Reported Preference vs. Revealed Preference: Evidence from the propensity to spend tax rebates*

Jonathan A. Parker
MIT and NBER

Nicholas S. Souleles
University of Pennsylvania and NBER

June 2017
First Draft: October 2016

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, counerfactually, reported preferences estimates are not higher for households with lower liquidity.

JEL Classifications: C83, C22, B40, H31, D14, D91, E21, E62

*We thank Christian Broda, Erik Hurst, David S. Johnson, Robert McLelland, Claudia Sahm, and Matthew D. Shapiro, as well as participants at talks at UT Austin, Brandeis, Chicago, and the 2017 ASSA Meetings. Colin Gray 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 I thank 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. Parker: MIT Sloan School of Management, 77 Massachusetts Avenue, Building E62-642, Cambridge MA 02139–4307; http://japarker.scripts.mit.edu/ ; Souleles: The Wharton School, University of Pennsylvania, 2300 Steinberg Hall - Dietrich Hall, Philadelphia, PA 19104-6367; http://finance.wharton.upenn.edu/~souleles/ .

1 Introduction

The foundation of testing and estimation in economics is the revealed preference approach in which inferences are drawn from observed behaviors and outcomes in different situations. This approach is closely linked to, and motivated by, revealed preference theory which posits hypothetical economic environments in which agents face the same choice multiple times with all factors beyond those of interest unchanged. In reality such situations are impossible to observe. Inference based on observation either compares different agents at the same time or the same agents but at different times. The alternative to the revealed-preference approach is to draw inference from *hypothetical* rather than *observed* behavior, using reports of intended choices in different hypothetical situations of interest. Not only is this reported-preference approach prevalent in other fields, such as history or psychology, where people's reports of what they would have done in a different situation are more commonly used as evidence, but it has always had a presence in economic theory and is becoming more widespread in empirical economics.

These two methodologies involve different assumptions and inputs. The revealed preference approach requires data on behavior and variation in the causal factors of interest, and also requires the assumption that differences over time or across people do not confound inference. In contrast, the reported preference approach requires comparatively less data and econometric ingenuity, but instead requires careful design of survey instruments and the assumption that the reported hypothetical behavior corresponds to the actual behavior that would occur were the scenarios of interest not just hypothetical, an assumption that raises numerous questions. Do people fully understand the alternative scenario(s) in the question, do they know or figure out accurately what their own behavior would be (or have been), and do they then respond honestly and accurately? Friedman and Wallis (1942, p. 179-80) discussing an early application of the reported preference methodology (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."

In this paper, we evaluate the consistency of estimates based on ex post reported hypothetical behavior and estimates from a randomized field experiment in the domain of consumption smoothing.[1] Specifically, we study the propensity to spend lump-sum tax rebates, an example of interest not only for its importance in distinguishing among models of consumer behavior and choosing among alternative policy options, but also because there has been a boom of recent research on consumption behavior using the reported preference approach.[2]

[1] The reported preference measures in Friedman and Wallis (1942, p. 179-80) and discussed by Friedman and Wallis (1942, p. 179-80) are based on purely hypothetical choices. In contrast, our measures concern a past choice, so that people presumably know their past behavior and only have to conjecture their response to the alternative hypothetical. We discuss various types of reported preference data in Section 4.

[2] While large sample survey studies of reported preferences go back to at least Juster and Shay (1964), in addition to the papers highlighted in the main text, recent examples include Smeeding, Phillips, and

We find that reported preference measures of spending are highly informative, perform well quantitatively on average, but fail to match the importance of liquidity found in the revealed preference analysis of spending.[3] More precisely, the extent to which households report that they spent their Economic Stimulus Payments in 2008 predicts well the extent to which their spending reveals that they spent their payments. Households reporting that they mostly spent their payments spend roughly twice as much as those reporting that they used their payments mostly to save or pay down debt. Further, the quantitative estimate of the average reported propensity to spend in each dataset is close to the average revealed propensity in each dataset. However, reported-spending propensities do not vary across levels of household liquidity, while revealed-preference propensities show larger spending responses for households with low liquidity. We conclude that reported preference data do not reliably measure quantitative spending directly, but that they are highly informative and so can be usefully used in a predictive context, such as potentially to measure changes in average behavior over time.

To conduct our analysis, we use two sets of questions – about receipt of stimulus payments and about perceived spending effect – added to two surveys of consumer spending: *i*) the Consumer Expenditure (CE) Survey, which contains comprehensive measures of household-level expenditures for a stratified random sample of U.S. households, and *ii*) the Nielsen Consumer Panel (NCP), which has a much larger sample and more accurate measure of spending but for a smaller set of goods. We worked with survey administrators at the Bureau of Labor Statistics and at Nielsen to design and field survey instruments consistent with the format of each survey. These surveys first ask about the policy variation, specifically whether the household has received a payment, and, if so, the amount and when it was received. Second, the survey asks people to self-report the difference between their actual spending and their spending in a hypothetical alternative scenario in which they did not receive a tax payment. With only minor customization for the particulars of each survey, we employ the survey instrument that Shapiro and Slemrod (1995, 2003a, 2003b, 2009) have used to measure spending responses to various tax policies. Additionally, in the NCP survey only, our survey instrument asks households to report how much the policy caused them to spend in dollars on both NCP goods and other goods and services.

We construct quantitative, reported-preference estimates of the propensity to spend based on these numerical responses, the discrete responses to the Shapiro-Slemrod questions, and numerical spending propensities associated with different subjective responses

O'Connor (2000), Coronado, Lupton, and Sheiner (2005), Leigh (2012), Jappelli and Pistaferri (2014), National Highway Traffic Safety Administration U.S. Department of Transportation (2009), Crump, Eusepi, Tambalotti, and Topa (2015), Graziani, van der Klaauw, and Zafar (2016), Auclert (2015), Ameriks, Briggs, Caplin, and Shapiro (2016), Bunn, Le Roux, Reinold, and Surico (2017), and Kan, Peng, and Wang (2017).

[3] Throughout the paper, our use of the term 'reported preference' data does not encompass survey data on decision inputs, such as preferences, expectations, or beliefs, but rather refers to survey responses about the difference between people's actual (or hypothetical) behavior and their behavior in a hypothetical alternative scenario.

used in the previous literature.

Our revealed-preference methodology follows analysis by Johnson, Parker, and Souleles (2006), Parker, Souleles, Johnson, and McClelland (2013) and Broda and Parker (2014) and identifies the spending effect from the fact that the disbursement of payments was effectively randomly timed across groups of households. More precisely, we estimate the spending effect of the receipt of a payment by comparing the spending of households that received payments in a given period to the spending of households that received payments in other periods, all relative to household spending prior to the payment disbursements.

It is important to note that these two methods of estimating the propensity to spend do not measure exactly the same concept. Most notably: *i*) the reported preference measures are unclear about the horizon over which spending is measured and *ii*) the revealed preference measure only spending caused by the arrival of the payment (e.g. omitting any spending that occurred at the time the law was passed). Nevertheless, as we discuss in Section 6, both theory and empirical evidence suggests that these differences are minor, at least relative to the uncertainty that exists around the quantitative estimate of the propensity to spend from each method.

We establish three main results.

First, reported spending is highly informative about revealed spending. 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 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, each method delivers a similar population average spending propensity. We show this in two ways. We follow the existing reported-preference literature and map our reported spending categories into quantitative propensities to spend. We also conduct this mapping using the quantitative spending responses from our survey of NCP households. In both cases, aggregating across households, reported preference and revealed preference methodologies produce similar estimates. Revealed-preference measures of spending are slightly larger in both datasets, but this difference is not large relative to standard errors or relative to the uncertainty in the appropriate scaling parameter used to make the two methods both measure total spending. These first two findings lend credence to the use of reported preference to evaluate the effects of policy or to estimate model parameter.

However, our third findings is inconsistent with accurate revealed-preference measurement of spending: people report spending at rates that are unrelated to whether or not they have low liquidity or low income. While we find that households with different levels of liquidity and, in the NCP sample, income have significantly different revealed-preference estimated propensities to spend, they have almost no differences in reported propensity to spend, either measured qualitatively or quantitatively. Reported preference measures are

still informative conditional on liquidity: for low or high liquidity, people reporting that they mostly spent their payments have significantly larger revealed-preference estimated propensities to spend than those reporting not spending. But the share of households reporting that they mostly spent their payments is unrelated to liquidity, as are the dollar amounts reported as spent.

Our findings imply that reported preference measures are informative – people who spent know they spent and report it – so reported spending data is likely to be valuable for predicting behavior. But reported spending is not robustly related to quantitative spending behavior across households, so that reported preference data should not be viewed as reliably measuring quantitative spending directly. Our findings, and these two conclusions, are consistent with people reporting spending or saving relative to their usual behavior given their persistent low or high level of liquidity, as we discuss in Section 10.

Despite the increasingly widespread use of reported preference evidence, there is little prior research evaluating survey reports of behavioral responses. To the best of our knowledge, there is only one other economic research paper that studies reported and revealed behavioral responses to the same variation in the same sample. Karlan, Osman, and Zinman (2016) compares two methods for inferring how borrowers use microcredit loan proceeds and finds that borrowers understate spending and purchases caused by loan receipt. The paper concludes that reported spending is biased relative to revealed preference estimates due to strategic concerns in responses that are specific to the setting. In particular, respondents may have believed that their survey responses would affect future access to micro credit.

This paper is closely related to several larger literatures. Environmental economics has studied reported measures of willingness to pay, mainly related to non-market based outcomes where prices are not observable like environment degradation. Reported willingness to pay in subjective questions is typically much larger than assumed realistic levels of willingness to pay (see Diamond and Hausman, 1994; List and Shogren, 1998; Harrison and Rutström, 2008).

Another body of related research asks households to report preferences or valuations rather than choice. For example, Manski (2004, forthcoming) evaluates the consistency of individual reports about probabilistic beliefs over outcomes (beliefs being part of the specification of preferences); the most relevant work in this area considers prediction about an individual's own future behavior. This literature concludes that survey-based intentions data has significant content, but the content is entirely related to prediction and not estimation of causal responses (as is the focus of the present paper).[4] This conclusion matches the widespread use of subjective reported preferences in the marketing literature – called choice-based conjoint analysis – to predict behavior (see Juster, 1966; and Rao, 2014).[5] This approach, active since at least the 1940's, is viewed as quite successful

[4] A second but distinct use of survey data on beliefs is to separate beliefs from preferences given estimated behavioral responses (such a separation cannot be done with revealed preference data alone absent ad hoc distributional or functional form assumptions). This method has found recent application in finance as for example Vissing-Jorgensen (2003) and Gennaioli, Ma, and Shleifer (2015).

[5] In choice based conjoint analysis researchers specify a set of attributes, describe products by the

in predicting individual choice among goods and largely concludes that well-designed questions are informative, particular when using self-explication (e.g. "how important is feature x in your choice of good y").[6]

Finally, and least related to the present work, there is a growing body of research on randomized surveys which infer behavioral responses from differences in subjective reported responses across randomly different hypothetical environments. For example, Hainmueller, Hangartner, and Yamamoto (2014) compares inference from randomized surveys to typical non-random reported preference estimates.

2 The Economic Stimulus Payments of 2008

Our analysis studies tax payments send to households as part of the Economic Stimulus Act of 2008 signed into law on February 13. The Act called for the U.S. Internal Revenue Service to distribute $100 billion in economic stimulus payments to about 130 million eligible taxpayers (about 85% of tax units) in the spring and summer of 2008.

Each stimulus payment consisted of a basic payment of $600 for individual filers or $1,200 for joint filers, increased by $300 per child that qualified for the child tax credit. To be eligible, a taxpayer had to have either positive income tax liability or at least $3,000 in qualified earnings. For households with low earnings, the payment amount was reduced but not below $300 for individuals or $600 for joint filers. For households with high earnings, the payment amount was reduced by five percent of the amount by which adjusted gross income exceeded a threshold of $75,000 of for individuals and $150,000 for joint filers.[7]

Within each of 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.[8] For recipients that had provided the IRS with their personal 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.[9] The

attributes, create an experimental design of hypothetical products, and have consumers choose from the hypothetical products.

[6] Despite being informative for prediction, probabilistic intensions to purchase a given good are often significantly biased, with reported probabilities much larger than actual probabilities. This literature has further investigated the question of whether one can observe hypothetical choices by using biometric responses to hypothetical questions rather than survey responses (e.g. Smith, Bernheim, Camerer, and Rangel, 2014). And it has compared reported hypothetical choices to actual choices in laboratory settings (e.g. Maximiano, 2012).

[7] All income and dependent information was based on tax returns for year 2007. If subsequently a household's tax year 2008 data implied a larger payment, the household could claim the difference on its 2008 return filed in 2009. However, if the 2008 data implied a smaller payment, the household did not have to return the difference.

[8] The last four digits of a Social Security number (SSN) are assigned sequentially to applicants within geographic areas (which determine the first three digits of the SSN) and a "group" (the middle two digits of the SSN).

[9] The payment was directly deposited only to a personal bank account, a debit card, or a "stored value

IRS mailed a statement to these recipients informing them about the deposit a couple of business days before the electronic transfer of funds. The on-line Appendix contains an example of this letter. 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. For these recipients, the IRS sent a notification letter one week before the check was mailed. Table 1 shows the weekly disbursement schedule in terms of the latest date by which the payments are supposed to have been received by different households.

For a number of reasons – primarily filing a late tax return – a small share of payments were distributed later than the schedule dictated.[10] We are interested in comparing cleanly-identified causal estimates to reported causal estimates. We therefore exclude households with late payments from the samples that we analyze because such delays in payment are non-random and so might introduce bias in our inferences based on variation in timing.

3 The two datasets

We conduct this study using two different household-level panel datasets. The two datasets have different sampling frames, different survey methods, and different recall periods. In each survey, we employ similar econometric techniques to estimate revealed spending, but adapted for the particulars of that survey. In each survey, we employ a similar survey instrument (delivered by different means) for households to report whether a payment was spent, saved, or used to pay down debt. The on-line Appendices contain additional information about our supplemental surveys and our samples of the CE and NCP.

3.1 Consumer Expenditure Survey

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. CE households are up to four times, at three month intervals, about spending over the prior three months (the "reference period"). New households are added to the survey every month, so the data can be used to identify spending effects from payments disbursed in different months. The CE survey also gathers some limited information about wealth in the final interview and how it has changed over the previous year.

card" from a personal tax preparer. The payment was mailed for any tax return for which the IRS had the tax preparer's routing number, as for example would occur as part of taking out a 'refund anticipation loan' or paying a tax preparation fee from a refund. These situations represent about a third of the tax refunds (not rebates) delivered via direct deposit in 2007.

[10] Taxpayers who filed their tax returns after April 15 and before October 15 received payments either in their allotted time based on their SSN, or as soon as possible after this date (about two weeks after they would receive a refund).

We worked with the Bureau of Labor Statistics to add two special modules of questions about the 2008 stimulus payments to the CE survey in interviews conducted between June 2008 and March 2009, which covers the crucial time during which the payments were disbursed. The 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. The question was phrased to be consistent with the style of other CE questions. We follow Parker, Souleles, Johnson, and McClelland (2013) and use this first module of questions to measure the revealed-preference spending effect of the arrival of an economic stimulus payment.

The second module was asked at most once, and only of households that had previously reported a payment in that survey. These households were asked whether the payment led them "mostly to increase spending, mostly to increase savings, or mostly to pay off debt." The next section discusses the survey methodology, but the wording of this question almost exactly follows the main question in the Michigan Survey of Consumers analyzed by Shapiro and Slemrod (2003b). As with many CE questions, respondents can answer that they do not know, and this response is flagged in the Survey.

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, and *ii*) total spending which adds durable expenditures such as home furnishings, entertainment equipment, and auto purchases.

On-line Appendix B contains the language of the CE survey questions and Appendix C contains more details about our use of the CE data.

3.2 Nielsen Consumer Panel

The second survey dataset is Nielsen's Consumer Panel (NCP) available through the Kilts-Nielsen Data Center 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"). Participants are given barcode scanners to use at the conclusion of every shopping trip for households items to input the total amount spent and then to scan the items they purchase.[11] While respondents may not report every trip, the spending that is reported is likely to be accurate because the recall period is short and household use receipts when reporting.

The spending measured in the NCP is concentrated in grocery, drugstore, and mass-merchandise sectors, and covers goods such as food and drug products, small appliances and electronic goods, and mass merchandise products largely excluding apparel. Nielsen

[11] Participants get newsletters and personalized tips and reminders via email and/or mail to upload spending information and to answer occasional surveys. For regularly uploading information, participants are entered in prize drawings and receive Nielsen points that can be accumulated and used to purchase gifts from a catalogue or prizes.

selects a 'static sample' of actively reporting households in each calendar year and produces sampling weights that are used to make the sample representative of the U.S. population along 10 demographic dimensions (including income). We use data at the weekly level on the trip-level spending on household goods for each household for the year 2008. Participants are surveyed when the begin participating in the survey and at the end of each calendar year about their demographic characteristics and previous year's income, and these answers are used for the following calendar year. We use income for 2007, as reported in the 2009 NCP data files.

We merge the NCP data from the KILTS Center with data from a supplemental, multi-wave survey of NCP households that was run while the payments were being distributed and that collected information about economic stimulus payments and additional household characteristics. This supplemental survey was run by Nielsen, using the methods that they typically use to run surveys of their panelists, and was administered in two parts. Part I contains characteristics questions pertaining to the household's liquid assets and typical behavior. Part II first describes the program of economic stimulus payments and then asks "Has your household received a tax rebate (stimulus payment) this year?" Households that respond 'yes,' are then asked about the amount and date of arrival of their stimulus payment, whether it was received by check or direct deposit, the extent to which the amount was expected. Households that respond 'no' (and not that they are certain that they are not getting one) are re-surveyed up to two more times until they report a payment. We follow Broda and Parker (2014) and use this supplemental survey to measure the revealed-preference spending effect of the arrival of an economic stimulus payment.

Following the questions about the actual receipt of the payment, the survey asks whether the payment has lead the household to 'mostly to increase spending,' 'mostly to increase saving,' 'mostly to pay off debt,' or 'not sure/don't know,' using a question again designed to be very similar to those of earlier work as we discuss in detail in Section 4. Unpublished Appendix D contains more information about the survey including information on the multiple waves, the survey instruments, the contact letter and E-mail, and response rates.[12]

3.3 Comparison of the CE and the NCP

Table 2 presents summary statistics from our sample for each survey. Relative to previous research, our sample is limited to households that provide both valid payment information and reported preference information.[13] The NCP spending data is weekly while the CE

[12] The survey was administered to all households meeting a Nielsen static reporting requirement for January through April 2008, which amounted to 46,620 households by email/web and 13,243 by mail/barcode scanner. For both types of survey, the response rates were 72% to the first wave, and 80% after all waves, giving 48,409 survey responses (of which some are invalid).

[13] Our final CE sample thus starts with interviews in September 2007 (when period t in equation (1) below covers expenditures in June to August 2007) and runs through interviews in March 2009 (when period t+1 covers December 2008 to February 2009). Our final NCP sample includes all weeks in 2008.

spending data covers three month periods. Adjusting for frequency, the CE nondurable spending covers about three times the amount of spending that the NCP data does. Nondurable spending in the CE is about half the total amount spent by households (which adds durable and other expenditures such as home furnishings, tuition, rent and mortgage payments, and auto purchases). The NCP sample has 5 times more households. The NCP sample reports lower average payments.[14] Two factors contribute to this difference. First, the CE measure of payments is the sum of all payments during a three-month period, while the NCP measure is only the first payment received. Second, the NCP sample, despite similar average family size, has a slightly fewer number of children per household (not shown). The incomes in the two surveys are similar, but the NCP has less cross-sectional dispersion in reported income.

The net section describes how we measure reported spending propensities and reports our first estimates, and Section 5 does the same for revealed spending propensities. We compare these two measures of spending conceptually in Section 6 and quantitatively in Section 3.

4 The reported propensity to spend

One can distinguish three types of revealed-preference data. Following Manski (1990)'s terminology, the first type involves forced choice in which the individual is asked to choose among purely hypothetical scenarios that they are unlikely to face. This approach is more common in economic theory (e.g. Allais, 1953) and in measurement closely tied to theory in which the hypothetical scenarios are most idealized (e.g. Barsky, Juster, Kimball, and Shapiro, 1997).

The second type of data, planned choice, comes from choices that people will have to make in the future. The reported preference data are based on reports about how peoples' intended behavior depends on some variation in the future scenario, one of which will actually be experienced. Because the individual will have to face one of the scenarios in the future and make a decision, it seems likely that individuals will have or will generate a more accurate understanding of what is causing their own response than for a choice that they will never face. Ceteris paribus then, we expect responses to this second type of question to be (perhaps weakly) more accurate than to the first.

The third type of scenario, and the one to which the present paper pertains, is past choice in which the individual has made a choice and is asked to consider how this choice would have been different under a hypothetical alternative. For past choice questions, the individual knows how he or she behaved under the actual past scenario and has only to construct and describe their behavior in the unobserved alternative scenario to report their response. As such, we expect this type of reported preference to be the most accurate.

[14] Parker, Souleles, Johnson, and McClelland (2013) and Broda and Parker (2014) document that the reported payments have distributions of amounts, temporal distributions, and patterns across households that are consistent with what is known from other sources and with each other.

Not only are our reported preference measures likely to be more accurate because they pertain to past choice, but we also expect that people have a reasonable understanding of their behavior under the hypothetical alternative because most of the time people do not receive stimulus tax payments. Thus, our survey responses about the spending caused by these stimulus payments are likely to be more accurate than reported preference measures in most other contexts that are not as favorable.

To measure reported spending, we use two different questions. The first, based on the survey instrument that is the basis for most existing research, asks the respondent to choose from three main uses of the tax payment –mostly spend, save or pay off debt. These choices have remained the same. The wording immediately preceding the choices has varied to reflect the particulars of the tax change and whether it will happen, is happening, or has already happened.

Our reported-preference question in the CE, which follows the reporting of payment amount, date of receipt, and method of disbursement, 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 answer that they do not know, and this response is flagged in the Survey. The question in the NCP 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 these three options or "Not sure/don't know."

In previous use, this question has always been accompanied by a lead-in to ensure that the respondent understands the policy that is being considered. We follow a similar procedure. Our question is preceded either 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 question, asked only in the NCP, is quantitative. We ask households to report how much the policy caused them to spend in dollars, which gives a direct, reported-preference estimate of the propensity to spend. In contrast to the discrete spending questions, we only loosely follow questions of this type employed in previous research, such as those in the 2010 Italian Survey of Household Income and Wealth (SHIW) studied by Jappelli and Pistaferri (2014). The SHIW questions ask about hypothetical receipt of future income, whereas we are interested in the response to an existing past payment.[15]

[15] The question in the SHIW is "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).

As detailed in on-line Appendix E, we ask respondents to report the in dollar amounts:

> 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. The next three categories are designed to capture spending in areas that would not be reported to Nielsen: entertainment and services, durable goods, and clothing. A final catchall asks about spending on "all other types of purchases."

What does revealed-preference spending measure? Critically, people interpret the question as applying to some implicit horizon after which the payment has arrived. At one extreme, the instant the payment arrived, presumably spending did not instantly respond so that over a short time interval it was mostly saved. At the other extreme, all payments are (ultimately) spent as long as budget constraints are satisfied with equality (or the alternative should be a reduction in labor supply not saving). However, 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 interpretation is also 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 the weeks and months following payment receipt. We return to more differences and similarities in the concepts measured by each methodology in Section 6.

All of our questions are only asked of households that have reported receiving a payment. In answering these questions about past behavior, a respondent only has to imagine the situation without the payment rather than also to hypothesize about how he or she will behave when receiving the payment (or might behave if the policy is hypothetical or they may not be subject to it). Previous research suggests that ex post questions capture more spending and that they are more accurate than ex ante questions. First, there is a tendency for smaller spending responses to be reported for anticipated tax reductions or payments relative to past tax reductions or payments (Sahm, Shapiro, and Slemrod, 2010). Second, reports of ex ante expected spending behavior correlate significantly but imperfectly with ex post reported spending behavior (Manski, 1990). Shapiro and Slemrod (2003b) find a correlation of 0.44 between spending responses of the same households before and after the stimulus payments, suggesting either that households estimate their future behavior imperfectly or that they change their behavior as their circumstances change (or both). In either case, ex post questions provide a more accurate measure of spending.[16]

[16] A common claim that reported preference is the only method for estimating individual-specific treatment effects. This claim relies upon the dubious assumption that the individual can estimate their treatment effect without error. The evidence reported in Shapiro and Slemrod (2003b) is consistent with substantial error at the individual level.

Table 3 reports the responses to the reported-spending questions in each of our datasets. We find that roughly a third of households report that they mostly spent the payments in the CE, and roughly a fifth report that they mostly spent in the NCP. Previous surveys have typically found that about 19 – 24 percent of households report that they mostly saved their payments, which is consistent with the share reporting spending in the NCP, but less than that reported in the CE.

In terms of the quantitative reported spending, 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. Despite a number of differences, this estimate is extremely close to the quantitative, national survey, reported-preference average propensity of 48% in the SHIW (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 the time lag between payment receipt and the subjective question is greater in the CE than in the NCP.[17] Reported-preference studies of observed spending behavior tend to find that cumulative spending rises over months following the arrival of funds.[18] Thus, since more time has elapsed between payment and survey in the CE, more spending has occurred during this additional time, and so more households report having mostly spent their payments. And existing evidence is qualitatively consistent with this reason for the difference. Sahm, Shapiro, and Slemrod (2012) show that the share of households reporting that they will mostly spend rises from 19 percent before the payments are disbursed, to 22 percent a few months afterwards, to 25 percent a year afterwards.[19]

But there are two reasons to think that the timing of the surveys is not the cause of the difference in reported spending. First, quantitatively, we find a much larger difference – 13 percent – over a few weeks (on average) than Sahm, Shapiro, and Slemrod (2012) find – 6 percent – over more than a year. Second, and we think more important, we considered the issue of time delay in the survey design, and used different wording in each survey. The CE survey asks "Did the tax rebate lead you or someone in your CU [consumer unit] mostly to . . . [emphasis added]" while the NCP instrument asks "Thinking about your household's financial situation this year, is the tax rebate leading you mostly to . . .

[17] In the CE, we are surveying households up to four months after receipt, with the time distance roughly evenly distributed over the four months, while in the NCP households are surveyed at most 7 weeks after, with most households reporting less than three weeks after receipt.

[18] This is particularly true in studies using credit card data (see Agarwal, Liu, and Souleles, 2007), and to a lesser extent for studies using CE data (see Parker, Souleles, Johnson, and McClelland, 2013). It is also true for reported spending. Sahm, Shapiro, and Slemrod (2010) show 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" (Table 3).

[19] Additional evidence is provided in Graziani, van der Klaauw, and Zafar (2016) which shows much larger increases in the share of households reporting spending from before the 2011 payroll tax cut (12 percent) to after it (35 percent). However, a payroll tax cut a different policy than the one studied here, and the paper attributes this large change in reported behavior to the fact that is the policy is a tax cut not a one-time payment.

[emphasis added]." In sum, while it appears that while some of the differences in reported spending may be due to differences in the timing of the questionnaire, the majority of the difference in reported rates is likely due to differences in sample or other differences in the surveys.

Do these issues of timing and sample affect how we compare reported and revealed spending? With respect to the timing of the survey instrument, we measure revealed-preference spending over a shorter horizon in the NCP than in the CE, consistent with the different horizons over which reported-preference spending appears to be reported in each survey. With respect to the sample, while non-representativeness of surveys is a concern for both methodologies' ability to estimate an unbiased measure of the average propensity to spend the payments in the population, this is not our main focus. Non-representativeness does not hinder the evaluation of whether the two different methodologies provide mutually consistent estimates of the average propensity to spend for any given sample of households.

5 The revealed propensity to spend

This section presents our revealed preference methodology and estimates for each dataset. The next section discusses what the revealed and reported preference estimates each measure in theory, and contrasts and compares them in practice.

To calculate the revealed-preference propensity to spend, we use the randomized timing of the disbursement of payments to estimate the causal effect of the receipt of a payment on spending. We follow the previous research and use slightly different estimating equations in each dataset (motivated by the different frequency of observed spending in each dataset).[20] In the CE, we estimate the following regression equation to measure the average impact of the receipt of a payment on spending:

$$C_{t,t} - C_{t,t-1} = \alpha\left(L\right) P_{i,t} + \tau_t + \boldsymbol{\theta}'\mathbf{X}_{i,t} + \eta_{i,t} \tag{1}$$

where $C_{i,t}$ is either the dollar amount of spending or the log of the same, $P_{i,t}$ is the key stimulus payment variable, which is either a dummy variable indicating whether any payment was received by household i in three-month period t or that dummy variable times the amount of the payment received, the $\mathbf{X}_{i,t}$ are change in the number of adults and change in the number of children in the household, included only to reduce unexplained variation in spending, and τ_t is a period-specific intercept. Finally, $\eta_{i,t}$ captures all expenditures unexplained by the previous factors. The parameters of interest are the elements of the lag polynomial $\alpha(L)$ which measure the changes in spending in the period of and periods following receipt relative to the period before the receipt. The polynomial includes all possible lags which is necessary for consistent estimation given possibly long-lived spending effects. We present impulse responses in levels by summing the estimates

[20] We have also used the levels specification that we use in the NCP in our analysis of the CE and we find similar results.

of changes in spending, assuming no spending prior to receipt.[21]

In the NCP, we have weekly data and do not have access to demographic variables that change during the year so that we estimate monthly (four week) impulse responses in levels directly. To keep the regression models comparable, we include a household-specific intercept, μ_i, to capture differences in the average level of spending across households and estimate:

$$C_{t,t} = \alpha(L) P_{i,t} + \mu_i + \tau_t + \eta_{i,t} \tag{2}$$

where $P_{i,t}$ is an indicator (or dollar amount) for whether a payment was received in that week of any of the previous three weeks.

For each dataset, we want to investigate heterogeneity in treatment effect – how spending responses differ by a household's reported responses, liquidity, or level of income. To do this, we simply estimate equations 1 and 2 on subsamples of each dataset.

We construct standard errors and statistical tests clustering residuals by household to allow for arbitrary heteroscedasticity and within-household serial correlation.

It is important to note that estimation based on randomization in the timing of receipt by construction omits any spending effects that are independent of the timing of receipt. One such type of spending is any spending that households might do as the stimulus payment policy is considered, announced, and enacted. Another is the possibility that households choose to spend some of their additional funds after receipt, but in a way unrelated to the timing of disbursement, such as during an August vacation. A final type of spending that we do not measure is that induced by the macroeconomic effects of the program, effects which are presumably uncorrelated across households with the timing of payment receipt. That is, our revealed-preference estimates are partial-equilibrium spending effects, they measure only the change in spending caused by the receipt of the individual payments, and not by the payment program directly or indirectly. We return to these issues in the next section when we discuss the comparability of these measures to those based on reported propensities to spend.

We conclude our discussion of methodology with two comments about the consistency of our estimates. First, consistent identification of the key parameters of interest requires that the variation in $P_{i,t}$ be uncorrelated with all other factors that might influence household expenditure besides the receipt-driven variation of interest. While the timing of payment mailing and payment direct deposit are each effectively random, households are not randomly assigned to different methods of disbursement. However, the spending effects are very similar (relative to standard errors) when we treat the two different disbursement methods as two separate experiments and estimate a common treatment effect. That is, it appears that any differences in baseline spending patterns between households receiving stimulus payments by mail and by direct deposit is relatively small. And so, in the interests of statistical power (and lacking evidence to the contrary), we treat all variation in timing as valid for identifying the spending effect.

[21] We also find similar but weaker results if we allow for some anticipatory spending by including one lead of P in the lag polynial (ie. the lead picks up very little spending in anticipation of receipt). We find similar spending short run spending effects if we instead estimate in levels.

Second, the ultimate sample used in each dataset is not truly a (stratified) random sample of the population for three reasons: the policy was not applied to the population, households can choose not to participate in the surveys, and households may have different propensities to attrit or provide invalid responses. But, as we mentioned at the end of Section 4, such non-random sampling does not invalidate the comparison of revealed and reported spending within the samples that we observe. For example, if we selected a sample of households that on average spent less of their payments than the average household, then each of the two methods should estimate the same spending propensity as the other for the observed sample if each method is equally valid.[22]

Tables 4 and 5 report estimates of the average spending responses in the CE and NCP data respectively. Each table displays result for three different combinations of spending measure ($C_{i,t}$) and payment measure ($P_{i,t}$). In the first column, spending in dollars is regressed on an indicator variable for payment receipt so that the estimated coefficients measure the dollar increase in spending on payment arrival and in the periods following. Panels A and B show that in the CE we find significant and persistently high spending concurrent and following receipt. The average spending response of nondurable goods is \$298 (highly statistically significant) in the three-month period of receipt and \$269 (statistically insignificant) over the next three months. The cumulative amount, \$567, has a standard error of \$280. Panel B shows larger dollar spending on CE total expenditures with similar levels of statistical significance. The first column of Table 5 shows that in the NCP we find a spending response on the smaller subset of goods measure in the NCP of \$40 (highly statistically significant) in the month of and following receipt, and \$6-7 (statistically insignificant) in each of the following months. Comparing the measures in the two datasets, nondurable spending measured in the CE is 2.8 times larger than that in the NCP and the spending response is larger by roughly a factor of 6, so that the spending response appears to be lower in the NCP than in the CE.[23]

The second columns of Tables 4 and 5 measure the average percent increase in spending upon arrival and shortly following.[24] In percent terms, the estimated responses are quite similar across datasets. Spending rises by 3.8% in the three months in which the payment arrives in the CE, which is quite comparable to the average of the increases of 6.9% and 1.8% that we find in the first two months in the NCP, and with the average increase of

[22] This approach answers the question as to whether the two approaches are mutually consistent in the datasets that we have and for the sample of people to which this policy applied. It is not impossible that this answer might differ from the answer for a representative sample of the population. This difference could arise if the households treated by the policy and/or selected into our samples have a different relationship between revealed and reported spending than that in the population. This issue seems unlikely to be a first-order concern, and by analyzing two datasets and reaching the same conclusion in each, we provide evidence that selection of this type is at least not operating differentially across the two datasets.

[23] While this difference is consistent with the fact that a larger share of spending in the NCP is on categories of goods like food at home that are less responsive to tax rebates, these cross-category differences are insufficient to account for the size of the difference in the point estimates.

[24] In the CE data, we use the dependent variable change in log spending, while in the NCP we use spending divided by average weekly spending in the first quarter of 2008.

3.6% over the first three months (last row).

The third columns of Tables 4 and 5 report the propensity to spend or percent of the payment spent, which is our main specification for the rest of the paper. We use the dollar amount of any payment as the key endogenous regressor, and an indicator of payment receipt in place of this amount in the instrument set. Inference accounts for the two-step estimation procedure (and still clustered by household).

Table 4, Panel A shows that households spent 31% of their payments on nondurable goods in the three month period of arrival; Panel B shows 74% spent on all CE-measured consumption goods and services. Both measures show some continued but imprecisely-measured spending. In the NCP, Table 5 shows that we find 4.3 percent of the payment spent in the month of and following receipt on goods measured in the NCP.[25] Given average payments amounts, these propensities to consume are consistent with the dollar spending in the first columns of both Tables, so that we again estimate a slightly lower rate of spending in the NCP than in the CE after adjusting for less spending measured in the NCP.[26] Although this difference between surveys is not statistically strong, it is consistent with the difference in reported propensities presented in Table 3 where we show that 17 percent of the NCP sample reports mostly spending their payments as compared to 30 percent in the CE.

In the CE, the revealed propensity to spend is 31% on nondurable goods during the three months of arrival and about a third of households report mostly spending their payments, which implies the two methods are broadly consistent with each other, as interpreted in the early literature (e.g. Shapiro and Slemrod, 2003a). But this is a weak standard. Further, the estimate of spending in the CE that includes durable goods is substantially larger, and the estimates of spending in the NCP that include only household items are also not readily comparable. In Section 3 we will make a more precise comparisons, but there are differences between what each method measures. Are our revealed preference spending responses comparable to our reported spending responses?

6 Are reported and revealed spending comparable?

This section discusses the four main ways in which the two measures may potentially differ, and presents evidence that these differences are likely to be of little quantitative importance with one important exception.

[25] 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. For example, Parker, Souleles, Johnson, and McClelland (2013) finds that the arrival of a payment caused an increase in total spending amount of 78% (Table 4, Panel C) or 91% (Table 3, Panel C) of the payments, whereas we find only 74% in this sample.

[26] To be more quantitative, and ignoring possible differences due to differing spending propensities across goods, the revealed propensity to spend in the NCP would be about 14 percent if measured spending were scaled up by 2.8 times to CE nondurable goods larger. Scaling by 9.4 to total CE spending would instead give a propensity of 47 percent. The corresponding numbers in the CE are 31 and 74 percent.

First, consider spending done in anticipation of payment receipt, potentially included in the reported spending response. The revealed-preference measure captures only spending related to the timing of the payment. What is missed by the revealed-preference measure is only spending that occurs at a calendar date that is common (or uncorrelated with data of payment receipt). This type of spending is likely to be a very small part of total spending, both in theory and based on evidence.

In theory, the households that are most likely to respond to the announcement are households that are able to smooth spending over time and that do so with sophistication, such as households whose behavior most closely follows the predictions of the rational-expectations permanent-income/life-cycle model of consumer behavior without (relevant or binding) liquidity constraints. The quantitative implication of such a model is that households should increase spending by roughly 5% per annum of the increase in lifetime wealth, or less than one half of one percent of the payment per month. Such a spending effect is small relative to our revealed-preference estimates, and is likely to lead to a reported spending response of "mostly save."

Existing evidence from revealed-preference studies finds very small anticipatory spending effects. Broda and Parker (2014) use the variation in timing across the month in which households learn about their payments to estimate households' spending responses to the arrival of the information about the payments. The estimated spending propensity upon learning about the payment is trivially small relative to the average response to arrival, even for the subset of households with significant liquidity (Table 6).

In sum, any spending responses upon announcement will be measured as part of reported spending propensities and not revealed-preference propensities, but this type of spending is likely to be trivial relative to the estimated responses, and so cause little difference between the measures.

Second, the indirect or general equilibrium channels through which the payments program affected spending – such as through changes in output, wages, and future taxes – are likely omitted from both programs and so pose no problems for our comparison of methodologies. As discussed, the revealed-preference methodology only measures the partial-equilibrium effect of the payments on spending. Similarly, the reported-preference questions ask households how they respond to their payments not to all payments or to the stimulus program as a whole. In the NCP survey, the spending question immediately follows detailed questions about the household's payment and in the CE Survey the lead-in refers back to the detailed questions about the household's payment. The question in each case then asks whether "the tax rebate [singular] is leading you to . . . " (or ". . . did. . . lead you to. . . " in the CE), a phrasing which seems to directly refer to the household's own response, not to the fiscal implications of the program nor to the indirect effects on the economy.[27]

[27]Shapiro and Slemrod (2003a, 2003b) also provides some indirect evidence consistent with this conclusion. The percent of households that report that they mostly spent the 2001 tax rebate was not economically or statistically significantly different between: i) households who thought the 2001 tax cut would make their personal finances better off and those who thought it would make them worse off; ii)

Third, purchases that use credit may cause a difference between our measures. The CE, data record total expenditures, such as the price of a new car rather than just the down payment if it is financed. Similarly, though less extreme, spending is measured in the NCP by receipts for household goods, and so also measures expenditures regardless of the use of credit cards debt or store credit. Thus, revealed preference propensities to spend can exceed 100 percent. And in fact Parker et. al. (2013) documents that the payments caused economically significant spending on new cars in the CE.

In contrast, the concept of spending measured by the reported preference methodology is less clear. At least for large purchases, households probably report outlays rather than purchase prices. For small purchases financed with credit card debt or (possibly) store credit, the wording of the questions is such that the purchase price is likely to determine the response. 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 a mute issue for the discrete response questions, since a large financed purchase with a substantial downpayment will lead to an answer of "mostly spend" whether the household responds about the entire purchase price or the down payment. But this issue definitely matters for mapping the discrete answers into quantitative measures of spending. And this issue matters for interpreting the answers to the dollar spending response reported in the NCP. These questions share this same phrasing and immediately follow the discrete question in the NCP, so very likely only measure out of pocket expenditures for large financed purchases involving new loans.[28]

In sum, due to likely differences in the reporting of the amount financed for large purchases, we need to be careful in comparing estimates from the two methodologies. We are on the most solid ground when comparing spending defined as out of pocket expenses exclusive of debt-financed purchases of large durable goods.

The fourth issue is whether the two methods both measure extra spending over a similar time horizon. Revealed-preference-based spending propensities can only measure spending (shortly before and) shortly after the payment receipt because the statistical power to measure spending responses declines with the time distance from the payment. As noted, reported spending propensities apply to some implicit short horizon following the receipt of the payment. There is no way to know whether these horizons are identical. However there are a number of reasons that is not as significant a limitation as one might at first fear.

The two measures are exactly the ones used in the literature and by policymakers to get at the same concept of interest – "additional" spending right around payment

those who thought the tax cut would improve the economy and those who thought it would worsen it; and iii) those who thought the tax cuts would increase government spending and those who thought it would decrease it.

[28] Additionally, the reported spending amounts are not allowed to exceed $9,999, which also probably implicitly leads the household to respond about out of pocket spending rather than total price. This point applies even more strongly to the SHIW reported spending responses, which are asked in terms of percent of payment spent, and propensities to spend are required to lie between 0 and 1 (Jappelli and Pistaferri, 2014.)

disbursement and not spending ultimately done years later in response to the payment. Thus, we are comparing and judging the measures as they are currently used, imperfect though they are. If the measures provide similar answers, that is at least consistent with a small effect of any differences in spending horizon.

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. As noted in Section 4, the measured effect of delay on reported spending found in previous research is quantitatively small. And the effect of longer horizon on the revealed-preference spending behavior also appears to be quite small. As Tables 5 shows, the additional spending caused by the arrival of a payment declines quite rapidly in the weeks after the payment, so that there appears to be little additional spending after the first month or two.[29]

We also try to compare similar horizons within each survey. Households are asked to report spending sooner after payment receipt in the NCP than in the CE survey, and we also measure revealed-preference spending over a shorter time interval in the NCP than in the CE. There is of course no exact rule for making consistent horizon choices across methodologies, and our choices are in part choices of expedience based on survey technologies and statistical power.

Finally, since there is some uncertainty over what the correct horizon is, we also investigate and report the revealed-preference propensity to spend at different horizons in each survey.

7 Revealed spending and reported mostly spending

In this section we show that revealed spending is consistent with qualitative measures of reported spending: households that report spending their payments have larger estimated revealed spending propensities than those that do not. In the next section, we find that quantitative estimates of the average propensity to spend are similar in each methods. Finally, in Section 9, we contrast the methods ability to measure differences in propensities to spend across different liquidity and income levels and find no consistency between the methods.

To compare revealed propensities by reported behavior, let R_i denote the reported spending response of household i, so $R_i \in \{Save, Spend, PayDebt\}$ where i indexes households. We estimate an expanded version of equations (1) and (2) in which we replace the impulse response to the receipt of the payment, $\alpha(L)P_{i,t}$, with three different impulses responses, one for each value of R_i,

$$\beta(L)P_{i,t}1[R_i = Spend] + \gamma^S(L)P_{i,t}1[R_i = Save] + \gamma^D(L)Pi,t1[R_i = PayDebt] \qquad (3)$$

[29] This is not the case for other datasets and methodologies and samples. The subset of households with credit cards that are near to their credit limits first pay off credit card debt, and then spend at steadily higher rates over a 9 month period until their credit card utilization returns to where it was initially (Agarwal, Liu, and Souleles, 2007).

where $1[\cdot]$ is the indicator function which equals one if its argument is true and zero otherwise. The coefficients $\beta(L)$ measure the dynamic spending responds of households reporting that they mostly spent the payments, and $\gamma^S(L)$ and $\gamma^D(L)$ measure the response for households reporting 'mostly save' and 'mostly pay down debt' respectively. When comparing those responding "mostly spend" to those giving either other response, we impose $\gamma^S(L) = \gamma^D(L)$. Table 6 (for the CE) and Table 7 (for the NCP) show our findings.

First, reported preferences are highly informative about revealed preference measures of the propensity to spend. Our revealed-preference methodology finds that households that reported that they mostly spent their payments did indeed spend at large rates. For total spending, Panel B in Table 6 shows that the payments caused more spending than the amount of the payments (the third row show 122% of the payments are spent on average). As can be seen in comparison to the Panel A (nondurable spending), only a third of this spending is on CE nondurable goods. Table 7 shows substantial spending also, most clearly in the second row of each table where spending increases by 13 percent the month after arrival and rises to 17 percent after three months.

Second, these large spending responses are greater than we observe for households reporting that they mostly saved their payments or used them to pay down debt. While not all such differences are statistically significant, they are economically large. According to the first row of Panel A of Table 6, in the three months during which the payment arrives, those reporting 'mostly spend' in the CE spend an imprecisely-estimated \$100 more on nondurable goods than those reporting they mostly saved or paid down debt. They increase their spending by a statistically significant 3.3 percentage points more, and they spent 11 percent more of their payments.[30] For total expenditures (Panel B), these differences are \$600, 5.7 percentage points, and 70 percent and are all statistically significant. Table 7 shows a similar mostly significant pattern in the NCP data. Spending responses for households reporting that they mostly spent their stimulus payments are about double the spending responses of other households in each specification in the NCP.

The third main point that these tables display is that there is still economically significant spending by households that report that they mostly saved their payments or used them to mostly pay down debt. The estimates of the contemporaneous spending in the CE data (Table 6) are only statistically significant for spending on nondurable goods and services; the estimates of total spending by households not reporting mostly spend are economically larger but are generally not statistically significant at the 95% level. In the NCP (Table 7), the spending of households not reporting that the mostly spent their payments is highly statistically significant, but as noted, only half that of those reporting mostly spend.

Finally, but statistically weakly, relative to those reporting that they used their payments mostly to pay down debt, those reporting they mostly saved their payments had a slightly lower propensity to spend immediately on receipt (Panel A in Tables 6 and 7)

[30] There are similar higher levels of continued spending estimated in the subsequent three month period, although none of these differences are statistically significant.

and a slightly higher propensity to spend later and on durable goods (Panel B in Tables 6 and 7).

8 Revealed and reported spending propensities

In this section, we quantify the reported spending propensities and compare them directly to the revealed propensities. Average reported and revealed propensities are quite similar in each dataset and for each discrete reported spending response. There is some tendency for households that report mostly saving to have larger revealed spending, a pattern not present for households reporting mostly paying down debt.

To show these points, we must quantify spending associated with each discrete reported spending response (mostly spend, save, or pay debt). We do this in three different ways. First, we follow the literature on reported spending and Shapiro and Slemrod (2003b) in particular. It is important to evaluate this method because, while ". . . this is just one of many possible reasonable methodologies for constructing an estimate of the MPC [marginal propensity to consume] from the survey responses. . ." (Sahm, Shapiro, and Slemrod, 2010), it is the one used for inferring the (partial-equilibrium) aggregate effect of the tax policies on the economy in existing work. This method derives a household-level distribution of propensities to consume based on the share of households reporting 'mostly spend' and the following assumptions: that a report of 'mostly spend' corresponds to a propensity to spend of one-half or greater, and that the density of the propensity to spend in the population is piecewise linear. Specifically the density consists of two lines with a kink at the share of households reporting mostly spending, a height of A at zero propensity, and a maximum propensity to spend of χ. Given that the density must integrate to one, the density is unique given a share of mostly spend and χ.

Our first quantification follows Sahm, Shapiro, and Slemrod (2010) by assuming $\chi = 1$, which implies a propensity to spend of 2/3 for households reporting mostly spend in each dataset. Our second calibration instead assumes $\chi = 1.2$ – consistent with some households purchasing durable goods and spending more than the payment amount. We also assume that the density is not continuous, but instead, the propensity to spend of households reporting that they did not mostly spend is uniformly distributed.[31] In our alternative calibration, the average spending by households reporting mostly spend is roughly 80% and the average propensity of the rest of the population is 25% (by assumption).

Our third quantification, most purely reported preference, uses the dollar amounts that people report spending of their payments 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

[31] We make this second assumption because without it the population average propensity to spend is decreasing in the assumed maximal propensity to spend. Further, without this second assumption, given our shares reporting 'mostly spend,' the distribution of propensities has a mode at the origin rather than at the share of households reporting 'mostly spend' (as intended by Shapiro and Slemrod, 2003b).

a reported-preference average propensity to spend for each reported spending response, which we can then aggregate in either the CE or the NCP.

The first set of results in Panel A (for the CE Survey) and Panel B (for the NCP) of Table 8 show the reported spending propensities (as percent of payment) for the Shapiro-Slemrod calibration, for our alternative calibration, and for the average reported spending divided by the average payment for each group in the NCP. The propensities for the calibrations depend on the share reporting mostly spend, but are nonetheless quite similar across the two datasets.

The average propensity to spend based on reported preferences range from 40 to 58 in the CE Survey and from 27 to 50 in the NCP (the first set of results in the last column of each Panel). These are calculated by multiplying the propensities to spend associated with each reported level of spending times the share of households giving each response (the first row in each Panel, from Table 3). In each dataset, the Shapiro-Slemrod calibration leads to the lowest estimate of the average spending propensity and the NCP reported dollar spending answers lead to the largest estimates.

The revealed preference propensities are estimated to be 57 or 67 percent, above the reported preference estimates which range from 40 to 58 percent. In order to calculate a revealed spending propensity for nondurable goods that is comparable to our reported spending on all goods, we scale up the propensities to spend on not-durable goods by the ratio of average total spending to average studied spending (1.94). The difference between reported and revealed propensities are not economically trivial, roughly 12 percent of the payment (giving equal weight to the calibration methods and reported spending in the NCP). However the differences lie well within the bands of statistical uncertainty surrounding both methods.

It is also interesting to note that the pattern of differences is consistent with differences in the treatment of debt-financed purchases. First, a household that increased borrowing by purchasing goods on credit is unlikely to report using its payment to mostly pay off debt. And looking at this column, the reported and revealed estimates are most similar for households that report mostly paying down debt. Second, debt-financed expenditures are most likely to be measured in CE total expenditures and so, if reported preference measures report only out of pocket spending, we would expect to find the largest differences between reported and revealed spending for CE total spending (and for households not reporting mostly paid down debt).[32] This is what Panel A of Table 8 shows: total revealed spending and reported spending are very different, non-durable scaled revealed spending and reported spending are more similar, mainly for the columns besides 'mostly pay down debt.'

In the NCP, the propensities to spend estimated by revealed and reported preference methods are even closer than in the CE. Panel B of Table 8 shows that the revealed preference estimates of spending are 33, 38, 40 and 48 percent of the payment in the

[32] In fact, our NCP survey does not allow reported spending on any category to exceed \$9,999 so that a household cannot report a payment causing it to increase spending by the price of a new car.

NCP.[33] The reported estimates of spending span a larger range, but one that covers all of these revealed estimates. The average revealed preference estimates is 40 percent, while the average reported preference estimate is 41 percent. The most direct reported preference measure (based on reported dollar spending) is 50%.

In sum, the overall quantification of spending is broadly similar in the two methods in each dataset. Second, while the two methods deliver very similar estimates in the NCP, the revealed preference estimates are slightly larger than the reported estimates in the CE. This difference is consistent with the presence of reporting issues related to the presence of some financed durable purchases. Third, fewer households report spending in the NCP sample implying a lower propensity to spend in this sample, and the revealed preference measures match this pattern, with an estimated spending (economically not statistically) significantly lower in the NCP than in the CE sample. While the methods move together across datasets, the next section show that they do not move together across sub-samples defined by liquidity within each dataset.

9 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, characteristics that are associated with different spending responses both in leading theories and in previous empirical research. We look at three pieces of evidence: whether households with low income/liquidity are more likely to report 'mostly spend', whether they report more dollar spending, and finally whether they have higher revealed-preference propensities to spend.

We measure income as family income over the previous 12 months before taxes (in the CE), and as income 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 yes or no responses to 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, Table 9 shows that the share of households reporting 'mostly spend' does not differ by income or liquidity in the CE (Panel A) or NCP (Panel B). The only difference of note is in the CE, where 37 percent of households with higher liquidity report that they mostly spent their payments compared to only 29 percent of households with low liquidity. This finding is the opposite of the higher propensities to spend among households with

[33] To construct these estimates, we scale the estimated propensities in the NCP to total spending in two ways, one revealed and one reported. The scale factor derived from revealed spending is the ratio of the estimated propensity to spend on NCP-type goods in the CE to the propensity to spend on all categories of spending (9.4). The scale factor derived from reported-spending is the ratio of reported propensity to spend on NCP goods to the propensity to spend on all categories of spending from the NCP supplemental survey.

low income or liquidity as is commonly found in revealed preference studies. Qualitative reported spending responses are unrelated to income and liquidity.

Second, we find a similar lack of relationship between *quantitative* reported spending propensities and either income or liquidity. Panel C of Table 9 shows that reported-spending propensities are very similar between levels of liquidity. In fact, they are actually slightly higher for more liquid households. In contrast, there is some evidence of slightly higher quantitative reported spending by the lowest income group (below $35,000) relative to the highest income group (above $70,000), 77 percent relative to 66 percent. But this different is mostly driven by greater rates of spending for people reporting that they mostly saved or paid down debt. We return to this difference by income group subsequently. But, on balance, unlike found in most revealed-preference studies, reported spending propensities are similar for households with different levels of income and particularly liquidity. But maybe this is correct? For this event and at this time, was there actually no stronger spending response by low-liquidity or low-income households? The answer is no.

Our third result is that revealed preference imply higher spending by households with low liquidity and, in the NCP, also those with lower income. Table 10 shows the results for income. Statistical power is limited in the CE, and despite collapsing to two income groups, we find no statistically significant difference in spending propensities by income. The point estimates show higher spending on nondurable goods by low income households (Panel A), but lower total spending (and larger standard errors) by low income households (Panel B). Spending responses in the NCP are more precisely estimated and imply much larger spending by low-income households (by a factor of nearly 3) than implied by the quantitative reported-spending methodology (a factor of 1.17), and certainly than implied by the discrete reported-spending methodology which estimated no difference.

For liquidity, Table 11 shows that low liquidity households spent at much greater rates according to revealed preference analysis of spending on CE non-durable goods and spending on NCP goods. There is only weak evidence of higher total spending in the CE. The statistical weakness of this last finding follows from the large variance in total spending for households over time. Certainly, in no case do we rule out the common previous finding of greater spending by households with low income or liquidity.

This failure of reported preference data to capture differences in revealed response across households is consistent with some pre-existing evidence on reported preferences. In particular, Sahm, Shapiro, and Slemrod (2010)'s analysis of the 2008 tax stimulus finds that the lowest income group and the group with no stock ownership have slightly higher mostly spend rates than the higher-income groups and stock owners, but the differences are small and statistically insignificant. The paper's final conclusion is that less-well-off households are not more likely than rich households to spend the tax payment. Similarly, Shapiro and Slemrod (2003a) studying the 2001 tax rebates find that the rate of reporting 'mostly spend' increases with both household income and households stock wealth.[34]

[34]There is also evidence that people report spending as occurring more rapidly when they are liquidity constrained. Sahm, Shapiro, and Slemrod (2010) show that 48% of low income and 57% of low asset

In contrast, Jappelli and Pistaferri (2014) shows that there is a clear relationship between cash on hand and quantitative reported spending propensities in the 2010 Italian SHIW. There are (at least) four possibly important differences with the quantitative measures in the present study. First, we relate spending to liquidity while Jappelli and Pistaferri (2014) relate to cash on hand defined as all financial assets plus annual income. We reanalyzed Jappelli and Pistaferri (2014)'s data and find that households with more financial assets alone also have higher propensities to spend, so that their finding is not simply driven by income in their measure of cash on hand. But we also find that there is no relationship in the SHIW between reported spending propensities and financial assets less debt. Could different measures of assets and liquidity be causing the difference between findings? It is possible that our data would show a difference in reported spending with a different measure of assets, or that the SHIW would for a measure of liquidity. However, it seems more likely that one of the remaining differences is the culprit.

The second major difference is that households in the NCP and SHIW may have different survey response rates and selection, so that they cover different types of households (in addition to drawing from different national populations). Third, 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). The final difference is the way the quantitative spending questions are asked. 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. Any of these may have made a significant difference, and if so, suggest ways of improving the performance of revealed preference surveys in the US in the future.

Before concluding, we note that while our reported-preference measures fail to capture the greater rates of spending by low income or low liquidity households generally found by revealed preference analyses, they continue to accurately reflect the pattern of revealed spending within each income or liquidity group, just as in Section 6 for the average household. For any sub-group, the estimated spending response is larger for households that report mostly spending their payments, in all 8 Panels of Tables 10 and 11.

In sum, and the main point of this section, households with different levels of liquidity, and to some extent income, have significantly different revealed-preference estimated propensities to spend, but almost no differences in reported propensity to spend. On the other hand, reported preference measures are still highly informative within any group: conditional on an income or liquidity level, households with high propensities to spend report that they mostly spent their payments. We discuss this conflicting evidence in the next and final section of the paper.

---

households report doing what spending they do within a few weeks, while only 36% of the entire sample did. However, the average share reporting mostly spend was 21% for those with low income, 20% for those with low assets, both similar to the 22% overall rate.

10 Discussion

Our analysis finds both strengths and weaknesses of the reported spending methodology. On the one hand, on average people report spending roughly the same fraction of their 2007 tax payments as their behavior reveals. And 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. However, people with low liquidity do not report spending more, as their behavior implies that they do.

These results are consistent with two conclusions about reported preference data, at least as it has been collected and used to measure people's propensities to spend. First, reported preference measures are highly informative – people who spent know they spent and report it – so reported spending data is likely to be valuable for predicting behavior. But second, because reported spending is not robustly related to quantitative spending behavior across households, the data should not be viewed as reliably measuring quantitative spending.

A theory of reporting that can reconcile these findings is that people 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. Thus, comparing people reporting 'mostly save' to those reporting 'mostly spend' in the population is informative, but there is also lots of variation in spending that these self-reports do not capture. And there is no difference in the reporting of 'mostly spend' across levels of liquidity since liquidity and so typical spending behavior are persistent, and people report relative to their typical spending behavior. In our data, baseline behavior seems to even drive perceptions of dollar spending (the NCP data) not just qualitative reports.

While our analysis provides no direct evidence on the value of reported preference data in other domains, our surveys and policy experiment represent an almost ideal situation for the reported preference methodology. We employed ex post questions about a policy that the household 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. It is an open question how accurate answers to forward-looking or purely hypothetical questions about spending will turn out to be in other settings.

In future research in the area of consumption spending, our findings suggest several avenues for improvement in the design and collection of reported spending data. First, reported preference questions about spending have typically been unclear about the horizon over which households are being asked to report additional spending. In the present paper, it seems reasonable to suppose that the implicit time period is between the receipt of the payment and the survey, at least for the analysis of the CE data. And indeed, the answers seem to capture most of the boom in consumption that very shortly follows

the receipt of a payment (given the time pattern we observe in the revealed-preference spending response). Nevertheless, it would be useful to make the horizon explicit to aid use of the data and with the comparison to revealed-preference measures.

Second, it would be useful to field reported-preference questions that clarify whether spending refers to purchases or expenditures, and so are more explicit about the purchase of durable items and the treatment of debt. Our analysis finds that the differences between the methods are most pronounced in the CE and so possibly related to the financing of durable goods. As an example, a survey instrument could ask not only how much spending was caused, but how the spending was financed and how much new debt the payment caused the household to take on (as well as how much existing debt it caused the household to pay off). Such data is necessary if one wants to distinguish expenditures from consumption. To date, it is not clear how people answer the current questions.

Finally, while one might naturally expect that it would be possible to improve measures of reported spending by incentivizing unbiasedness or accuracy, this may not actually be feasible. Revealed preference estimates are measured with error at the individual level. Presumably reported preference measures are also. An incentive to align reported spending and revealed spending at the individual level could then improve reported spending by making it more accurate but then would also bias it by rewarding reports that includes the error contained in the revealed preference measure. An alternative would be to reward correctly reporting a group-average spending response. But a group-average approach would incorrectly incentivize reports based on inferences about other people's behavior not just one's own. The most promising, but costly and difficult approach, would be an incentive that rewarded matching the average reported and revealed spending responses of an individual across a sequence of policy interventions.

References

Agarwal, S., C. Liu, and N. S. Souleles (2007): "The Response of Consumer Spending and Debt to Tax Rebates – Evidence from Consumer Credit Data," *Journal of Political Economy*, 115(6), 986–1019.

Allais, P. M. (1953): "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine," *Econometrica*, 21(4), 503–554.

Ameriks, J., J. Briggs, A. Caplin, and M. D. Shapiro (2016): "Late-in-Life Risks and the Under-Insurance Puzzle," *NBER Working Papers*.

Auclert, A. (2015): "Monetary Policy and the Redistribution Channel," Discussion paper, Manuscript, Stanford University.

Barsky, R., T. Juster, M. Kimball, and M. 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, C., and J. A. Parker (2014): "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption," *Journal of Monetary Economics*, 68(S20-36).

Bunn, P., J. Le Roux, K. Reinold, and P. Surico (2017): "The consumption response to positive and negative income changes," Bank of England working papers 645, Bank of England.

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

Crump, R. K., S. Eusepi, A. Tambalotti, and G. Topa (2015): "Subjective intertemporal substitution," Staff Reports 734, Federal Reserve Bank of New York.

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

Friedman, M., and W. A. Wallis (1942): "The Empirical Derivation of Indifference Functions," in *Studies in Mathematical Economics and Economics in Memory of Henry Schultz*, ed. by O. Lange, F. McIntyre, and T. Yntema, pp. 177–183. Chicago University Press.

Gennaioli, N., Y. Ma, and A. Shleifer (2015): "Expectations and Investment," in *NBER Macroeconomics Annual*, ed. by M. Eichenbaum, and J. A. Parker, vol. 30, pp. 379–442. Chicago University Press.

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

Hainmueller, J., D. Hangartner, and T. Yamamoto (2014): "Do Survey Experiments Capture Real-World Behavior? External Validation of Conjoint and Vignette Analyses with a Natural Experiment," *Working Papers*.

Harrison, G. W., and E. E. Rutström (2008): "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods," vol. 1 of *Handbook of Experimental Economics Results*, pp. 752–767. Elsevier.

Internal Revenue Service U. S. Department of the Treasury (2008): "IRS Announces Economic Stimulus Payment Schedules, Provides Online Payment Calculator," .

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

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

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

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

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

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

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

——— (forthcoming): "Survey Measurement Of Probabilistic Economic Expectations:Progress And Promise," in *NBER Macroeconomics Annual*, ed. by M. Eichenbaum, and J. A. Parker, vol. 32. Chicago University Press.

Maximiano, S. (2012): "Measuring reciprocity: Do survey and experimental data correlate," *Working paper*.

National Highway Traffic Safety Administration U.S. Department of Transportation (2009): "Consumer Assistance to Recycle and Save Act of 2009: Report to Congress," .

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103(6), 2530–53.

Rao, V. R. (2014): *Applied Conjoint Analysis*. Springer, New York.

Sahm, C., M. D. Shapiro, and J. 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.

Sahm, C. R., M. D. Shapiro, and J. Slemrod (2010): "Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications," in *Tax Policy and the Economy, Volume 24*, NBER Chapters, pp. 69–110. National Bureau of Economic Research, Inc.

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

——— (2003a): "Consumer Response to Tax Rebates," *American Economic Review*, 93(1), 381–396.

——— (2003b): "Did the 2001 Tax Rebate Stimulate Spending? Evidence from Taxpayer Surveys," *Tax Policy and the Economy*, 17, 83–109.

——— (2009): "Did the 2008 Tax Rebates Stimulate Spending?," *American Economic Review*, 99(2), 374–79.

Smeeding, T. M., K. R. Phillips, and M. O'Connor (2000): "The EITC: Expectation, Knowledge, Use, and Economic and Social Mobility," *National Tax Journal*, 53(4), 1187–1210.

Smith, A., D. B. Bernheim, C. F. Camerer, and A. Rangel (2014): "Neural Activity Reveals Preferences without Choices," *American Economic Journal: Microeconomics*, 6(2), 1–36.

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

VISSING-JORGENSEN, A. (2003): "Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth? Evidence from Expectations and Actions," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, vol. 18, pp. 139–194. Chicago University Press.

Table 1: The timing of the economic stimulus payments

Panel A: Payments by transfer of electronic funds		*Panel B: Payments by paper check*	
Last two digits of taxpayer SSN	Date by which payment funds deposited	Last two digits of taxpayer SSN	Date by which payment check in mail
00 – 20	May 2	00 – 09	May 16
21 – 75	May 9	10 – 18	May 23
76 – 99	May 16	19 – 25	May 30
		26 – 38	June 6
		39 – 51	June 13
		52 – 63	June 20
		64 – 75	June 27
		76 – 87	July 4
		88 – 99	July 11

Source: Internal Revenue Service (2008)

Table 2: Sample Summary Statistics

	CE Data, Three-month periods			NCP Data, Weekly		
	Mean	Median	Std Dev	Mean	Median	Std Dev
Panel A: Observations						
Number of Observations		10,353			995,748	
Spending type:	CE nondurable goods & services			NCP household goods		
Spending amount	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
Observations w/ Spending > 0		10,353			845,487	
Spending type:		CE total				
Spending amount	10,410	8,646	7,195			
ESP	259	0	498	70	0	282
ESP | *ESP* > 0	951	900	504	910	600	521
Panel B: Households						
Number of Unique 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	58,707	48,800	41,611			
Indicator (Income <= $20K)	0.16	0	0.36	0.15	0	0.35
Indicator ($20K < Income <= $50K)	0.36	0	0.48	0.37	0	0.48
Indicator ($50K < Income <= $100K)	0.33	0	0.47	0.36	0	0.48
Indicator (Income > $100K)	0.16	0	0.36	0.13	0	0.33
[Households w/ Income]		3,427			15,449	
Liquidity	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]		1,819			19,149	

Note: Sample statistics cover the baseline samples used for the rest of the paper. For the CE, statistics are based on the first-differenced dataset and so drop the first observation per households in levels. For the CE, we weight means, medians, and standard deviations by the CE household weight, and for the NCP by the NCP projection factor for 2008. In the NCP, income and liquidity are categorical variables so statistics are omitted; and all NCP calculations are based on data from The Nielsen Company (US) LLC and provided by the Marketing Data Center at the University of Chicago Booth School of Business.

Table 3: Spending Reported by Households

	CE Data, Three-month periods			NCP Data, Weekly		
	Mean	Median	Std Dev	Mean	Median	Std Dev
Used Payment Mostly to Spend	0.32	0	0.47	0.19	0	0.39
Used Payment Mostly to Save	0.17	0	0.38	0.27	0	0.44
Used Payment Mostly to Pay Debt	0.51	1	0.50	0.54	1	0.50
[Households w/ Reported Behavior]		4,076			19,149	
Reported Spending				452	300	627
Reported Spending on Household Items				62	0	182
Reported Spending on Other Items				390	200	564
[Households w/ Reported Spending]					19,059	
ESP amount				910	600	521
Implied Reported Spending Propensity						
(Reported spending/ESP amount in percent)				50		
Ratio of Spending on All Items to NCP Items				6.32		

Note: Sample statistics cover the baseline samples used for the rest of the paper. For the CE, statistics are based on the first-differenced dataset and so drop the first observation per households in levels. For the CE, we weight means, medians, and standard deviations by the CE household weight, and for the NCP by the NCP projection factor for 2008. In the NCP, income and liquidity are categorical variables so statistics are omitted; and all NCP calculations are based on data from The Nielsen Company (US) LLC and provided by the Marketing Data Center at the University of Chicago Booth School of Business.

Table 4: Revealed Spending Response to Receipt of Stimulus Payment, CE

Specification: (Intepretation)	Dollar Spending on Payment (Dollars spent during period)	Log Spending on Payment Indicator (Avg. percent increase in spending during period)	Dollar Spending on Payment, 2SLS (Percent of payment spent during period)
		Panel A: CE Spending on Non-Durable Goods	
Increase in spending during Three Month Period of Receipt	298	4.72	31.1
	(106)	(1.65)	(11.9)
First Three Month Period After Period of Receipt	269	4.79	26.5
	(189)	(2.93)	(22.4)
Second Three-Month Period After Period of	172	3.73	15.8
	(267)	(4.24)	(31.5)
Effect over Period of Receipt and Period After	567	4.76	57.6
	(280)	(2.19)	(32.9)
		Panel B: CE Total Spending	
Increase in spending during Three Month Period of Receipt	730	4.25	74.1
	(340)	(2.11)	(38.6)
First Three Month Period After Period of Receipt	473	0.64	41.6
	(588)	(3.79)	(69.9)
Second Three-Month Period After Period of	170	-2.60	8.1
	(870)	(5.53)	(103.0)
Effect over Period of Receipt and Period After	1,204	2.45	115.7
	(891)	(2.83)	(105.0)

Note: Results are based on regressions in which the dependent variable is the change in household spending (or the log-change) and the regressors are the distributed lag of the stimulus payment (or an indicator of payment receipt), period fixed effects, age, change in the number of children and change in the number of adults. The third column is based on instrumental variables estimaton in which the amount of the payment is excluded and the indicator of payment receipt is used in the instrument set. Data covers 10,353 observations on 4,296 households. CE weights are used. Parentheses contain standard errors that are robust to arbitrary heteroskedasticity and within-household correlation.

Table 5: Revealed Spending Response to Receipt of Stimulus Payment, NCP

Specification: (Intepretation)	Dollar Spending on Payment Indicator (Dollars spent during period)	Spending as Pct. Of Monthly Q1 Spending on Indicator (Avg. Percent increase in spending during period)	Dollar Spending on Payment, 2SLS (Percent of payment spent during period)
Increase in spending during			
Month Following Receipt	39.94 (5.61)	6.89 (1.26)	4.31 (0.62)
Second Month After Receipt	6.95 (7.01)	1.80 (1.60)	0.50 (0.70)
Third Month After Receipt	6.08 (9.35)	2.00 (2.00)	0.30 (1.00)
Effect Over First Three Months	52.97 (19.57)	3.57 (1.50)	5.06 (2.02)

Note: Parentheses contain standard errors that are robust to arbitrary heteroskedasticity and within-household correlation. In addition to the reported leads and lags of the ESP variable, all regressions include household fixed effects and lags of the ESP variable covering the entire sample period. NPC observations are weighted by Nielsen's projection factor. All NCP calculations are based on data from The Nielsen Company (US) LLC and provided by the Marketing Data Center at the University of Chicago Booth School of Business.

Table 6: Revealed Spending by Reported Spending in the CE

Household Report:	Mostly Spend	Mostly Save or Pay Down Debt	Test of Equality, p-value	Mostly Save	Mostly Pay Down Debt
Panel A: 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: 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)

Note: Statistics based on regressions with dependent variable of dollar change (or log-change) in spending and explanatory variables are a distributed lag of payment amount, an indicator of payment receipt, or the latter used as an instrument for the former, as well as month effects, age, change in the number of children, and change in the number of adults. 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.

Table 7: Revealed Spending by Reported Spending in the NCP

Household Report:	Mostly Spend	Mostly Save or Pay Down Debt	Test of Equality, p-value	Mostly Save	Mostly Pay Down Debt
Panel A: 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 B: Cumulative 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	17.10	9.23	0.08	9.86	8.84
	(5.97)	(4.53)		(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: Statistics based on regressions with dependent variable of dollar change (or log-change) in spending and explanatory variables are a distributed lag of payment amount, an indicator of payment receipt, or the latter used as an instrument for the former, as well as month effects, age, change in the number of children, and change in the number of adults. 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.

Table 8: Main Comparison of Reported and Revealed Spending Propensities (in Percent of Payment)

Reported Spending:	Mostly Spend	Mostly Save	Mostly Pay Down Debt	Implied Average Spending Propensity
		Panel A: Consumer Expenditure Survey		
Fraction of Sample (Table 3)	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 3)	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 Propenities, 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 Propenities, 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

Notes: Scale factor for CE nondurable goods and some services is the ratio of CE total spending to nondurable spending (Table 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. The details of the calibration methods are described in the text. See notes to other tables.

Table 9: Reported Spending Behavior by Income and Liquidity

Share Reporting:	Any Report	Mostly Spend	Mostly Save	Mostly Pay Down Debt	Share of total with each income or liquidity
Panel A: Share of households reporting each spending behavior in CE Data					
All Households with Income Data		0.32	0.18	0.50	1.00
Income < $35K		0.32	0.17	0.51	0.35
$35K <= Income < $70K		0.31	0.16	0.53	0.33
$70K >= Income		0.33	0.21	0.46	0.32
Number of households with valid income			3,277		
All Household with Wealth Data		0.33	0.16	0.51	1.00
Liquid Assets < $2K		0.29	0.08	0.63	0.46
Liquid Assets >= $2K		0.37	0.23	0.40	0.54
Number of households with liquid wealth			1,803		
Panel B: Share of households reporting each spending behavior in NCP Data					
All Household with Income Data		0.20	0.28	0.52	1.00
Income < $35K		0.20	0.23	0.57	0.35
$35K <= Income < $70K		0.19	0.29	0.52	0.35
$70K >= Income		0.20	0.33	0.47	0.31
Number of households with valid income			15,449		
All Household with Liquidity Data		0.19	0.27	0.54	1.00
Low Liquid Wealth		0.17	0.11	0.72	0.43
Sufficient Liquid Wealth		0.21	0.38	0.41	0.57
Households with valid liquidity			19,149		
Panel C: Reported Spending Propensities in Percent of Payment, NCP Data					
All Household with Valid Data	69	97	25	45	1.00
Income < $35K	77	100	33	58	0.35
$35K <= Income < $70K	67	95	26	43	0.35
$70K >= Income	66	98	21	39	0.31
Number of households with valid data			15,370		
All Household with Valid Data	69	98	25	44	1.00
Low Liquid Wealth	66	93	26	43	0.43
Sufficient Liquid Wealth	71	101	25	46	0.57
Number of households with valid data			19,059		

Note: In the first three columns, each entry is the share of households in that income or liquidity group that report that spending behavior, out of all households in the main sample that have income or liquidity recorded. Column 5 contains the share of each income or liquidity group. For panels B and D, we weight all NCP observations by the NCP projection factor for 2008.

Table 10: Revealed Spending Propensities (in Percent of Payment) by Reported Spending and Income

Reported Spending:	Any Response	Mostly Spend	Mostly Save or Pay Debt	Any Response	Mostly Spend	Mostly Save or Pay Debt
Household Income	*Panel A: MPC on CE Nondurable Goods During Three Month Period of Receipt*			*Panel B: MPC on CE Total Spending During Three Month Period of Receipt*		
Income < $50K	32.0	38.3	29.0	17.1	71.5	-8.5
	(14.5)	(22.3)	(18.6)	(38.1)	(53.9)	(50.0)
Income >= $50K	10.6	16.7	7.7	79.0	135.2	52.6
	(13.9)	(19.9)	(18.1)	(56.5)	(79.6)	(74.2)
	Panel C: MPC on NCP Household Goods During Month Following Receipt			*Panel D: MPC During Month Scaled to Total Using NCP Reported by Type of Good*		
Income < $35K	6.9	11.9	5.7	68.9	118.1	56.6
	(1.3)	(2.8)	(1.5)	(13.2)	(28.0)	(15.0)
Income > $35K and Income < $70K	4.5	7.7	3.6	44.2	76.3	36.1
	(1.1)	(2.1)	(1.2)	(10.5)	(20.5)	(12.0)
Income >$70K	2.4	5.1	1.8	24.1	50.9	17.4
	(1.1)	(2.0)	(1.3)	(10.9)	(20.3)	(12.7)

Note: Parentheses contain standard errors that are robust to arbitrary heteroskedasticity and within-household correlation. All regressions use the full set of lags (and one lead in the NCP panels) of a dummy for ESP receipt as instruments for the full set of lags (and one lead in the NCP panels) of ESP amounts. These dummies and amounts are interacted with dummies for "Mostly Spend" and "Pay Down Debt", and regression coefficients are summed to yield the coefficients above. In both samples, observations with blank or "Don't Know" answers are omitted from the sample. In addition to the reported leads and lags of the ESP variable, all regressions include period and household fixed effects and lags of the ESP variable covering the entire sample period (in addition to those reported). The CE regressions include controls for age, age squared, number of children in the household, and number of adults in the household. NPC regressions include one lead of the ESP variable. NPC observations are weighted by Nielsen's projection factor for 2008.

Table 11: Revealed Spending Propensities (in Percent of Payment) by Reported Spending and Liquidity

Reported Spending:	Any Response	Mostly Spend	Mostly Save or Pay Debt	Any Response	Mostly Spend	Mostly Save or Pay Debt
Household Liquidity	*Panel A: MPC on CE Nondurable Goods During Three Month Period of Receipt*			*Panel B: MPC on CE Total Spending During Three Month Period of Receipt*		
Liquidity < $2,000	50.2	74.1	39.0	54.3	129.8	18.8
	(20.4)	(29.9)	(26.5)	(52.0)	(92.2)	(62.9)
Liquidity > $2,000	18.5	29.0	13.6	43.3	84.2	24.0
	(17.3)	(24.0)	(22.8)	(67.6)	(91.2)	(89.6)
	Panel C: MPC on NCP Household Goods During Month Following Receipt			*Panel D: MPC on NCP Scaled to Total Using Reported by Type of Good*		
Insufficient Liquidity	7.5	11.7	6.4	74.4	116.1	63.9
	(0.9)	(2.1)	(1.1)	(9.4)	(20.4)	(10.6)
Sufficient Liquidity	2.1	5.9	1.2	21.0	58.4	11.6
	(0.7)	(1.4)	(0.8)	(7.0)	(14.3)	(7.9)

Note: Parentheses contain standard errors that are robust to arbitrary heteroskedasticity and within-household correlation. All regressions use the full set of lags (and one lead in the NCP panels) of a dummy for ESP receipt as instruments for the full set of lags (and one lead in the NCP panels) of ESP amounts. These dummies and amounts are interacted with dummies for "Mostly Spend" and "Pay Down Debt", and regression coefficients are summed to yield the coefficients above. In both samples, observations with blank or "Don't Know" answers are omitted from the sample. In addition to the reported leads and lags of the ESP variable, all regressions include period and household fixed effects and lags of the ESP variable covering the entire sample period (in addition to those reported). The CE regressions include controls for age, age squared, number of children in the household, and number of adults in the household. NPC regressions include one lead of the ESP variable. NPC observations are weighted by Nielsen's projection factor for 2008.