

SONDERFORSCHUNGSBEREICH 504

Rationalitätskonzepte,
Entscheidungsverhalten und
ökonomische Modellierung

No. 02-34

**Design effects in survey-based measures of
household consumption**

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

Design effects in survey-based measures of household consumption*

Joachim K. Winter**

University of Mannheim

First draft: August 2001

This version: May 2002

Abstract: An important issue in designing survey questions on household consumption is the level of aggregation of expenditure categories. In this paper, I present experimental evidence on how the choice of expenditure categories influences measures of household consumption. The survey experiment, conducted in the Netherlands using the CentERpanel, compares responses to a “one shot” question that asks households to report total expenditures on non-durables with responses to a more disaggregated design that uses 35 expenditure categories. The experiment reveals significant underreporting in the one-shot question. Moreover, underreporting is also present in the sub-categories (which are still relatively broad). This is illustrated by comparing responses to one of the 35 categories, “food and beverages”, to an external measure obtained from a detailed budget survey. The response bias in the measurement of consumption in household surveys is sufficiently large to affect subsequent substantive economic analysis such as the estimation of Engel curves.

Keywords: survey methodology, consumption

JEL classification: C81

* I am grateful to Rob Alessie, Marcel Das, Mike Hurd, Arie Kapteyn, Dan McFadden, Arthur van Soest, Guglielmo Weber, and to seminar participants at the TMR workshop at Evian-les-Bains, August 21-23, 2001, for comments and suggestions at various stages of this project. 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

Important issues in public policy are how consumption levels and patterns change over the life cycle and particularly after retirement, whether such changes are expected or not, and how they are related to financial planning in pre-retirement years. The empirical analysis of these questions, and of many other aspects of household behavior, requires reliable micro data on households' consumption expenditure.

In most developed countries, expenditure data are available at the household level from budget surveys which use detailed expenditure diaries. Examples are the *Consumer Expenditure Survey* (CEX) in the United States, the *Family Expenditure Survey* (FES) in the United Kingdom, and the *Einkommens- und Verbrauchsstichprobe* (EVS) in Germany. There is some evidence that the consumption measures obtained from budget surveys are quite reliable, based on the comparison of aggregated consumption measures with national account data, see Banks and Johnson (1998).

While budget surveys are reliable instruments for measuring household consumption, they usually provide much less complete data on other aspects of household behavior. This limits their usefulness in the analysis of many research questions. For instance, the German EVS consists of repeated cross sections that are administered only every five years, hence no information on the income dynamics faced by individual households is available. The scope for the empirical analysis of life-cycle saving decisions is therefore quite limited, see Börsch-Supan *et al.* (2001) for the kind of restricted analysis that is possible. In contrast to the EVS, the *German Socio-Economic Panel* (GSOEP), a panel study that currently provides up to 17 years of household data, has detailed information on employment histories and income dynamics, but lacks detailed consumption data. In other countries, the situation is similar. For instance, Skinner (1987) discusses the lack of consumption data in the *Panel Study of Income Dynamics* (PSID) in the United States.

Many researchers feel that adding reliable measures of consumption to existing household panel surveys or including such measures in newly designed surveys such as the *Survey of Health, Aging and Retirement in Europe* (SHARE) is important. However, since time and space in general purpose surveys are restricted, obtaining reliable consumption measures is a challenging task. Browning *et al.* (2002) provide a detailed discussion of these problems and of alternative survey designs that have been suggested for obtaining consumption measures in general purpose surveys. The present paper belongs to this line of research: Using survey experiments that were administered in the Dutch CentERpanel in June 2001, I evaluate design effects associated with two alternative designs of questions on households' non-durables expenditure.¹

¹ A detailed description of the CentERpanel is provided in the appendix.

Specifically, this paper investigates the effects of the level of aggregation of expenditure categories in survey questions on responses. The choice of the aggregation level and of the number of categories to be included are important issues in designing survey questions on household consumption. Existing research suggests that using an extensive list of expenditure items yields the most reliable consumption data, see the review of the literature in section 2 below for details. In practice, time and space constraints do not allow for more than a few questions on consumption expenditure, so there arises the well-known trade-off between data quality and survey cost (see Groves, 1989). A practical solution is to rely on a small number of expenditure items, and then to infer total consumption from these data. However, there exists very little systematic evidence on how variations in survey design affect measures of consumption in household surveys.

A reliable methodological approach towards studying design effects in household surveys is the use of controlled experiments. In the experiment reported in this paper, I compare responses to a “one shot” question that asks households to report total expenditures on non-durables with responses to a more disaggregated design that uses 35 expenditure categories. The experiment reveals significant underreporting in the one-shot question. Moreover, underreporting is also present in the sub-categories (which are still relatively broad). This is illustrated by comparing responses to one of the 35 categories, “food and beverages”, to an external measure obtained from a very detailed budget survey based on diaries. The response bias in the measurement of consumption in household surveys is sufficiently large to affect subsequent substantive economic analysis such as the estimation of Engel curves.

The remainder of this paper is structured as follows. In section 2, I briefly review the existing literature on measures of household consumption in household surveys. Section 3 describes the design of the survey experiment. Results are presented in Section 4. The consequences of design effects in applied research are illustrated in section 5. Section 6 draws conclusions.

2 Measurement of consumption in household surveys

Obtaining a measure of household consumption that directly corresponds to abstract models of life-cycle behavior is no easy task.² To get a complete picture of households’ consumption, data on three broad classes of consumption goods and services are required – housing services, consumption flows from durables, and consumption of non-durables. Each of these classes has its own measurement problems. The focus of the present paper is on measuring expenditures for non-durables in household surveys.³

² The following discussion draws on Browning *et al.* (2002).

³ In the sequel, “expenditures” are understood to be expenditures on non-durable consumption goods and services.

There are basically three survey designs that can be used to obtain retrospective data on expenditure; these approaches vary in the aggregation level of the expenditure items and in the number of items asked.⁴

- At one extreme, there is an exhaustive list of expenditure items, similar to the expenditure diaries in budget surveys. Space and time constraints rule this option out in most surveys.
- At the other extreme, there is a single (“one shot”) question on total expenditure.⁵ While such a question is easy to administer, with low cost in terms of valuable survey space, the danger that respondents miss important components of their expenditure is quite high, and underreporting appears to be very likely.
- Between these extremes, there is a third option, using a list of selected expenditure items that can be employed to impute total expenditure.

The subject of the present paper is to analyze response behavior in the second and third options, in particular with respect to underreporting in the open-ended question. Before I present the experimental design and the results, the existing literature on survey designs for measuring household consumption is reviewed in this section.

Skinner (1987) is an early example for the third option (imputing total consumption expenditure from a small list of expenditure items). Skinner’s focus is on the PSID which has detailed information on household income and wealth. Therefore, the PSID is well suited for life-cycle analysis. However, data on household consumption in the PSID are restricted to very few items such as food expenditure, the market value of owned homes, rents, utilities, and the number of automobiles. To impute total consumption of PSID households, Skinner uses detailed and reliable consumption data from the CEX. Just using those CEX consumption variables that are available in the PSID as well, he estimates a linear prediction model of total consumption. Based on the estimated coefficients of that model, total consumption can be predicted for PSID households. Since Skinner’s work, much more sophisticated approaches have been developed for predicting total expenditure across surveys. Such approaches use sophisticated matching procedures adopted from the literature on treatment effects and inverse Engel curves regressions rather than Skinner’s more restricted model; see Ziliak (1998) and Battistin *et al.* (2001) for recent applications and Browning *et al.* (2002) for a review.

The issue of aggregation in consumption questions has received particular attention in development economics. In developing countries, consumption is the main measure used to assess

⁴ There is not necessarily a one-to-one correspondence between aggregation level and number of items since only subsets of the complete list of highly disaggregated items could be used.

⁵ This question format was invented in 1991 by Martin Browning, see Browning *et al.* (2002).

welfare, poverty, and economic development more generally. At the same time, almost no household-level data are available, so that the need to construct survey instruments that are easy to administer yet provide reliable measures of consumption is evident; see Deaton (1997), pp. 26–29, and Lanjouw and Lanjouw (2001). Empirical evidence on aggregation bias in consumption surveys in developing countries is discussed by Joliffe and Scott (1995) and Pradhan (2001). Based on a pseudo-experimental design used in a household survey conducted in Indonesia, Pradhan (2001) compares consumption measures obtained from two different levels of aggregation (23 questions vs. 320 categories). His main results are that (i) a higher level of aggregation yields a lower consumption measure, and (ii) the degree of underestimation is higher for higher levels of total consumption.

For industrialized countries, the most comprehensive study of alternative measures of consumption is Battistin *et al.* (2001) who compare data from two Italian household surveys. The first is the *Survey on Household Income and Wealth* (SHIW), a panel study administered by the Bank of Italy. Like the PSID and the GSOEP, the SHIW provides detailed longitudinal data on income, savings, and wealth, but only broad measures of consumption. One question asks for average monthly consumption expenditure, explicitly excluding only a few durables; the other question asks for monthly expenditure on food. Both questions use open-ended question formats. The other survey Battistin *et al.* use is the *Survey on Family Budgets* (SFB) run by the Italian Institute for Statistics. The SFB is a typical budget survey that uses expenditure diaries with highly disaggregated categories. Battistin *et al.* show that recall consumption data obtained from the broad consumption questions in the SHIW are heavily affected by heaping and rounding problems (which can be taken as an indication of uncertainty about the quantities asked). Food expenditure data are of comparable quality and informational content across the two surveys. For the broad measure of expenditure on non-durables, Battistin *et al.* show that there are major differences across the surveys.

Similar evidence is available for a sample of relatively old individuals from the United States. In a methodological study of a particular form of survey response bias which is not of interest for the present discussion (namely, anchoring effects in unfolding brackets sequences), Hurd *et al.* (1998) compare responses to a one-shot question on consumption expenditure administered in 1995 in an experimental module that was part of the second wave of the *Asset and Health Dynamics of the Oldest Old* (AHEAD) survey with data from the CEX. While this is not the main focus of their paper, Hurd *et al.* report that the consumption measure obtained from the one-shot question in AHEAD is significantly downward biased relative to CEX data which are based on detailed expenditure diaries.

Taken together, these studies suggest that measures of household consumption that are based on just one broad question are problematic – they are typically subject to substantial downward bias. This result does not necessarily imply that data from one-shot questions are useless.

For instance, biased responses to a one-shot question might still have valid covariance with demographic variables and could therefore be of some use in regression models. I come back to circumstances under which data from a one-shot question might be useful in the concluding section 6.

Summarizing their discussion of the literature, Browning *et al.* (2002) state that it is “unlikely that we will ever be able to design a set of questions that is succinct enough to be included in most surveys and also comprehensive enough to meet most research needs.” At the same time, they recommend that extensive pre-testing is recommended for non-standard questions, and the elicitation of expenditure measures with small numbers of questions or even a one-shot question definitely belongs to the class of non-standard questions in household surveys. The following analysis is an example of how experimental surveys can be used in the process of finding better designs for survey questions.

3 Design of the survey experiment

The experiment reported in this paper was administered over the Dutch CentERpanel in June 2001.⁶ The experiment compares an extreme option in designing a survey on household consumption – a one-shot question that asks for total household non-durable expenditure in the previous month – with a more disaggregated design. Accordingly, the experiment has two treatments which were randomly assigned to CentERpanel households. 1117 households received the one-shot question, 116 households received the disaggregated question. The difference in the number of households in the treatment groups results from the overall experimental design. The one-shot question was also used as a control group in other experiments conducted in the same survey, hence this group was sampled with a larger target sample size. One group of households received a one-shot open-ended question, taken *verbatim* from an experimental module for the *Health and Retirement Study* (HRS). The question reads 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?

⁶ Appendix 6 provides a detailed description of the CentERpanel.

The other group received a somewhat more disaggregated design that comprises 35 different expenditure categories; a detailed list of these categories can be found in Table 1.⁷ It should be noted that these categories are still much broader than the hundreds of categories contained in expenditure surveys.

In both the one-shot question and the disaggregate questions, it was made explicit that they referred to the previous month (i. e., May 2001). Asking retrospective expenditure questions for short, recent periods (as opposed to, say, the previous year) should make it easier for respondents to come up with an estimate, as pointed out by Browning *et al.* (2002). All questions – the one-shot question and all 35 questions on expenditure items – were open-ended. In the questions on the 35 expenditure items, respondents were not given “don’t know” or “refuse to say” options since they should be induced to provide their best estimates. In the one-shot question, this approach was less likely to work, hence the “don’t know” option was provided for the open-ended question.

Since the purpose of this experiment was to compare alternative measures of non-durables expenditure (as reflected by the choice of the 35 categories), it was important to insure that households did not include durables expenditure in their response to the one-shot question. Therefore, all households were asked about expenditures in six categories of durable consumption goods before entering the experimental module. First, household were asked whether they had expenditures on each of the six durables in the previous month, and if so, they were asked to state the exact amount in that category. These categories were as follows: automobile, refrigerator, washing machine and/or dryer, dishwasher, television, personal computer.

For comparison purposes, the analysis in the following sections uses data on household expenditure from the Dutch Budget Survey, a detailed diary-based expenditure survey that should provide reliable expenditure data. The latest data from that survey available when this paper was written are from 1998. All values are converted to 2001 prices using the consumer price index. Also, these data are converted from annual to monthly values since the experiments focused on monthly expenditures. This raises the issue of seasonal fluctuations in the composition of household expenditure which might distort comparisons with the experimental data that were collected for one specific month. In summary, these evaluation data might not be very accurate (I return to this issue below). Most of the subsequent analysis is based on comparison of the experimental treatment groups and is not affected by the quality of the control data.

⁷ The choice of these categories followed roughly a design used in exploratory work related to an upcoming revision of the HRS (the so-called 2001 HRS mail-out), made available by Michael Hurd in private communication.

4 Results

Table 1 contains descriptive statistics for the 35 disaggregated consumption categories. Respondents received open-ended questions and were not given “don’t know” or “refuse to say” options, so for each category, there are 116 observations. Two results are worth noting. First, for all categories, there are some households that report zero expenditure. This is even the case for expenditure categories such as “food and beverages” (five households), which might be surprising but not entirely implausible. More strikingly, there are three households with a total non-durables expenditure (obtained by summing the responses to the 35 expenditure items) of zero. This is an indication that zero responses might also reflect uncertainty about responses.⁸ Second, almost all median values are multiples of 10 or 100 guilders (if not zero). This reflects the fact that in these open-ended consumption questions, households have a very strong tendency to report focal values, so the medians tend to be focal values as well. The role of focal values is discussed below.

In the one-shot question on total monthly non-durables, the “don’t know” option was chosen by 364 (32.6%) of the 1117 respondents. This non-response rate is somewhat smaller than the non-response rate of 35.8% reported by Hurd *et al.* (1998) for a very similar one-shot question on total expenditure that was administered in an experimental module of the AHEAD survey (see section 2). Non-response rates of about one third raise the question of whether the remaining responses are biased due to selection effects. To check for selection effects, I use a probit regression with a binary indicator of non-response as the dependent variable. Explanatory variables were the available personal and household characteristics of the subject.⁹ Results are reported in table 2.¹⁰

Rather surprisingly, the demographic and household characteristics are jointly insignificant ($p = 0.441$ for the likelihood ratio test). This finding is in line with similar regressions of non-response indicators in other survey experiments administered over the CentERpanel (Winter (2002)). In contrast, Hurd *et al.* (1998) find significant effects of demographic and household characteristics in a similar non-response regression for the AHEAD one-shot expenditure question. This difference could have several reasons. For instance, members of the CentERpanel answer surveys on a regular basis which might change their response behavior

⁸ Based on previous experience with the CentERpanel, it is unlikely that consistent zero reports are the results of respondents’ unwillingness to answer such surveys.

⁹ 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.

¹⁰ The number of observations in the non-response regression is slightly smaller than the number of subjects in the treatment group because of missing covariates.

(in the sense of making some groups who are, in other surveys, typical non-respondents as likely to respond as the other panel members). Also, members of the AHEAD sample are of course much older than those of the CentERpanel. In any case, the fact that respondents' demographic and household characteristics are jointly insignificant justifies the assumption of random non-response used in the subsequent analysis – i. e., the responses to the one-shot question are used without further adjustment for selection effects.

Table 3 contains descriptive statistics for the responses to the one-shot expenditure question. Note that the minimum of reported total monthly non-durables expenditure was 0 guilders (6 observations) which is implausible and suggests that these households either misunderstood the question entirely or deliberately gave an incorrect answer. The next smallest responses were 300, 420, and 500 guilders, which seems more plausible. Next, note the high proportion of focal point responses. The presence of focal points implies that the responses are subject to error from heaping. Substantively, 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.

Table 4 compares descriptive statistics for alternative measures of household consumption. The one-shot open-ended question yields significantly lower estimates of consumption than the disaggregated question with 35 categories. The data from the 1998 expenditure survey, which are interpreted as exact measures of household consumption, suggest that the difference between the two measures is due to underreporting in the one-shot question rather than to overreporting in the disaggregated question.¹¹ Overall, the underreporting ratio (the consumption measure based on the open-ended question as a fraction of the measure based on the 35 disaggregated categories) is about 85%, evaluated at the sample means.

As can be seen from the non-parametric estimates¹² of the distribution functions of the three consumption measures in figure 1, both disaggregated measures (budget survey and experiment) indicate that consumption is more dispersed than a measure based on a one-shot question would suggest.

At the household level, the degree of underreporting cannot be assessed directly since every household answered only one of the two alternative question designs. A heuristic way to deal with this problem is to match households by other variables and to compare measures within these matched groups. Figure 2 contains an illustration of this approach. In constructing this

¹¹ As noted before, the 1998 expenditure data have been adjusted for inflation and converted into monthly values, so the “true” consumption measure is potentially subject to some error itself.

¹² These non-parametric estimates were obtained with a kernel density estimation method, using a Gaussian kernel and optimal bandwidth selection.

figure, households are matched based on monthly net household income, and the underreporting ratio is evaluated at the means within income quintiles. Underreporting is high for the middle income groups and decreases with income. This finding is similar to results reported by Pradhan (2001).

A similar analysis for age can be found in figure 3. Again, each point corresponds to the mean underreporting ratio within an age quintile. The age effect is rather interesting: Underreporting appears to be most severe for middle-aged respondents. For the second and fifth age quintiles, the mean underreporting ratio is very close to one. The latter effect could be explained by the fact that older households have non-durables expenditures that are concentrated on few items, and they are therefore easier to recall.

It would be extremely interesting to explore the effects of respondents' demographic and household characteristics further. However, the scope for such an analysis is limited, with only about 100 subjects in one of the treatment groups.

Aggregation bias might also be present at lower levels of aggregation. Since this experiment uses only one level of disaggregated questions, aggregation bias at lower levels can only be recovered by comparing a specific category with its counterpart in the budget survey. For illustration, we focus on food and beverage consumption, a category that receives special attention in applied studies of household behavior and welfare. Again, seasonal effects and imperfect adjustment for inflation reduce the power of such a comparison. However, if one is willing to take the number constructed from the 1998 budget survey as a measure of true food consumption, the differences in means and medians are significant. Results are reported in table 5. The shapes of the distribution functions of the food expenditure measures obtained from one experimental survey question and multiple sub-categories in the budget survey, respectively, are also different (figure 4). The distribution functions once again reveal that the data from the highly disaggregated budget survey are more dispersed.

The last result indicates that even responses to a more disaggregated design (in this case, using 35 categories) might be subject to aggregation bias since each of these 35 items could still be underestimated.

5 Implications for applied demand analysis

To illustrate the effects of measurement problems in consumption questions, I use the experimental data to estimate an Engel curve for the food and beverages expenditure share. This Engel curve can then be compared with an Engel curve estimated on the total expenditure and food and beverages expenditure data obtained from the 1998 budget survey. The rationale for this approach is that the problem of inflation adjustment is much less severe in this analysis than in the direct comparisons of descriptive statistics presented in the previous section.

At the same time, the results of such an exercise illustrate potential effects of response bias originating from question design on substantive analysis.

Figures 5 and 6 show Engel curves for the share of food and beverage expenditure in total non-durables expenditure, estimated using detailed budget survey data and one-shot experimental survey data, respectively. The figures contain both a linear model and a nonparametric (kernel regression) estimate. As is found in many studies (a recent example is Battistin *et al.*, 2001), the nonparametric Engel curves are almost linear. In addition to the linear model, I also estimated polynomial Engel curve specifications; linearity cannot be rejected in these parametric models. In the linear model, the slope coefficients obtained from the experimental and budget survey data are significantly different from each other. This fact is another indication that the source of consumption data can affect substantive analysis.

While the approach of comparing expenditure data obtained from two different sources using Engel curves is interesting, in this particular instance it is plagued by an unexpected problem. It appears that the food and beverages and total expenditure data obtained from the Dutch Budget Survey might be subject to systematic error – the Engel curve for the food share estimated on budget survey data appears to be implausibly flat. The source of this problem warrants further investigation.¹³

The result that not only the one-shot question on total expenditure is subject to underreporting, but also the experimental survey question on food expenditure (the latter relative to budget survey data), should be interpreted with caution as long as the 1998 budget data have not been validated. This caveat does not apply to the direct experimental comparison of the two total expenditure measures obtained from the one-shot question and a list of 35 expenditure items, respectively. In this comparison, significant underreporting is found, and this result does not depend on the quality of the budget survey control data.

6 Conclusions

In this paper, I presented experimental evidence on how the choice of expenditure categories influences measures of household consumption. Comparing responses to a one shot question on total monthly non-durables expenditure with responses to a more disaggregated design based on 35 expenditure categories reveals significant underreporting. The same holds true when responses to a one-shot question on one of these 35 still relatively broadly defined categories (namely, “food and beverage” expenditure) are compared with measures from a budget survey

¹³ After this paper was written, I was informed by Rob Alessie in personal communication that the 1998 wave of the Dutch Budget Survey was reported to have some data problems that might affect substantive analysis. Future revisions of this paper will replace the 1998 wave with more reliable waves.

that is based on a very detailed expenditure diary, with the caveat that there are some doubts about the quality of these budget survey data.

It might not come as a surprise that aggregation bias arises in responses to a one-shot question, relative to measures obtained from a very detailed budget diary with many hundreds of items. More interestingly, the results presented in this paper suggest that this measurement problem can be significantly reduced by using relatively few categories – in this case, 35 – even though underreporting is still likely to be present. The distribution of the total expenditure variable constructed from 35 expenditure categories shows similar shape and dispersion as total expenditure in the budget survey. In practice, including a list of 35 items might, however, be difficult in general purpose household surveys.

From a practical perspective, it appears to be impossible to obtain reliable measures of total consumption in a household survey if space restrictions allow questions only on a few expenditure items. It seems that rather than trying to obtain measures of total consumption, attention should be restricted to a few expenditure items that are of particular interest, such as health care expenditure in surveys like the HRS that focus on older households.

Moreover, results ranging from Skinner (1987) to Ziliak (1998) and Battistin *et al.* (2001) suggest that it is possible to obtain measures of total expenditure from a small number of key expenditure items measured in a general purpose household survey, provided that auxiliary data from a detailed expenditure survey are available. Browning *et al.* (2002) argue that “asking for just three or four sub-items of expenditure recovers a reasonable amount of information to impute non-durable consumption accurately.” They recommend asking for food expenditure at home, food expenditure outside home, and utility and communication expenditure, based on their analysis of the predictive power of these variables for total expenditure.

What do the results reported in this paper, confirming underreporting of on average about 15%, mean for the one-shot question on total non-durables expenditure? Browning *et al.* (2002) suggest that such a question should still be asked, despite underreporting. To the extent that underreporting is stable, at least conditional on covariates, such data would still be useful. However, a method for optimally combining information from a one-shot question and a small number of expenditure items has not yet been found. This should be the subject of future research. Controlled survey experiments such as the one presented in this paper (but with somewhat larger sample size) will be very helpful in this process.

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.¹⁴

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>.

¹⁴ 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

- Banks, J. and P. Johnson (1998):** *How Reliable is the Family Expenditure Survey?* London, UK: Institute for Fiscal Studies.
- Battistin, E., R. Miniaci, and G. Weber (2001):** What do we learn from recall consumption data? Unpublished manuscript, Padua University.
- Börsch-Supan, A., A. Reil-Held, R. Rodepeter, R. Schnabel, and J. K. Winter (2001):** The German savings puzzle. *Research in Economics*, 55(1), 15–38.
- Browning, M., T. F. Crossley, and G. Weber (2002):** Asking consumption questions in general purpose surveys. Unpublished manuscript.
- Deaton, A. (1997):** *The Analysis of Household Surveys*. Baltimore, MD and London, UK: Johns Hopkins University Press.
- Groves, R. M. (1989):** *Survey Errors and Survey Costs*. New York, NY: Wiley.
- 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.
- 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.
- Joliffe, D. and K. Scott (1995):** The sensitivity of measures of household consumption to survey design: Results from an experiment in El Salvador. Unpublished manuscript, Policy Research Department, The World Bank, Washington, DC.
- Lanjouw, J. O. and P. Lanjouw (2001):** How to compare apples and oranges: Poverty measurement based on different definitions of consumption. *Review of Income and Wealth*, 47(1), 25–42.
- Philipson, T. (2001):** Data markets, missing data, and incentive pay. *Econometrica*, 69(4), 1099–1111.
- Pradhan, M. (2001):** Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia. Unpublished manuscript, Economic and Social Institute, Free University, Amsterdam.
- Skinner, J. (1987):** A superior measure of consumption from the Panel Study of Income Dynamics. *Economics Letters*, 23, 213–216.

Winter, J. K. (2002): Bracketing effects in categorized survey questions and the measurement of economic quantities. Discussion Paper No. 02-35, Sonderforschungsbereich 504, University of Mannheim.

Ziliak, J. P. (1998): Does the choice of consumption measure matter? An application to the permanent income hypothesis. *Journal of Monetary Economics*, 41, 201–216.

Table 1: List of items used to elicit total monthly non-durables expenditure

consumption categories, last month	varname	zero reports (%)	mean	st. dev.	median
housing mortgages	house1	40.5	558.7	601.4	480.0
homeowners insurance	house2	63.8	43.4	185.2	0.0
property taxes	house3	46.6	93.8	166.1	32.5
rent	house4	68.1	279.0	815.2	0.0
electricity, water, fuel	energy1	6.9	324.9	915.6	250.0
energy	energy2	5.2	144.3	97.7	120.0
housekeeping supplies	homesup1	6.9	65.9	160.0	30.0
garden and lawn supplies	homesup2	43.1	59.0	98.4	20.0
home repair and maintenance	homesup3	68.1	421.7	2196.4	0.0
domestic services	homesup4	87.1	14.6	57.4	0.0
food and beverages	food1	4.3	569.4	461.4	500.0
eating and drinking out	food2	14.7	148.3	150.3	100.0
clothing and apparel	food3	18.1	238.9	223.2	200.0
personal care products	food4	12.1	52.3	47.5	50.0
personal care services	food5	99.1	2.8	29.7	0.0
vehicle finance charges	vehicle1	35.3	66.6	108.6	43.5
gasoline	vehicle2	25.0	162.7	146.2	150.0
vehicle maintenance	vehicle3	69.0	113.2	373.1	0.0
vehicle insurance	vehicle4	39.7	96.3	222.8	42.5
prescription and non-prescription drugs	health1	56.9	34.8	141.1	0.0
health care services	health2	81.0	30.4	91.2	0.0
medical supplies	health3	93.1	5.8	33.2	0.0
health insurance	health4	20.7	166.2	191.3	81.0
trips and vacations	entert1	39.7	482.6	988.8	100.0
tickets	entert2	76.7	23.5	65.1	0.0
membership to health / social clubs	entert3	60.3	20.8	34.8	0.0
video and audio entertainment	entert4	67.2	26.1	70.2	0.0
hobbies	entert5	58.6	51.8	161.9	0.0
computer equipment	entert6	80.2	14.7	41.4	0.0
reading	entert7	19.8	53.2	59.1	40.0
pet food, products and expenses	entert8	61.2	24.6	55.5	0.0
tobacco products	entert9	64.7	36.0	73.0	0.0
contributions	contri1	58.6	22.5	34.3	0.0
gifts	contri2	71.6	30.7	88.4	0.0
life and liability insurance	insur	37.1	79.7	127.7	20.0
total consumption (35 categories)		2.6	4559.2	3329.1	3995.5

Source: Experiments conducted as part of the CentER Panel, June 2001.

Notes: Respondents received open-ended questions and were not given “don’t know” or “refuse to say” options.

Table 2: Non-response regression for the one-shot question on total non-durables expenditure

Variable	Coefficient	Standard error	p-value
Age	-0.005	0.018	0.789
Age squared	0.000	0.000	0.726
Female (D)	-0.014	0.107	0.895
Low secondary education (D)	0.262	0.151	0.084
High secondary education (D)	0.129	0.127	0.308
High education (D)	-0.012	0.148	0.936
Root of household size	-0.119	0.112	0.285
Housekeeper (D)	0.266	0.163	0.102
Retired (D)	0.025	0.173	0.887
Household head (D)	0.132	0.131	0.316
Log net household income	0.040	0.044	0.372
Constant	0.175	0.645	0.787
Number of observations	1014		
Log likelihood	-629.0		
Pseudo R^2	0.009		
LR test statistic	11.03		0.441

Source: Experiments conducted as part of the CentER Panel, June 2001.

Notes: Probit regression of a binary indicator for non-response to an open-ended, one-shot question on total monthly non-durables expenditure. Explanatory variables are for the respondents or his/her household. “D” denotes a dummy variable. The reference category for education is “primary education”.

Table 3: Responses to the one-shot question on total monthly non-durables expenditure

	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 4: Responses to alternative questions on total monthly non-durables expenditure

Total non-durables consumption, last month	N	mean	st. dev.	median
Expenditure survey (adjusted)	1989	4828.5	2252.2	4555.6
(1) open-ended one-shot question	753	3890.6	3741.9	3000.0
(2) aggregated from 35 open-ended consumption categories	116	4559.2	3329.1	3995.5
Measure (1) as percent of measure (2)		85.3%		75.1%

Source: Experiments conducted as part of the CentER Panel, June 2001, and 1998 expenditure survey.

Notes: The 1998 expenditure survey reports annual values. The measure reported here is converted to monthly values and adjusted for inflation using the Dutch consumer price index.

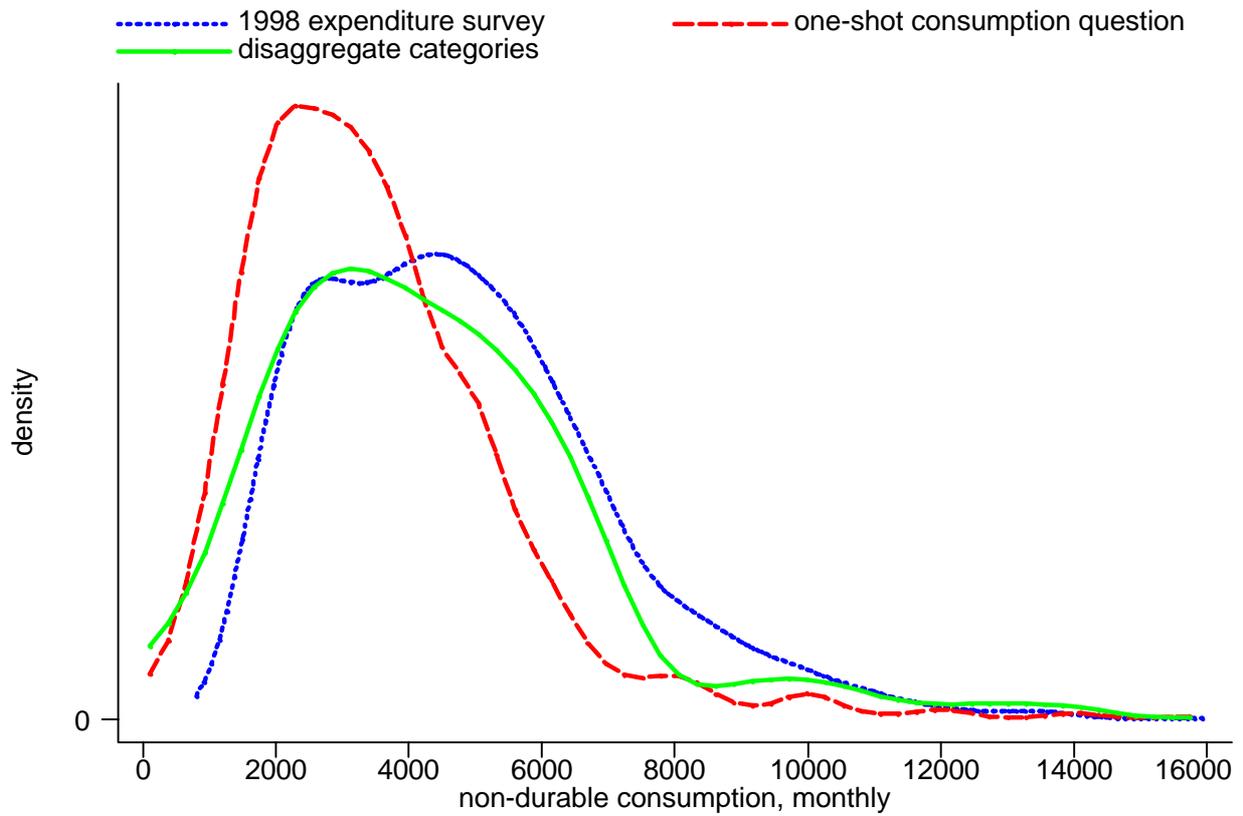
Table 5: Responses to alternative questions on total monthly food expenditure

Food expenditure, last month	N	mean	st. dev.	median
expenditure survey (adjusted)	1989	824.8	471.7	755.4
open-ended one-shot question	116	569.4	461.4	500.0

Source: Experiments conducted as part of the CentER Panel, June 2001, and 1998 expenditure survey.

Notes: The 1998 expenditure survey reports annual values. The measure reported here is converted to monthly values and adjusted for inflation using the Dutch consumer price index.

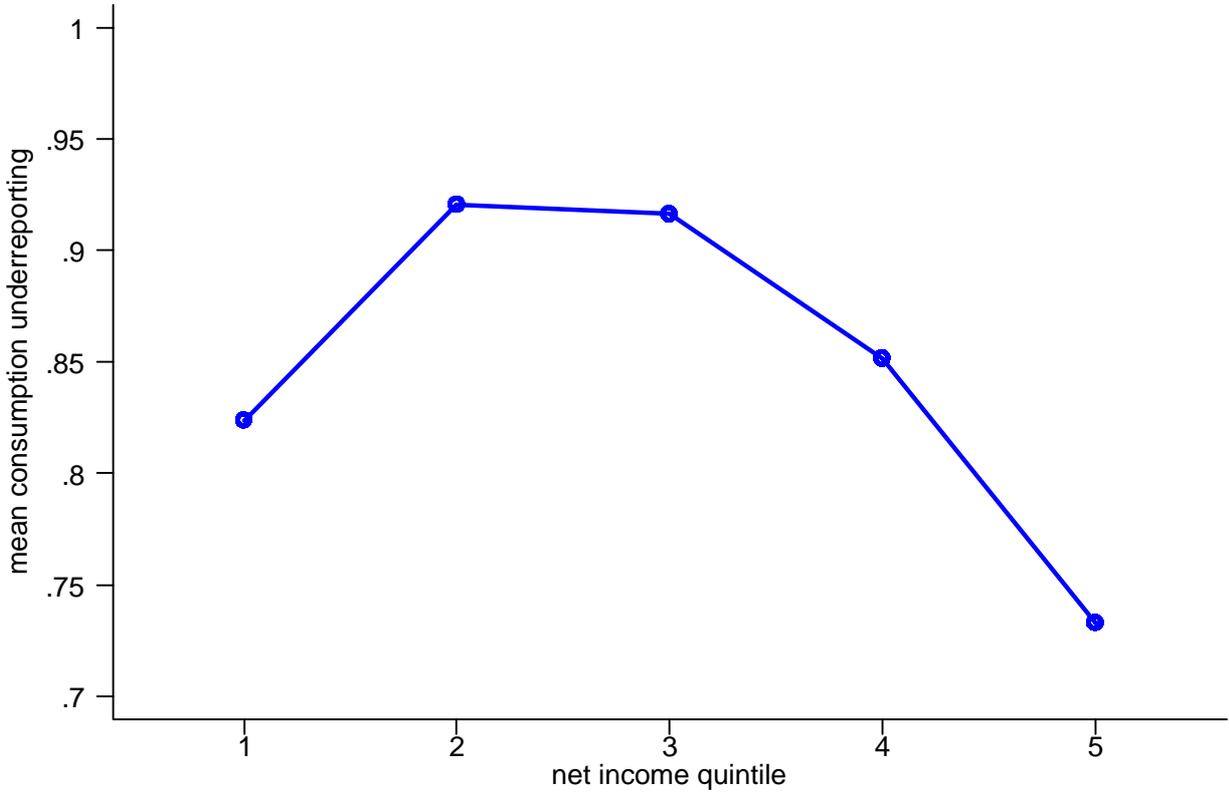
Figure 1: Distributions of alternative total monthly non-durables expenditure measures



Source: Experiments conducted as part of the CentER Panel, June 2001, and 1998 expenditure survey.

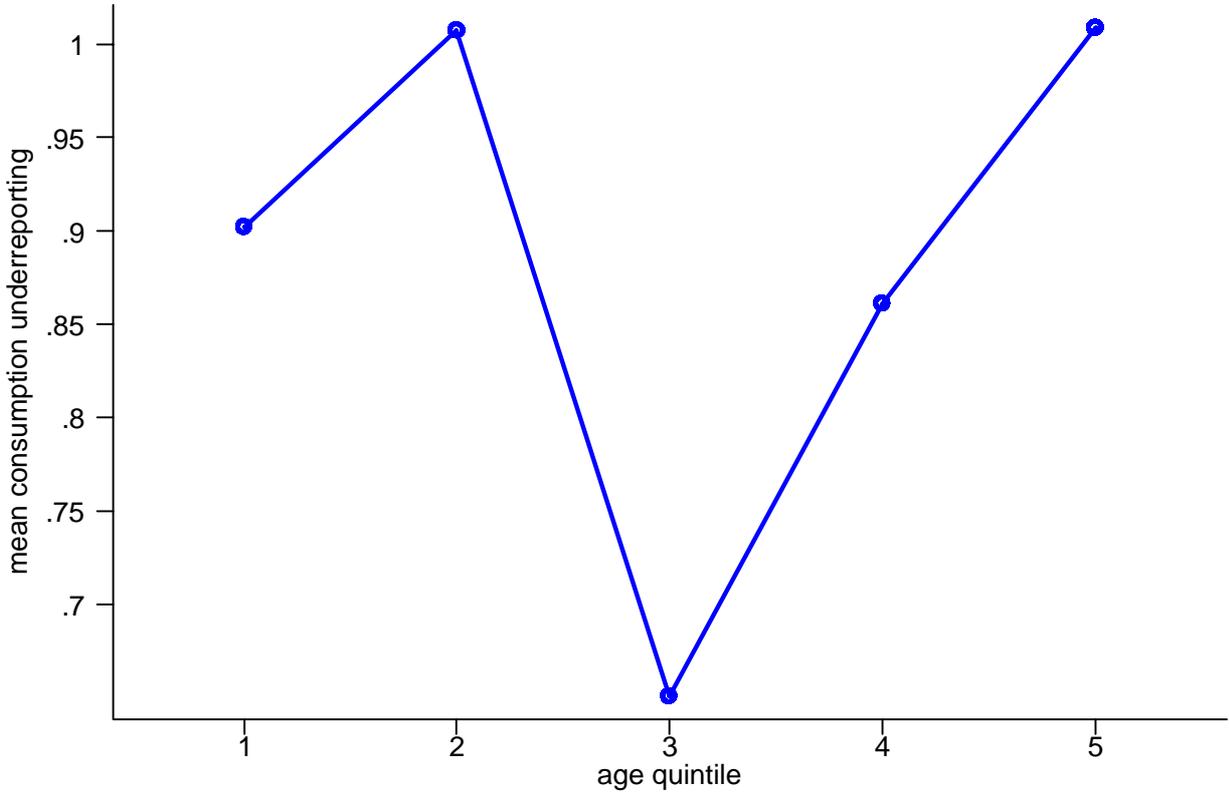
Notes: The 1998 expenditure survey reports annual values. The measure reported here is converted to monthly values and adjusted for inflation using the Dutch consumer price index.

Figure 2: Underreporting in the one-shot question as a function of household income



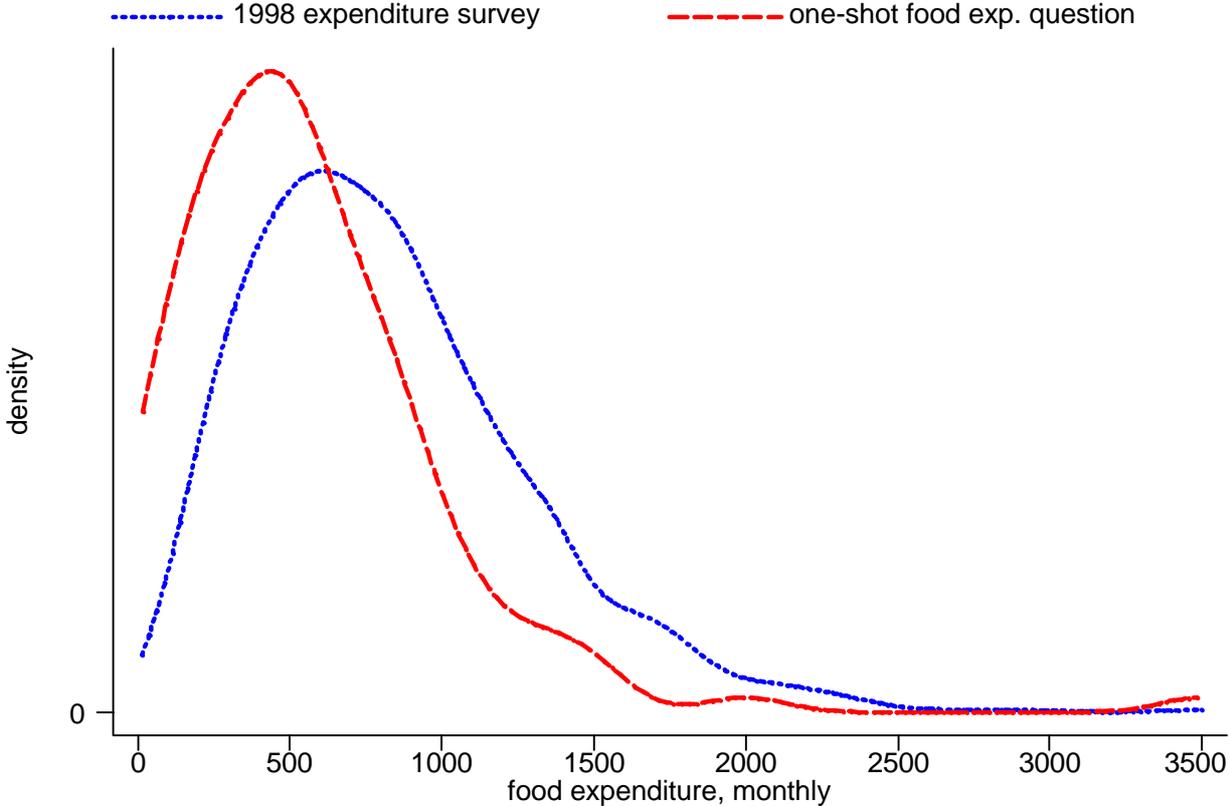
Source: Experiments conducted as part of the CentER Panel, June 2001.

Figure 3: Underreporting in the one-shot question as a function of age



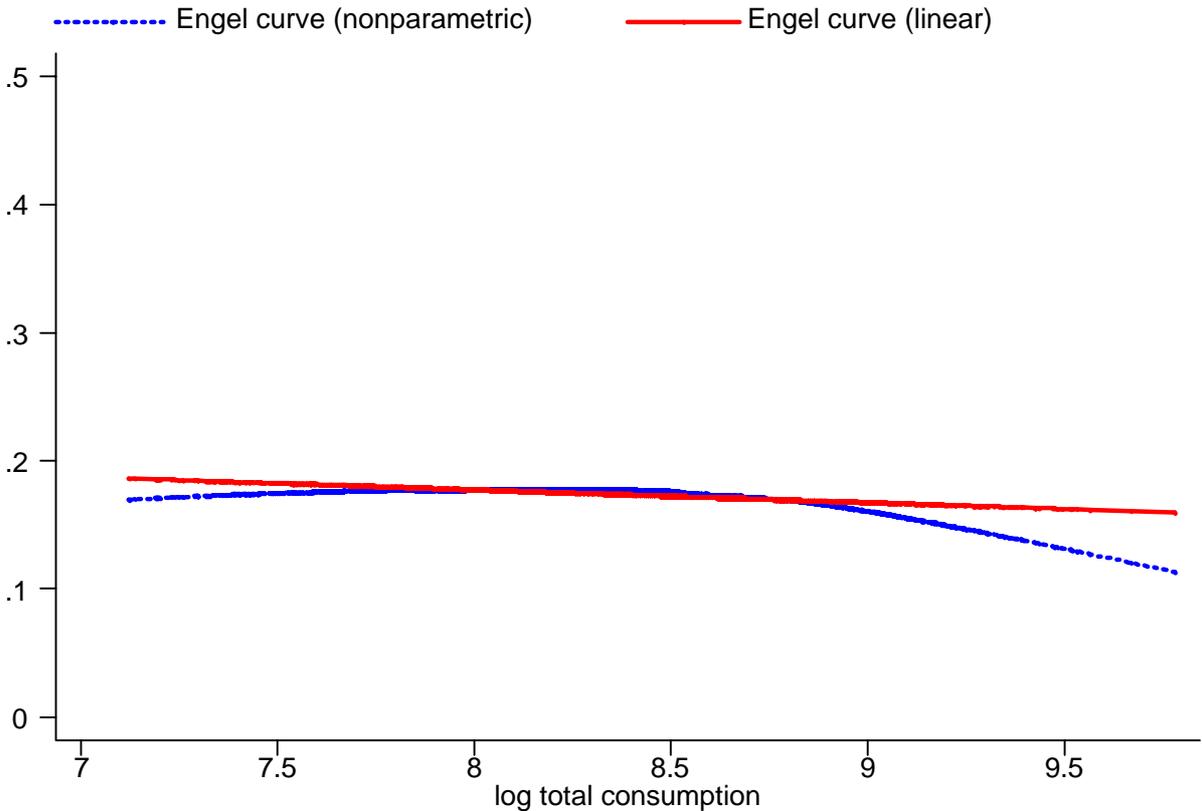
Source: Experiments conducted as part of the CentER Panel, June 2001.

Figure 4: Distributions of alternative food expenditure measures



Source: Experiments conducted as part of the CentER Panel, June 2001, and 1998 expenditure survey.

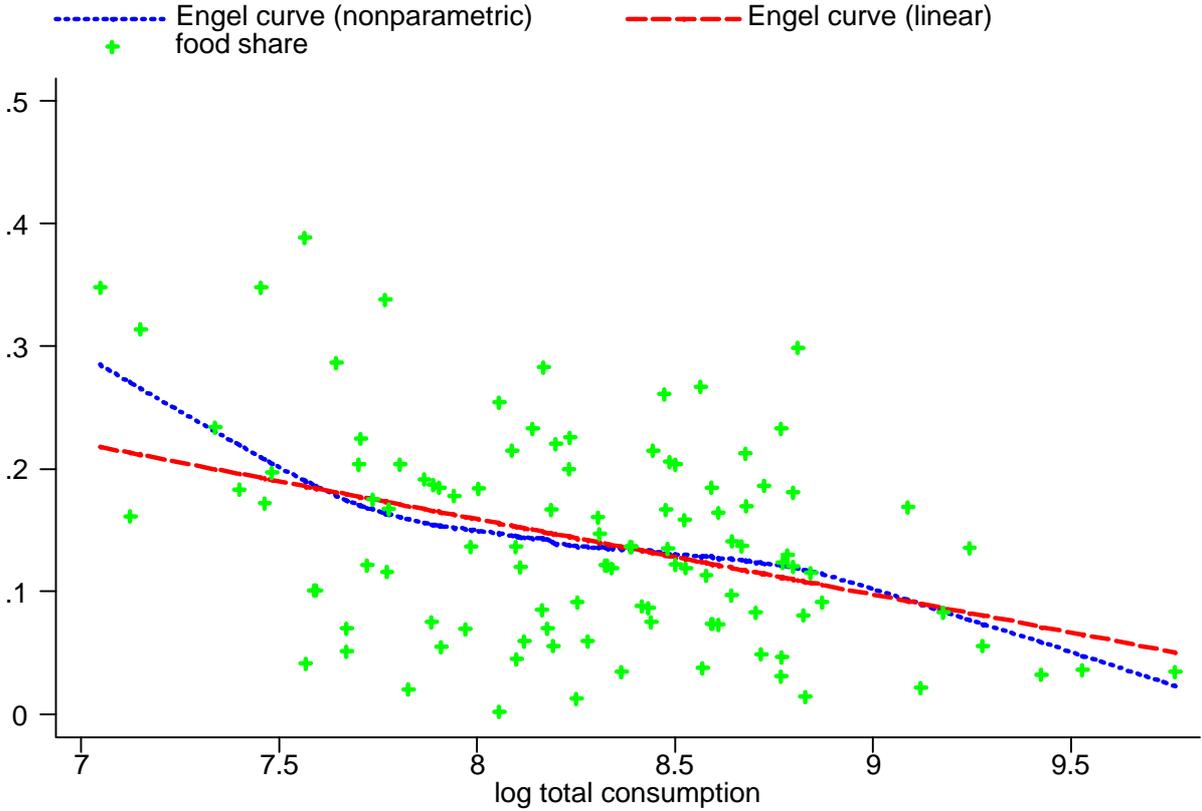
Figure 5: Estimated Engel curves for food expenditure (budget survey data)



Source: 1998 expenditure survey.

Notes: The 1998 expenditure survey reports annual values. The measure reported here is converted to monthly values and adjusted for inflation using the Dutch consumer price index.

Figure 6: Estimated Engel curves for food expenditure (experimental survey question)



Source: Experiments conducted as part of the CentER Panel, June 2001

SONDERFORSCHUNGSBereich 504 WORKING PAPER SERIES

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