Institute for
Fiscal Studies

The measurement of household consumption expenditures

IFS Working Paper W14/07

Martin Browning
Thomas Crossley
Joachim Winter

ESRC
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

The Measurement of Household Consumption Expenditures

Martin Browning,[1] Thomas. F. Crossley,[2,3] and Joachim Winter[4]

Prepared for: The Annual Review of Economics

[1]Department of Economics, University of Oxford, Oxford OX1 3UQ, United Kingdom

[2]Department of Economics, University of Essex, Colchester CO4 3SQ, United Kingdom; email: tcross@essex.ac.uk

[3]Institute for Fiscal Studies, London WC1E 7AE, United Kingdom

[4]Department of Economics, University of Munich, Munich 80539, Germany

Abstract:

Household-level data on consumer expenditures underpins a wide range of empirical research in modern economics, spanning micro- and macroeconomics. This research includes work on consumption and saving, on poverty and inequality, and on risk sharing and insurance. We review different ways in which such data can be collected or captured: traditional detailed budget surveys, less onerous survey procedures that might be included in more general surveys, and administrative or process data. We discuss the advantages and difficulties of each approach and suggest directions for future investigation.

1. INTRODUCTION

Consider the follow research questions. Do tax-favored savings accounts generate net new savings or simply result in the reallocation of portfolios (Attanasio & Deleire 2002)? How persistent, and how heterogeneous, are the earnings processes that individuals and households face (Guvenen 2007)? How well insured are households against job loss (Stephens 2001, Browning & Crossley 2008)? More generally, how well insured are households against income shocks and changes in the economic environment (Attanasio & Davis 1996, Blundell et al. 2008)? How has the well-being of the poor evolved over decades (Meyer & Sullivan 2012)? How do consumption and saving respond to interest rates (Attanasio & Weber 1993)? Do house price movements have a causal effect on consumer spending (Attanasio & Weber 1994, Campbell & Cocco 2007)? How effective are tax rebates and economic stimulus payments for fiscal stimulus (Parker et al. 2013)? How can we have confidence in the heterogeneous agent equilibrium models we use for macro policy analysis (Krueger et al. 2010)?

What these questions have in common is that attempts to answer them have rested onhousehold-level data on consumption expenditure. They are critical questions for our understanding of the economy and for the formulation of economic policy, and many more such questions could be added to this list. In this review, we consider how data on consumption expenditure can be obtained.

The traditional source of data for measuring household consumption is a national budget survey. Examples include the Consumer Expenditure Survey (CE) in the United States or the Family Expenditure Survey (FES; now called the Living Standards and Food Survey) in the United Kingdom. However, there is a growing sentiment that such surveys, at least in their current form, do not fully meet researchers' needs. There are several aspects of this. First, budget surveys have a high respondent burden and so typically have a limited (if any) longitudinal component and collect limited information in other domains. Much of the research that requires household consumption expenditure data is greatly facilitated if we observe the consumption of the same household over time and in conjunction with health, labor supply, wealth, or other covariates. Second, evidence suggests that the quality of the data produced by national budget surveys is declining. In some (but not all) countries, response rates are falling, and the correspondence between national budget survey data and national accounts data is deteriorating (Krueger et al. 2010, Barrett et al. 2014). For example, for the period 1986–2007, Barrett et al. (2014) report that the response rate to the CE fell by 11 percentage points (to 74%) and to the FES by 17 percentage points (to 53%) and that the

coverage of a national accounts–based measure of consumption fell by 13 percentage points for the CE (to 71%) and by 16 percentage points (to 67%) for the FES. Recent research on the CE (Sabelhaus et al. 2014) has demonstrated that nonresponse in that survey is differential by income, with very high-income households being underrepresented. This is one plausible source of the declining correspondence to national accounts data.[1]

Clearly, enormous social, economic, and technological changes have altered the ways in which households spend and consume. Examples include changes in the financial autonomy of household members, changes in the organization of the retail sector, growing prosperity, and enormous innovation in the goods and services available to households. It is therefore perhaps not surprising that data-collection methods designed 30 or more years ago exhibit changing performance. On the positive side, new technologies have opened up possibilities for alternative data-collection methods. Examples include Internet panel surveys, product scanners, and electronic banking.

In the past few years, these considerations have motivated a great deal of research and reflection regarding the collection of household expenditure data. In 2009, the US Bureau of Labor Statistics (BLS) initiated the Gemini Project with a five-year mission to redesign its budget survey (the CE). As part of the Gemini Project, the BLS asked the US National Research Council's Committee on National Statistics (CNSTAT) to convene an expert panel to further consider the collection of household expenditure data and to propose redesign options for the CE. That panel reported in 2013 (Natl. Res. Counc. 2013). Similar redesigns or reconsiderations of national budget surveys have been, or are being, undertaken in other countries. In December 2011, the Conference on Research in Income and Wealth considered ways of improving the measurement of consumer expenditures, with the conference papers drawn from around the world collected in an NBER volume (Carroll et al. 2014At the same time, there has been a rapid expansion in the collection and use of consumption data from other sources. Expenditure questions have been included in general purpose household surveys, and the expenditure questions in surveys that already had such content, such as the Panel Study of Income Dynamics (PSID), have been expanded (See Li et al, 2010). Experiments have been conducted with Internet-based surveys and with the construction of household consumption measures from administrative data. There is an increasing use of data from store or home scanners.

How consumption should be measured in the future remains an important open question. In this article, we selectively review the current state of the literature. Naturally, an article-length review imposes limits of scope, and we focus primarily on the measurement of

consumption expenditures at the household level in developed economies. This implies important exclusions. The first regards the level of aggregation. We do not discuss the measurement of aggregate consumption in national accounts. At the same time, it is often natural to think of consumption as fundamentally an individual, rather than a household, activity. The literature on intrahousehold allocation has made very significant progress in recent years; we are unable to cover it here.

A second point is that consumption expenditure is not consumption. We define household expenditure as the nominal money outlay of the household. Household consumption expenditure is then those money outlays directed toward consumption and excludes outlays for investment purposes, as well as gifts to other households and charities. Household consumption is the quantity of goods and services that the household enjoys in a given period. A measure of household consumption expenditure is a primary input into the construction of a measure of household consumption, but that construction also requires, for example, adjustments for the different prices faced by different households and the replacement of expenditures on durable goods with the flows of services from those goods.[2]

Finally, whereas the consumption of particular goods and services is of interest for some kinds of research, our primary focus is the measurement of total household consumption expenditure (on all goods and services) as an input into the measurement of total household consumption.

In what follows, we first focus on the traditional approach to measuring consumption expenditure at the household level, namely budget surveys. We summarize the measurement problems and design choices they pose in Section 2. We then discuss in Section 3 alternative survey-based approaches that try to obtain a measure of total consumption expenditure with many fewer survey items. These methods have lower respondent burden and so are particularly useful with general purpose surveys that collect data in multiple domains (income, health, wealth) or longitudinally. In Section 4, we explore how the emerging availability of administrative and other process-generated data opens up new approaches for expenditure measurement. Section 5 concludes.

Again, our review is necessarily selective. Several complementary reviews are available. Deaton & Grosh (2000) review the methodology of budget surveys with a particular focus on developing countries. Browning et al. (2003) focus particularly on the measurement of consumption in general purpose surveys Crossley & Winter (2014) review research on the methodology of budget surveys in developed countries, highlighting particularly the role of the survey response process, and the potential biases it induces, in

expenditure measurement. Natl. Res. Counc. (2013) surveys the same literature with a particular focus on the CE.

2. BUDGET SURVEYS

Budget surveys are dedicated surveys that attempt to collect a detailed picture of each household's spending in a reference period. Total household consumption expenditure is measured as the sum of reported expenditures across fine categories of consumption expenditure. In this section, we survey the design choices and response problems encountered in such surveys.[3]

2.1. Survey Method and Mode

There are two broad survey strategies for collecting detailed household expenditure data. Either respondents are asked to report how much they spent on different categories of consumption goods in a certain period (the recall approach), or they are asked to keep a diary over a certain period of time in which they record every single expense (the diary approach). Well-known budget surveys use one or both of these methods. Until recently, the national budget survey in Canada was based entirely on recall. In the United States, the CE comprises two surveys (with different respondents), one based on recall (the Quarterly Interview Survey) and one based on diaries (the Diary Survey). The UK national budget survey uses a mix of diaries and recall to collect information from a single set of respondents.

There is a widely held view that diaries are the superior methodology and that recall should only be used (if at all) to collect information on infrequently purchased goods and services. This view is based on significant concerns about recall. The literature on survey response behavior has long recognized that questions that require recalling quantities from memory are difficult to answer (Gray 1955). There is substantial evidence of forgetting: Memory declines with the length of the recall period, leading to underestimation (see Sudman et al. 1996 for a review). An important development in the design of budget surveys was the identification of telescoping as a significant problem with recall expenditure questions (Neter & Waksberg 1964; see also Jacobs & Shipp 1993). Respondents erroneously include in their response expenditures that occurred before the specified recall period, leading to an overestimation of expenditures in that period. Neter & Waksberg (1964) document this phenomenon in the CE (particularly for home alterations and repairs), and the process has been formally modeled (Rubin & Baddeley 1989). The combination of forgetting and telescoping implies that recall answers could lead to either over- or underestimates.

Neter & Waksberg (1964) propose "bounded" recall as a way of minimizing telescoping problems, and this methodology is currently employed in the CE Quarterly Interview Survey. The recall sample is interviewed five times, with data from the first interview discarded; the first interview serves to bound the beginning of the first recall period. For some categories of expenditure, the (normally discarded) data from the first interview suggest significantly higher rates of expenditure for some categories of goods (Silberstein 1990), consistent with telescoping.

Another potential problem with recall surveys of expenditure is the potential for mode effects. The possible modes for a recall expenditure survey include personal (face-to-face) interviews, telephone interviews, and self-administered questionnaires, whereas diaries are by their nature self-administered. All three modes could be based on a paper questionnaire or computer assisted. Recall budget surveys typically involve an interviewer (in either personal or telephone mode) although there have been some experiments with self-completion of recall expenditure surveys, for example, the Consumption and Activities Mail Survey fielded as part of the US Health and Retirement Study (HRS) (Hurd & Rohwedder 2009).

It is well known that survey mode influences response behavior via various channels; Tourangeau et al. (2000) provide a good overview. In the context of budget surveys, a key consideration is the sensitivity or confidentiality of the target quantities (as the presence of an interviewer might increase such concerns). It is also thought that self-administered surveys make it easier for respondents to look up information on hard-to-recall quantities on bills or bank statements. Against this, a present interviewer may aid with the comprehension of survey questions (if the survey protocol allows for this).

Evidence on mode effects in budget surveys is scant. Silberstein & Scott (1991) report evidence that personal interviews result in lower estimates on potentially sensitive goods, such as alcohol. Safir & Goldenberg (2008) analyze variation in the mode of administration of the CE Quarterly Interview Survey. They argue that although telephone interviewing may impact the quality of CE data relative to that obtained by personal visit interviewing, mode effects can be mitigated by using recall aids in both modes, for example, through a mailed information booklet and user-friendly checklists for records and receipts.

Despite these concerns with recall methods, there are reasons to be skeptical of the widespread view that diary methods are superior to recall methods. In principle, perfect compliance with diaries covering a long period would indeed give very good expenditure data. In practice, diaries have a very high response burden, and compliance is far from perfect.

Recognizing that careful completion of an expenditure diary involves high respondent burden, modern budget surveys have only asked respondents to complete diaries covering relatively short periods (one or two weeks). This causes infrequency problems for goods that are purchased irregularly or at regular intervals that exceed the diary period. Reported expenditure will over- or underestimate a household's true rate of expenditure on such categories of goods, depending on whether the diary period captures a purchase. This kind of measurement error may average out in the estimation of population means, but it causes severe problems in the estimation of measures based on dispersion (inequality or poverty) or in situations in which a researcher wants to use consumption as a conditioning variable. It is exactly because of this limitation of diaries that many budget surveys (including the current CE) use a mix of diary and recall questions, with the recall questions intended to capture less frequently purchased categories of goods. Researchers have developed methods to deal with purchase infrequency, particularly in the context of demand estimation, by jointly modeling demands and purchase frequency (see Meghir & Robin 1992, and references therein). Such an approach to modeling, which takes explicit account of the limitations of the data at hand, has much to recommend it (a point we return to in Section 5). However, these models do make assumptions of varying plausibility: regarding the form of demands, that all zeros reflect infrequency and never corner solutions, or the conditional independence of the level of demand and purchase frequency. The difficulties posed by purchase infrequency have not been entirely overcome.

Moreover, even with one- or two-week diaries, there is considerable evidence that expenditure diaries suffer from a great deal of noncompliance. Silberstein & Scott (1991) report that in as many as one-quarter of CE diaries, additional expenses are recorded at the time of diary pickup (interviewers collecting the diary check for completeness). In such cases, the diary survey relies on recall.

In addition, compliance with diaries appears to decline with the duration of recording. Several well-known budget surveys use two week-long diaries. Recorded rates of expenditure are significantly lower in the second week and, within weeks, lower toward the end of the week. Such patterns have been reported in the CE (Silberstein & Scott 1991, Stephens 2003), the Canadian Food Expenditure Survey (Stat. Can. 1999, Ahmed et al. 2006), and the UK FES (Tanner 1998). These patterns are usually attributed to diary fatigue: Respondents stop recording some part of their expenditures during the diary period to reduce their burden (e.g., Silberstein & Scott 1991, Stat. Can. 1999), although another possibility is that completion of the diary affects behavior (see the sidebar, Does Survey Participation Affect Behaviour?).

DOES SURVEY PARTICIPATION AFFECT BEHAVIOR?

The decline in recorded expenditures in diary surveys is widely interpreted as a decline in the accuracy of recording (diary fatigue). However, an alternative possibility is that it is consumption behavior, rather than reporting behavior, that is changing. That detailed recording of behavior (spending) might change behavior is not surprising: Self-monitoring is often advocated as a means to behavioral change (e.g., food diaries for weight loss, expenditure diaries for saving), and there is some evidence that such methods are effective (Burke et al. 2011). If diary surveys affect the behavior of respondents, then the resulting data are not an inaccurate measure of the consumption expenditures of the surveyed households, but those households are no longer representative of the population.
Disentangling changes in response behavior from changes in the underlying behavior is difficult. Kemsley et al. (1980) and McKenzie (1983) consider this issue in the context of the UK FES with inconclusive results. We are aware of no other work on this question in the context of budget surveys particularly, but in more recent work, Zwane et al. (2011) show that survey participation can alter behavior in a number of domains (including health-related behaviors), and Crossley et al. (2014) show that participation in a detailed survey on retirement income needs can change saving behavior.

Thus, in practice, both diary and recall methods suffer from response problems, and it is an empirical question which method provides better data. An obvious issue is how to determine whether data are "better." Recall measures typically have a longer reference period and so suffer less from infrequency and have lower variance. Most categories of expenditure are thought to be underreported, so the most common criterion of comparison is the rate of expenditure, and higher rates of expenditure have been taken to be evidence of less error.[4]

Several studies have shown that recall measures of food expenditure are on average higher than diary measures (Gieseman 1987, Ahmed et al. 2006, Bee et al. 2014). Reports of food expenditure are unlikely to suffer from telescoping (which affects primarily large and irregular purchases). The most likely explanation for this common finding is that diary fatigue and noncompliance are more severe than any forgetting in the recall responses. Silberstein & Scott (1991) make comparisons for a number of items that are collected in both the interview and diary components of the CE. They report that the diary method produces higher expenditure estimates for some categories (apparel, home furnishings), whereas the interview produces higher expenditures for others (entertainment and hobbies). Recently, Bee et al. (2014) show that for many large categories of expenditure, aggregates from the interview survey of the CE come close to national accounts aggregates, and the correspondence between interview survey data and the national accounts has been quite stable over time. They report that the same is not true of diary data of the CE. Based on these

and other considerations, they conclude that the interview survey is preferable to the diary survey in the CE.

Our reading of the evidence is that, as currently implemented in national budget surveys, recall methods work as well as, or better than, diary methods. The recent CNSTAT report (Natl. Res. Counc. 2013) takes the view that there is an insufficient evidence base to compare diary and recall methods. There was a dissent to the main report, and a significant aspect of that dissent is to emphasize the evidence against diary methods. The main CNSTAT report does conclude that the current design of the CE diary method is flawed. However, the conclusion the panel seems to have drawn is that the way forward is to improve diary collection methods rather than to shift emphasis to recall methods. Self-administered data collection, essentially diaries, is an important component of the redesign options put forward in the proposal. However, the proposed diaries are much different from current practice. Considerable use of technology (tablet computers, the Internet, bar-code scanners) is envisioned so that these are "supported" diaries. Implicitly, the panel is suggesting that diaries (or self-administered data collection, more generally), in conjunction with new technology, offer the best prospect for improved data collection. The BLS's own redesign proposal (Edgar et al. 2013) suggests a mix of recall and diary data collection and, similar to the CNSTAT panel report, envisions radically redesigned diaries employing new technology (in particular, the diaries would be web-based).

2.2. Further Design Issues

With either diary- or recall-based data collection, there are numerous additional design details that determine the effectiveness of a budget survey. In this section, we very selectively review some of the issues and evidence. The reviews listed in Section 1 provide further detail.

2.2.1. Response unit.

Even where the focus is on measuring expenditures at the household level, there is a decision to be made regarding whom to collect data from. In a recall survey, data might be collected from one individual, and an attempt might be made to identify the person most knowledgeable about the expenditures of the entire household. This can lead to a number of difficulties. When purchases are made by other individuals, this is a kind of proxy reporting, which may be plagued by inaccuracies and nonresponse. Comerford et al. (2009) demonstrate that respondents often report individual expenditures, even when asked to report household expenditures. Browning et al. (2003) report that nonresponse to household expenditure questions is higher among complex households: those with unrelated adults (sharers) or

multiple generations of adults. Focus group results suggest that proxy reporting problems may be more severe with more finely disaggregated expenditure categories: Respondents may be able to estimate the total spending of other household members but do not know how that spending is allocated across different goods and services (d'Ardenne & Blake 2012). The same focus group research identified another problem. Respondents may interpret "household spending" or even "spending of your household" as meaning only shared expenses or expenditures on those goods and services necessary to run the household.[5] These problems might be overcome by asking multiple members of each household to report expenditures. However, in addition to the obvious issue of cost, such a procedure risks double counting. Many household expenditures are made collectively (e.g., from a joint account); even when one individual makes a transaction, it may be made on behalf of the whole household (particularly for items such as rent or utilities). Thus, survey instructions and fieldwork protocols need to be carefully designed and tested.

The parallel issue with diaries is whether there should be a single diary for the household or individual diaries. Both practices are currently employed with, for example, the current CE Diary Survey using a single household diary and the UK national budget survey using individual diaries for each household member above an age minimum. Evidence on this design choice is mixed (Kemsley & Nicholson 19601, Grootaert 1986, Goldenberg & Ryan 2009). Multiple diaries give higher totals, suggesting that some expenditure is missed with a single diary, but multiple diaries lead to a higher incidence of noncooperation.[6] The BLS's proposal for a redesigned CE includes a move to multiple (individual) diaries.

2.2.2. Reporting period.

Whether data are collected by recall or diary, the measurement of any flow variable, such as expenditure or consumption, requires a choice of reporting period. There are difficult trade-offs here. Longer reporting periods may lead to underreporting (through more forgetting, in the case of recall questions, or greater diary fatigue, in the case of diaries). Against this, shorter reporting periods exacerbate problems arising from purchase infrequency. The severity of these problems varies across expenditure categories with different frequency and salience of purchases (Silberstein & Jacob 1989, Hurd & Rohwedder 2009), suggesting that the optimal reporting period likely varies by category of expenditure. Hurd & Rohwedder (2009) also note that when offered a choice, respondents tend to choose a longer reference period for less frequently purchased items. Nevertheless, several studies have suggested that common reporting periods work quite well. Experiments undertaken by the Indian National Sample Survey Organization and reported in Deaton & Kozel (2005) suggest that, in that

country, a common 30-day recall period is not dominated by other designs. Early experiments in Canada (McWhinney & Champion 1974) suggest that a design with an annual recall of all categories of expenditure in conjunction with a balance edit (discussed below) works well. The Canadian national budget survey had this design until quite recently. It is also important to bear in mind that using different reporting periods for different goods and services complicates the construction of aggregates such as total household expenditure.

With a recall design, a further issue that arises is whether the respondents should be asked about a specific period (e.g., last month or last week) or about a typical or usual period of a given length (diaries of course record expenditures in a specific period). The motivation for the typical or usual formulation is that it may diminish inaccuracies arising from purchase infrequency. At the same time, it also likely affects the way in which respondents construct answers (e.g., the mix of recall and estimation or the method of estimation). Cognitive interviews suggest that respondents are heterogeneous both with respect to their estimation strategies and with respect to their understanding of the term typical (Edgar 2009).

Overall, the literature on length and nature (usual versus specific) does not yet give strong guidance on practice. What the literature does establish is that these design details can have significant effects on reporting behavior and very significant follow-on effects on objects of interest, such as reporting rates (Deaton & Kozel 2005, Angrisani & Kapteyn 2014).

2.2.3. How many categories?

The design of budget surveys is to collect information on expenditure in a comprehensive set of categories of goods and services and then, if an aggregate (total expenditure or total consumption expenditure) is required, to aggregate across categories for each household. An obvious question is, How many categories should there be (or, equivalently, how fine should the categories be)? It is worth noting that historically, the design of national budget surveys has been very much driven by the requirements of price-index construction. So the level of disaggregation was not determined by what might be necessary or optimal for constructing a measure of household consumption.

There is good experimental evidence that finer disaggregation leads to higher totals for household expenditure (see Jolliffe 2001 and Pradhan 2009 for developing countries and Winter 2004 for evidence from a developed country). Against this, focus groups indicate that respondents find finer disaggregation not only a greater response burden but also more intrusive (d'Ardenne & Blake 2012). More disaggregated designs may lead to higher totals for responding units but also, in some contexts, to more refusals or noncompletion.

11

It is not only the absolute number of expenditure categories that matters but also the structure of those questions. Expenditure surveys sometimes have many repetitive blocks of identical questions, each with a filter question. Respondents might be asked first, "Have you spent any money on [x] in the last month?" If they answer no, the survey proceeds to the next category of expenditure, whereas if they answer yes, follow-up questions probe further details of the expenditure. Recent research (Kreuter et al. 2011, 2012) shows that such designs can lead to motivated underreporting: Through the course of many expenditure categories, respondents learn to answer no to the filter question to avoid burdensome follow-up questions. This is thought to be a significant problem with the current CE design (Natl. Res. Counc. 2013).

Finally, data from less disaggregated designs have been shown to capture well the important life-cycle and time-series properties of consumer expenditure (Hurd & Rohwedder 2014). It may be that for research purposes, a less disaggregated design is desirable. This is another area in which more evidence would be welcome.

2.2.4. Open or closed response formats.

Other aspects of response format also matter. The literature has considered whether expenditure questions should be open-ended (inviting a numerical response) or closed-response formats (ranges or brackets). The arguments against open formats are that they have a higher respondent burden and typically yield rounded or heaped responses (Pudney 2007). This coarsening of the data is likely heterogeneous and nonrandom. Against this, closed formats such as brackets may induce respondents to use certain response heuristics that lead to systematic biases, similar to the well-known anchoring bias (Winter 2002), and in any case, the data are less informative (Manski & Tamer 2002). Systematic biases related to anchoring occur when item nonresponse (to open-ended questions) is follow-up by bracketing questions (sometimes known as unfolding brackets) (see, e.g., Hurd et al. 1998, van Soest & Hurd 2008). It is probably fair to say that no consensus has been reached on this issue, although our own preference is for open-ended expenditure questions rather than ranges or brackets.

2.2.5. In-survey correction of response error.

The final budget survey design issue we consider is whether response errors might be corrected in the field, in particular by using the logic of the budget constraint. This is particularly attractive for computer-assisted surveys, whether those surveys be face-to-face, telephone, or Internet based.

Until recently, the national budget survey in Canada has been based on annual recall with a field editing procedure called the balance edit.[7] Total reported expenditures are compared to reported net income and changes in money balances. Households that are too far out of balance are asked to review their responses. Brzozowski & Crossley (2011) demonstrate that the main effect of this procedure is improved income reports, particularly among low-income households.

Fricker et al. (2014) report on a small experiment exploring the feasibility of a balance edit in a revised CE. In their experiment, this cash flow reconciliation improved the balance but did not eliminate imbalance (although, in their experiment, expenditures and income were reported over different intervals, so the budget constraint did not need to hold exactly). Follow-up cognitive interviews revealed very heterogeneous reactions to the balance edit among subjects. Some found the procedure helpful, whereas others—particularly those whose expenditures exceeded their incomes—found it threatening.

Hurd & Rohwedder (2010) report on the use of a procedure similar to a balance edit in an Internet panel. After reporting expenditures in a set of categories, respondents were presented with a reconciliation screen in which they were given the total of their responses and asked to review and correct the information they had provided. The authors report that approximately 3% of entries were corrected, and the corrections reduced the number of outliers in the data. A similar procedure has been implemented to good effect in the Dutch Internet panel module fielded by Cherchye et al. (2012).

These pieces of evidence suggest that the use of a budget constraint to improve expenditure data quality is promising and deserves further investigation. More generally, although much has been learned about the design and conduct of budget surveys, it is clear that a great deal remains to be investigated.

3. OTHER SURVEY-BASED APPROACHES

Even if redesign and the incorporation of new technologies can make budget surveys much more effective, such surveys will continue to have high respondent burden. As a consequence, they will likely have limited information from other domains and limited longitudinal information.[8] To be able to study the dynamics of consumption, including responses to income and wealth shocks, and the interaction between consumption and labor supply, health, and other variables, we need to be able to collect good measures of consumption expenditure in general household surveys (particularly longitudinal surveys). Several approaches based on small sets of survey items have been pursued. We first consider

attempts to get at total consumption expenditure either through a direct question or through the logic of the budget constraint, by asking about income and savings. We then consider approaches based on collecting information about expenditure on a subset of goods and services. We note that many of the issues discussed in the previous section—proxy reporting, the definition of the household, optimal reference periods—matter here too, but we do not repeat the discussion.

3.1. One-Shot Questions

The simplest approach is to simply ask households for their total consumption expenditure in a given reporting period with a single question (the one-shot question). As one example, the Italian Survey on Household Income and Wealth has asked the following:

> What was your family's <u>average monthly</u> expenditure in 1995
>
> for all <u>consumption</u> items?
>
> Consider all expenses, including food, but excluding those for:
>
> housing maintenance; mortgage installments; purchases of
>
> valuables, automobiles, home durables and furniture; housing
>
> rent; insurance premiums.

A similar question has been asked in the Canadian Out of Employment Panel Survey, a survey of job leavers in the 1990s (Browning et al. 2003); in the Spanish Survey of Household Finances (Velilla 2014); in pilot studies in the US Asset and Health Dynamics among the Oldest Old (AHEAD) sample (Hurd et al. 1998); the CentER Data Panel in the Netherlands (Winter 2004); and Understanding Society, the main UK longitudinal household study (Bottazzi et al. 2008, Gray et al. 2008). The intellectual underpinning for such a question derives from the assumption of intertemporal separability made in much (but not all) modeling of consumption and saving behavior. If preferences are separable across periods, then one way to implement an optimal plan is to use two-stage budgeting. In the top stage, expenditure is allocated to the period (total expenditure), and in the bottom stage, total expenditure is allocated to individual goods. If a household uses such a mechanism to implement the optimal allocation under separability, then total expenditure should be a salient variable for them.

The experience with this strategy is mixed. In all but one case, one-shot questions have generated very high response rates: Respondents are able and willing to provide an answer.[9] Browning et al. (2003) suggest that one aspect of this is that respondents view questions about broad categories of expenditure as being less sensitive than comparable

income questions. This interpretation has been corroborated in focus group research (d'Ardenne & Blake 2012).

One-shot questions have also generated usable data. Relationships between total expenditure reports and household demographic characteristics lined up well with patterns in budget survey data (Browning et al. 2003, Bottazzi et al. 2008), and data from one-shot questions have been successfully employed in a number of research papers (e.g., Browning & Crossley 2001, 2008). These data contain a significant and useful signal. Against this, in all of the examples cited above, one-shot questions gave significantly lower estimates of total consumption expenditure than more disaggregated data collection. Focus groups and cognitive interviews have also documented problems with one-shot questions (Gray et al. 2008, d'Ardenne & Blake 2012). These lines of evidence suggest that recall of total expenditure is challenging for many respondents, although they appear to use a variety of methods for estimation, depending on their personal circumstances (more on this below). Our assessment is that the jury is still out on one-shot questions. With collaborators in the United Kingdom, we are awaiting data from a new survey experiment with one-shot questions in the Innovation Panel of the main UK longitudinal household survey.

3.2. Income Minus Saving

An alternative approach is to not collect expenditure data at all, but rather to use the logic of budget constraints to impute consumption expenditure from data on income and saving, or changes in wealth. It is an identity that

$$x_{t,h} \equiv y_{t,h} - s_{t,h},$$

where x is expenditure, y is net income, s is saving (a flow, possibly negative in the case of dissaving), and t and h index time and households. Thus, data on income and saving for a common period allow the calculation of total expenditure. Saving flows (acquisition and sales of assets, borrowing and retirement of debt) are not often measured in surveys. Longitudinal data on wealth levels might be used instead. A simple intertemporal budget constraint is

$$w_{t+1,h} = (w_{t,h} + y_{t,h} - x_{t,h})(1 + r_{t,h}),$$

where w is wealth and r is the interest rate. Inverting the budget constraint gives

$$x_{t,h} = y_{t,h} - [(1 + r_{t,h})^{-1} w_{t+1,h} - w_{t,h}].$$

This might be approximated by

$$x_{t,h} \approx y_{t,h} - [w_{t+1,h} - w_{t,h}].$$

We note that with this approximation, capital gains are misclassified as active saving so that expenditure is underestimated if there are gains (and overestimated if there are losses).

Ziliak (1998) constructs a measure of $x_{t,h} \approx y_{t,h} - [w_{t+1,h} - w_{t,h}]$ in the PSID. As wealth is measured only every five years in the PSID, he further imputes annual wealth levels from capital income. Several studies have implemented this approach with administrative data, beginning with Browning & Leth-Petersen (2003). We return to these papers when we take up the use of administrative data more generally in the next section.

Recent UK focus group evidence (d'Ardenne & Blake 2012) suggests that, when asked a one-shot question on total expenditure, many (but by no means all) respondents follow a heuristic based on the budget constraint: They actually work out an answer by beginning with income and adjusting for savings. This finding suggests the possibility of designing a small number of questions about net income and saving flows as an effective way to elicit total consumption expenditure in a survey.

However, there are several problems with a strategy of using survey questions on income and saving or wealth changes to get total consumption expenditure. First, even if total expenditure can be identified in this way, it may contain expenditures (e.g., investments or gifts) that do not belong in consumption expenditure, so it may be necessary to collect additional information to make appropriate adjustments to total expenditure.[10]

Second, whereas the income minus savings heuristic is very natural for some respondents (e.g., those with high and stable incomes), it is not natural for others (including lower-income households with unstable and varied income sources). Although in principle it might be possible to target different kinds of questions for different respondents, focus group evidence (d'Ardenne & Blake 2012) suggests that it may be difficult to identify these groups quickly during a survey. The same focus group evidence also suggests that respondents whose expenditures exceed their incomes find questions about saving flows or changes in wealth intrusive (they seem to find it uncomfortable to report that their outlays exceed their incomes).

Finally, an important use of longitudinal information about consumption expenditure is research on the response of consumption expenditure to income and wealth shocks. Consider a regression of the form

$$\Delta x_{t,h} = \alpha + \beta \Delta y_{t,h} + \gamma \Delta w_{t,h} + u_{t,h},$$

where h indexes households, and $u_{t,h}$ is an error term. If consumption expenditure is derived from information on income and changes in wealth, this becomes

$$\Delta[y_{t,h} - (w_{t+1,h} - w_{t,h})] = \alpha + \beta \Delta y_{t,h} + \gamma \Delta w_{t,h} + u_{t,h},$$
$$\Delta y_{t,h} - \Delta w_{t+1,h} + \Delta w_{t,h} = \alpha + \beta \Delta y_{t,h} + \gamma \Delta w_{t,h} + u_{t,h}.$$

Measurement error in income or wealth will generate spurious correlations between the left- and right-hand sides of this equation and seriously bias naïve estimates of such an equation.

We conclude that the $y_{t,h} - \Delta w_{t,h}$ approach may be useful with administrative data, for which concerns about the intrusiveness of questions and about measurement error in income may be mitigated, but it is less likely to find a wide application in surveys.

An alternative set of approaches asks a short list of questions about a subset of expenditure categories and then uses this information to impute total consumption expenditure or to estimate moments of total consumption expenditure, or even to directly estimate preference parameters or other objects of interest. These approaches are motivated by the notion that there are some goods for which households seem better able to estimate or recall expenditure. Food expenditure is often thought to be well reported. Other categories of expenditure might be well reported if records can be consulted (e.g., rent or utilities). Several longitudinal household surveys, including the PSID and the British Household Panel Survey, have long collected information on a small number of expenditure categories, including food. A common feature of these approaches is that they require an external source of information on the relationships between total consumption expenditure and expenditure on particular goods and services. The natural source of such information is a traditional budget survey.

3.3. Imputing Total Expenditure from a Subset of Categories

Skinner (1987) proposes imputing the total consumption expenditure of PSID respondent households, on the basis of the limited expenditure questions in the PSID and information from the CE. His procedure is essentially statistical matching. Observations on total consumption expenditure from the CE are imputed to the PSID using a linear regression:

$$\text{CE}: \; x_{t,h} = \beta_0 + \sum_j x_{t,h}^j \beta_j + v_{t,h},$$
$$\text{PSID}: \; \hat{x}_{t,h} = \hat{\beta}_0 + \sum_j x_{t,h}^j \hat{\beta}_j.$$

where j indexes goods in the small subset used for predition. Skinner's predictors are food at home, food out, utilities, house value, rent (imputed rent based on house value in the case of homeowners), and number of vehicles owned. He assesses the stability of this relationship in the CE over time and suggests that a single set of coefficients is adequate.

The procedure has been much applied, both with the PSID and with longitudinal data from other countries, and has sometimes been extended to allow for more flexible functional forms (Palumbo 1999, Battistin et al. 2003). Ziliak (1998) compares this measure to the income minus saving ($y_{t,h} - \Delta w_{t,h}$) measure he developed for the PSID (described above). He reports evidence that the measurement error in the $y_{t,h} - \Delta w_{t,h}$ measure is more complicated but argues that the advantage of this measure is that it is entirely internal to the PSID and does not depend on relationships that must be estimated in the CE and that might change over time as, for example, relative prices change. He further shows that both the Skinner measure and the $y_{t,h} - \Delta w_{t,h}$ measure lead to rejections of the permanent income hypothesis in the PSID, whereas tests employing food expenditure alone as a proxy for consumption do not.

Blundell et al. (2004, 2008) develop and apply a very similar method. The key difference is that Blundell et al. start from an explicit recognition that the relationship between total consumer expenditure and expenditure on a particular good is an inverse Engel curve. That is, they reinterpret the imputation equation as a well-studied structural economic relationship. From theory and accumulated experience with modeling consumer demands, we know that the coefficients of the imputation equation must depend on relative prices and likely also vary with household demographics.

Blundell et al. (2004, 2008) start with a general food Engel curve of the form

$$\tau(x_{t,h}^{f}) = Z'\alpha + \gamma\phi(x_{t,h}) + e_{t,h},$$

where $\tau(\cdot)$ and $\phi(\cdot)$ are known monotonic functions, and $x_{t,h}^{f}$ is food expenditure. We note that $\phi(x_{t,h})$ rather than $x_{t,h}$ may be the object of interest (e.g., modeling may require a measure of the logarithm of total consumption expenditure). Z is a matrix of additional variables, including relative prices and household demographics. With estimates of the Engel curve from the CE in hand, one can estimate $\phi(x_{t,h})$ and $x_{t,h}$ as

$$\hat{\phi} = \frac{1}{\hat{\gamma}}(\tau(x_{t,h}^{f}) - Z'\hat{\alpha}),$$

$$\hat{x}_{t,h} = \phi^{-1}[\frac{1}{\hat{\gamma}}(\tau(x_{t,h}^{f}) - Z'\hat{\alpha})].$$

Blundell et al. (2004) provide a careful assessment of this procedure. They ask, in particular, when are the mean ($M[\hat{x}]$) and variance ($V[\hat{x}]$) of $\hat{x}_{t,h}$ consistent estimators of the mean ($M[x]$) and variance ($V[x]$) of $x_{t,h}$ (and similarly for monotonic functions of $\hat{x}_{t,h}$ and $x_{t,h}$)? A well-known consideration in the estimation of Engel curves is that total consumption

expenditure may be endogenous (because it is jointly determined with demands) or measured with error. Either problem means that ordinary least squares (OLS) estimation of the Engel curve will result in inconsistent parameter estimates. Instrumental variables (IV) estimation (with a suitable instrument for total consumption expenditure) can mitigate these problems. An important point that Blundell et al. make is that $M[\hat{x}]$ is a consistent estimator of $M[x]$, even if $\hat{x}_{t,h}$ is based on (inconsistent) OLS estimates of the Engel curve, but the same is not true for variances. Blundell et al. show that $V[\hat{x}]$ is an inconsistent estimator of $V[x]$, whether the Engel curve is estimated by OLS or IV, but when IV is employed, $V[\hat{x}]$ converges to $V[x]$ plus an additive term. If that additive term is constant over time, then $V[\hat{x}]$ may provide a good guide to the evolution of $V[x]$ over time. This does not occur if the Engel curve is estimated by OLS and suffers from endogeneity or measurement-error bias. In those cases, $V[\hat{x}]$ and $V[x]$ will grow at different rates. Thus, there is a good case for careful estimation of the Engel curve, with attention to the usual problems of empirical demand analysis, even though it is prediction, rather than the parameters of the Engel curve, that we are interested in.

Attanasio and Pistaferri (2014) use a version of this procedure to impute total consumption expenditure in earlier years of the PSID using, not the CE, but the more complete measure of consumption expenditure available in the PSID after 1997. An advantage of this is that they can directly compare imputed and observed measures over the post 1997 period. They show that the movements in the standard deviation of the logarithm of consumption based on the imputed measure match well with movements in the standard deviation of the logarithm of consumption based on in the observed measure.

What the analysis in Blundell et al. (2004) highlights is that the best linear predictor of $x_{t,h}$ may not be sufficient if what we are interested in is higher moments (or other nonlinear functions) of total consumption expenditure in a sample. This leads naturally to alternative approaches that seek consistent estimates of the objects of interest directly.

3.4. Estimating Moments of the Total Expenditure Distribution

The above methods all attempt to estimate total consumption expenditure for each household, as an input to further analysis. If a feature of the distribution of total consumption expenditure is the object of interest, then it may be possible to estimate this well from error-ridden measures of household total expenditure (without producing a "good" estimate of total expenditure for any particular household). For example, studies of inequality are often interested in the variance of the logarithm of total consumption expenditure.

Suppose data are available on expenditure in some subset of goods and services. Consider an approximately linear-in-logs Engel curve that relates expenditure on a specific good ($x_{t,h}^i$) to the target variable, total consumption expenditure ($x_{t,h}$):

$$\log x_{t,h}^i = \alpha_i \log x_{t,h} + \eta_i(x_{t,h}) + e_{t,h}^i.$$

The parameter α_i captures the income elasticity of good i if the double log form is correct; luxuries have $\alpha_i > 1$, and necessities have $\alpha_i < 1$. The variable $\eta_i(x_{t,h})$ is the approximation error from using the log-log form. The variable $e_{t,h}^i$ captures heterogeneity in tastes. Define the measurement error for good i (as a measure of total consumption expenditure) as

$$\varepsilon_{t,h}^i \equiv x_{t,h}^i - x_{t,h} = (a_i - 1)x_{t,h} + \eta_i(x_{t,h}) + e_{t,h}^i$$

so that

$$x_{t,h}^i \equiv x_{t,h} + \varepsilon_{t,h}^i.$$

Then the covariance of expenditure (in logarithms) on two goods, $C[x^1 x^2]$, can be used to estimate the variance of the logarithm of total consumption expenditure, $V[x]$:

$$C[x^1 x^2] = V[x] + C[x\varepsilon^1] + C[x\varepsilon^2] + C[\varepsilon^1 \varepsilon^2].$$

Browning & Crossley (2009) suggest that cognitive theories of response behavior and especially economic theory can be informative about the sources of bias in this estimate. In particular, $C[x\varepsilon^i]$ will be small for goods with unit income elasticity and a small log-log approximation error [$a_i \approx 1$ and $\eta_i(x_{t,h}) \approx 0$], and we would expect $C[x\varepsilon^i] > 0$ for luxuries and $C[x\varepsilon^i] < 0$ for necessities. This suggests using one luxury and one necessity rather than either two luxuries or two necessities. In terms of $C[\varepsilon^1 \varepsilon^2]$, complementary goods (coffee and cream) will tend to have $C[\varepsilon^1 \varepsilon^2] > 1$, and substitutes (coffee and tea) will tend to have $C[\varepsilon^1 \varepsilon^2] < 0$. Adding up implies $C[\varepsilon^1 \varepsilon^2] < 0$ on average, especially for highly aggregated goods. Browning & Crossley suggest therefore choosing two goods that (*a*) respondents can readily report, (*b*) have close to unit income elasticities (or a luxury and a necessity), (*c*) do not have too much approximation error (the double-logarithm form well approximates the Engel curves), and (*d*) are not strong complements or substitutes. It is perhaps surprising what is not on list: Large budget shares are not necessary (it is not necessary to capture a large fraction of total expenditure), and the reliability (low variance for measurement error) of individual measures is not critical. In this sense, two bad measures may be superior to one good measure.

Browning & Crossley (2009) report an illustrative application of this method that shows some promise. Attanasio et al. (2014) implement a similar approach. Such methods merit further consideration. However, as with imputation, these methods, if based on data on good-specific spending, have the limitation that some more comprehensive and good-quality source of expenditure data must be available at least periodically (to provide information on Engel curves). An alternative might be to use them in conjunction with a combination of one-shot and ($y_{t,h} - \Delta w_{t,h}$) measures. To our knowledge, this has not been tested.

With complementary information from a budget survey in hand, it may also be possible to estimate preference parameters directly from expenditures on an incomplete set of goods. For example, good-specific elasticities of intertemporal substitution are related to the overall elasticity of intertemporal substitution by the income elasticity of the good in question (Browning & Crossley 2000). Thus, estimates of the intertemporal substitution elasticity for food from a longitudinal survey might be combined with an estimate of the food income elasticity from a budget survey to give an estimate of the overall elasticity of intertemporal substitution. Again, this would not require household-level measures of total consumption expenditure to be imputed to the longitudinal data.

Overall, it remains an open question how best to measure consumption in general surveys. Part of the challenge is the great heterogeneity among households in their income sources, spending patterns, and approaches to financial management. It is not surprising that different data-collection strategies work with different people. As noted above, it has proven difficult to screen households and target different data-collection strategies effectively.

4. ADMINISTRATIVE AND PROCESS DATA

Recent years have seen a surge in the availability of administrative or process data for empirical research in economics and the social sciences more generally. Administrative data are generated by some process within a private firm or public administration, and as such, they contrast with other kinds of data used in the social sciences, which researchers collect purposefully by way of observation, experiment, or survey.[11] Examples of administrative data are scanner data produced at checkout counters in retail stores; tax records; eligibility and claims data from social assistance programs; medical insurance claims; banking and credit card records; and data generated by users of e-commerce websites, such as data on bidding in online auctions. In a white paper written for the US National Science Foundation, Card et al. (2010) stress that "administrative data offer much larger sample sizes and have far fewer problems with attrition, non-response, and measurement error than traditional survey data

sources." Many of the measurement problems associated with survey response behavior we discuss in Section 2 do not exist in administrative data, although other challenges arise, which we discuss below. We begin by considering the different types of administrative data used to measure household consumption expenditure.

4.1. Home Scanner Data

The use of retail scanner data is now widespread, particularly in industrial organization, in which the focus is often on narrow categories of expenditure and on markets rather than households. However, home (as opposed to store) scanner data may be useful in situations in which household consumption expenditure is the object of research interest. Leicester (2014) provides a discussion of the application of home scanner data to research on consumer behavior as well as a detailed comparison of home scanner and budget survey data in the United Kingdom. Potential advantages of home scanner data include accuracy, commodity detail, and the longitudinal and high-frequency nature of such data. However, the key feature of these data that has been exploited by research to date is the recording of purchase-specific price information.

Survey data on consumption expenditure are typically matched to price indices by time and geography. At best, survey data may contain both quantities and expenditures, allowing for the calculation of unit values at the household level. But unit values are not prices, and variation in unit values across time or households might reflect quality differences rather than differences in prices paid for an identical commodity. The commodity detail in scanner data means that true prices can be recovered. Aguiar & Hurst (2007) use such data to show that part of the decline in consumption expenditure around retirement results from older households shopping more and paying lower prices. Broda et al. (2009) use home scanner data to show that, in contrast to a widely held view, the poor pay less than rich household for identical goods, a finding with potentially important implications for poverty measurement. Griffith et al. (2009) employ home scanner data to characterize the savings available and realized through buying sale items, buying in bulk, buying generic brands, and buying in outlet stores. They show that the savings available through sales and bulk buying are of the same magnitude as the savings available through outlet stores, which is an important finding for the construction of price indices.

These papers illustrate how the price and purchase information in home scanner data can be exploited to give a much richer picture of the opportunities consumers face and their consequent behaviors. There are of course important potential limitations to these data. Representativeness is one concern, although the commercial firms that collect these data are

expending considerable effort to make their data sets population representative. The burden of participation in a home scanner panel probably means that they will always be short rotating panels, suitable for studying the dynamics of consumption expenditure at higher frequencies. As Leicester (2014) points out, the UK home scanner data he studies have much less rich demographic information than the UK national budget survey. In principle, more demographic information could be collected. Perhaps more importantly, home scanner data typically do not provide a complete picture of total consumption expenditure. The data that Leicester studies cover food and grocery products. Finally, it is a common finding (see Leicester 2014, and references therein) that average expenditure levels are significantly lower in scanner data than in budget survey data (for goods captured by both). Despite these concerns, there is much further potential for the use of home scanner data in consumption research.

4.2. Tax Data on Income and Wealth

Tax records typically contain information on net income, and in some tax systems, they record household-level wealth information as well. When they do, tax records allow for the construction of the ($y_{t,h} - \Delta w_{t,h}$) measure of consumption expenditure introduced in Section 3. Construction of such a measure from tax records, rather than survey data, has a number of possible advantages. Tax records typically offer large samples and the ability to follow households longitudinally through time. Income and wealth information may be very accurate, eliminating the concerns about measurement error in income and wealth raised in Section 3. Tax records should be representative of the population, which is important given the evidence in Sabelhaus et al. (2014) that budget survey nonresponse may be related to income. Browning & Leth-Petersen (2003), Kreiner et al. (2014) , and Koijen et al. (2014) construct this measure in Danish and Swedish tax data and compare it to budget survey data, whereas Browning et al. (2013) use this measure to study consumer behavior in Denmark.

Browning et al. (2013) use Danish tax record data to construct a ($y_{t,h} - \Delta w_{t,h}$) measure of consumption expenditure and then to test for causal effects of housing wealth shocks on consumer expenditure. Their main substantive finding is that there does not appear to be a direct house wealth effect, although there is some evidence that housing wealth increases drive the consumption expenditure of young households through a collateral channel. The analysis also highlights both strengths and weaknesses of this kind of data.

An important advantage of the administrative panel data employed by Browning et al. (2013) is that these data have a considerable longitudinal dimension. Much of the previous

literature on housing wealth and consumption (e.g., Attanasio & Weber 1994, Campbell & Cocco 2007, Attanasio et al. 2009) has employed panels of synthetic cohorts, constructed from repeated cross-sectional budget surveys (in the case of the papers mentioned, the UK FES). A disadvantage of the synthetic cohort approach is that cohorts can be defined only with respect to time-invariant variables. Specifically, it is not possible to compare the reaction to house price changes of owners and renters because the composition of renters and owners changes over time. With a long true panel derived from tax records, Browning et al. are able to compare the consumption responses of long-term renters to those of long-term owners.

However, even with Browning et al.'s (2013) high-quality administrative data, the expenditure measure given by the difference between income and the change in wealth is very noisy, which echoes Ziliak's (1998) finding for the same measure constructed with PSID survey data (see Section 3). This means that even though Browning et al. have a very large sample (10% of the Danish population), their wealth effect estimates are not very precise. As they admit, although their central estimate is for no effect, the confidence intervals include values that might be interpreted as a substantial effect.

Although tax data on income and wealth may be very accurate, as noted above, the $(y_{t,h} - \Delta w_{t,h})$ measure will underestimate consumption expenditures when there are capital gains. This is likely a key source of the noise in the expenditure measure used by Browning et al. (2013). Correcting for this requires data on household-level portfolio returns or on quantities of assets held. Neither are typically available, although the Swedish tax data studied by Koijen et al. (2014) are an exception. Browning & Leth-Petersen (2003) evaluate the $(y_{t,h} - \Delta w_{t,h})$ measure by linking their tax record data to the Danish budget survey. They find that the worst fit is for households with significant bond or equity holdings. Koijen et al. (2014) exploit that Swedish data do allow the calculation of household-specific portfolio returns to assess the consequences of ignoring capital gains or losses. They find that this induces substantial errors in the consumption expenditure measure and that those measurement errors are increasing in wealth (and hence in consumption).

Despite these concerns, the research potential of long panels on household consumption expenditure data and the other advantages noted above suggest that this approach to measuring consumption expenditure will see growing use. It is also striking to note the changing perspective on administrative data reflected in these papers: Browning & Leth-Petersen (2003) use budget survey data to validate the measure derived from tax

records, whereas Kreiner et al. (2014) and Koijen et al. (2014) use administrative data to assess budget survey data.

4.3. Data from Online Financial Services

The above papers focus on constructing measures of total expenditure from administrative data, which then could be used in a variety of research contexts. An emerging complementary research agenda aims at testing specific implications of models of economic behavior using more specific administrative data. For testing models of household consumption, transaction-level data obtained from online financial services are attractive. Of course, important privacy issues arise when using such data, an issue to which we return below.

The objective of Kuchler (2013) is to test models of intertemporal choice that allow for present bias and impatience, with hyperbolic discounting as the leading example. Such models have testable implications with respect to the sensitivity of consumption to paycheck receipt. (The consumption of an individual with standard preferences, including exponential discounting, would in similar circumstances not react to the receipt of an income payment.) Kuchler obtained a data set containing the daily balances and transactions on all bank accounts of a sample of individuals from an online debt management service provider. The data allow her to construct high-frequency data on expenditures paid for by credit or check cards (classified in 50 different categories). To test for present bias and impatience, she relates the high-frequency patterns of spending on goods that are immediately consumed, such as restaurant meals, to the timing of income receipt as well as planned and realized debt repayment.

Although the sample is clearly selective, from a measurement perspective, the administrative data set Kuchler (2013) uses has an important advantage: Income, financial balances, debt repayments, and certain expenditure categories are measured precisely and at a high frequency in a single data set. Earlier research on the sensitivity of consumption expenditure to the timing of income receipt faced more restrictive data sets. For example, Stephens (2006) uses detailed expenditure data collected via diaries as part of the UK FES together with information on when a paycheck was received. In the data he can use, expenditures are aggregated by week, whereas for paycheck receipt, the exact day is observed. That expenditures are recorded only by week is a technical restriction of how this particular expenditure data are processed, but in principle, budget surveys would deliver daily data. Kuchler's administrative data have a more important advantage: She can use information on financial balances, observed at the same daily frequency as the expenditure

data, to rule out other possible explanations of consumption expenditure such as short-term credit constraints.

Baker (2013) and Baugh et al. (2014) obtained data sets from online personal financial aggregator services. These websites connect their users' financial accounts, allowing users to see summaries of their income, spending, debt, and financial investments in a single location. Data sets obtained from the providers of these services come close to the ideal of observing a complete accounting system of private households—if a user registers all his or her accounts, all noncash transactions (financial flows) are covered, and stocks of financial investments are measured precisely as well. In both papers, detailed disaggregated expenditure measures are constructed from the payment transaction data, the general approach being similar to that used by Kuchler (2013). Both Baker (2013) and Baugh et al. (2014) address the question of whether and by how much households react to income shocks. The papers differ in how they construct income shocks.

Baker (2013) enriches the financial aggregator data set by linking users' financial information to financial data of their employers (which are identified using textual descriptions included in the direct deposit records). Thus, he can link the spending data with measures of firm-level income shocks, which, as he argues, are exogenous for the individual users. Baker illustrates nicely how administrative data can be enriched by linking them with outside information, which in this case is publicly available, using automated matching algorithms.

Baugh et al. (2014) consider a different income shock, cash flows from tax returns. They distinguish both in a theoretical model and in the data between the information event (when tax returns are filed using tax preparation software so that the amount is known) and the cash flow event (when the tax refund is received). The exact date when the tax return was filed is constructed by identifying in the financial account data when a payment to a tax filing service such as TurboTax was made. The date of the tax refund receipt can be observed in the financial account data as well. With these data, Baugh et al. (2014) can study households' spending reactions to the information and cash flow events separately. The key feature of the administrative data here is that they not only provide precise high-frequency observations on expenditures, but also allow researchers to precisely date various cash flow events.

4.4. Challenges in the Use of Administrative Data

As the above examples show, the use of administrative data to study consumer behavior has considerable promise. There are, however, several significant challenges that need to be

addressed if this promise is to be realized.[12] Some are common to most research uses of administrative data sets, whereas others are specific to consumer expenditure data.[13]

The use of administrative data raises important privacy concerns, many of which are not fully addressed by standard approaches to (pseudo-)anonymization. Heffetz & Ligett (2013) provide an extensive review of the problems that arise, in particular when several data sources are linked.

Administrative and process data are designed to allow the conduct of private or public programs. As such, they can be more difficult to work with than survey data that are deliberately collected for research purposes.

Many administrative data sets cover only specific samples that are not representative of the general population, so the generalizability of findings is a concern. The selection process might be difficult to model. At the other extreme, some administrative samples cover entire populations; although selection is not an issue in these cases, classical tests of statistical significance are meaningless, but they are routinely applied, and there does not appear to be a widely accepted alternative.

In the case of expenditure measurement, a feature of most administrative data sets is that they cover only subsets of households' expenditures, perhaps just a few nondurable consumption items. For some purposes, such limited data might be sufficient (as in Baker 2013, Kuchler 2013, and Baugh et al. 2014). Deriving a measure of, for example, total household expenditure from administrative data is possible, as the Danish and Swedish examples show, but the comprehensive data that are needed exist only in a few countries. It might also be possible to apply some of the imputation approaches discussed in Section 3 to administrative data sets.

Many administrative data sets lack sociodemographic information and additional covariates, such as measures of income or wealth, that are needed for answering key research questions related to consumption. One solution to this problem is the linkage of administrative and survey data. We are not aware of studies that link administrative data on expenditure with detailed survey data on other household-level variables, but several examples exist in which data from social benefit programs have been linked with survey data, for instance, from the US HRS. Although data linkage can potentially enrich administrative data tremendously, it also exacerbates concerns of data privacy (see Heffetz & Ligett 2013). Moreover, typically not all respondents give consent to linking their survey responses to administrative data. The resulting linked data sets might be highly selective, as suggested by several recent studies of respondents' consent to data linkage requests (see, e.g., Sakshaug &

Kreuter 2012, Sakshaug et al. 2012). Even leaving the issue of selective consent aside, the linkage of survey and administrative data can be difficult at a purely technical level. Kapteyn & Ypma (2007) and Meijer et al. (2011) discuss the statistical implications of imperfect record linkage. An important insight from these papers is that when mismatches in the linkage are allowed for, a variable from an administrative data set linked to survey data might be subject to measurement error, even if one maintains that the administrative data as such are error free.

Despite these challenges, deriving measures of household consumption expenditure from administrative and process data is a very promising direction of data development.

5. CONCLUSION

An article such as this one naturally highlights the challenges and problems in the area under discussion. Measuring consumption expenditure is not easy, but despite the difficulties, we strongly believe it is worth the effort. Two points bear emphasis. First, it is important not to lose sight of what has been accomplished. As the list of questions and citations that began Section 1 reminds us, the past collection of household expenditure data has been enormously fruitful.

Second, good measurement is difficult in many, perhaps most, domains, and measuring household expenditure is not obviously more difficult than other economic measurements. Household income is an important example. It is widely presumed that, at least in developed countries, household income is easier to measure that household expenditure. Probably for this reason, and despite the strong conceptual arguments for measuring household well-being with consumption rather than income, official poverty measures and much research on poverty and inequality remain income based. Yet in an important series of papers, Meyer & Sullivan (2003, 2004, 2008, 2011, 2012) argue convincingly against this presumption. At least at the bottom of the resource distribution, consumption expenditure appears to be better measured than income.[14]

Nevertheless, the ways in which households organize their finances and make purchases have changed dramatically. At the same time, the technological change has opened up new possibilities for the collection of consumer expenditure data. There are many outstanding questions in the fields of household finance and consumer behavior. The CNSTAT report (Natl. Res. Counc. 2013) calls for a program of continuous research on consumption expenditure data collection, and we echo that sentiment. A number of directions stand out as particularly ripe for investigation. They are as follows.

5.1. Can New Technology Significantly Improve Self-Administered Survey Collection of Expenditure Data?

The traditional view of expenditure surveys has been that diaries are the gold standard for data collection. This view reflects a deep concern with recall error. Accumulating evidence (see Section 2) questions this view. Diary data on household consumption expenditures suffer from significant problems, and by many measures, recall data appear to be of higher quality. Nevertheless, the CNSTAT panel's prototypes for future versions of the CE put a lot of emphasis on self-administered data collection—diaries. This choice seems to reflect a belief that self-administered data collection has the most scope for improvement, particularly by incorporating new technologies: Internet surveys, mobile devices such as cell phones and tablet computers, and home scanners. Moving forward, a key area for research on the collection of consumption expenditure data will be to test this proposition: Can diaries (or self-administered data collection, more generally) be very significantly improved by redesign and the incorporation of new technologies?

5.2. Can Process or Administrative Data Replace Survey Data Collection in the Measurement of Household Expenditures?

Administrative and process data offer great opportunities for the construction of measures of household consumption; there are also significant challenges to be overcome if such data are to become the basis for research in household finance and consumer behavior. Some are common to most research uses of administrative data sets, whereas others are specific to consumer expenditure data. Data privacy concerns are an important challenge when using administrative data, and even more so when several sources of data are linked. We believe that the profession will be increasingly concerned about these issues. For instance, there is a certain tension between data privacy, on the one hand, and the profession's push toward replication of empirical studies, on the other. Because of privacy concerns, many administrative data sets can be analyzed only under very restrictive conditions, and it might be impossible to make them widely accessible for replication. The data sets obtained from online financial services reviewed above are a case in point. Other challenges related to the use of administrative data discussed include the selectivity of many process-generated samples and the limited availability of expenditure items in many administrative data sets. With respect to the latter, some of the statistical techniques developed for the analysis of consumption behavior with limited data from surveys (see Section 3) might be applied.

5.3. Can a Good Measure of Household Consumption Expenditure Be Obtained Quickly in General Surveys?

As Section 3 recounted, several approaches have been proposed or further developed in the decade since Browning et al. (2003) considered this question. No consensus has emerged. Despite the encouraging prospects for administrative and process data, large longitudinal surveys have important advantages, many of which follow from the fact that they are designed for research. They seem likely to remain a major source of research data for some time. Thus the need to establish the best way to collect household consumption expenditure data in such surveys remains.

5.4. Can We Better Integrate Measurement and Analysis?

Finally, we would argue against a division of labor in which measurement and econometric analysis are treated as distinct activities. Econometric analysis should take careful account of the way in which consumption expenditure data—or any data—are collected, and survey design must consider how the data will be analyzed. One positive example is the work on infrequency discussed in Section 2. Meghir & Robin (1992) show how the expenditure data from short diaries can be analyzed while taking proper account of the inherent infrequency, and at the same time, they identify a key additional piece of information—the number of purchases—which could be recorded in surveys and which would much aid the analysis of the data. Battistin & Padula (2013) provide another example of an analysis that takes careful account of survey design; Browning & Crossley (2009) discuss how survey design choices might follow from the estimation strategy. A further integration of measurement and econometric analysis could improve both.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Aguiar, M, Hurst, E, 2007. Life-Cycle Prices and Production. *American Economic Review*. 97(5):1533-1559.

Ahmed N, Brzozowski M, Crossley TF. 2006. *Measurement errors in recall food consumption data*. IFS Work. Pap. W06/21, Inst. Fisc. Stud., London

Angrisani M, Kapteyn A. 2014. Measuring household spending and payment habits: the role of "typical" and "specific" time frames in survey questions. See Carroll et al. 2014

Attanasio O, Blow L, Hamilton R, Leicester A. 2009. Booms and busts: consumption, house prices and expectations. *Economica* 71:20–50Attanasio, O, Davis S. 1996. Relative Wage Movements and the Distribution of Consumption., *Journal of Political Economy* 104(6):1227-62

Attanasio O, Deleire T. 2002. The effect of Individual Retirement Accounts on household consumption and national saving. *Econ. J.* 112:504–38

Attanasio O, Hurst E, Pistaferri L. 2014. The evolution of income, consumption, and leisure inequality in the US, 1980–2010. See Carroll et al. 2014

Attanasio O, Pistaferri L. 2014. Consumption Inequality Over the Last Half Century. Some Evidence using the New PSID Consumption Measure. Forthcoming, *American Economic Review Papers and Proceedings.*

Attanasio O, Weber G. 1993. Consumption growth, the interest rate and aggregation. *Rev. Econ. Stud.* 60:631–50

Attanasio O, Weber G. 1994. The UK consumption boom of the late 1980s: aggregate implications of microeconomic evidence. *Econ. J.* 104:1269–302

Baker SR. 2013. *Debt and the consumption response to household income shocks.* Unpublished manuscript, Stanford Univ., Stanford, CA

Barrett G, Levell P, Milligan K. 2014. A comparison of micro and macro expenditure measures across countries using differing survey methods. See Carroll et al. 2014

Battistin E, Miniaci R, Weber G. 2003. What do we learn from recall consumption data? *J. Hum. Resour.* 38:354–85

Battistin E, Padula M. 2013. *Survey instruments and the reports of consumption expenditures: evidence from the Consumer Expenditure Surveys.* Unpublished manuscript, Univ. Padova, Univ. Venice

Baugh B, Ben-David I, Park H. 2014. *Disentangling financial constraints, precautionary savings, and myopia: household behavior surrounding federal tax returns.* NBER Work. Pap. 19783

Bee A, Meyer BD, Sullivan JX. 2014. The validity of consumption data: Are the consumer expenditure interview and diary surveys informative? See Carroll et al. 2014

Blundell R, Pistaferri L, Preston I. 2004. *Imputing consumption in the PSID using demand estimates from the CEX.* Work. Pap. W04/27, Inst. Fisc. Stud., London

Blundell R, Pistaferri L, Preston I. 2008. Consumption inequality and partial insurance. *Am. Econ. Rev.* 98:1887–921

Bonke J, Browning M. 2009. The allocation of expenditures within the household: a new survey. *Fisc. Stud.* 30:461–81

Bottazzi R, Crossley TF, O'Dea C. 2008. *UKHLS Innovation Panel consumption questions: preliminary analysis.* Unpublished manuscript, Inst. Fisc. Stud., London

Brewer M, Etheridge B, O'Dea C. 2013. *Why are households that report the lowest incomes so well-off?* Discuss. Pap. 736, Econ. Dep., Univ. Essex

Broda C, Leibtag E, Weinstein DE. 2009. The role of prices in measuring the poor's living standards. *J. Econ. Perspect.* 23(2):77–97

Browning M, Crossley TF. 2000. Luxuries are easier to postpone: a proof. *J. Polit. Econ.* 108:1022–26

Browning M, Crossley TF. 2001. Unemployment insurance levels and consumption changes. *J. Public Econ.* 80:1–23

Browning M, Crossley TF. 2008. The long-run costs of job loss as measured by consumption changes. *J. Econom.* 145:109–20

Browning M, Crossley TF. 2009. Are two cheap, noisy measures better than one expensive, accurate one? *Am. Econ. Rev.* 99:99–103

Browning M, Crossley TF, Weber G. 2003. Asking consumption questions in general purpose surveys. *Econ. J.* 113:F540–67

Browning M, Gørtz M, Leth-Petersen S. 2013. Housing wealth and consumption: a micro panel study. *Econ. J.* 123:401–28

Browning M, Leth-Petersen S. 2003. Imputing consumption from income and wealth information. *Econ. J.* 113:F282–301

Brzozowski M, Crossley TF. 2011. Measuring the well-being of the poor with income or consumption: a Canadian perspective. *Can. J. Econ.* 44:88–106

Burke LE, Wang J, Sevick MA. 2011. Self-monitoring in weight loss: a systematic review of the literature. *J. Am. Diet. Assoc.* 111:92–102

Campbell JY, Cocco JF. 2007. How do house prices affect consumption? Evidence from micro data. *J. Monet. Econ.* 3:591–621

Card D, Chetty R, Feldstein M, Saez E. 2010. *Expanding access to administrative data for research in the United States*. White Pap., Natl. Sci. Found., Arlington, VA

Carroll C, Crossley TF, Sabelhaus J, eds. 2014. *Improving the Measurement of Consumer Expenditures*. NBER Ser. Stud. Income Wealth. Cambridge, MA: Natl. Bur. Econ. Res. In press Cherchye, L., De Rock, B, Vermeulen, F., 2012, Married with children. A collective labor supply model with detailed time use and intrahousehold expenditure information. *American Economic Review*, 102:3377-3405

Comerford D, Delaney L, Harmon C. 2009. Experimental tests of survey responses to expenditure questions. *Fisc. Stud.* 30:419–33

Crossley TF, de Bresser J, Delaney L, Winter J. 2014. *Can survey participation alter household saving behaviour?* Unpublished manuscript, Inst. Fisc. Stud., London Manuscript in progress.

Crossley TF, Winter JK. 2014. Asking households about expenditures: What have we learned? See Carroll et al. 2014

d'Ardenne J, Blake M. 2012. *Developing expenditure questions: findings from focus groups*. Work. Pap. W12/18, Inst. Fisc. Stud., London

Deaton A. 1992. *Understanding Consumption*. New York: Oxford Univ. Press

Deaton A, Grosh M. 2000. Consumption. In *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*, ed. M Grosh, P Glewwe, pp.91-133. Washington, DC: World Bank

Deaton A, Kozel V. 2005. Data and dogma: the great Indian poverty debate. *World Bank Res. Obs.* 20:177–99

Edgar J. 2009. *What does 'usual' usually mean?* Unpublished manuscript, Bur. Labor Stat., Washington, DC

Edgar J, Nelson DV, Paszkiewicz L, Safir A. 2013. *The Gemini Project to redesign the Consumer Expenditure Survey: redesign proposal*. Rep., Bur. Labor Stat., Washington, DC

Einav L, Levin JD. 2013. *The data revolution and economic analysis*. NBER Work. Pap. 19035

Fricker S, Kopp B, To N. 2014. Exploring a balance edit approach in the Consumer Expenditure Quarterly Interview Survey. See Carroll et al. 2014

Gieseman R. 1987. The Consumer Expenditure Survey: quality control by comparative analysis. *Mon. Labor Rev.* 1987(3):8–14

Goldenberg K, Ryan J. 2009. *Evolution and change in the Consumer Expenditure Surveys: adapting to meet changing needs*. Unpublished manuscript, **Bureau of Labour Statistics (BLS).**

Gray M, Constantine R, d'Ardenne J, Blake M, Uhrig SCN. 2008. *Cognitive testing of Understanding Society: the UK Household Longitudinal Study Questionnaire*. Underst. Soc. Work. Pap. Ser. 2008-04, Univ. Essex

Gray PG. 1955. The memory factor in social surveys. *J. Am. Stat. Assoc.* 50:344–63

Griffith R, Leibtag E, Leicester A, Nevo A. 2009. Consumer shopping behavior: How much do consumers save? *J. Econ. Perspect.* 23(2):99–120

Grootaert C. 1986. Use of multiple diaries in a household expenditure survey in Hong Kong. *J. Am. Stat. Assoc.* 81:938–44

Guvenen F. 2007. Learning your earning: Are labor income shocks really very persistent? *Am. Econ. Rev.* 97:687–712

Heffetz O, Ligett K. 2013. *Privacy and data-based research*. NBER Work. Pap. 19433

Hurd M, McFadden D, Chand H, Gan L, Merrill A, Roberts M. 1998. Consumption and savings balances of the elderly: experimental evidence on survey response bias. In *Frontiers in the Economics of Aging*, ed. D Wise, pp. 353–87. Chicago: Univ. Chicago Press

Hurd M, Rohwedder S. 2009. Methodological innovations in collecting spending data: the HRS Consumption and Activities Mail Survey. *Fisc. Stud.* 30:435–59

Hurd M, Rohwedder S. 2010. *The effects of the financial crisis and the great recession on American household*. NBER Work. Pap. 16407

Hurd MD, Rohwedder S. 2014. Measuring total household spending in a monthly Internet survey: evidence from American Life Panel. See Carroll et al. 2014**.**

Jacobs E, Shipp S. 1993. A history of the U.S. Consumer Expenditure Survey: 1935–36 to 1988–89. *J. Econ. Soc. Meas.* **19:** 59–96

Jolliffe D. 2001. Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. *J. Econ. Soc. Meas.* 27:1–23

Kapteyn A, Ypma JY. 2007. Measurement error and misclassification: a comparison of survey and administrative data. *J. Labor Econ.* 25:513–51

Kemsley WFF, Nicholson, JL. 1960. Some experiments in Methods of Conducting Family Expenditure Surveys," *Journal of the Royal Statistical Society*, 123(307):307-328.

Kemsley, WFF, Redpath, RU Holms, M. 1980. Family *Expenditure Survey Handbook*, Office of Populaton Surveys, London, Her Majesty's Stationary Office, London.

Koijen R, Nieuwerburgh SV, Vestman R. 2014. Judging the quality of survey data by comparison with "truth" as measured by administrative records: evidence from Sweden. See Carroll et al. 2014

Kreiner CT, Lassen DD, Leth-Petersen S. 2014. Measuring the accuracy of survey responses using administrative register data: evidence from Denmark. See Carroll et al. 2014

Krueger D, Perri F, Pistaferri L, Violante GL. 2010. Cross-sectional facts for macroeconomists. *Rev. Econ. Dyn.* 13:1–14

Kreuter F, Eckman S, Jaeckle A, Kirchner A, Presser S, Tourangeau R. 2012. *Mechanisms of misreporting to filter questions*. Presented at AAOPR Annu. Conf., 67th, Orlando, FL

Kreuter F, McCulloch S, Presser S, Tourangeau R. 2011. The effects of asking filter questions in interleafed versus grouped format. *Sociol. Methods Res.* 40:88–104

Kuchler T. 2013. *Sticking to your plan: hyperbolic discounting and credit card debt paydown*. Unpublished manuscript, Stanford University.

Leicester A. 2014. Using scanner data to construct detailed weights for certain categories of spending. See Carroll et al. 2014

Li, G, Schoeni, RF, Danziger, S, Charles KK. 2010. New expenditure data in the PSID: comparisons with the CE. Monthly Labor Review, February, 29-39.

Manski CF, Tamer E. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70:519–46

McKenzie J. 1983. The accuracy of telephone call data collected by diary methods. *J. Mark. Res.* 20:417–27

McWhinney I, Champion H. 1974. The Canadian experience with recall and diary methods in consumer expenditure surveys. In *Annals of Economic and Social Measurement*, Vol. 3, ed. SV Berg, pp. 113–140. Cambridge, MA: Natl. Bur. Econ. Res.Meghir, C, Robin, JM, 1992. Frequency of purchase and the estimation of demand systems. *Journal of Econometrics*, 53(1-3):53-85

Meijer E, Rohwedder S, Wansbeek T. 2011. Measurement error in earnings data: using a mixture model approach to combine survey and register data. *J. Bus. Econ. Stat.* 30:191–201

Meyer BD, Sullivan JX. 2003. Measuring the well-being of the poor using income and consumption. *J. Hum. Resour.* 38:S1180–220

Meyer BD, Sullivan JX. 2004. The effects of welfare and tax reform: the material well-being of single mothers in the 1980s and 1990s. *J. Public Econ.* 88:1387–420

Meyer BD, Sullivan JX. 2008. Changes in the consumption, income, and well-being of single mother headed families. *Am. Econ. Rev.* 98:2221–41

Meyer BD, Sullivan JX. 2011. Viewpoint: further results on measuring the well-being of the poor using income and consumption. *Can. J. Econ.* 44:52–87

Meyer B, Sullivan J. 2012. Winning the war: poverty from the great society to the great recession. *Brookings Pap. Econ. Act.* 45:133–200

Natl. Res. Counc. 2013. *Measuring What We Spend: Toward a New Consumer Expenditure Survey*. Washington, DC: Natl. Acad. Press **.**

Neter J, Waksberg J. 1964. A study of response errors in expenditures data from household interviews. *J. Am. Stat. Assoc.* 59:18–55Palumbo, MG, 1999. Uncertain Medical Expenses and Precautionary Saving Near the End of the Life Cycle. *Review of Economic Studies,* 66(2):395-421

Parker J, Souleles NS, Johnson DS, McClelland R. 2013. Consumer spending and the economic stimulus payments of 2008. *Am. Econ. Rev.* 103:2530–53

Pradhan M. 2009. Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia. *Fisc. Stud.* 30:391–417

Pudney S. 2007. *Heaping and leaping: survey response behavior and the dynamics of self-reported consumption expenditure*. Unpublished manuscript, Univ. Essex

Rubin DC, Baddeley AD. 1989. Telescoping is not time compression: a model of the dating of autobiographical events. *Mem. Cognit.* 17:653–61

Sabelhaus J, Groen JA. 2000. Can permanent-income theory explain cross-sectional consumption patterns? *Rev. Econ. Stat.* 82:431–38

Sabelhaus J, Johnson D, Ash S, Swanson D, Garner T, et al. 2014. Is the Consumer Expenditure Survey representative by income? See Carroll et al. 2014

Safir A, Goldenberg KL. 2008. *Mode effects in a survey of consumer expenditures*. Unpublished manuscript, Bur. Labor Stat., Washington, DC. **http://www.bls.gov/cex/cesrvymethssafir1.pdf**

Sakshaug JW, Kreuter F. 2012. Assessing the magnitude of non-consent biases in linked survey and administrative data. *Surv. Res. Methods* 6:113–22

Sakshaug JW, Couper MP, Ofstedal MB, Weir DR. 2012. Linking survey and administrative records: mechanisms of consent. *Sociol. Methods Res.* 41:535–69

Silberstein AR. 1990. First wave effects in the U.S. Consumer Expenditure Interview Survey. *Surv. Methodol.* 16:293–304

Silberstein AR, Jacob CA. 1989. Symptoms of repeated interview effects in the Consumer Expenditure Interview Survey. In *Panel Surveys*, ed. D Kasprzyk, G Duncan, G Kalton, MP Singh, pp.289-303. New York: Wiley

Silberstein AR, Scott S. 1991. Expenditure diary surveys and their associated errors. In *Measurement Errors in Surveys*, ed. PP Biermer, RM Groves, LE Lyberg, NA Mathiowetz, S Sudman, pp. 303-326 New York: Wiley

Skinner J. 1987. A superior measure of consumption from the Panel Study of Income Dynamics. *Econ. Lett.* 23:213–16

Stat. Can. 1999. *1996 Food Expenditure Survey, public-use microdata files*. Income Stat. Div., Stat. Can., Ottawa

Stephens M. 2001. The long-run consumption effects of earning shocks. *Rev. Econ. Stat.* 83:28–36

Stephens M. 2003. "3rd of tha month": Do Social Security recipients smooth consumption between checks? *Am. Econ. Rev.* 93:406–22

Stephens M. 2006. Paycheque Receipt and the Timing of Consumption. *Economic Journal*, 116(513):680-701. Sudman S, Bradburn NM, Schwarz N. 1996. *Thinking About Answers*. San Francisco: Jossey-Bass

Tanner S. 1998. How much do consumers spend? Comparing the FES and National Accounts. In *How Reliable Is the Family Expenditure Survey?*, ed. J Banks, P Johnson, pp. 67-121 London: Inst. Fisc. Stud.

Tourangeau R, Rips LJ, Rasinski K. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge Univ. Press

van Soest A, Hurd M. 2008. A test for anchoring and yea-saying in experimental consumption data. *J. Am. Stat. Assoc.* 103:126–36

Velilla P. 2014. *Micro and macro data on household income and wealth: mapping the Spanish Survey of Household Finances (EFF) with other statistical sources*. Doc. Ocas., Banco Esp., Madrid

Winter J. 2002. *Bracketing effects in categorized survey questions and the measurement of economic quantities*. Discuss. Pap. 02-35, Univ. Mannheim

Winter J. 2004. Response bias in survey-based measures of household consumption. *Econ. Bull.* 3(9):1–12Zilliak, JP. 1998. Does the chioice of consumption measure matter? An application to the permanent-income hypothesis. *Journal of Monetary Economics*, 41:201-216

Zwane A, Zinman J, Dusen EV, Pariente W, Null C, et al. 2011. Being surveyed can change later behavior and related parameter estimates. *Proc. Natl. Acad. Sci. USA* 108:1821–26

NOTES

[1]Again, this experience has not been shared by all national budget surveys. For example, Barrett et al. (2014) report that over a similar period, the Australian and Canadian budget surveys also experienced a fall in response rates but that those falls started later and were not accompanied by a decline in the coverage of national accounts consumption.

[2]Readers are referred to Deaton (1992) for a discussion of some of the issues in constructing a measure of consumption. Meyer & Sullivan (2012) illustrate current best practices.

[3]This section draws on the longer review in Crossley & Winter (2014).

[4]Another line of research has used situations in which both a diary and recall measure of food expenditure are available to try to infer characteristics of the measurement error in expenditure data. This work is motivated by the idea that if measurement error can be characterized, then it might be possible to mitigate its effects through econometric modeling. The results, however, are not encouraging. Ahmed et al. (2006) show that the differences between recall and diary responses are correlated with both the level of expenditure and observable characteristics of the households. This implies that there are errors in one or both measures that do not have the classical properties assumed by simple econometric models of measurement error. Worse, Battistin & Padula (2013) show that recall and diary food expenditure measures are not rank preserving (they do not order households identically). Rank preservation is among the weakest identification conditions required by econometric models of measurement error.

[5]This finding illustrates a general danger: Concepts that are quite natural or obvious to researchers may be understood differently by survey respondents.

[6]Although it is beyond the scope of this article, there is much interest in attempting to measure individual consumption. Whereas most of the literature on intrahousehold allocation attempts to infer individual consumption (or at least features of allocation rules) from essentially household-level data, surveys could be designed to aid in the identification. Bonke & Browning (2009) and Cherchye et al. (2012) report progress in this direction.

[7]This cash flow reconciliation was used even before the introduction of computer-assisted personal interviewing in the Canadian survey. Early predecessors of the CE also had a balance edit, but this was dropped when the survey was redesigned in 1972 (Jacobs & Shipp 1993).

[8]For example, the BLS's own proposal for the redesign of the CE involves a reduced longitudinal component.

[9]The exception is the AHEAD pilot, reported by Hurd et al. (1998), which studies an elderly population.

[10]Of course, the same may be true of a one-shot expenditure question, unless the question is accompanied by very detailed information on inclusions and exclusions.

[11]Administrative data are sometimes also referred to as process-generated data or naturally occurring data.

[12]The authors of the studies reviewed above are clearly aware of these issues, and they address many of them explicitly.

[13]Readers are referred to Einav & Levin (2013) for a general review of methodological issues arising with administrative data.

[14]Their argument rests on multiple lines of evidence, with expenditure data appearing superior in terms of survey and item nonresponse and the degree of imputation, correspondence to external data sources, and correlation with other measures of material being. These findings appear to hold not just in the United States but also in other countries (see Brewer et al. 2013 for the United Kingdom and Brzozowski & Crossley 2011 for Canada). An important part of the argument concerns underreporting. Households at the bottom of the income distribution have expenditures greatly in excess of reported incomes. Sabelhaus & Groen (2000) and Brewer et al. (2013) use structural life-cycle/permanent income models to show that it is implausible that this is dissaving (to smooth temporarily low income). Instead, it seems that income is significantly underreported at the bottom of the distribution. The quasi-experiment in survey design studied by Brzozowski & Crossley (2011) is also consistent with this interpretation.