

Statistical Analysis of Choice Experiments and Surveys

Daniel L. McFadden (University of California, Berkeley) *

Albert C. Bemmaor (ESSEC, Cergy-Pontoise)

Francis G. Caro (University of Massachusetts, Boston)

Jeff Dominitz (Carnegie Mellon University)

Byung-hill Jun (University of California, Berkeley)

Arthur Lewbel (Boston College)

Rosa L. Matzkin (Northwestern University)

Francesca Molinari (Cornell University)

Norbert Schwarz (University of Michigan)

Robert J. Willis (University of Michigan)

Joachim K. Winter (University of Munich) **

Abstract: Measures of households' past behavior, their expectations with respect to future events and contingencies, and their intentions with respect to future behavior are frequently collected using household surveys. These questions are conceptually difficult. Answering them requires elaborate cognitive and social processes, and often respondents report only their "best" guesses and/or estimates, using more or less sophisticated heuristics. A large body of literature in psychology and survey research shows that as a result, responses to such questions may be severely biased. In this paper, (1) we describe some of the problems that are typically encountered, (2) provide some empirical illustrations of these biases, and (3) develop a framework for conceptualizing survey response behavior and for integrating structural models of response behavior into the statistical analysis of the underlying economic behavior.

* *Workshop chair.* Address: Econometrics Laboratory, 549 Evans Hall, University of California, Berkeley, CA 94720-3880; E-mail: mcfadden@econ.berkeley.edu

** *Corresponding author.* Address: Department of Economics, University of Munich, Ludwigstr. 28 (RG), 80539 Munich, Germany; E-mail: joachim.winter@lrz.uni-muenchen.de

Acknowledgements: We would like to thank Dipankar Chakravarti and two anonymous reviewers for their helpful comments. Financial support by the National Institute on Aging (Grant No. 263-MD-317352) is gratefully acknowledged.

Introduction

Surveys are an important source of data for the empirical analysis of household behavior. Unfortunately, data problems such as sample selection, item non-response, and measurement error are the rule rather than the exception in survey data. Well-designed studies using household survey data are careful to detect outliers, to impute missing values, and to correct for selection bias caused by missing observations.

Economists and econometricians have traditionally addressed such data problems using statistical *ex post* approaches. These approaches have reached a high level of sophistication, as summarized for instance by Wansbeek and Meijer (2000) and Bound, Brown, and Mathiowetz (2001). However, these approaches generally neglect the vast literature on the sources of response errors and biases that has accumulated in psychology and survey research. The main motivation for the research program outlined in this paper is to make better use of this information, with the ultimate aim of integrating structural models of survey response behavior into statistical and econometric models of behavior in which economists and market researchers are ultimately interested. Conversely, advanced statistical and econometric techniques can benefit from some experimental variation in survey design, and we argue that the flow of information might also go in the other direction, from statisticians and econometricians to survey designers.

Survey response behavior

Surveys are sometimes described as “structured conversations between strangers”. Questions can be misheard, or interpreted within the context of these conversations, and responses can be designed, consciously or unconsciously, to “solve” the cognitive tasks that the questions apparently pose, to please or to impress the interviewer, to support and to justify the self-image of the re-

spondent, to minimize effort and to hasten conclusion of the interview, or to provide a forum for opinions and grievances. While the most straightforward factual questions (age, gender, marital status, occupation) seem immune to these response effects, even here there can be genuine confusion (e.g., when stating occupation, should one refer to lifetime profession or to the most recent job?), misrepresentation (e.g., the understatement of age or weight), or imperfect recall (e.g., the number of doctor's visits in the past six months); for reviews and further discussion see Dex (1995), Gardiner and Richardson-Klavehn (2000), Tourangeau, Rips, and Rasinski (2000), and Schwarz (2003).

Questions such as that on current monthly income can leave the definition of the quantity requested to the judgment of the respondent, or require an elaborate and lengthy preamble (i.e., "Please state income before taxes, including fringes such as employer-paid health insurance, excluding income from sale of household goods or automobiles, excluding bonuses, ...") which can be imperfectly attended. Other questions can require the reconstructions of factual and affective memory and the integration of information that are cognitively difficult (e.g., net worth, monthly consumption, pain level in a medical episode). Finally, many questions, in particular in market research, refer to hypothetical choice scenarios, and these scenarios typically cannot specify all the aspects the respondents may need to produce an answer, so the respondents may fill in that missing information in ways that cannot be controlled by the question design (Manski, 1999).

In all these situations, when respondents cannot recall or construct the exact answer to a question, they can interpret the task as one of constructing a "correct" response, much like responding to a teacher's questions in school. Analogous problem-solving cognitive responses can be invoked, including item non-response or artificial "safe" or "edited" responses. This motive appears to underlie many of the phenomena associated with reconstructed memory, anchoring, or-

der effects, and focal responses in household surveys. Tourangeau, Rips, and Rasinski (2000) provide a general conceptual framework for analyzing these issues.

A related phenomenon that is of particular relevance in market research is that answering a survey may affect future behavior. Various channels have been discussed in the literature. In an influential early contribution, Feldman and Lynch (1988) argue that answering questions on beliefs, attitudes, intentions, and behaviors may result in self-generated validity effects. Such “mere measurement” effects change the correlations among the responses to survey questions and may also affect responses in subsequent surveys and may even influence market behavior. For instance, answering a detailed survey about purchase intentions for some good may increase the proportion of realized purchase in comparison to a control group that has not answered the survey. These effects have received much interest in market research, see Chandon, Morwitz, and Reinartz (2005) for a recent contribution.

In many cases, response effects can be systematic, so that the effects of question context and format are predictable. Survey researchers have built considerable experience and accumulated knowledge on response effects, and practical rules for the design of surveys to minimize response errors. For example, focus groups and verbal protocols are often used to determine if questions are understood by the respondents. Unfortunately, there appears no tradition of systematically conducting and documenting pre-survey design studies, and little cumulative learning outside the relatively small circle of survey design professionals. Detection, control, and compensation for response errors in economic and health data is possible when experimental treatments and/or validation procedures are built into the survey process (e.g., by asking for health conditions using different question treatments, or validating the results through matching Medi-

care records), and the systematic study of response errors makes possible the construction of quantitative models that can accurately recover “true” quantities.

There are obvious risks in both scientific and policy analyses that use potentially inaccurate survey measurements. In one classic example, Cowing and McFadden (1984) found that in successive waves of the American Housing Survey, five percent of houses reported to have basements in one wave were reported to have no basement in the following wave. This study also found in validation audits that a substantial number of respondents reporting that they had electric heat did so because their heating system was controlled by a thermostat. In a second example, Poterba and Summers (1986) found through the audits of employment surveys that the correction of reported employment status can change the estimated duration of unemployment by a factor of two. Clearly, inaccurate data can mislead energy conservation or unemployment insurance policies. In surveys of health and economic status of the elderly, the inaccurate measurements of health conditions, assets, use and cost of health services, or expectations can lead to faulty inference on causal links between health and wealth, and also to misleading conclusions on the need for, or the effectiveness of, policy interventions.

Examples of variables in surveys of aging where response issues are important are self-rated health status (SRHS), economic assets, health utilization variables such as the number of doctor visits, and expectations regarding future need for health services. Quality control problems include item non-response, confusion about questions, imperfect recall, and anomalies in the formation of judgments, beliefs, choices, and reports. There are issues of the comparability of responses across time, individuals, and cultures. These issues are present in most survey studies, but are particularly important in the longitudinal studies of aging when they interact with the aging process and confound more fundamental consequences of aging. The usual method for meas-

uring SRHS is a response on a five or seven-point scale (from “poor” to “excellent”), used without standardization. The anecdote that illustrates the difficulty of interpersonal comparisons using this measure is that 62 percent of Danish men, but only 14 percent of French men, give a SRHS of “excellent”, yet French men have a life expectancy two years greater. Clearly, a prerequisite for interpersonal comparisons, or even intertemporal comparisons for one individual, is some form of scaling. Recent attention has focused on the use of health vignettes as a method for scaling (e.g., Banks, *et al.*, 2004). While these methods show promise in initial studies, further research is needed to explore their properties and optimize their use.

Sampling and survey design issues run through many survey measurement problems, and cross disciplinary lines: The interaction of sample frame and questionnaire design with the selection effects of subject and item non-response or uninformative response, the use of multiple measurements and validation procedures to reduce and correct for measurement error, the use of experiments embedded within surveys, as well as more conventional pilot studies, to assess comprehension and response behavior, the lessons from cognitive and affective psychology regarding attention, recall, and task processing, and the lessons from econometrics regarding the analysis of treatment effects in experiments.

The relevance of getting a better understanding of survey response behavior seems obvious. While there is a large body of research in cognitive and social psychology and in survey research on survey response behavior, as summarized by Tourangeau, Rips, and Rasinski (2000), very few studies have focused on the questions that are typically asked by economists and market researchers. Moreover, researchers in econometrics and marketing have developed quite sophisticated methods for analyzing the data from expectations and hypothetical choice questions (e.g., Ben-Akiva, *et al.*, 1997, 1999, 2002; Dominitz and Manski, 1997; Manski, 2004). It might, at

least conceptually, be possible to integrate structural models of survey response behavior that are informed by research in psychology into structural models of the underlying economic behavior (expectations, intentions, and choices).

Two examples: Asking for subjective health status and for expectations

In order to structure the discussion of the challenges and opportunities in the area of survey response behavior, it is useful to focus on specific examples. First, consider a question on self-rated health status. Such a question may read: “Overall, how would you rate your health condition today?”, with responses being collected on a five or seven-point scale that ranges from “excellent” to “poor”. One of the problems that arises is that different respondents can interpret the response scales differently, and these differences can be related to the “true” health status, to other factors such as education, gender, socio-economic status, and social, reference groups, and possibly country-specific differences (as the discrepancy between subjective health reports and actual life expectancy in Denmark and France mentioned above). Given such apparently inconsistent results, it seems obvious that cross-country comparisons of responses to questions on self-rated health can be difficult, and similar arguments can be made for inter- and intra-individual comparisons as well.

A second example of questions that may be difficult to answer and thus are likely to be subject to response effects concern expectations. Expectations regarding future events, and the closely related subject of choice under uncertainty, have been the focus of study since Allais (1953) and Davidson, Suppes, and Siegel (1957). The Allais Paradox and the classic experiment by Tversky and Kahneman (1974) on the effects of framing on choice among lotteries demonstrate that the formation and reporting of risk perceptions display systematic cognitive (apparent) anomalies.

How should questions on expectations be implemented in surveys? Recent research by Dominitz and Manski (1997, 2004), Gan, Hurd, and McFadden (2003), and others shows that responses to questions of the form “On a scale of 0 to 100, where 0 means no chance, and 100 means certainty, what would you say is the probability of ...?” yield responses that are quite predictive of subsequent behavior. For example, the question asked about mortality risk is generally predictive of subsequent mortality experience (Hurd and McGarry, 1995, 2002; Smith, Taylor, and Sloan, 2001) and more predictive of subsequent savings behavior than life table hazard rates (Hurd, *et al.*, 1998).

For instance, using data from the Health and Retirement Study (HRS), McGarry (2004) finds that subjective reports of health have important effects on the age of retirement; these effects are arguably stronger than those of financial variables. However, survey response effects such as item nonresponse or focal responses (at 0, 50, 100 percent) are common, and may be concomitant of uncertainty and the construction of perceptions. In some cases, more than three quarters of the responses to expectations questions have been documented to be 50 percent (Bruine de Bruin, *et al.*, 2002). Understanding these effects can improve the predictive accuracy of subjective expectations data. One possible way to explain these effects is that at times, researchers may ask questions that some respondents had not thought about previously; there exists some evidence that shows that the concern for the future varies across individuals; see, e.g., Bergadaa (1990). Individual heterogeneity with respect to response effects such as rounding or providing a 50 percent response therefore requires special attention.

Questions on expectations with respect to future events and contingencies and hypothetical choice questions that are prevalent in economics and market research can at times be harder to answer than questions on self-rated health or subjective survival. In market studies for new prod-

ucts, many questions focus on planned purchase frequencies (see, e.g., Kalwani and Silk, 1982; Infosino, 1986; and Buchanan and Morrison, 1987). The responses to such questions are often subject to substantial upward biases (e.g., Klein and Sherman, 1997). Further evidence of the biases in a choice situation and means to handle them are reported in Hsiao and Sun (1999). Still, using an extensive number of data sets on stated intents and subsequent behavior for a variety of focused actions, Bemmaor (1995) shows that stated intents can provide a valuable basis for predicting subsequent behavior. Exceptions can occur when a significant unexpected event takes place after the survey (i.e., a sudden promotion for deep quantity discounts).

Themes in the analysis of survey response behavior

Starting from a question like that on self-rated health, several themes can be developed. At the most general level, how one deals with data problems depends very much on the substantive issue that is analyzed. In the context of the self-rated health question, a researcher can be interested in the average health status of an entire society. This health measure would then be a measure of the output of the health sector, and the public may be interested in comparing such measures across countries and political systems, or over time (say, before and after a major reform in health financing). Alternatively, the measure of self-rated health can be used as a regressor in an analysis in which the dependent variable is something else – say, socio-economic status (Adams, *et al.*, 2003). The requirements for data accuracy and the statistical treatment of errors and biases can be completely different depending on which of these issues is investigated. Correspondingly, the econometric analysis of response errors can have different goals. A researcher may be interested in correcting for the biases in a measure such as self-rated health at the individual level, or he/she may wish to recover the features of the population distribution such as the average or the

median health status. These two tasks require different approaches involving different assumptions about response behavior and response accuracy.

A related theme in the analysis is how survey response errors and biases can be conceptualized. Much of the existing literature on measurement error, as summarized by Bound, Brown, and Mathiowetz (2001) assumes that errors are classical (loosely speaking, that they are random variables conditioned upon observed covariates). However, the literature in psychology and survey research suggests that in many cases survey response errors do not confirm to the classical measurement assumption – for instance, they can be correlated with unobservable factors such as cognitive ability. Again, how this issue can be dealt with depends on the application. In some large household panels, such as the Health and Retirement Study (HRS), measures of cognitive ability are available and can be used as instruments in an econometric model that takes survey response errors into account. In many other household surveys, including most that are used in market research, such measures do not exist at the individual respondent's level

Strategies for the design of survey questions

Once researchers have realized that responses to questions such as those on self-rated health are potentially problematic, several approaches to deal with the problem at the survey design stage have been developed.

Using vignettes

One of the most prominent approaches is the use of vignettes (Alexander and Becker, 1978). The elicitation of subjective probabilities of various end-of-life vents, particularly long-term care in a nursing home, is facilitated by the use of vignettes. Vignette surveys have been defined as “Use of stories which represent hypothetical situations to elicit preferences, judgments, or anticipated

behavior” (Wason, Polonsky, and Hyman, 2002). Vignette surveys have been used successfully in such fields as marketing, environmental economics, transportation, ethics, and professional decision making to determine the forces that influence multidimensional judgments. A premise in the use of these methods is that choices are often made by individuals through the consideration of the characteristics of specific objects or situations as they occur rather than through a systematic application of principles. These techniques use judgments made about vignettes as a basis for inferring the principles that drive individual choices. In particular, vignettes help to describe hypothetical outcomes and to assess how they affect behavior that is subject to social norms, such as the ability to work (Banks, *et al.*, 2004). Also, vignette designs such as the factorial survey methods are related to the broader class of hypothetical choice models discussed by Louviere, Hensher, and Swait (2000).

Several measurement issues with respect to vignette surveys are unresolved. We will experiment with fractional factorial surveys as a means of obtaining an improved understanding of the social norms that underlie both individual behavior and public support for aging policy. It is as yet unclear whether vignettes offer a satisfactory solution to the measurement problems in questions such as that on self-rated health. Future research needs to focus on the assumptions about response behavior under which vignettes provide identification of unobserved variables such as “true” health. Corresponding econometric techniques need to be developed, and it needs to be established which weaker assumptions on response behavior can be used when the strong assumptions used in existing studies such as King, *et al.*, (2004) and Banks, *et al.*, (2004) fail.

Obtaining auxiliary information

An alternative approach to the measurement error problem in variables such as self-rated health that is probably more akin to strategies that are traditionally used in applied econometrics is the use of multiple measurements (see Crossley and Kennedy, 2002, for a study that exploits repeated measurements of self-rated health status). Under certain conditions, repeated observation of a mismeasured variable such as health can provide identification of the underlying parameters of a structural econometric or statistical model. Unfortunately, repeated measurements that use the same survey instrument and that are not too far apart in time are rare. In panel surveys such as the Health and Retirement Study (HRS), such repeated measures can be available, but refer to different points in time (which requires the specification of a full model of health dynamics).

A study by Molinari (2005) illustrates a related approach: In the 2000 wave of HRS, questions on health status for one member of a couple were administered with both spouses so that two measurements were obtained. (The other spouse is often used as a proxy respondent in surveys if the actual respondent is unable to answer, but only very rarely are the same questions administered to both spouses.) Molinari finds that in a question on self-rated health with a five-point scale, the self report and the proxy report were identical in about half of the cases, in 34% of the cases, the proxy respondent reported worse health than the self respondent, and in 17.5%, the proxy report was better than the self report.

These findings confirm an empirical regularity that has been documented in other studies of proxy responses: On the average, proxy reports tend to deliver lower health ratings than self reports. Psychologists argue that individuals use different information sets when they describe their own behavior or state versus that of the others. Unfortunately, it is unclear which of the two reports is more accurate (on the average), because self reports may show over-reporting, proxy-

reports may show under-reporting, or both can show the same direction of bias. Future research on measurement error can nevertheless exploit proxy responses fruitfully: Conceptually, one can model both self and proxy reports as two measures of the same quantity. Both reports are subject to error, but under certain conditions they can nevertheless be used as instruments for one another and thus can provide identification of the parameters of interest in a structural model of health-related behavior.

Future research needs to consider also whether proxy reports and double measurement can be used in other surveys; for example, when it comes to purchase intentions for durable goods such as cars or family homes, it seems natural to ask both spouses and to interpret both reports as indicators of an underlying joint intention (Granbois and Summers, 1975). One obstacle to using such approaches in applied work is that at times structural models used to analyze behavior and intentions can be highly nonlinear. While instrumental variable techniques for linear models are well established, there are many open research questions when it comes to using instruments in nonlinear models. We return to this issue below.

Using embedded survey experiments

A related technique that can become an important ingredient of survey design that takes responses errors into account is the use of embedded experiments. The idea is to generate instruments for various sources of error as part of the survey design. At a more modest level, a survey experiment can simply vary the wording of a question or the design of the response alternatives randomly so that a comparison of the response distribution indicates whether these design features affect responses. This can be particularly useful in pre-tests, but it can also be useful to have some experimental variation even in “production” surveys. New ways of survey administra-

tion such as computer-assisted personal interviews and internet surveys facilitate the use of embedded experiments considerably.

The power of embedded survey experiments is illustrated by a recent study by McFadden, Schwarz, and Winter (2004). In order to explore response effects known from survey research and psychology in a survey with economic focus, they designed the Retirement Perspectives Survey (RPS), a combined mail and internet survey of older Americans. The first wave of the RPS was fielded in the spring of 2003 in collaboration with AARP. Participants were recruited from a random sample of AARP members by regular mail; the resulting sample size was about 4000 individuals aged 50 and older, about evenly split between the mail and internet versions of the survey. (Understanding self-selection into internet surveys is one of the main research questions addressed by this project, in addition to conducting experiments on survey response behavior.) The Retirement Perspectives Survey 2003 contained several embedded survey experiments.

One of these experiments focused on how older Americans perceive the risk of being admitted to a nursing home and whether they make special provisions for nursing home expenses, such as purchasing nursing home insurance. It consisted of a sequence of questions. The first question contained an experimental manipulation that induced different expectations about future events (in particular, the population probability of being admitted to a nursing home). It asked for a population quantity, the “number of Americans that will eventually be admitted to a nursing home”. The experimental manipulation concerned the format of the response options which were given as pre-specified brackets. In the “low bracket” treatment, these brackets were “Less than 2%”, “Between 2% and 5%”, “Between 6% and 10%”, and “More than 10%”. In the “high bracket” treatment, they were “Less than 10%”, “Between 10% and 25%”, “Between 26% and 50%”, and “More than 50%”. The subsequent questions did not contain any experimental varia-

tions; they asked for the respondents' subjective probability of being admitted to a nursing home themselves, and whether he or she had made financial provisions for nursing home care. The purpose of the manipulation of the first question was to test whether variations in the respondent's perception of his or her individual risk of being admitted to a nursing home can be induced by variations in a question on a corresponding population quantity.

The idea behind the low vs. high bracket range manipulation is to exploit the "bracketing effect", a phenomenon that is well established in psychology and survey research. In the original study, Schwarz, *et al.* (1985) 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; see Schwarz, Groves, and Schuman (1998) for a review. They have also been shown to exist in questions on quantities that are relevant in surveys on households' economic situation; see Menon, Raghubir, and Schwarz (1997) and Winter (2002).

In the RPS 2003 experiment on nursing home expectations, the manipulation of response scales in the first question allows us to test whether respondents' subjective expectations, behavior, and intentions reported in subsequent questions are stable or whether they are affected by information contained in the first question. A preliminary analysis, reported in McFadden, Schwarz, and Winter (2004), shows that respondents indeed use information contained in the first question to update their priors about how likely it is they will be admitted to a nursing home, confirming results from the earlier literature. More interestingly, their reports on the factual question of

whether they have already purchased nursing home insurance are also affected by survey design. These effects resemble the mere measurement effects discussed above.

Even though the response effects are small, they indicate that much more research is needed on how survey design and response behavior interact. From a methodological viewpoint, this experiment illustrates how established knowledge from psychology and survey research can be used to design experiments that allow econometricians to identify aspects of survey response behavior and of the underlying expectations, intentions, and subsequent behavior that would otherwise be hard to analyze.

Strategies for the econometric analysis of survey response behavior

Complementary to the development of new survey and question designs such as vignettes, proxy reports, and embedded experiments is the development of econometric techniques. There exist a number of studies that combine econometric models of the survey response behavior and of the underlying intentions or market behaviors; examples are Klein and Sherman (1997), Harris and Keane (1998), Hsiao and Sun (1999), and McFadden, Schwarz, and Winter (2004).

One approach to capture response effects that is already used in many applications is to allow for individual heterogeneity by making the parameters of the model random, as in the mixed logit model by McFadden and Train (2000). For instance, one could test whether respondents who are less certain about their response and who thus are more likely affected by features of the survey questionnaire show greater variability of preferences. Such a test could be implemented either in a between-person setting or within subjects using repeated tasks in stated preference experiments that vary in difficulty and design. The approach of allowing for parameter heterogeneity is very

flexible and can capture a variety of survey response effects; it can also be combined with a latent model of the survey response process; see also Louviere, *et al.*, (2003).

At a more fundamental level, econometric research will focus on the assumptions that provide identification of the parameters of interest in the presence of certain forms of response errors, in particular in nonlinear models (Lewbel, 2004; Matzkin, 2004). Lewbel (2004) shows how instruments can be used in the semi-parametric estimation of nonlinear econometric models. From an applied perspective, the crucial issue here is whether such instruments exist. Here, knowledge from psychology and survey research on the process of providing responses to survey question can be exploited fruitfully. In the nonparametric approach developed by Matzkin (2004), the main assumption is not the existence of an (observed) instrument, but the existence of some (potentially unobserved) exogenous variation that can be used in ways similar to using an observed instrument. Again, the key practical issue is whether such exogenous variation exists – or, from a constructive perspective, whether it can be induced by a clever survey design as the use of embedded experimental variation. Much research is needed in these areas, but the potential for cooperation between econometricians, market researchers and survey researchers seems large.

A final issue of high relevance for applied research is whether the biases generated by the survey response process affect primarily the mean of the outcome distribution or whether higher moments may be affected as well.

Conclusion

Survey response errors need to be taken into account in applied econometric studies of consumer behavior, in particular since many of the phenomena that have been documented in the literature cannot be easily captured within the classical measurement error framework. Taking survey re-

response errors into account means that we may need to adjust the objective of the analysis: In many cases, it may simply be impossible to recover the “correct” responses at the individual level when all we have are poorly measured survey data. However, econometric or statistical techniques can still allow us to estimate structural parameters that can be of interest, such as the features of the distribution of a quantity across the sample. In many cases, this is all the information that is needed for policy analysis or marketing-decision making (i.e., whether to launch a new product or not).

We have shown that future research needs to focus on two major issues. First, we need to refine the way(s) how potentially difficult survey questions can be asked. It is unlikely that some difficult concepts or sensitive attitudes will ever be measured without error, but techniques such as repeated measurement and/or proxy reports, the use of vignettes and other experimental designs with hypothetical choice situations, and the use of embedded survey experiments more generally can provide additional variation and instruments to be used in structural econometric models. Second, to the extent that errors cannot be avoided with better questions, psychologists and survey researchers can inform econometricians and marketing analysts about the underlying cognitive processes, resulting in joint structural models of both response behavior and the underlying economic behavior. Third, while the development of semiparametric and nonparametric estimators is proceeding rapidly, obtaining instruments and sources of exogenous variation in survey data seems crucial. Here, econometricians can inform survey designers about which sources of variation seem to be particularly helpful in the subsequent statistical analysis. In summary, the analysis of survey response behavior will benefit most from interdisciplinary research by psychologists, survey researchers, market researchers, economists, and econometricians.

References

- Adams, P., M. D. Hurd, D. McFadden, A. Merrill, and T. Ribeiro (2003). "Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status," *Journal of Econometrics*, 112(1), 3–56.
- Alexander, C. S., and H. J. Becker (1978). "The use of vignettes in survey research," *Public Opinion Quarterly*, 42(1), 93–104.
- Allais, M. (1953). "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine," *Econometrica*, 21(4), 503–546.
- Banks, J., A. Kapteyn, J. P. Smith, and A. van Soest (2004). "International comparisons of work disability," Working Paper No. 155, Labor and Population Group, RAND, Santa Monica, CA.
- Bemmaor, A. C. (1995). "Predicting behavior from intention-to-buy measures: The parametric case," *Journal of Marketing Research*, 32(May), 176–191.
- Ben-Akiva, M., D. McFadden, M. Abe, U. Böckenholt, D. Bolduc, D. Gopinath, T. Morikawa, V. Ramaswamy, V. Rao, D. Revelt, and D. Steinberg (1997). "Modeling methods for discrete choice analysis," *Marketing Letters*, 8(3), 273–286.
- Ben-Akiva, M., D. McFadden, T. Gärling, D. Gopinath, D. Bolduc, A. Börsch-Supan, P. Delquié, O. Larichev, T. Morikawa, A. Polydoropoulou, and V. Rao (1999). "Extended framework for modeling choice behavior," *Marketing Letters*, 10(3), 187–203.
- Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Börsch-Supan, D. Brownstone, D. S. Bunch, A. Daly, A. de Palma, D. Gopinath, A. Karlstrom, and M. Munizaga (2002). "Hybrid choice models: Progress and challenges," *Marketing Letters*, 13(3), 163–175.

- Bergadaà, M. M. (1990). "The role of time in the action of the consumer," *Journal of Consumer Research*, 17(3), 289–302.
- Bound, J., C. Brown, and N. Mathiowetz (2001). "Measurement error in survey data," In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Volume 5, 3705–3843. Amsterdam: Elsevier.
- Bruine de Bruin, W., P. S. Fischbeck, N. A. Stiber, and B. Fischhoff (2002). "What number is Fifty-Fifty? Redistributing excessive 50% responses in elicited probabilities," *Risk Analysis*, 22(4), 713–723.
- Buchanan, B., and G. Morrison (1987). "Sampling properties of rate questions with implications for survey research," *Marketing Science*, 16(3), 286–298.
- Chandon, P., V. G. Morwitz, and W. J. Reinartz (2005). "Do intentions really predict behavior? Self-generated validity effects in survey research," *Journal of Marketing*, 69(2), 1–14.
- Cowing, T., and D. McFadden (1984). *Microeconomic modeling and policy analysis: Studies in residential energy demand*. Academic Press: New York.
- Crossley, T. F., and S. Kennedy (2002). "The reliability of self-assessed health status," *Journal of Health Economics*, 21(4), 643–658.
- Davidson, D., P. Suppes, and S. Siegel (1956). *Decision Making: An Experimental Approach*. Stanford, CA: Stanford University Press.
- Dex, S. (1995). "The reliability of recall data: A literature review," *Bulletin de Méthodologie Sociologique*, No. 49, December, 58–89.
- Dominitz, J., and C. F. Manski (1997). "Using expectations data to study subjective income expectations," *Journal of the American Statistical Association*, 92(439), 855–867.

- Dominitz, J., and C. F. Manski (2004). "How should we measure consumer confidence?" *Journal of Economic Perspectives*, 18(2), 51–66.
- Feldman, J. M., and J. G. Lynch Jr. (1988). "Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior," *Journal of Applied Psychology*, 73(3), 421–435.
- Gan, L., M. D. Hurd, and D. McFadden (2003). "Individual subjective survival curves," Working Paper No. 9480, National Bureau of Economic Research (NBER), Cambridge, MA.
- Gardiner, J. M., and A. Richardson-Klavehn (2000). "Remembering and Knowing," in E. Tulving and F. I. M. Craik (Eds.): *Handbook of Memory*, 229–244. New York: Oxford University Press.
- Granbois, D. H., and J. O. Summers (1975). "Primary and secondary validity of consumer purchase probabilities," *Journal of Consumer Research*, 1(4), 31–38.
- Harris, K. M., and M. P. Keane (1999). "A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data," *Journal of Econometrics*, 89(1–2), 131–157.
- Hill, D., M. Perry, and R. J. Willis (2004). "Estimating Knightian uncertainty from probability questions on the HRS," unpublished manuscript, University of Michigan.
- Hsiao, C., and B.-H. Sun (1999), "Modeling response bias – with an analysis of the demand for an advanced electronic device," *Journal of Econometrics*, 89(1–2), 15–39.
- 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.

- Hurd, M. D., and K. McGarry (1995). "Evaluation of the subjective probabilities of survival in the Health and Retirement Study," *Journal of Human Resources*, 30, Supplement, S268–S292.
- Hurd, M. D., and K. McGarry (2002). "The predictive validity of subjective probabilities of survival," *Economic Journal*, 112(482), 966–985
- Infosino, W. J. (1986). "Forecasting new product sales from likelihood of purchase ratings," *Marketing Science*, 5(4), 372–384
- Kalwani, M. U., and A. J. Silk (1982). "On the reliability and predictive validity of purchase intention measures," *Marketing Science*, 1(3), 243–286.
- King, G., C. Murray, J.A. Salomon, and A. Tandon (2004). "Enhancing the validity and cross-cultural comparability of measurement in survey research," *American Political Science Review*, 98(1), 191–207.
- Klein, R., and R. Sherman (1997). "Estimating new product demand from biased survey data," *Journal of Econometrics*, 76(1-2), 53–76.
- Lewbel, A. (2004). "Simple estimators for hard problems: Endogeneity in binary choice and other nonlinear models," unpublished manuscript, Boston College.
- Louviere, J. J., D. A. Hensher, and J. D. Swait (2000). *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge University Press.
- Louviere, J. J., D. Street, R. Carson, A. Ainslie, J. R. Deshazo, T. Cameron, D. Hensher, R. Kohn, and T. Marley (2002). "Dissecting the random component of utility", *Marketing Letters*, 13(3), 163–176.
- Manski, C. F. (1999). "Analysis of choice expectations in incomplete scenarios," *Journal of Risk and Uncertainty*, 19(1-3), 49–65.

- Manski, C. F. (2004). "Measuring expectations," *Econometrica*, 72(5), 1329–1376.
- Matzkin, R. (2004). "Unobservable instruments," unpublished manuscript, Northwestern University.
- McFadden, D. (2004). "Foundations of economic survey research," unpublished manuscript, University of California, Berkeley.
- McFadden, D., N. Schwarz, and J. Winter (2004), "Measuring perceptions and behavior in household surveys," paper presented at the Annual Meeting of the American Economic Association, San Diego, CA.
- McFadden, D., and K. Train (2000), "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15(5), 447–470.
- McGarry, K. (2004). "Health and retirement: Do changes in health affect retirement expectations?" *Journal of Human Resources*, 39(3), 624–648.
- Menon, G., O. Raghurir, and N. Schwarz (1997). "How much will I spend? Factors affecting consumers' estimates of future expense." *Journal of Consumer Psychology*, 6(2), 141–164.
- Morwitz, V. G., and G. J. Fitzsimons (2004). "The mere measurement effect: Why does measuring intentions change actual behavior?" *Journal of Consumer Psychology*, 14(1–2), 64–74.
- Molinari, F. (2005): "Cognitive processes and self/proxy response status: Evidence from HRS 2000," unpublished manuscript, Cornell University.
- Poterba, J. and L. Summers (1986). "Reporting errors and labor market dynamics," *Econometrica*, 54(6), 1319–1338.
- Schwarz, N. (2003). "Self-reports in consumer research: The challenge of comparing cohorts and cultures," *Journal of Consumer Research*, 29(4), 588–594.

- Schwarz, N., R. M. Groves, and H. Schuman (1998). "Survey methods," in N. Schwarz, R. Groves, and H. Schuman (Eds.), *Handbook of Social Psychology*, 4th Edition, 143–179, New York: McGraw-Hill.
- Schwarz, N., H. J. Hippler, B. Deutsch, and F. Strack (1985). „Response categories: Effects on behavioral reports and comparative judgements,” *Public Opinion Quarterly*, 49(3), 388–395.
- Smith, V. K., D. H. Taylor, and F. A. Sloan (2001). "Longevity expectations and death: Can people predict their own demise?" *American Economic Review*, 91(4), 1126–1134.
- Tourangeau, R., L. J. Rips, and K. Rasinski (2000). *The Psychology of Survey Response*. New York and Cambridge, UK: Cambridge University Press.
- Tversky, A. and D. Kahneman (1974). "Judgement under uncertainty: Heuristics and biases," *Science*, 185(4157), 1124–1131.
- Wansbeek, T., and E. Meijer (2000). *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier.
- Wason, K., Polonsky, M., and Hyman, M. (2002). "Designing Vignette Studies in Marketing," *Australasian Marketing Journal*, 10(3), 41–58.
- Winter, J. (2002). "Bracketing effects in categorized survey questions and the measurement of economic quantities," Discussion Paper No. 02-35, Sonderforschungsbereich 504, University of Mannheim.