NBER WORKING PAPER SERIES

REPORTED PREFERENCE VS. REVEALED PREFERENCE: EVIDENCE FROM THE PROPENSITY TO SPEND TAX REBATES

Jonathan A. Parker
Nicholas S. Souleles

Working Paper 23920
http://www.nber.org/papers/w23920

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2017, Revised February 2019

We thank Christian Broda, Erik Hurst, David S. Johnson, Robert McLelland, Scott Nelson, Claudia Sahm, and Matthew D. Shapiro, as well as participants at talks at UT Austin, Brandeis, Chicago, and the 2017 ASSA Meetings. Colin Gray and Fangzhou Lu provided excellent research assistance. We also thank the staff of the Division of Consumer Expenditure Surveys at the Bureau of Labor Statistics for their work in getting the economic stimulus payment questions added to the Consumer Expenditure Survey, and Ed Grove, Matt Knain, and Molly Hagen at Nielsen for their work on the survey and their careful explanation of the Nielsen Consumer Panel. Parker thanks the MIT Sloan School of Management, the Kellogg School of Management at Northwestern University, the Initiative for Global Markets at the University of Chicago, and the Zell Center at the Kellogg School of Management for funding. The results of this paper are calculated based on data from The Nielsen Company (U.S.) LLC and provided by the Marketing Data Center and the University of Chicago Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w23920.ack

Reported Preference vs. Revealed Preference: Evidence from the Propensity to Spend Tax Rebates
Jonathan A. Parker and Nicholas S. Souleles
NBER Working Paper No. 23920
October 2017, Revised February 2019
JEL No. B40,C42,D14,E21,E62,H31

ABSTRACT

We evaluate the consistency of two methods for estimating the effect of an economic policy: i) surveying people to report the change in their behavior caused by the policy, ii) inferring this change using (reported) actual behavior and differences in treatment across people. Both methods have been widely used to measure propensities to spend. Using Federal stimulus payments disbursed quasi-randomly over time in 2008, we find greater revealed-preference estimates of spending by households reporting greater spending and the two methods produce similar estimates of average spending. But, counterfactually, reported-preferences estimates are not higher for households with lower liquidity.

Jonathan A. Parker
MIT Sloan School of Management
100 Main Street, E62-642
Cambridge, MA 02142-1347
and NBER
JAParker@MIT.edu

Nicholas S. Souleles
Finance Department
The Wharton School
2300 SH-DH
University of Pennsylvania
Philadelphia, PA 19104-6367
and NBER
souleles@wharton.upenn.edu

A data appendix is available at http://www.nber.org/data-appendix/w23920

The foundation of testing and estimation in economics is the revealed-preference approach in which inferences about causal effects are drawn from people's actions in different situations. The alternative to the revealed-preference approach is to rely on people to make the causal inferences themselves by asking them to report their choices in different hypothetical situations. This alternative is prevalent in other fields, such as history and psychology, where people's reports of what they would have done in counterfactual situations are commonly used as evidence, and it has always had a presence in economic theory. It is also becoming more widespread in empirical economics: major survey datasets contain estimates of preference parameters derived from this approach, structural models are fitted to reported changes in behavior, and researchers are using reported responses to evaluate policy.[1] Most closely related to our analysis, there has been a recent boom in research on consumption and saving behavior using reported causal effects, with the dual goals of distinguishing among models of household saving behavior and better understanding stabilization policy.[2]

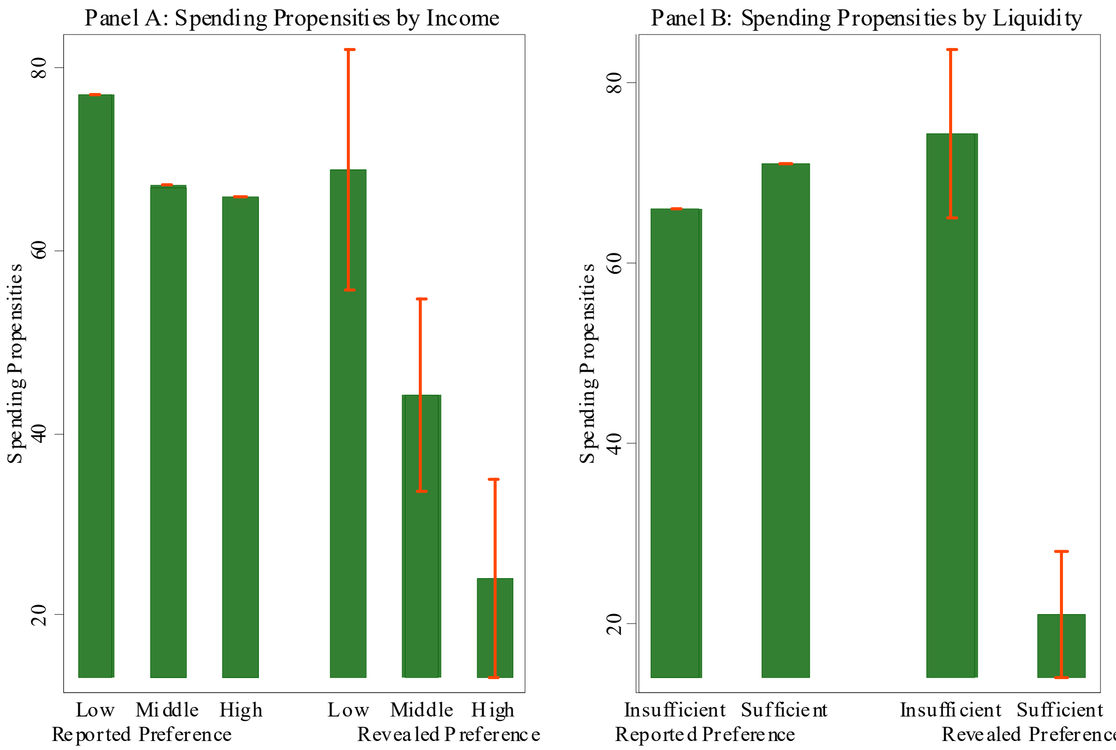
Traditional revealed-preference estimation requires the economist to observe differences in circumstances that are either plausibly exogenous or well-enough understood as to be used in structural estimation. Surveying people to directly report causal effects appears comparatively simple. Further, the survey approach can elicit behavioral responses to idealized *ceteris paribus* experiments through the careful design of survey instruments. The pivotal question is whether the elicited behavior corresponds to the actual behavior that would occur were the counterfactual of interest not just hypothetical. Do people fully understand the scenario(s) in the question, do they accurately determine what their own behavior would have been, and do they respond honestly? Friedman and Wallis (1942, p. 179-80), discussing an early application of this methodology in Thurstone (1931), famously wrote "Questionnaires or other devices based on conjectural responses to hypothetical stimuli . . . are valueless because the subject cannot know how he would react."

This paper evaluates the use of surveys to directly elicit causal effects by analyzing whether the spending that people *report* as caused by lump-sum tax rebates – specifically the Economic Stimulus Payments of 2008 – matches the spending that the same people are *estimated* to have done in response to the same rebates using revealed-preference methods. We estimate the revealed-preference propensity to spend following Johnson, Parker and Souleles (2006), Parker et al. (2013) and Broda and Parker (2014), identifying the spending propensity from the quasi-randomly-timed disbursement of the rebates across groups of households. We measure reported spending effects using both the Shapiro and Slemrod (1995, 2003*a*, 2003*b*, 2009) survey instrument and our own instrument designed

[1] See for example Barsky et al. (1997), Ameriks et al. (2016), and D'Acunto, Hoang and Weber (2016) respectively.

[2] Questions asking respondents to report spending propensities have recently been added to surveys run by the Federal Reserve Bank of New York, the Bank of England, and the Bank of Italy. While large sample survey studies of reported preferences go back to at least Juster and Shay (1964), in addition to papers discussed later, the boom in research is exemplified by Smeeding, Phillips and O'Connor (2000), Coronado, Lupton and Sheiner (2005), Leigh (2012), Crump et al. (2015), Graziani, van der Klaauw and Zafar (2016), Auclert (2017), Bunn et al. (2017), Kan, Peng and Wang (2017), Ameriks et al. (2018), and Fuster, Kaplan and Zafar (2018).

Figure 1: Reported and Revealed Estimated Spending Propensities in NCP Dataset



Note: Based on Tables 9, 10, and 11, see notes to these Tables. Whiskers display one standard error bands.

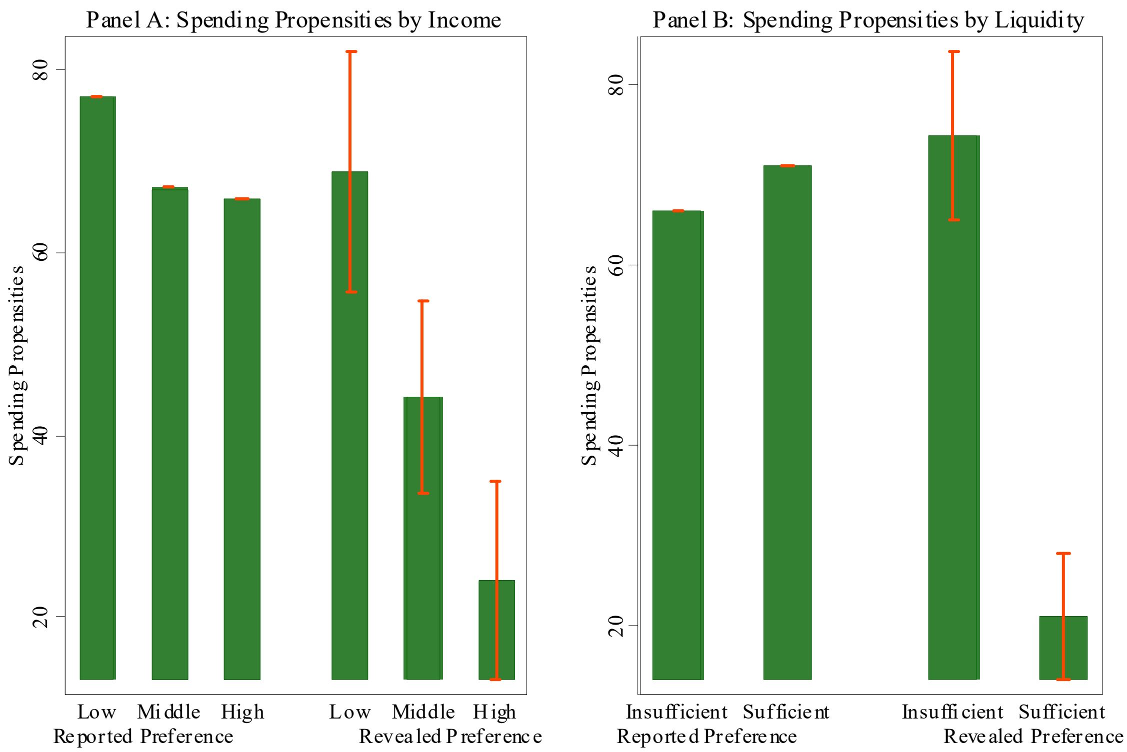
to elicit the dollar spending caused by the rebates. We worked with survey administrators to design and field survey instruments in both the Consumer Expenditure (CE) Survey and the Nielsen Consumer Panel (NCP) during 2008, and we perform our comparison separately within each dataset.

While the Friedman and Wallis (1942) quote above refers to purely hypothetical choices (or 'forced choice' situations), our analysis instead involves the elicitation of causal effects in past choices, a situation in which reported effects are likely to be more accurate (Manski, 1990). Respondents need only infer how their spending would have been different in the hypothetical counterfactual situation in which a payment was not received, a situation they typically encounter.

For our specific case involving survey elicitation methods applied to past spending, we establish three main results.

First, reported spending is highly informative about revealed-preference propensities to spend. As shown in Figure 1, we find large revealed-preference estimates of spending by households that report that they mostly spent their stimulus payments. These revealed-preference spending responses are economically much larger than the spending responses of households that report mostly saving their payments or mostly using them to pay

Figure 2: Reported and Revealed Estimated Spending Propensities in NCP Dataset



Note: Based on Tables 9, 10, and 11, see notes to these Tables. Whiskers display one standard error bands.

down debt. This is true in both the CE survey sample and the NCP sample. While we find statistically and economically significant revealed-preference estimates of spending by households that report that they mostly used their payments to save or pay down debt, these estimates of spending are not inconsistent with the quantification of the reported-preference answers in previous research. In sum, reported spending captures economically large differences in spending.

Second, in each dataset, the average propensities to spend estimated from the two methods are similar. The one difference of note seems likely related to the use of debt for the purchase of durable goods, and the resulting difference between expenditures (measured by our revealed-preference propensities) and out of pocket expenses (likely measured by survey reports of spending effects). These first two findings lend credence to the use of reported causal effects to measure the effect of economic policy or to estimate model parameters.

Third, we find that, in both datasets, reported propensities to spend show almost no variation across households with different levels of income or across those with different levels of liquidity. This is inconsistent with our reading of the large literature on consumption smoothing (e.g. Fuchs-Schündeln and Hassan, 2015), which is that households with little liquidity have higher propensities to spend. And this is also inconsistent with our revealed-preference estimates for NCP goods as shows in Figure 2. And we also find greater spending by less liquid households on broadly-defined nondurable goods in the CE but the difference by liquidity is not statistically significant. So one might conclude that the survey methodology for inferring causation fails to capture an important difference in the population. However, this would be premature. We find no difference in revealed-preference propensities to spend on total goods in the CE, consistent with what we find with our survey reports. Thus, it is possible that less liquid households do not spend at greater rates. We discuss possible interpretations of these results in Section VIII, including that these findings could be driven by shortcomings in either measure or be an artifact of statistical imprecision.

Two caveats apply to all of our findings. First, current survey questions about spending are not clear about the time horizon over which additional spending is to be reported. Our evaluation of these measures uses the horizons typically assumed by other researchers and by policymakers. Second, our revealed-preference propensities only capture spending caused by the arrival of the payment (e.g. omitting any spending that occurred as the policy was developed and when the payments were announced) while reported propensities (presumably) include spending unrelated to the timing of the payment's arrival. Nevertheless, as we discuss in Section V, both theory and empirical evidence suggests that these differences are minor, at least relative to the statistical uncertainty that exists around the quantitative estimate of the propensity to spend from each method.

Our paper's closest predecessors are Hainmueller, Hangartner and Yamamoto (2014) and Karlan, Osman and Zinman (2016). Karlan, Osman and Zinman (2016) finds that reported spending is biased relative to revealed-preference estimates and interprets this finding as due to strategic responses that are specific to the incentives in the micro-credit setting of the study. Our evaluation of reported behavior also relates to the analysis of reported probabilities (as in Manski, 2004, 2017), to the use of reported willingness to pay in environmental economics (see Diamond and Hausman, 1994; List and Shogren, 1998; Harrison and Rutström, 2008), and to conjoint analysis in marketing (see Juster, 1966, and Rao, 2014). These last two literatures conclude that survey-based intentions data are informative for prediction but do not provide precise quantitative responses.

I The economic stimulus payments of 2008

The economic stimulus payments of 2008 were distributed to about 130 million eligible taxpayers (about 85% of tax units) in the spring and summer of 2008. Conditional on a taxpayer meeting a minimum earnings test for 2007, the stimulus payment consisted of \$600 (\$1,200) for individual (joint) filers plus \$300 per child that qualified for the child tax credit. The payment was phased out for high income households. Most importantly, the distribution of the payments involved some quasi-random variation over time. For recipients that had provided the IRS with their pfersonal bank routing number (i.e., for direct deposit of tax refunds), the stimulus payments were disbursed electronically over three one-week periods ranging from late April to mid-May. For recipients that did not provide a personal bank routing number, the payments were mailed (using paper checks) in one of nine one-week periods ranging from mid-May to mid-July. Within each of these two groups, the timing of the payment was determined by the last two digits of the recipient's Social Security number, digits which are effectively randomly assigned. The on-line Appendix contains further details.

II The two datasets

A strength of our study is that we conduct our analysis separately on two different datasets – the CE survey and the NCP – that have different sampling time-frames, survey methods, and recall periods.

We use the 2007 and 2008 waves of the CE interview survey data which contains detailed measures of the expenditures of a stratified random sample of U.S. households. We worked with the Bureau of Labor Statistics to add questions about the 2008 stimulus payments to the survey for interviews conducted between June 2008 and March 2009. A first module of questions asked households whether they received any "economic stimulus payments... also called a tax rebate" since the beginning of the reference period for the interview and, if so, the amount of each payment, the date it was received, and whether it was received by check or direct deposit. A second module, described in the next section, asked households to report spending from the payments.

In our analysis, we focus on two measures of spending: *i*) nondurable spending (and some services) which includes CE categories like food, utilities, household operations, gas, personal care, and tobacco as well as semi-durable categories like apparel, health, and reading materials (average spending $5,400 over three months), and *ii*) total CE spending which adds durable expenditures such as home furnishings, entertainment equipment, and auto purchases (average $10,410 over three months).

Our second dataset is the NCP, available through the Kilts Center for Marketing at the University of Chicago Booth School of Business. The NCP is a panel survey of U.S. households in 52 metropolitan areas that tracks spending mainly on household goods with Universal Product Codes (UPCs, or "barcodes"). At the conclusion of every shopping trip for household items, participants use barcode scanners to manually input the total amount spent and scan the items they purchase.[3] These procedures imply that spending is reported accurately for the goods (and trips) that are reported. On the other hand, the NCP only measures a small share of consumption spending (average spending $601 per month). NCP spending primarily covers purchases at grocery, drugstore, and mass-merchandise stores, and primarily goods such as food and drug products, small appliances and electronic goods, and mass-merchandise products largely excluding apparel. We aggregate total spending reported for each trip to the household-week level for the year 2008.

We merge these NCP data with data from a two-part, multi-wave survey that we designed and was administered by Nielsen while the payments were being distributed. Part I asked questions pertaining to the household's liquid assets and typical behavior.

[3] Participants get newsletters and personalized reminders via email and/or mail to upload spending information and to answer occasional surveys. Households that participate regularly are entered in prize drawings or receive Nielsen points that can be used to purchase items from a catalogue.

Table 1: Sample Summary Statistics

	CE Data, Three-month periods			NCP Data, Weekly		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
	Panel A: Observations					
Number of Observations		10,353			995,748	
Spending type:	CE nondurables & services			NCP household goods		
Spending (dollars)	5,436	4,867	3,017	150.25	99.51	186.24
Spending | Spending > 0	5,436	4,867	3,017	180.13	126.74	190.26
Obs. w/ Spending > 0		10,353			845,487	
Spending type:		CE total				
Spending (dollars)	10,410	8,646	7,195			
ESP (dollars)	259	0	498	70	0	282
ESP | ESP > 0	951	900	504	910	600	521
	Panel B: Households					
Number of Households		4,296			19,149	
Household Size				2.65	2	1.48
Number of Adults	1.96	2	0.82			
Number of Children	0.67	0	1.06			
Income (dollars)	58,707	48,800	41,611			
Indicators: Income ≤ \$20K	0.16	0	0.36	0.15	0	0.35
\$20K < Income ≤ \$50K	0.36	0	0.48	0.37	0	0.48
\$50K < Income ≤ \$100K	0.33	0	0.47	0.36	0	0.48
Income > \$100K	0.16	0	0.36	0.13	0	0.33
Households w/ income data		3,427			15,449	
Liquidity (dollars)	9,172	2,100	19,347			
Indicator: Liquidity > \$2,000	0.50	1	0.50			
"Yes" to "Enough Liquidity"?				0.57	1	0.49
Households w/ liquidity data		1,819			19,149	

Note: For the CE, statistics are based on the first-differenced dataset and so drop the first observation per household in levels. We calculate means, medians, and standard deviations using CE household weight and the NCP projection factor for 2008. In the NCP, income and liquidity are categorical variables. The researchers own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data at The University of Chicago Booth School of Business.

Part II described the program of economic stimulus payments, asked the household to report the details of any stimulus payment received, and then asked the household to report how much of their payments they spent.

Table 1 presents summary statistics for the two datasets, and the on-line Appendix contains our supplemental surveys.

III Survey methodology for reported spending

To measure reported spending, we use two different survey instruments.

The first is that used by Shapiro and Slemrod (2003*b*) in the Michigan Survey of Consumers, which is the basis for most existing research. In the CE this question is:

> Earlier, you or someone in your CU [consumer unit] reported receiving a one-time tax rebate that was part of the Federal government's economic stimulus package. Did the tax rebate lead you or someone in your CU [consumer unit] mostly to increase spending, mostly to increase savings, or mostly to pay off debt?

As with many CE questions, the respondent can respond that they do not know.

In the NCP, this question is:

> Thinking about your household's financial situation this year, is the tax rebate leading you mostly to increase spending, mostly to increase savings, or mostly to pay off debt?

The respondent can then choose one of the same three responses or 'Not sure/don't know.' Consistent with previous use, our question is preceded by a lead-in to remind the respondent about earlier questions about the stimulus payment (the CE) or by questions about the stimulus payment (the NCP).

Our second survey instrument, asked only in the NCP, asks respondents to provide a continuous estimate of the dollar amount that the payment caused them to increase their spending:[4]

> For the following questions, please think about the extra amount you are spending because of this rebate. How much (in dollars rounded to the nearest dollar) are you spending on each of the following:

The first category is 'Food, health & beauty aids, and household products' which is designed to capture spending on household items reported as spending in the NCP. There

[4] There is no standard phrasing for quantitative questions. For example, the 2010 Italian Survey of Household Income and Wealth (SHIW) asks about hypothetical receipt of future income with the phrasing: "Imagine you unexpectedly receive a reimbursement equal to the amount your household earns in a month. How much of it would you save and how much would you spend? Please give the percentage you would save and the percentage you would spend" (as translated in Jappelli and Pistaferri, 2014).

Table 2: Reported-Preference Spending Propensities

	CE Data			NCP Data		
Reported Use of Payment	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
Mostly to Spend	0.32	0	0.47	0.19	0	0.39
Mostly to Save	0.17	0	0.38	0.27	0	0.44
Mostly to Pay Debt	0.51	1	0.50	0.54	1	0.50
Households w/ Response		4,076			19,149	
Reported Spending ($)						
All Spending (A)				452	300	627
Household Items (B)				62	0	182
Non-Household Items				390	200	564
Households w/ Responses					19,059	
Stimulus Payment Amount ($) (C)	951	900	504	910	600	521
Spending Propensity (A/C)				50		
Spending Ratio: All Items to Household Items (A/B)				7.32		

Note: For the CE, statistics are calculated using household weights. For the NCP, statistics are calculated using the NCP projection factor for 2008. The last two rows are calculated as ratios of means. See notes to previous table.

are four additional categories: 'entertainment and services,' 'durable goods,' 'clothing', and 'all other types of purchases.'

A critical question is: over what time horizon is spending reported? The present-tense wording of the question, the lag between payment receipt and survey, and the typical use of language all suggest that people report additional spending caused by the payment over a few weeks or months following the payment, and possibly also preceding it. This is exactly how the answers have been used in previous research. And we evaluate this interpretation of the answers by contrasting them to revealed-preference spending over one to three months following payment receipt.

Closely related, our questions are only asked of households that have reported receiving a payment. Previous research suggests that ex post questions capture more spending and that they are more accurate than ex ante questions.[5]

Table 2 presents the responses to the reported-spending questions in each of our datasets. We find that roughly a fifth of households report that they mostly spent their payments in the NCP and roughly a third report that they mostly spent in the CE. In

[5]First, there is a tendency for less spending to be reported in response to questions before tax reductions or payments as compared to ex post questions (Sahm, Shapiro and Slemrod, 2012). Second, ex ante expected behavior correlates significantly but imperfectly with ex post reported behavior (Manski, 1990). Shapiro and Slemrod (2003*b*) find spending responses of the same households before and after the stimulus payments have a correlation of 0.44.

terms of the quantitative spending amounts that people report, NCP households report spending \$452 on average, of which \$62 was on goods covered in the NCP and \$390 was on other goods and services. Given an average stimulus payment amount of \$910, the implied average propensity to spend is 50 percent. Both our discrete and continuous quantitative responses lie within the range found in the literature (e.g. Bureau of Labor Statistics, U.S. Department of Labor, 2009, and Jappelli and Pistaferri, 2014).

Why do we find a significantly higher share of households reporting that they will mostly spend the payment in the CE than in the NCP? One possibility is that this is due to the tendency for reported spending to rise with the amount of time between receipt and survey.[6] In the CE, we survey households up to four months after receipt while in the NCP we survey them at most 7 weeks after receipt. But as we argue in on-line Appendix A, existing evidence suggests that this effect is likely too small to explain the difference we observe.

The more likely possibility is that the reported spending propensities differ between the surveys because the samples in the surveys differ. This difference in samples does not impede our main analysis of whether, in each dataset, the two different methodologies estimate the same propensity to spend.

IV Revealed-preference methodology

We use the different timing of the disbursement of payments across households to estimate the causal effect of the receipt of a payment on spending, following Parker et al. (2013) for the CE and Broda and Parker (2014) for the NCP.

Letting R_i be the response to the discrete spending question, we estimate an equation of the form:

$$\begin{array}{c} \Delta C^{CE}_{i,t} \\ \text{or} \\ C^{NCP}_{i,t} \end{array} = \left\{ \begin{array}{c} \beta(L)P_{i,t}\mathbf{1}[R_i = \mathit{Mostly\ Spend}]+ \\ \gamma^S(L)P_{i,t}\mathbf{1}[R_i = \mathit{Mostly\ Save}]+ \\ \gamma^D(L)P_{i,t}\mathbf{1}[R_i = \mathit{Mostly\ Pay\ Debt}] \end{array} \right\} + \boldsymbol{\theta}'\mathbf{X}_{i,t} + \tau_t + \eta_{i,t} \tag{1}$$

where $C_{i,t}$ is the log or the dollar amount of spending, $P_{i,t}$ is an indicator of payment receipt or the dollar amount received by household i at time t, the $\mathbf{X}_{i,t}$ are control variables that include R_i and otherwise differ by survey, τ_t is a period-specific intercept. Because it is possible that households with large payments differ from those with small payments, we instrument the distributed lags of the dollar payment with distributed lags of the indicator of payment receipt. The lag polynomials, $\beta(L)$, $\gamma^S(L)$, and $\gamma^D(L)$, measure the changes in spending caused by receipt over time (relative to before receipt). When studying those reporting most spending vs. those not, we impose $\gamma^S(L) = \gamma^D(L)$.

We use a (slightly) different method in each dataset because of differences in the frequency of the data, differences in the available household-level information, and differences in the previous literature. In the CE, spending data is observed over three-month periods

[6] Sahm, Shapiro and Slemrod (2010) Table 3 shows that 36% of those who say that they will mostly spend say their spending rises “within a few weeks,” 50% report “within 1-3 months”, and 14% “more than 3 months.”

and we do not include individual effects but control for observable household characteristics: age, change in the number of children, and change in the number of adults. In the NCP, the frequency is weekly and, because the NCP lacks any covariates that change (during a year) and we estimate in levels, we include individual effects.

As we discuss in the next section, these measures omit any spending responses that are unrelated to the timing of receipt, such as response to aggregate news about the stimulus payment program or responses that occur at a common calendar time such as during an August vacation.

We first estimate variants of equation (1) without the interaction with R_i. We find in the CE that spending on nondurable goods in the three-month period of receipt rises by \$298, 4.7% of spending, or 31% of the payment. In the NCP, we estimate that spending on NCP goods in the month following receipt rises by \$40, 6.9% of spending, or 4.3% of the payment, with some continued spending in the next month or two.[7] Complete results are detailed in on-line Appendix A.6. The estimates imply a slightly lower rate of spending in the NCP than in the CE even after adjusting for less spending measured in the NCP. Although this difference between surveys is not statistically strong, it is consistent with the difference in reported propensities across surveys presented in Table 2.

V Differences in spending concepts in reported and revealed spending methodologies

First, as noted, the revealed-preference measure only captures changes in spending that are related to the timing of the arrival of the payment, while the reported spending measure may also capture anticipatory spending unrelated to this timing. Based on theory and evidence, this difference is likely to be a very small part of the total spending response. In theory, households that respond to news about the stimulus program in advance of receipt are those that optimally smooth consumption over time. Such households should have small spending responses to the news because it represents a very small change in lifetime wealth. In practice, revealed-preference studies find small anticipatory spending effects (e.g. see Table 6, Broda and Parker, 2014).

Second, purchases that use credit may cause a significant difference between our measures. Revealed-preference spending propensities are based on measured expenditures. The CE measures total expenditures, and spending is the price of a new car not just the down payment if it is financed. Similarly, the NCP measures spending using receipt totals, regardless of whether the household paid with a debit card or credit card.

In contrast, the concept of spending measured by the survey-based self-reports is less clear. It seems likely that for larger purchases, such as a financed purchase of a car, the out-of-pocket expense (down payment) is likely to determine the response. To some extent this is not a concern for the discrete response questions, since a large financed purchase with a substantial down payment will lead to an answer of "mostly spend" even if the household responds only based on a down payment. But it is not at all clear whether

[7]These estimates are generally consistent with those of the earlier literature, but are different due to the fact that we omit households that do not respond to reported spending questions.

the reported additional dollar spending is based on total purchase price or out-of-pocket payment.

The final issue is whether the two methods both measure the spending caused by the payment over a similar time horizon. While there is no clear answer, the two measures are used by the previous literature and by policymakers to quantify the same concept of interest: additional spending around the time of payment disbursement and not spending done years later in response to the payment. Additionally, for both measures, the effect of horizon on spending propensity seems to be quantitatively relatively minor for the variation we observe and are interested in (see the discussion on time horizon in Section III and on-line Appendix A).

VI Revealed spending and reported use of payment

Our first main finding is that self-reports of 'mostly spending' are highly informative about the revealed-preference propensity to spend.

The first column of Table 3 shows that households reporting that they mostly spent their payments did indeed spend at large rates. For the CE, during the three month period of receipt, spending on non-durable goods rose by \$366 (Panel A) and on total goods rose by \$1,167 (Panel B) which is more than the average payment amount of \$910. For the NCP, Panels C and D show substantial spending by households reporting that they mostly spent – 13% higher spending in spending over the month following receipt and 5.7% higher spending on average over the three months following receipt.

While not always statistically significant, for all specifications, these spending responses are quantitatively much larger than those of households reporting that they mostly saved their payments or used them to pay down debt (comparing the first and second columns of Table 3). Except for Panel A, spending responses for households reporting that they mostly spent their stimulus payments are nearly double the spending responses of other households. The third column shows that these differences are mostly statistically significant.

Note that, while not always statistically significant, there is economically significant spending by households reporting either that they mostly saved or mostly paid down debt (Table 3, column 2). The final two columns of Table 3 break down this difference between households that report mostly saving and those that report paying down debt.

In sum, the people's discrete reports of their propensity to spend are highly predictive of their behavior. People who report mostly spending are econometrically estimated to have marginal propensities to consume roughly double those of households reporting mostly saving or paying down debt.

VII Average spending propensities

We quantify spending associated with each discrete reported spending response (mostly spend, save, or pay debt) in three different ways.

First, we follow the literature on reported spending and Shapiro and Slemrod (2003*b*) in particular. This approach assumes that the distribution of spending propensities is

Table 3: Revealed Spending by Qualitative Reported Spending

Report:	Mostly Spend	Mostly Save or Pay Down Debt	Equality Test p-value	Mostly Save	Mostly Pay Down Debt
Panel A:CE Survey, Nondurable Spending During Three-Month Period of Receipt					
Spending in dollars	366	267	0.29	221	255
	(120)	(111)		(133)	(103)
Log-percent increase	7.02	3.71	0.02	2.79	3.87
	(1.89)	(1.72)		(2.08)	(1.61)
Percent of payment spent	39.0	27.5	0.24	22.5	25.9
	(13.5)	(12.4)		(13.9)	(10.9)
Panel B: CE Survey, Total Spending During Three-Month Period of Receipt					
Spending in dollars	1167	539	0.03	645	357
	(400)	(350)		(393)	(327)
Log-percent increase	8.19	2.52	0.00	2.72	2.51
	(2.52)	(2.17)		(2.63)	(2.01)
Percent of payment spent	122.4	52.9	0.03	63.9	32.9
	(45.1)	(39.4)		(41.5)	(34.7)
Panel C: NCP, Spending on Household Goods in Month Following Receipt					
Spending in dollars	76.46	31.33	0.00	27.95	32.84
	(10.80)	(5.85)		(7.83)	(6.76)
Percent increase	13.30	5.39	0.00	3.79	6.15
	(2.34)	(1.29)		(1.80)	(1.43)
Percent of payment spent	8.16	3.39	0.00	2.99	3.58
	(1.19)	(0.64)		(0.85)	(0.75)
Panel D: NCP, Spending on Household Goods Over Three Months Following Receipt					
Spending in dollars	89.34	44.65	0.03	72.85	30.89
	(25.79)	(20.03)		(24.18)	(20.81)
Percent increase	5.70	3.08	0.08	9.86	8.84
	(1.99)	(1.51)		(5.62)	(4.61)
Percent of payment spent	8.67	4.23	0.06	6.93	2.90
	(2.74)	(2.08)		(2.48)	(2.20)

Note: In Panels A and B, each row comes from a regression of a distributed lag of an indicator of payment receipt, or the amount instrumented with the indicator, as well as month effects, age, change in the number of children, and change in the number of adults on dollar change, or log change, in spending. In Panels C and D, each row comes from a regression of a complete distributed lag of an indicator of payment receipt or the indicator used as an instrument for the amount, as well as month and household effects on dollar spending or that normalized by average monthly spending during 2008Q1. The final two columns are from a different regression than the first three. Parentheses contain standard errors that are robust to arbitrary heteroskedasticity and within-household correlation. See notes to Table 1.

triangular and calibrates the parameters from the data and by assumption. Given the share of households that report 'mostly spend,' this method implies a propensity to spend of 2/3 for these households. While this method is somewhat arbitrary, it is important to evaluate this method because it is the one used for inferring the (partial-equilibrium) aggregate effect of the tax policies on the economy in existing work.

Our second calibration instead assumes a higher upper bound on the distribution of spending propensities to account for the possibility that some households purchase durable goods and so spend more than the payment amount.[8] In this calibration, the average propensity to spend among households reporting 'mostly spend' is roughly 80% and the average propensity of the rest of the population is 25% (by assumption).

Our third quantification is based on the dollar amounts that people report spending in the NCP. For households reporting 'mostly spend,' we divide average total spending by the average payment amount. We do the same for households reporting 'mostly save' and 'mostly pay down debt.' This gives a self-reported average propensity to spend for each reported spending response that we can use in each survey.

In order to calculate revealed-preference spending propensities for all expenditures from CE nondurable goods and from NCP goods, we scale up the propensities by the ratios of goods covered. For example, in the CE, we scale up the propensities to spend on non-durable goods by the ratio of average total spending to average non-durable spending (1.94).

Our main finding is that the two methods – reported and revealed preference – deliver broadly similar estimates of the sample-average propensity to spend (Table 4, final column). The average propensity to spend based on reported spending range from 40 to 58 percent in the CE Survey and from 27 to 50 percent in the NCP. The revealed-preference propensities range from 57 to 67 percent in the CE, and from 33 to 48 percent in the NCP.

Second, while the two methods deliver broadly similar estimates, the revealed-preference estimates are slightly larger than those derived from household reports of spending in the CE, but not in the NCP. This difference is consistent with differences between methods in the measurement of debt-financed purchases. If people report only out of pocket payment as the amount of the payment that they spent, since the CE instead records total expenditures, then we would expect the CE revealed-preference measure to exceed the CE reported measure. This is what Panel A of Table 4 shows: total revealed spending and reported spending are quite different, while non-durable scaled revealed spending and reported spending are more similar (and mainly for the columns besides 'mostly pay down debt'). Since the NCP does not cover large durable goods, we would not expect as much of a difference in the NCP, consistent with what we find.

VIII Spending propensities by liquidity and income

In this section, we test the ability of the methodologies to produce mutually consistent estimates of the propensity to spend for subsamples defined by income and liquidity, char-

[8] Further, to maintain a reasonable distribution, we also assume that the spending propensity is uniformly distributed for households that report not mostly spending. See On-line Appendix $A.7$ for details.

Table 4: Reported and Revealed Spending Propensities (percent of payment)

Qualitative Reported Spending:	Mostly Spend	Mostly Save	Mostly Pay Debt	Weighted Average
Panel A: Consumer Expenditure Survey				
Fraction of Sample (Table 1)	0.32	0.17	0.51	
Reported Spending Propensities				
Shapiro-Slemrod Calibration	67	27	27	40
Alternative Calibration	80	25	25	42
Reported Spending in NCP	98	25	44	58
Revealed Spending Propensities, Three Months of Arrival				
Total Scaled from Nondurable (1.94)	76	44	50	57
Total Spending	122	64	33	67
Panel B: Nielsen Consumer Panel				
Fraction of Sample (Table 1)	0.19	0.27	0.54	
Reported Spending Propensities				
Shapiro-Slemrod Calibration	67	18	18	27
Alternative Calibration	80	25	25	36
Reported Spending in NCP	98	25	44	50
Revealed Spending Propensities, Month After Arrival				
Scaled up by CE Revealed by Category (9.4)	77	28	34	40
Scaled up by NCP Reported (9.9, 6.5, 6.2)	81	20	22	33
Revealed Spending Propensities, Three Months After Arrival				
Scaled up by CE Revealed by Category (9.4)	81	65	27	48
Scaled up from NCP Reported (9.9, 6.5, 6.2)	86	45	18	38

Note: Scale factor for CE nondurable goods and some services is the ratio of CE total spending to nondurable spending (See on-line Appendix Section A.7 and Table A.2). The first scale factor for NCP data is the ratio of the revealed propensity to spend on all goods in the CE relative to the spending on NCP-type goods in the CE (method 3 in Table 5 of Broda Parker, 2014, which is the intermediate scale factor of the three considered). The second scale factor is the average reported payment spent on all goods and services in the NCP divided by the average reported payment spent on NCP goods for each discrete reported spending response. Further details and the calibrations are described in the on-line appendix. See notes to previous tables.

acteristics that are associated with different spending responses both in leading theories and, at least in our reading, much previous empirical research.

We measure income as family pre-tax income over the previous 12 months (in the CE) and during the previous calendar year in the NCP. In each dataset, we divide the sample roughly into thirds. We measure liquidity as the sum of balances in checking and saving accounts prior to the first interview in the CE. In the NCP, we measure liquidity from our survey question "In case of an unexpected decline in income or increase in expenses, do you have at least two months of income available in cash, bank accounts, or easily accessible funds?"

First, the first two columns of Table 5 show that the discrete reported spending responses are unrelated to income (Panel A) and liquidity (Panel B). The share of households reporting 'mostly spend' does not differ by income or liquidity in the CE or NCP. The only difference of note is in the CE, where we find evidence inconsistent with most revealed-preference studies: 37 percent of households with high liquidity report that they mostly spent their payments compared to only 29 percent of households with low liquidity.

Second, column 3 of Table 5 shows some evidence of somewhat higher quantitative reported spending by the lowest income group (below $35,000): an average propensity of 59 percent relative to 48 or 45 percent for middle and high income groups.[9] But panel B shows that reported spending propensities are very similar between levels of liquidity. Thus, as with the discrete responses, we find little relationship between reported spending propensities and liquidity.

Canonical theory and our reading of the literature on consumption smoothing would lead us to expect larger spending propensities by households with low income or low liquidity. Although statistically weak, we find this pattern in revealed-preference spending propensities on nondurable goods in the CE (the fourth column of Table 5). Similarly, in the NCP (last column of Table 5), the spending propensities estimated by revealed-preference imply significantly larger spending responses by lower income or liquidity groups (by factors of nearly 3) than do the propensities reported by households, or than do the reports of 'mostly spend' (which do not differ across income groups).

However, there is no evidence that spending propensities on total spending as measured in the CE are higher for households with low income or liquidity (column 5). Total spending propensities reported by households and propensities estimated from behavior are both uncorrelated with income and liquidity, although these revealed-preference estimates have very low statistical power.

We see three reasonable explanations for this pattern. The revealed-preference spending propensities for NCP goods and for nondurable goods in the CE could be misleading because they measure only spending on a subset of expenditure categories, missing services and vehicle purchases for example. If tax rebates cause households with significant liquidity to spend disproportionately more on goods and services outside of these measures, then the observed differences in revealed-preference non-durable spending propensities across liquidity or income would be overstated. If this bias were strong enough, propensities to spend from the payment could be uncorrelated with income or liquidity and the reported

[9] This difference is driven by lower spending by households who report that they mostly saved or paid down debt.

Table 5: Propensities to Spend by Income and Liquidity

	Reported Spending			*Revealed-Preference Propensities*		
	Share Reporting		*Reported*	*Three Months*		*One Month*
	Mostly Spend		*Propensity*	Nondurable	Total	NCP Goods
	CE	NCP	NCP	CE	CE	NCP
			Panel A: By Household Income			
Income < \$35K	0.32	0.20	59	52.6	70.0	6.9
				(23.5)	(45.4)	(1.3)
\$35K ≤ Inc. < \$70K	0.31	0.19	48	-11.2	-25.4	4.5
				(21.5)	(88.5)	(1.1)
\$70K ≤ Income	0.33	0.20	45	17.0	84.1	2.4
				(20.1)	(82.0)	(1.1)
Low less middle	0.01	0.01	11	63.8	95.4	2.4
				(31.9)	(99.5)	(1.7)
Low less high	-0.01	0.00	13	35.6	-14.1	4.5
				(30.9)	(93.7)	(1.7)
			Panel B: By Household Liquidity			
Liquid Assets < \$2K or Insufficient Liquidity	0.29	0.17	48	48.3	44.6	7.5
				(26.5)	(66.5)	(0.9)
Liquid Assets ≥ \$2K or Sufficient Liquidity	0.37	0.21	51	19.1	48.3	2.1
				(21.7)	(86.0)	(0.7)
Low liquidity less high	-0.08	-0.04	-3	29.2	-3.7	5.4
				(34.3)	(108.7)	(1.1)

Note: The first three columns report the share of households in each income or liquidity group that reports 'mostly spend.' Each income group contains roughly a third of observations and each liquidity group roughly half. The last three columns are from estimation of equation (1) imposing $\gamma^S(L) = \gamma^D(L)$ separately on each liquidity or income group. See notes to other tables.

propensities could be correct. Alternatively, it may be that reported causal estimates are biased. One possibility is that high-income or high-liquidity households report purchases that they would have made anyway as spending caused by the payment. A final possibility is that people may report spending or saving relative to their usual behavior. People who are persistently constrained and so spend money as it arrives may perceive and report 'mostly saving' when they spend more slowly than they typically do; similarly people who typically save may perceive and report 'mostly spend' when they spend some of their payment more rapidly than usual.

Does other research also find that reported propensities are unrelated to income or liquidity? The lack of difference in reported spending responses across households is consistent with the pre-existing evidence on discrete reported propensities. Sahm, Shapiro and Slemrod (2010) analyzes reported spending responses to these same 2008 stimulus payments and concludes that less-well-off households were not more likely than rich households to spend the tax payment. Similarly, the Shapiro and Slemrod (2003*a*) study of the 2001 tax rebates finds that the rate of reporting 'mostly spend' actually increases with household income and with households stock wealth.

In contrast, Jappelli and Pistaferri (2014) finds that the continuous quantitative reported spending propensities rise with cash-on-hand in the 2010 Italian SHIW. There are (at least) three possible reasons for this different finding. First, households in the NCP and SHIW may have different survey response rates and sample selection, so that they cover different types of households (in addition to drawing from the populations of different nations). Second, the questions in the SHIW are purely hypothetical, referring to an abstract increase in income, and to an amount significantly larger than the typical stimulus payment we study (relative to income). Finally, the questions are asked differently. The SHIW asks the questions as propensities, bounded between 0 and 100 percent. Our questions in NCP asked about spending in dollars in five different categories of expenditures, with spending bounded by 0 and $2,999 in each category.

IX Discussion

Our analysis finds that reported spending propensities are highly informative. The people who report 'mostly spending' their payments spent at roughly twice the rate as people reporting that they used their payments to save or pay down debt. Further, on average, people report spending roughly the same fraction of their 2008 tax payments as their behavior reveals. Thus, estimates of the spending caused by the 2008 economic stimulus payments based on survey reports of spending propensities appear quite similar to those based on traditional revealed-preference measures.

Our analysis of whether reported propensities accurately capture differences in spending across households is less conclusive. In both surveys, we find that people do not report different spending propensities at different levels of liquidity. This finding is consistent with the (noisy) estimated spending propensities on total spending in the CE, which like the reported propensities are uncorrelated with liquidity. But this pattern is inconsistent with estimated spending propensities on NCP goods and on CE broadly-defined non-durable goods, both of which are lower at higher levels of liquidity and income. Previous research on reported propensities is also mixed in that the share of people reporting 'mostly spending' payments has generally been found to be uncorrelated with liquidity, while reported continuous propensities to spend from hypothetical income increases have generally been found to be much higher among households with low levels of liquidity.

Our surveys and policy experiment represent an almost ideal situation for the approach of asking people to infer and report the causal effect of a given policy. We employed ex post questions about a policy that the households had recently experienced, so that recall was relatively straightforward. People also had experience with the counterfactual of no payment (most of the rest of their lives). Thus, respondents were familiar with their behavior in both the baseline event and the hypothetical alternative.

We conclude by suggesting several avenues for improvement in the design and collection of self-reported data on spending propensities. First, reported spending questions could be improved by being explicit about the time period that they are asking about. Second, reported spending questions could clarify whether spending refers to purchase price or out-of-pocket payment, and so be more explicit about the purchase of durable items and the treatment of debt. Finally, the accuracy of reported spending might be improved by incentivizing unbiasedness or accuracy. But given idiosyncratic differences in propensities over time, providing such incentives is far from simple.

References

Ameriks, John, Gábor Kézdi, Minjoon Lee, and Matthew D. Shapiro. 2018. "Heterogeneity in Expectations, Risk Tolerance, and Household Stock Shares: The Attenuation Puzzle." *NBER Working Paper No. 25269.*

Ameriks, John, Joseph Briggs, Andrew Caplin, and Matthew D. Shapiro. 2016. "Late-in-Life Risks and the Under-Insurance Puzzle." *NBER Working Paper No. 22726.*

Auclert, Adrien. 2017. "Monetary Policy and the Redistribution Channel." NBER Working Paper No. 23451.

Barsky, Robert, Thomas Juster, Miles Kimball, and Matthew Shapiro. 1997. "Preference Parameters and Behavioral Heterogeneity: an Experimental Approach in the Health and Retirement Survey." *Quarterly Journal of Economics*, 112(2): 537–80.

Broda, Christian, and Jonathan A. Parker. 2014. "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption." *Journal of Monetary Economics*, 68(S20-36).

Bunn, Philip, Jeanne Le Roux, Kate Reinold, and Paolo Surico. 2017. "The consumption response to positive and negative income changes." Bank of England Bank of England working papers 645.

Bureau of Labor Statistics, U.S. Department of Labor. 2009. "Consumer Expenditure Survey Results on the 2008 Economic Stimulus Payments (Tax Rebates)." *http://www.bls.gov/cex/taxrebate.htm.*

Coronado, Julia Lynn, Joseph P. Lupton, and Louise Sheiner. 2005. "The household spending response to the 2003 tax cut: evidence from survey data." Board of Governors of the Federal Reserve System (U.S.) Finance and Economics Discussion Series 2005-32.

Crump, Richard K., Stefano Eusepi, Andrea Tambalotti, and Giorgio Topa. 2015. "Subjective intertemporal substitution." Federal Reserve Bank of New York Staff Reports 734.

D'Acunto, Francesco, Daniel Hoang, and Michael Weber. 2016. "The Effect of Unconventional Fiscal Policy on Consumption Expenditure." *NBER Working Paper No. 22563.*

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives*, 8(4): 45–64.

Friedman, Milton, and W. Allen Wallis. 1942. "The Empirical Derivation of Indifference Functions." 177–183. Chicago University Press.

Fuchs-Schündeln, Nicola, and Tarek Hassan. 2015. "Natural Experiments in Macroeconomics." In *The Handbook of Macroeconomics.* Elsevier.

Fuster, Andreas, Greg Kaplan, and Basit Zafar. 2018. "What woould you do with \$500? Spending Responses to Gains, Losses, News and Loans." *NBER Working Paper 24386.*

Graziani, Grant, Wilbert van der Klaauw, and Basit Zafar. 2016. "Workers' Spending Response to the 2011 Payroll Tax Cuts." *American Economic Journal: Economic Policy*, 8(4): 124–59.

Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2014. "Validating vignette and conjoint survey experiments against real-world behavior." *Proceedings of the National Academy of Sciences*, 112(8): 2395–2400.

Harrison, Glenn W., and E. Elisabet Rutström. 2008. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." Vol. 1 of *Handbook of Experimental Economics Results.* 1 ed., 752–767. Elsevier.

Internal Revenue Service U. S. Department of the Treasury. 2008. "IRS Announces Economic Stimulus Payment Schedules, Provides Online Payment Calculator." https://www.irs.gov/uac/irs-announces-economic-stimulus-payment-schedules-provides-online-payment-calculator.

Jappelli, Tullio, and Luigi Pistaferri. 2014. "Fiscal Policy and MPC Heterogeneity." *American Economic Journal: Macroeconomics*, 6(4): 107–36.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review*, 96(5): 1589–1610.

Juster, Thomas F. 1966. "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design." *Journal of the American Statistical Association*, 61(315): 658–696.

Juster, Thomas F., and Robert P. Shay. 1964. *Consumer Sensitivity to Finance Rates: An Empirical and Analytical Investigation.* National Bureau of Economic Research, Inc.

Kan, Kamhon, Shin-Kun Peng, and Ping Wang. 2017. "Understanding Consumption Behavior: Evidence from Consumers' Reaction to Shopping Vouchers." *American Economic Journal: Economic Policy*, 9(1): 137–53.

Karlan, Dean, Adam Osman, and Jonathan Zinman. 2016. "Follow the money not the cash: Comparing methods for identifying consumption and investment responses to a liquidity shock." *Journal of Development Economics*, 121(C): 11–23.

Leigh, Andrew. 2012. "How Much Did the 2009 Australian Fiscal Stimulus Boost Demand? Evidence from Household-Reported Spending Effects." *The B.E. Journal of Macroeconomics*, 12(1): 1–24.

List, John, and Jason Shogren. 1998. "Calibration of the difference between actual and hypothetical valuations in a field experiment." *Journal of Economic Behavior and Organization*, 37: 193–205.

Manski, Charles F. 1990. "The Use of Intentions Data to Predict Behavior: A Best-Case Analysis." *Journal of the American Statistical Association*, 85(412): 934–940.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica*, 72(5): 1329–1376.

Manski, Charles F. 2017. "Survey Measurement Of Probabilistic Economic Expectations:Progress And Promise." Vol. 32, 411–471. Chicago University Press.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, 103(6): 2530–53.

Rao, Vithala R. 2014. *Applied Conjoint Analysis.* New York:Springer.

Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod. 2010. "Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications." In *Tax Policy and the Economy, Volume 24. NBER Chapters*, 69–110. National Bureau of Economic Research, Inc.

Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod. 2012. "Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How It Is Delivered?" *American Economic Journal: Economic Policy*, 4(3): 216–50.

Shapiro, Matthew D., and Joel Slemrod. 1995. "Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding." *American Economic Review*, 85: 274–83.

Shapiro, Matthew D., and Joel Slemrod. 2003*a*. "Consumer Response to Tax Rebates." *American Economic Review*, 93(1): 381–396.

Shapiro, Matthew D., and Joel Slemrod. 2003*b*. "Did the 2001 Tax Rebate Stimulate Spending? Evidence from Taxpayer Surveys." *Tax Policy and the Economy*, 17: 83–109.

Shapiro, Matthew D., and Joel Slemrod. 2009. "Did the 2008 Tax Rebates Stimulate Spending?" *American Economic Review*, 99(2): 374–79.

Smeeding, Timothy M., Katherin Ross Phillips, and Michael O'Connor. 2000. "The EITC: Expectation, Knowledge, Use, and Economic and Social Mobility." *National Tax Journal*, 53(4): 1187–1210.

Thurstone, L. L. 1931. "The Indifference Function." *The Journal of Social Psychology*, 2(2): 139–167.