

Response bias in survey–based measures of household consumption

Joachim Winter
University of Mannheim

Abstract

An important aspect of household surveys is the design of consumption questions. A controlled experiment shows that a single question on total monthly nondurables expenditure and a design with 35 disaggregated categories produce different results. These differences vary with household characteristics.

I would like to thank Rob Alessie, Axel Börsch–Supan, Marcel Das, Bernd Fitzenberger, Mike Hurd, Arie Kapteyn, Melanie Lührmann, Arthur van Soest, Guglielmo Weber, and participants of a TMR workshop at Evian–les–Bains for helpful comments. I appreciate the hospitality of the CentER for Economic Research at Tilburg University where this research was conducted. 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).

Citation: Winter, Joachim, (2004) "Response bias in survey–based measures of household consumption." *Economics Bulletin*, Vol. 3, No. 9 pp. 1–12

Submitted: December 30, 2003. **Accepted:** March 15, 2004.

URL: <http://www.economicsbulletin.com/2004/volume3/EB–03C80001A.pdf>

1 Introduction

Answering many public policy questions requires knowledge of consumption levels and of how consumption changes over the life cycle. For instance, Meyer and Sullivan (2003) argue that consumption is the preferred measure of well-being of the poor. In a different application, Hurd and Rohwedder (2003) investigate whether consumption changes after retirement are expected or not, which in turn is relevant for assessing the empirical validity of standard life-cycle models of saving. The analysis of these questions, and of many other aspects of household behavior, requires reliable micro data on households' consumption.

In most developed countries, expenditure data are available at the household level from budget surveys which use diaries with several hundreds of expenditure categories. Examples are the Consumer Expenditure Survey (CEX) in the United States and the *Einkommens- und Verbrauchsstichprobe* (EVS) in Germany. Consumption measures obtained from these budget surveys are generally taken to be reliable. However, budget surveys provide much less detailed data on other aspects such as income and employment which limits their usefulness in the analysis of many research questions. In contrast, general-purpose panel studies such as the *German Socio-Economic Panel* (GSOEP) or the Panel Study of Income Dynamics (PSID) in the United States contain detailed information on employment histories and income dynamics, but lack sufficient data on consumption. Many researchers feel that adding reliable measures of consumption to general-purpose panel studies is important. However, since time and space in such surveys are restricted, obtaining reliable consumption measures is challenging, and there is an inherent trade-off between data quality and survey cost.

Experimental survey research conducted in developing countries suggests that using an extensive list of expenditure items yields the most reliable consumption data; see Deaton (1997), pp. 26–29, Lanjouw and Lanjouw (2001), Pradhan (2001), and Jolliffe (2002). Given restrictions on the number of questions that can be asked, such extensive lists are not practicable in industrialized countries. Interestingly, however, there exists no systematic experimental evidence on how variations in survey design affect measures of consumption in household surveys in industrialized countries.

The present paper investigates the effects of the level of aggregation of expenditure categories on measured nondurables consumption using a controlled survey experiment that was administered in the Netherlands in June 2001. Specifically, I compare responses to a “one-shot” question that asks households to report total expenditures on nondurables with responses to a more disaggregated design that uses 35 expenditure categories. The data reveal significant differences between the responses to the one-shot question and the measure obtained from the disaggregate design. I interpret these differences as underreporting in the one-shot question. The degree of underreporting is shown to vary with household characteristics.

The remainder of this paper is structured as follows. Section 2 describes the design of the survey experiment. Results are presented in Section 3. Section 4 concludes.

2 Description of the survey experiment

The experiment reported in this paper was administered over the Dutch CentERpanel in June 2001. The CentERpanel is an internet-based telepanel. It consists of some 2000 households in the Netherlands. Every week, the members of these households fill in a questionnaire at home, using computerized interfaces. Households may use their own computers or they are provided with PCs or set-top boxes by CentERdata, the agency running the panel. Each year, about fifty questionnaires of up to 30 minutes each are presented to the respondents, but panel households are not required to participate every week. The CentERpanel is representative of the Dutch population.¹

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 participating 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 during the same week, 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 of the 2000 wave of 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?

The other group received a more disaggregated design with 35 different expenditure categories; a detailed list of these categories can be found in Table 1. The choice of these categories followed the design used in the 2001 Consumption and Activities Mail Survey (CAMS), a

¹ Detailed tabulations of the distributions of key demographic variables (such as age, sex, education, region) in the CentERpanel and in population data can be found on CentERdata's website at <http://cdata4.uvt.nl/eng/representative>.

supplement to the HRS. These categories are still much broader than the hundreds of categories contained in household budget surveys; they should be taken as a practical compromise between the detailed data of budget surveys and the space restrictions that prevail in general-purpose 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 nondurables expenditure (as reflected by the choice of the 35 categories), it was important to insure that households did not include expenditure on durables in their response to the one-shot question. Therefore, all households were asked about expenditures in six categories of durable consumption goods (automobile, refrigerator, washing machine and/or dryer, dishwasher, television, personal computer) before entering the experimental module. The present paper uses only the experimental data on nondurables expenditure.

3 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 nondurables expenditure (obtained by summing the responses to the 35 expenditure items) of zero. Based on previous experience with response behavior the CentERpanel, it is unlikely that consistent zero reports are the result of respondents’ unwillingness to answer such questions. Rather, observations of zero expenditure seem to be an indication of uncertainty about the response. Second, almost all median values are multiples of 10 or 100 guilders (if not zero). In open-ended consumption questions, households have a strong tendency to report focal values, so the medians tend to be focal values as well. 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.* (2003).²

In the one-shot question on total monthly nondurables, 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 Assets and Health Dynamics Among the Oldest Old (AHEAD) survey. 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 (specifically, 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). 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. Results are reported in table 2.

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; see Winter (2002a). 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 (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, on average, 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 from nonresponse.

After nonresponse, we are left with 753 responses to the one-shot expenditure question. As in the responses to the disaggregate questions, focal values are frequent; 714 (about 95%) of respondents reported a multiple of 100 guilders, and still 352 (about 47%) a multiple of 1000 guilders. Again, this is an indication that respondents are quite uncertain about the exact amount spent on total nondurables consumption.

² Potential biases in subsequent analysis that arise from focal values could be addressed using the method proposed by Heitjan and Rubin (1990). This was not attempted in the present paper because of the small number of observations.

Table 3 compares descriptive statistics for the two alternative measures of household consumption. The one-shot open-ended question yields significantly lower estimates of consumption than the disaggregated question with 35 categories. In light of the results from prior research discussed in the introduction, I interpret these differences as the result of underreporting in the one-shot question. 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. The main explanation for such underreporting is that respondents forget to take account of some expenditure items when thinking about total nondurables consumption; naturally, a more detailed list mitigates this problem.

As can be seen from the non-parametric estimates of the distribution functions of the three consumption measures in figure 1, the measure obtained from disaggregated questions indicates that consumption is more dispersed than the measure based on a one-shot question would suggest. This result indicates that heterogeneity between households is greater for expenditure items that are typically left out in the reports to a one-shot question.

The degree of underreporting cannot be assessed directly since every household answered only one of the two alternative question designs. However, it is instructive to match households by their characteristics and to compare measures within the matched groups. Such pairwise comparisons are reported in table 4. Underreporting is high for the middle income groups and decreases with income. This finding is similar to results reported by Pradhan (2001). Also, 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 nondurables expenditures that are concentrated on few items and therefore easier to recall. The data also indicate that retired households tend to overestimate their nondurables expenditure when confronted with a one-shot question on the total amount. Finally, underreporting is smaller for respondents who list “housekeeper” as their occupation, which is not surprising since the housekeeper should have better knowledge of nondurable expenditure and is therefore less likely to underreport due to forgotten items.

4 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 nondurables expenditure with responses to a more disaggregated design based on 35 expenditure categories reveals significant differences. I interpret these differences as reflecting underreporting in the one-shot question.

It is, however, not clear whether the nondurables consumption measure constructed from 35 categories is itself still subject to underreporting. Addressing this issue would require a detailed comparison with data from a budget survey. However, for the Netherlands, such data are not yet available for the field period of this experiment, June 2001. Winter (2002b) contains a preliminary analysis using data from the Dutch 1998 budget survey; it suggests that the measure obtained from the 35 disaggregate items in this experiment is close to the measure constructed from the budget survey. A more detailed analysis is left to future research.

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 on only a few (typically, much less than 35) 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 that focus on older households. Moreover, results by Battistin *et al.* (2003) 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. Therefore, 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 as measured in a detailed budget survey.

Acknowledgements

I would like to thank Rob Alessie, Axel Börsch-Supan, Marcel Das, Bernd Fitzenberger, Mike Hurd, Arie Kapteyn, Melanie Lührmann, Arthur van Soest, Guglielmo Weber, and participants of a TMR workshop at Evian-les-Bains for helpful comments. I appreciate the hospitality of the CentER for Economic Research at Tilburg University where this research was conducted. 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).

References

- Battistin, E., R. Miniaci, and G. Weber (2003):** What do we learn from recall consumption data? *Journal of Human Resources*, 38(2), 354–385.
- Browning, M., T. F. Crossley, and G. Weber (2002):** Asking consumption questions in general purpose surveys. CAM Working Paper No. 2002-05, Institute of Economics, University of Copenhagen.
- Deaton, A. (1997):** *The Analysis of Household Surveys*. Baltimore, MD and London, UK: Johns Hopkins University Press.
- 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.
- Hurd, M. and S. Rohwedder (2003):** The retirement-consumption puzzle: Anticipated and actual decisions in spending at retirement. NBER Working Paper No. 9586.
- 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.
- Jolliffe, D. (2002):** Measuring absolute and relative poverty: The sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, forthcoming.
- 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.
- Meyer, B. D. and J. X. Sullivan (2003):** Measuring the well-being of the poor using income and consumption. NBER Working Paper No. 9760.
- Pradhan, M. (2001):** Welfare analysis with a proxy consumption measure: Evidence from a repeated experiment in Indonesia. Discussion Paper No. 01 - 092/2, Tinbergen Institute.
- Winter, J. (2002a):** Bracketing effects in categorized survey questions and the measurement of economic quantities. Discussion Paper No. 02-35, Sonderforschungsbereich 504, University of Mannheim.
- Winter, J. (2002b):** Design effects in survey-based measures of household consumption. Discussion Paper No. 02-34, Sonderforschungsbereich 504, University of Mannheim.

Table 1: List of items used to elicit total monthly nondurables expenditure

consumption categories, last month	zero reports (%)	mean	st. dev.	median
housing mortgages	40.5	558.7	601.4	480.0
homeowners insurance	63.8	43.4	185.2	0.0
property taxes	46.6	93.8	166.1	32.5
rent	68.1	279.0	815.2	0.0
electricity, water, fuel	6.9	324.9	915.6	250.0
energy	5.2	144.3	97.7	120.0
housekeeping supplies	6.9	65.9	160.0	30.0
garden and lawn supplies	43.1	59.0	98.4	20.0
home repair and maintenance	68.1	421.7	2196.4	0.0
domestic services	87.1	14.6	57.4	0.0
food and beverages	4.3	569.4	461.4	500.0
eating and drinking out	14.7	148.3	150.3	100.0
clothing and apparel	18.1	238.9	223.2	200.0
personal care products	12.1	52.3	47.5	50.0
personal care services	99.1	2.8	29.7	0.0
vehicle finance charges	35.3	66.6	108.6	43.5
gasoline	25.0	162.7	146.2	150.0
vehicle maintenance	69.0	113.2	373.1	0.0
vehicle insurance	39.7	96.3	222.8	42.5
prescription and non-prescription drugs	56.9	34.8	141.1	0.0
health care services	81.0	30.4	91.2	0.0
medical supplies	93.1	5.8	33.2	0.0
health insurance	20.7	166.2	191.3	81.0
trips and vacations	39.7	482.6	988.8	100.0
tickets	76.7	23.5	65.1	0.0
membership to health / social clubs	60.3	20.8	34.8	0.0
video and audio entertainment	67.2	26.1	70.2	0.0
hobbies	58.6	51.8	161.9	0.0
computer equipment	80.2	14.7	41.4	0.0
reading	19.8	53.2	59.1	40.0
pet food, products and expenses	61.2	24.6	55.5	0.0
tobacco products	64.7	36.0	73.0	0.0
contributions	58.6	22.5	34.3	0.0
gifts	71.6	30.7	88.4	0.0
life and liability insurance	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 nondurables 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 nondurables 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 alternative questions on total monthly nondurables expenditure

Total nondurables consumption, last month	N	mean	st. dev.	median
Response to the one-shot question	753	3890.6	3741.9	3000.0
Sum of responses to questions on 35 expenditure categories	116	4559.2	3329.1	3995.5
First line as percent of second line		85.3%		75.1%

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

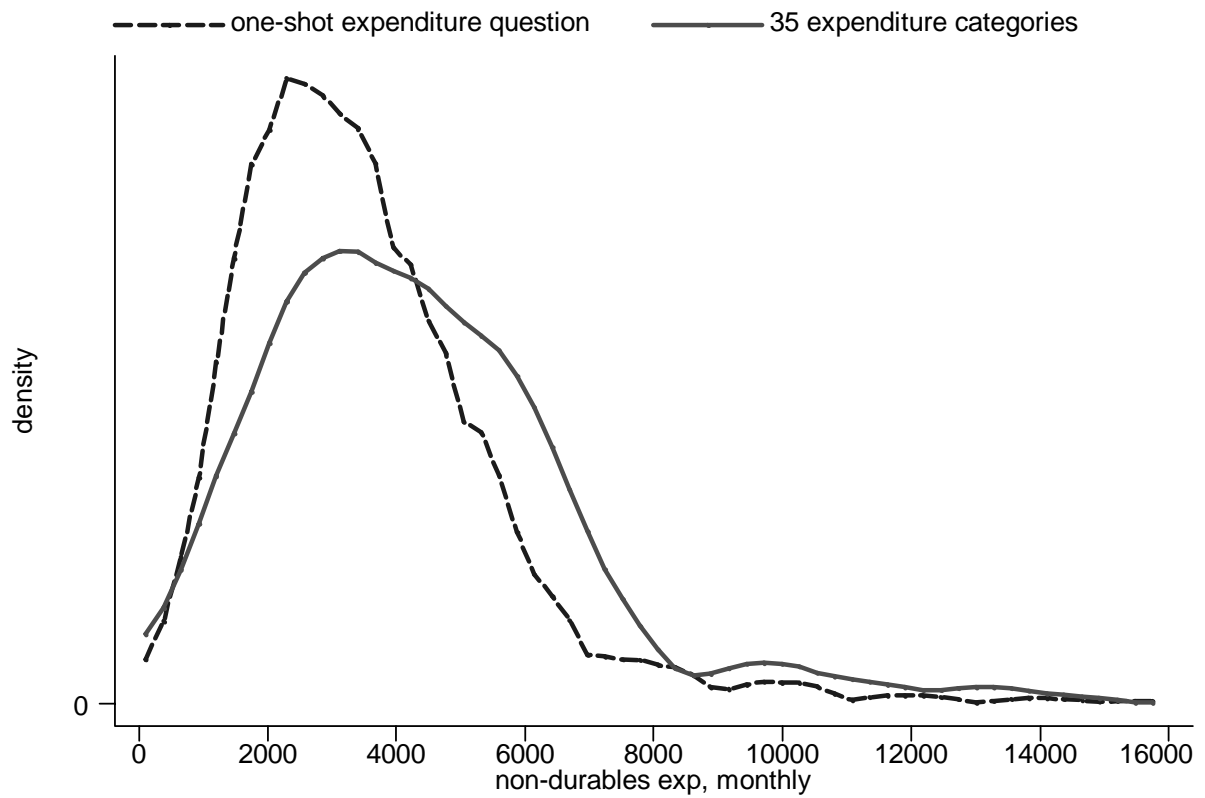
Table 4: Underreporting ratios for different household groups

Variable	Group	Underreporting ratio
Net household income	1 st quintile	82.4%
	2 nd quintile	92.1%
	3 rd quintile	91.6%
	4 th quintile	85.2%
	5 th quintile	73.3%
Age of respondent	1 st quintile	90.0%
	2 nd quintile	100.7%
	3 rd quintile	65.1%
	4 th quintile	86.1%
	5 th quintile	100.8%
Respondent is retired	No	82.6%
	Yes	104.2%
Respondent's occupation is "housekeeper"	No	84.4%
	Yes	96.1%

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

Note: Underreporting ratios are the ratios of the mean of the responses to the one-shot question and the mean of the sum of the responses to questions on 35 expenditure categories, with means computed for each household group.

Figure 1: Distributions of alternative total monthly nondurables expenditure measures



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

Note: Kernel density estimates, using the Epanechnikov kernel and optimal bandwidth selection.