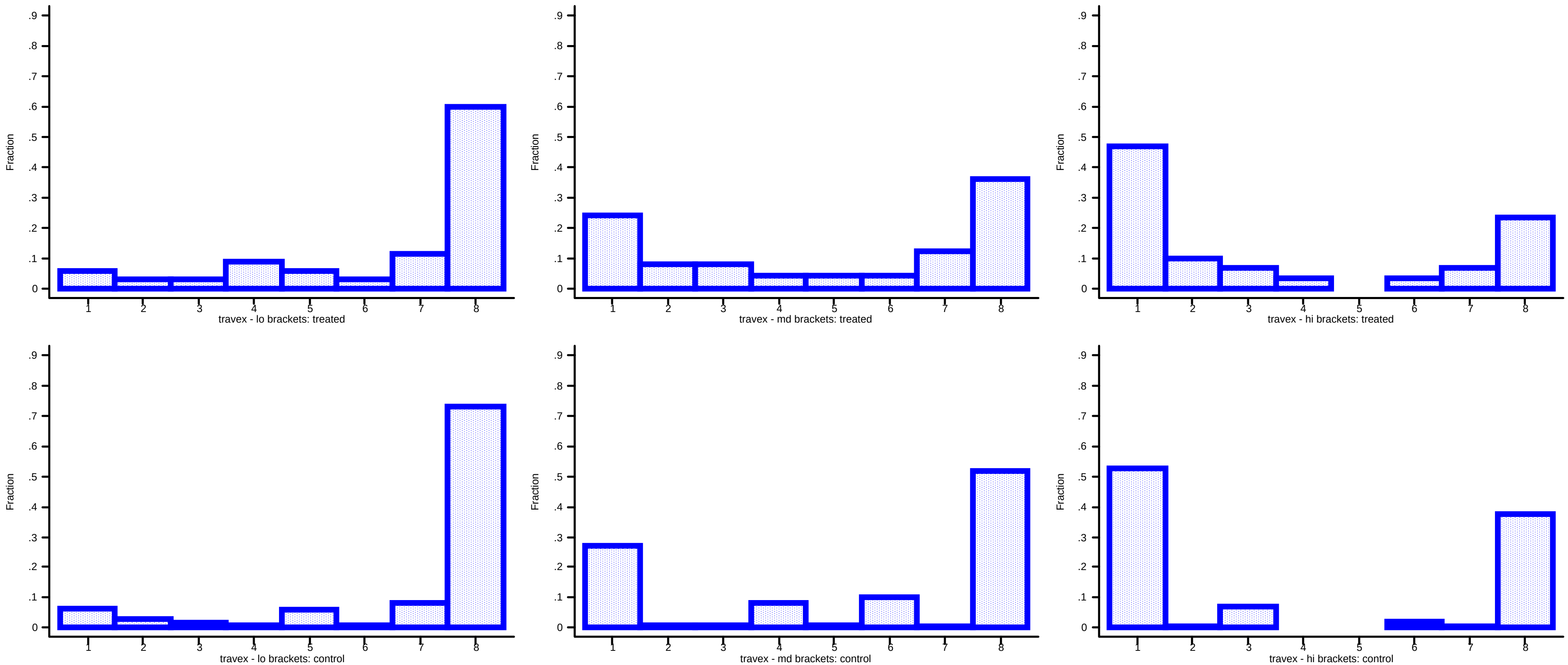


Figure 10: Experiment 1, cumulative distribution functions (travel expenditure, last year)

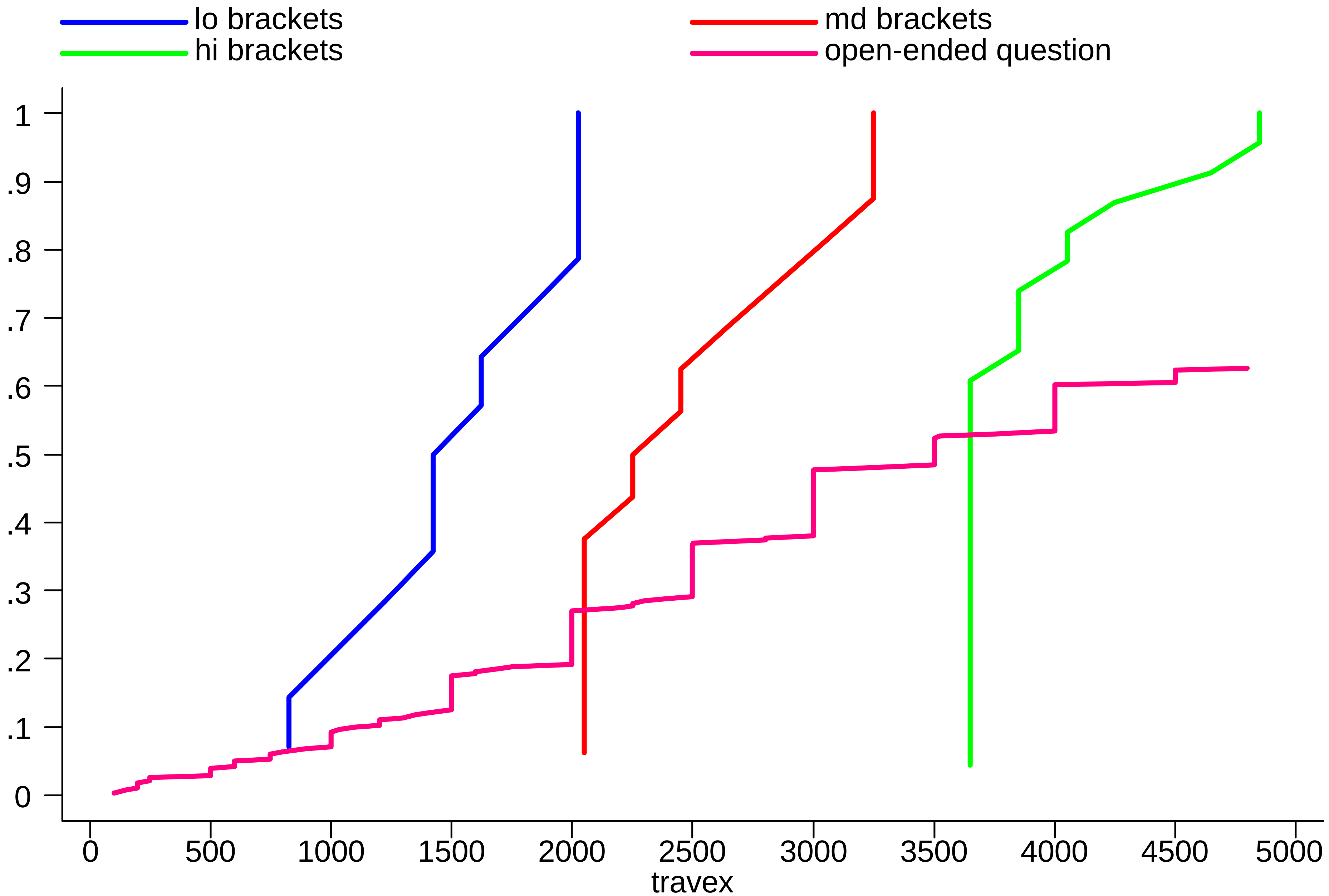


Figure 11: Experiment 1, response distributions (clothing expenditure, last year)

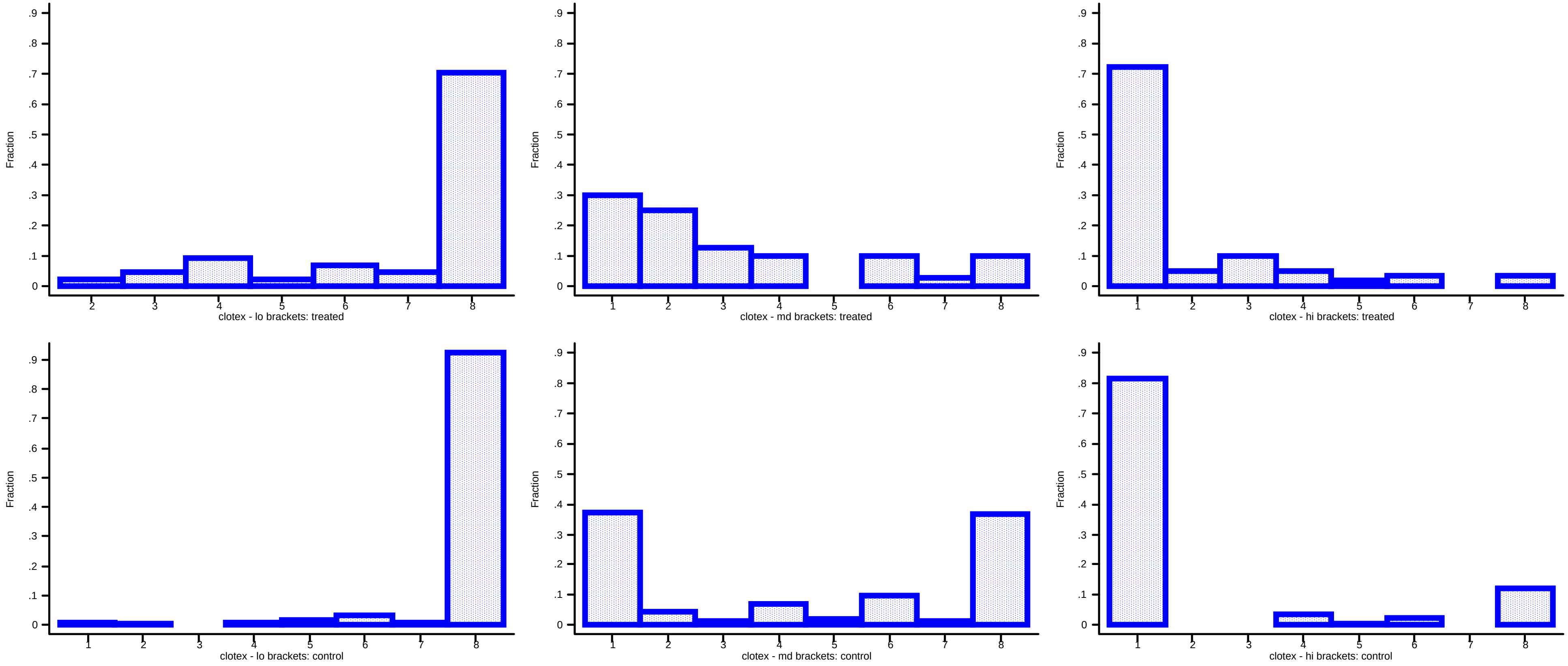


Figure 12: Experiment 1, cumulative distribution functions (clothing expenditure, last year)

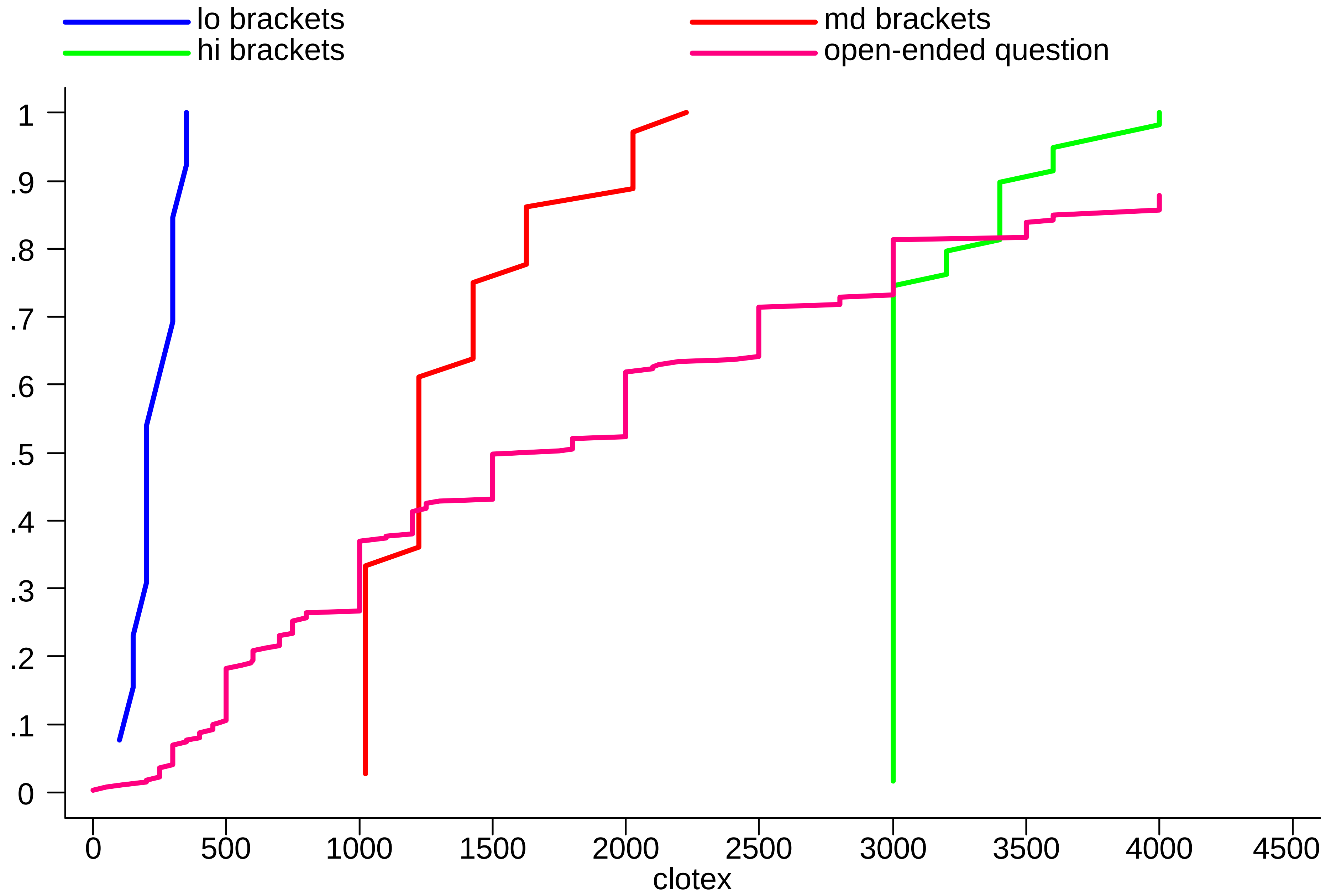


Figure 13: Experiment 1, response distributions (total household non-durables expenditure, last month)

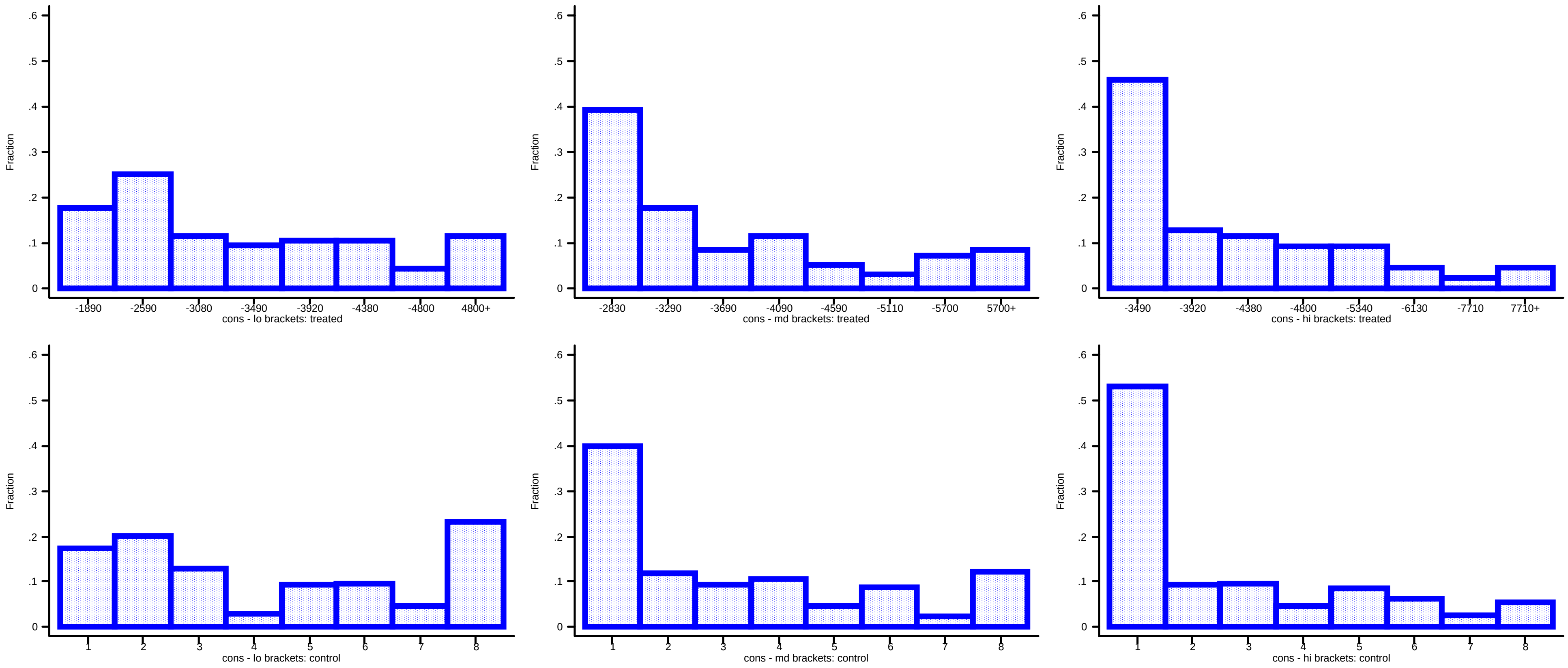
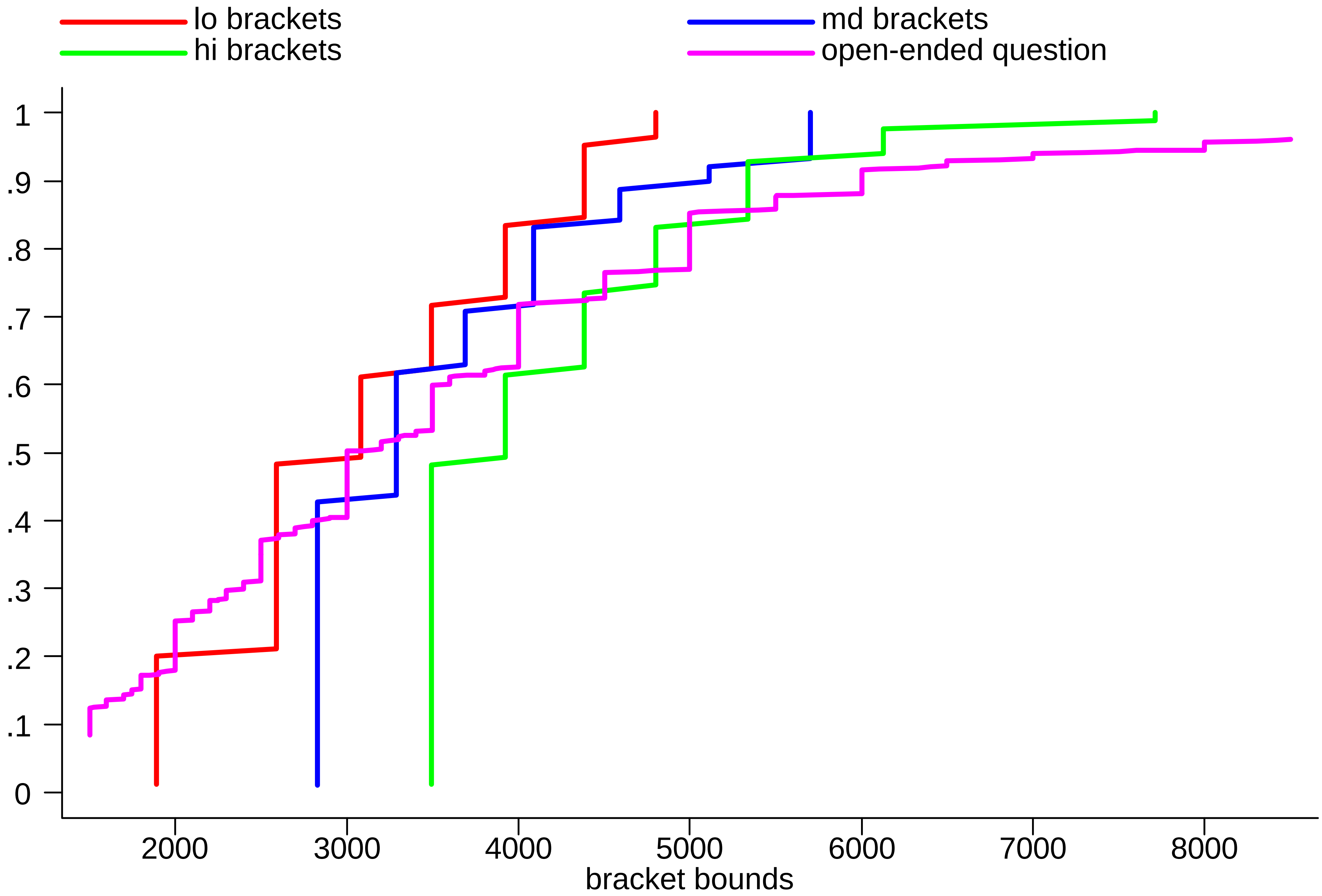


Figure 14: Experiment 2, cumulative distribution functions (total household non-durables expenditure, last month)



Nr.	Author	Title
02-44	Alexander Klos	Die Risikoprämie am deutschen Kapitalmarkt
02-43	Markus Glaser Martin Weber	Momentum and Turnover: Evidence from the German Stock Market
02-42	Mohammed Abdellaoui Frank Voßmann Martin Weber	An Experimental Analysis of Decision Weights in Cumulative Prospect Theory under Uncertainty
02-41	Carlo Kraemer Martin Weber	To buy or not to buy: Why do people buy too much information?
02-40	Nikolaus Beck	Kumulation und Verweildauerabhängigkeit von Regeländerungen
02-39	Eric Igou	The Role of Lay Theories of Affect Progressions in Affective Forecasting
02-38	Eric Igou Herbert Bless	My future emotions versus your future emotions: The self-other effect in affective forecasting
02-37	Stefan Schwarz Dagmar Stahlberg Sabine Sczesny	Denying the foreseeability of an event as a means of self-protection. The impact of self-threatening outcome information on the strength of the hindsight bias
02-36	Susanne Abele Herbert Bless Karl-Martin Ehrhart	Social Information Processing in Strategic Decision Making: Why Timing Matters
02-35	Joachim Winter	Bracketing effects in categorized survey questions and the measurement of economic quantities
02-34	Joachim Winter	Design effects in survey-based measures of household consumption
02-33	Stefan Schwarz Dagmar Stahlberg	Motivational influences on the strength of the hindsight bias
02-32	Stefan Schwarz Dagmar Stahlberg	Strength of hindsight bias as a consequence of meta-cognitions