NBER WORKING PAPER SERIES

HOW RESPONSIVE ARE DURABLES EXPENDITURES TO TRANSITORY INCOME SHOCKS?

Christoph E. Boehm
Changseok Ma

Working Paper 35026
http://www.nber.org/papers/w35026

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

The acknowledgement in the paper includes the sentence: Both authors declare that they have no relevant or material financial interests that relate to the research described in the paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Responsive are Durables Expenditures to Transitory Income Shocks?
Christoph E. Boehm and Changseok Ma
NBER Working Paper No. 35026
April 2026
JEL No. E21, E22, E62

ABSTRACT

We estimate how expenditures on durable goods respond to transitory income shocks using variation from the Economic Stimulus Act of 2008. The estimated responses are large: Households spent about 80 percent of their stimulus payment within three months of receiving the income transfer on durables alone. Most of this spending was on motor vehicles, either financed, used, or both, and about 20 percent of the total durables response was on smaller items. To purchase motor vehicles the average household levered up by 40 cents on the dollar using vehicle loans. Our findings suggest that durable goods and their financing are key to understanding the link between current income and spending.

Christoph E. Boehm
University of Texas at Austin
Department of Economics
and NBER
chris.e.boehm@gmail.com

Changseok Ma
changseok.ma@gmail.com

1 Introduction

Leading heterogeneous agents New Keynesian (or HANK) models are predicated on a strong link between spending and income. These models have been used to study topics related to business cycles and macroeconomic policy and have generated a range of important insights (e.g., Kaplan, Moll, and Violante, 2018; Auclert, Rognlie, and Straub, 2024, and the literature that followed). For tractability, however, the overwhelming majority of this literature abstracts from modeling durable goods explicitly. In this paper, we show that the spending response to transfers is particularly strong for durables and that there are other reasons why modeling durables is likely to be important.[1]

Following earlier work, we study the Economic Stimulus Act of 2008 and use the Consumer Expenditure (CE) survey in our analysis. Different from earlier work, we utilize a previously unexploited feature of this survey, namely that almost all durables expenditures are measured at the monthly – as opposed to quarterly – frequency. The monthly data allows us to estimate the effects on durables spending using an event study (or staggered difference-in-differences) design, thus permitting the transparent statement and assessment of the identification assumptions. The method we use addresses several estimation problems identified by prior work while allowing for an arbitrary shape of the expenditure response. Our main finding is that durables spending responds strongly to stimulus payments. We estimate a marginal propensity to expend (MPX) of around 0.80 on durable goods alone. Most of these expenditures were on motor vehicles, either financed, used, or both. To purchase motor vehicles, the average household took out a vehicle loan worth 40 cents for each dollar of stimulus, implying that stimulus payments led households to substantially lever up. About 20 percent of the expenditure response was on smaller items.

To inform our estimation framework and sample we begin with reviewing the Economic Stimulus Act (or ESA) of 2008. This stimulus program has been studied repeatedly because it has – from an econometric perspective – the appealing feature that the receipt timing of stimulus payments for most households was effectively randomized. We argue that care must be taken when using this variation nonetheless. First, households with high (and very low) incomes did not qualify for stimulus payments, so whether a household received a stimulus payment was not random. To avoid comparing households on potentially different spending trajectories we exclusively study samples with stimulus recipients only. Second, whether a household received a stimulus payment by paper check or direct deposit was determined by the household's mode of tax filing, and therefore was also not random. We follow Broda and Parker

[1] Other research has also shown that consumer durables and investment are important for understanding business cycles and policy, including Berger and Vavra (2015), Boehm (2020), and McKay and Wieland (2021), among many others.

(2014) and only use variation in receipt timing among households with the same disbursement method. Third, the receipt timing was only randomized for on-time tax filers. Households whose tax returns were filed or processed late were generally not subject to the randomized disbursement schedule of the Internal Revenue Service (IRS). As Parker et al. (2013) do for parts of their analysis, we therefore only study samples of households who received their stimulus payment approximately on time. Taken together, we exclusively use the variation in receipt timing of on-time recipients among households with the same disbursement method. Absent unplanned departures from the IRS' schedule to role out the ESA, this is precisely the variation that is as good as randomly assigned. Randomization, in turn, implies that the parallel trends assumption required for our event study analysis holds.

A separate but related feature of the ESA of 2008 is that the disbursement schedule was made public about two and a half months before disbursements began. It is therefore possible that households responded to the *expected* receipt of their stimulus payment prior to actually receiving it. We test for such behavior in the data and indeed detect a substantial – and in some specifications – significant expenditure increase one month prior to the reported receipt time. An alternative explanation for this early estimated response is recall error in the reported receipt month or in the reported month of durables expenditures, as we discuss below. In either case, however, is this early expenditure increase part of the causal response to the stimulus receipt and should be included in estimates of the MPX. We therefore define treatment as occurring one month before the receipt of the stimulus payment. In this case, no pre-trends are detectable in the data.

Our main finding is that the estimated expenditure responses are very large. Spending on durables rises to around \$260 one month before receipt. It continues to rise to approximately \$300 in the month of receipt before falling to about \$170 in the subsequent month. The cumulative expenditure response on durables over this three-month period is \$733, corresponding to an MPX of 0.81. This baseline estimate is significantly different from zero with a p-value of 0.03.

The large majority of spending was concentrated on motor vehicles. Other durables categories either received a small share of spending or their responses are imprecisely estimated. We decompose the expenditure response in two additional ways. First, we ask whether the durables response was driven by the extensive margin (i.e., whether to buy) or the intensive margin (spending conditional on buying). While the answer is somewhat sensitive to how these two margins are defined, it is unambiguous that the extensive margin of *large* durables accounted for most of the response. That is, households purchased more cars and trucks, instead of buying differently priced motor vehicles. At the same time, the extensive margin of large durables does not account for the entire response. The intensive margin and small adjustments, which are neither counted as the intensive nor the extensive margin, always account for at least around 30 percent of the total durables response. Second, we decompose expenditures by realized size. When conditioning on realized expenditures below $2,000 dollars, we estimate a significant response accounting for about 20 percent of the total durables response. We emphasize this finding because decompositions by CE survey expenditure category mask the importance of smaller purchases.

We then zoom in on motor vehicles. A perhaps puzzling feature of the estimated responses is that households purchased motor vehicles after receiving stimulus payments averaging around 900 dollars. The explanation for this behavior appears to be linked to the financing of motor vehicles and their resale markets. First, a substantial share of expenditures was on used motor vehicles, which naturally require lower expenditures conditional on a purchase. In addition, we document that households substantially reduce the cash outlays required for motor vehicle purchases by (i) trading in or selling an older motor vehicle and (ii) taking out large vehicle loans. While this is qualitatively no surprise, it is (to us) quantitatively: For instance, we estimate that the average cash outlay for a used and financed motor vehicle between 2007 and 2009 was $760 after netting out trade-ins, sales, and car loans – and hence *less* than the average stimulus payment. When estimating the expenditure responses for all four combinations of new/used and financed/not financed motor vehicles, we find that expenditures increased substantially in three of these categories – with the exception being new motor vehicle purchases that were not financed. This suggests that the amount of cash outlays required for purchases matters.

These findings highlight an adjustment margin, which to our knowledge has received insufficient attention in the literature: Households lever up after receiving stimulus payments by taking out loans. Our estimates imply that the average household borrowed 40 cents for each dollar of stimulus payment received. We call this object the MPLU (for levering up). Thus, instead of spending or saving the one dollar received from the government, the average household borrows an additional 40 cents and thus spends or saves $1.4. The implied heterogeneity in behavior across households is striking. While some households presumably save some or all of their stimulus payment, others lever up to spend more.

Literature Although economists have long estimated consumption expenditure responses to income changes, a major empirical breakthrough was made by Johnson, Parker, and Souleles (2006), who pioneered the use of quasi-experimental designs for estimating the marginal propensity to consume (or MPC). Exploiting the quasi-random receipt timing of the 2001 income tax rebates, Johnson, Parker, and Souleles (2006) estimate an increase in nondurables spending of 20 to 40 percent of the received rebate within the receipt quarter. Parker et al. (2013) estimate the expenditure responses to the ESA of 2008 using the same methodology, and obtain MPX estimates of 50 to 90 percent for total spending during the receipt quarter, the majority of which was concentrated on motor vehicles. Many studies, including ours, build on these two contributions.

Subsequent work, however, identified econometric problems with the estimation strategy. Kaplan and Violante (2014) pointed out that Johnson, Parker, and Souleles's (2006) specification does not generally identify the MPX, in part because households treated earlier are used as control observations for those treated later. Shortly thereafter, an applied econometrics literature uncovered that additional problems can arise with common implementations of event study analyses when the timing of treatment is staggered (e.g., de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021; Borusyak, Jaravel, and Spiess, 2024). The research designs by Borusyak, Jaravel, and Spiess (2024) and Orchard, Ramey, and Wieland (2025) address these econometric problems and estimate substantially smaller expenditure responses than prior work. Our estimation also addresses these problems, but we estimate a very large MPX of around 0.8 despite focusing on a durables spending alone. We show that the difference relative to Orchard, Ramey, and Wieland (2025) is explained by the inclusion of late stimulus recipients in their sample, who we show are on a different expenditure trajectory. Our results do not conflict with Borusyak, Jaravel, and Spiess's (2024), who mostly study nondurables spending. Although some of our findings are reminiscent of Parker et al.'s (2013), there are also notable differences. For instance, we estimate much larger increases in expenditures on used motor vehicles than on new vehicles, while Parker et al. (2013) find the opposite.

There are a number of additional approaches to estimating the MPC. One strand of literature estimates the MPC using survey questions on whether households use tax rebates to mostly spend, mostly save, or mostly pay off debt (Shapiro and Slemrod, 2003a,b, 2009; Sahm, Shapiro, and Slemrod, 2010, 2012). Since only about 20 percent of households report wanting to mostly spend their extra income, the general conclusion from this line of work is that the MPC is likely small. We note that the mapping from these survey questions to the MPC becomes more involved when some households lever up and spend more than the stimulus payment – as is implied by our findings. Fagereng, Holm, and Natvik (2021) use variation from Norwegian lottery prizes and estimate an average MPC of about one-half. They emphasize that their MPC estimates differ with various household characteristics including liquidity. Hsieh (2003) and Kueng (2018) study payments from the Alaska Permanent Fund, which are predictable and large. While Hsieh (2003) finds no sizable expenditure response, Kueng (2018) estimates a nondurables MPC of 0.25. Using historical data, Hausman (2016) studies the veteran's bonus of 1936 and estimates MPXs betweeen 0.6 and 0.75, primarily on durable goods. Based on a randomized experiment in France, Boehm, Fize, and Jaravel

(2025) estimate a one-month MPX of 0.23. Although not all of these studies are directly comparable, e.g., because they estimate MPXs for different types of spending or because they estimate them over different time horizons, it is clear from these examples that estimates vary widely (see also Havranek and Sokolova, 2020).[2]

Several other findings from prior work are related to ours. Aaronson, Agarwal, and French (2012) document that following minimum wage hikes minimum wage workers substantially raise spending on durables, mostly on new and financed vehicles. While also not directly comparable to our findings, because minimum wage hikes generate more persistent increases in income and a relatively selected group of households works at the minimum wage, the large spending response on financed motor vehicles is similar to ours. Misra and Surico (2014), Jappelli and Pistaferri (2014), and Lewis, Melcangi, and Pilossoph (forthcoming), among others, find that expenditure responses are very heterogeneous across households, including for durable goods spending. Asking respondents about hypothetical income gains in a survey, Fuster, Kaplan, and Zafar (2020) find that the extensive margin of adjustment is important. Both substantial MPX heterogeneity and the importance of the extensive margin is consistent with our findings.

What sets our study apart from most prior work is the exclusive focus on durable goods. This focus allows us to use the Consumer Expenditure Survey with detailed data on durables spending at the monthly frequency. The monthly data, in turn, makes an event study analysis possible, for which we can transparently state the identification assumptions and assess their plausibility either on *a priori* grounds – randomized disbursement timing implies that the parallel trends assumption holds – or through the analysis of pre-trends. In addition, we can estimate an expenditure response whose shape is not restricted by the econometric model. In comparison to prior work our estimates of the durables MPX are high.

On the theory side Abel (1990) provides an early treatment of the durables MPX. He shows that in a conventional class of models without adjustment costs the durables MPX is related to the nondurables MPC through a multiplicative factor, which depends on the user cost and the expenditure share on durables. Back-of-the-envelope calculations with this formula suggest that the durables MPX is substantially larger than the nondurables MPC (2.41 times by his calculations).[3] The closest theoretical analog to our study is Beraja and Zorzi (2025),

[2] Other important work on the estimation of expenditure responses to income shocks includes the following. Wilcox (1989) studies increases in social security benefits and finds a substantial response of durables expenditures. Souleles (1999) uses variation from income tax refunds and estimates a large MPX of about 0.65, with most of the responses being accounted for by durables expenditures. Parker et al. (2022) study the economic impact payments of the CARES Act following the outbreak of the pandemic and find relatively small expenditure responses. Ganong et al. (2024) study unemployment benefit expansions and find large spending responses.

[3] Building on Abel's (1990) formula, Laibson, Maxted, and Moll (forthcoming) construct a mapping between the MPX for total expenditures and a *notional* MPC – a theoretical construct in common models of consumption that do not explicitly feature durable goods. They conclude that the MPX for total expenditures is about 3 times as large as

who develop an incomplete markets model with durables to study the MPX quantitatively. Many of the features and predictions of their model align with our findings, for instance, the financing of durables and the importance of the extensive margin. Although their model can generate an average MPX for durables of above 0.6 (see their Figure 3.1), they calibrate the average to 16 percent (for a stimulus payment of $500), citing prior empirical work. Since our findings imply a tight link between income and spending on durables, we believe there is a strong case for quantitative work with HANK models to follow Beraja and Zorzi (2025) and model durables and their financing explicitly – though with a different calibration.

Roadmap The remainder of this paper proceeds as follows. Section 2 reviews important features of the ESA of 2008 and describes the data. We discuss the empirical strategy and identification in Section 3. Section 4 presents the main results. We provide robustness checks and extensions in Section 5. Section 6 concludes.

2 Policy Background and Data

2.1 Policy Background

The ESA of 2008 was aimed at stabilizing the US economy during the looming downturn. It directed the Treasury to send tax rebates or "economic stimulus payments" of between $300-$600 to most singles and between $600-$1200 to most couples as well as $300 per qualifying child. Throughout the paper, we will refer to these transfers as stimulus payments or SPs. Note that since payments were phased out for higher incomes, households with sufficiently high incomes did not receive any SPs. Households with very low incomes were also not eligible.

Several features of this stimulus program are important for estimating and interpreting households' expenditure responses to the SPs. First, the payments were spread over several months and the individual receipt time for most households was determined by the last two digits of their Social Security numbers (SSNs). As discussed in Johnson, Parker, and Souleles (2006), these numbers are essentially randomly assigned and should therefore be unrelated to household characteristics and other shocks that affect expenditures during this episode. The random timing of stimulus receipts is what this and other studies seek to exploit.

Second, the ESA of 2008 was signed into law on February 13, 2008, approximately two and a half months before the disbursements of stimulus payments began. Moreover, information about the ESA was available prior to that date; for instance, it was passed by Congress in January of 2008. Most research designs, including ours, are unable to capture expenditure

the notional MPC. With a quarterly notional MPC from Kaplan and Violante (2014) of 15 percent, the MPX for total expenditures according to their model is therefore 0.45. See also Auclert (2019, Appendix A.5).

responses that may have occurred following the announcement of these stimulus payments *per-se*—that is, without the actual disbursements.[4]

Third, the IRS published the full schedule of stimulus payments before beginning with the disbursements, including how the individual receipt time depended on the last two digits of a household's SSN (see Internal Revenue Service, 2008b). News outlets circulated this information widely.[5] In addition, households received advance notices from the IRS. A fact sheet released by the IRS in February 2008 states that: "Most taxpayers will receive two notices from the IRS. The first general notice from the IRS will explain the stimulus payment program. The second notice will confirm the recipients' eligibility, the payment amount and the approximate time table for the payment" (Internal Revenue Service, 2008a). It is therefore likely that at least some households were aware of their personal receipt time prior to receiving the stimulus payment – in addition to the fact that they qualified for the SP and would receive it at some point. We emphasize this feature because it opens the door to the possibility that households changed expenditures in anticipation of receiving their SP, and not only upon receipt. We therefore choose an empirical strategy that can detect and accommodate such behavior.

Fourth, the schedule according to which the receipt time was determined by a tax filer's last two digits of its SSN only applied to tax returns received and processed by April 15. For households whose tax return was received and processed late, the timing of stimulus receipts was generally not randomized. Specifically, an IRS release from February 17, 2008 states: "To accommodate people whose tax returns are processed after April 15, the IRS will continue sending weekly payments. People who file tax returns after April 15 and receive a refund can expect to receive their economic stimulus payments in about two weeks after receiving their tax refunds, but not before the date they would have received their payment if the return had been processed by April 15. To ensure taxpayers receive their stimulus payment this year, they must file a tax return by Oct. 15" (Internal Revenue Service, 2008b). Since there is no guarantee that households submitting their tax return on time or late are comparable to one another and since late filers have some control over the timing of their stimulus payment, we restrict our analysis to stimulus payments that were received roughly on time (Parker et al., 2013).

Lastly, we note that two additional sources of variation are not random and need to be handled with care. One is that recipients of SPs and non-recipients differ systematically with respect to their incomes, and presumably their jobs and other characteristics that could

[4]Some studies have been able to evaluate responses to anticipated income gains. Studying the ESA of 2008, Broda and Parker (2014), for instance, collect information on when households learned about their SP with a survey but find no significant increase in expenditures around those times. Fuster, Kaplan, and Zafar (2020) present hypothetical scenarios to respondents of a survey and also find weak responses to anticipated income gains.

[5]See, e.g., `https://money.cnn.com/2008/04/25/news/economy/rebate_update/`, last accessed 07/03/2023.

affect their spending path. To avoid identification problems that could arise from using non-recipients as control observations for SP recipients we exclusively study samples that drop non-recipients. In addition, whether a household's SP was received by direct deposit or paper check was determined by the household's mode of tax filing (see Internal Revenue Service, 2008b) and was therefore also not random. As Broda and Parker (2014) show, households receiving direct deposits tend to be younger, have larger families, and higher incomes. Since there is again no guarantee that, say, not-yet treated paper check recipients are a suitable control group for direct deposit recipients, we use a research design that avoids such comparisons.

In summary, the good variation we seek to exploit comes from the random receipt timing of on-time recipients and within the paper check and direct deposit groups. In addition, anticipation effects cannot be ruled out *a priori* and must be checked with pre-trend tests.

2.2 Data

2.2.1 The Consumer Expenditure Survey

We use data from the Consumer Expenditure (CE) survey of the Bureau of Labor Statistics (BLS) for our empirical analysis. The CE program provides detailed information on all household expenditures. Data on major items such as durable goods or summary categories such as food are collected in the Interview Module of the CE survey, which is collected as a rotating panel of households. Households are interviewed every three months over four (formerly five) consecutive quarters before they are replaced with a new cohort of households. In each survey, a household (or consumer unit) reports its expenditures over the preceding three months (the reference period). Hence, the time dimension of the CE survey is relatively short. For most households expenditure data is available for no longer than 12 consecutive months.

Expenditures are reported at the level of Universal Classification Codes (UCCs). Examples of UCCs include "food at home", "women's footwear", and "new auto purchases". Total consumption expenditures are the sum of expenditures on around 500 actively used UCCs in a given survey year. For our purposes, a notable feature of the CE survey is that for a subset of UCCs it contains information on the month(s) in which the expenditures occurred. That is, although interviews are conducted quarterly, if a household reports purchasing, say, a new car, the survey asks for the month of the purchase. For other UCCs the CE only asks for the total quarterly expenditures and assigns an equal share of one third of the reported expenditures to each month.

Since the CE survey does not include a variable that indicates whether spending is recorded monthly or smoothed over the reference quarter, we infer this information from the 2008 questionnaire and the data. In essence, we classify expenditures on a particular UCC as recorded monthly if (i) one or multiple survey question explicitly provide information on the month of purchase *and* (ii) if expenditures are not excessively smooth in the data. This approach is deliberately conservative in the sense that we will classify several UCCs as only providing smoothed spending information even though (some) monthly spending information exists. The reason for this conservatism is that smoothed expenditure data can lead to biases in the estimation (likely downward). As will become clear below, our estimation procedure allows for some degree of recall error of the month of purchase. For instance, a household purchasing a car in June but incorrectly reporting the expenditures in May is not generally a problem for our estimation. Details on the classification of monthly- versus quarterly-recorded spending are available in Appendix A.4.

Since all expenditures in the CE survey are reported in current dollars and our analysis combines data over time, we deflate all expenditures with the price indexes available from the BLS. Specifically, we use a concordance between the UCC-level expenditure categories and the categories of the Consumer Price Index (CPI) and deflate with the most detailed price data available.[6] We further deflate the dollar value of SPs with the CPI for all items. The base period for all dollar amounts reported in tables and figures is December 2007.

To ensure that our findings are not driven by extreme outliers, we winsorize expenditures at the 98th percentile of the absolute value of expenditures, conditional on a non-zero response. Importantly, we winsorize at the UCC-level, that is, prior to the aggregation of expenditures. Doing so contains the role of extreme outliers *within* narrow expenditure categories and thereby reduces the risk that we incorrectly cap correctly-recorded large purchases – such as cars. We prefer this winsorization because we believe that it makes our estimates more reliable. It also reduces the standard errors. We show in robustness checks that this winsorization does not drive our the main conclusions.

2.2.2 Nondurables, Semi-durables, and Durables

In order to estimate how expenditures on durable goods respond to transitory changes in income, we need to define what constitutes these expenditures. Unfortunately, there is some disagreement with regard to the definition of durability and this ambiguity creates discrepancies across official statistics as well as the literature. For instance, curtains are considered durable in the Bureau of Economic Analysis' (BEA) definition of consumption, but nondurable in the BLS'. On the other hand, pillows are categorized as durable by the BLS but as non-

[6] The most detailed price series available are typically at the *Item-Strata* level and often aggregate a handful of UCCs into a summary category. When items are not sampled or price data is missing for other reasons we use price series from the next higher level of disaggregation—typically from the level of *Expenditure Classes*.

durable by the BEA.[7] Further, some prior studies such as Berger and Vavra (2015) focus on relatively large durables such as expenditures on homes and cars. Browning and Crossley (2009), on the other hand, emphasize the importance of semi-durables such as clothing for expenditure responses to temporary income changes. Clothing and footwear are not considered durable in any of the official statistics.

To accommodate the possibility that relatively small and short-lived durables play an important role for our research question, we err on the side of inclusion and use a broad definition of expenditures on durable goods. Our main criterion for durability is that the item or service purchased does not have a trivially short service life, that is, it is not consumed in a single use. The second criterion is that it is not a payment associated with a rental or a lease. Although, say, the leasing of a car is clearly a form of consuming a durable good, the expenditures on vehicles that are subsequently leased to households constitute capital expenditures of the leasing company. For households, renting or leasing constitutes the consumption of a service. To separate short-lived durables such as clothing and footwear from longer-lived durables, we classify them into a third category, which we label semi-durables. We further classify repair services of durable goods as durables expenditures but demonstrate below that this choice is not important for our results.

Following both the BEA and the BLS, we group durable goods into four subcategories. These are 1) motor vehicles and parts, 2) household furnishings and equipment, 3) entertainment, and 4) miscellaneous durables. We classify all UCCs that enter the calculation of total consumption expenditures as defined by the CE survey but are neither classified as durable nor semi-durable as nondurable. These include nondurable goods and most services. To more closely match the analogue of net expenditures in economic models, we also subtract vehicle sales from purchases of vehicles when computing expenditures on durable goods.[8] This contrasts with the definition of consumption expenditures in the CE survey, which does not net out vehicle sales. Aside from the inclusion of vehicle sales our definition of total consumption expenditures coincides with the definition of total consumption expenditures of the CE survey. We provide details on our classification of durable and semi-durable goods in Appendix A.3.

Table 1 presents summary statistics for households' expenditures. A notable feature of the data is that almost all spending on durables and all spending on semi-durables is recorded

[7]Specifically, curtains (UCC 280210) are included in the category *window coverings* (DWCTRC) (see Bureau of Labor Statistics, 2017). Further, the category DWCTRC is a subcategory of durable goods for the BEA (see Bureau of Economic Analysis, 2022). On the other hand, curtains are part of the category *window and floor coverings and other linens* (SEHH) (see Bureau of Labor Statistics, 2022a). This category is considered nondurable by the BLS (see Bureau of Labor Statistics, 2022b). Pillows (UCC 280220) are included in the category *household linens* (DLINRC) (see Bureau of Labor Statistics, 2017). This category is considered nondurable by the BEA (see Bureau of Economic Analysis, 2022). The BLS includes pillows in the category *furniture and bedding* (SEHJ) (see Bureau of Labor Statistics, 2022a) and this category is a subcategory of durable goods (see Bureau of Labor Statistics, 2022b).

[8]This implies that the expenditures we measure are counted as value added. Specifically, the difference between used vehicle purchases and sales (including trade-ins) captures dealer services.

Table 1: Summary Statistics

	# of UCCs	Quarterly expenditures				Share recorded monthly	
		Mean	p10	p50	p90	UCCs	Expenditures
		(in 2007 m12 dollars)				*(percent)*	
Panel A. Total consumption expenditures							
Total consumption expenditures	515	11,230	3,516	8,989	21,431	51.8	17.4
Nondurables	242	9,469	3,304	8,017	17,311	11.2	2.9
Semi-durables	70	221	0	97	590	100.0	100.0
Durables net of vehicle sales	203	1,540	0	328	3,131	83.7	95.1
Panel B. Durables breakdown							
Motor vehicles and parts	31	814	0	25	785	93.5	99.7
Vehicles	9	659	0	0	0	100.0	100.0
Purchases	6	739	0	0	0	100.0	100.0
Sales	3	-80	0	0	0	100.0	100.0
Parts and repair services	22	155	0	21	485	90.9	98.6
Household furnishings and equipment	113	465	0	52	1,186	82.3	90.8
Entertainment	42	182	0	0	412	85.7	91.1
Miscellaneous durables	17	80	0	0	223	70.6	82.6

Notes: This table reports summary statistics for quarterly expenditures of households in the CE survey of the BLS. The sample includes all households interviewed between 2007q1 and 2009q4. For our definition of monthly-recorded expenditures, see Appendix A.4.

at the monthly frequency. This contrasts with nondurables for which only 2.9 percent of expenditures are recorded monthly.[9] The main dependent variable of interest below is the 95.1 percent of durables spending that is recorded monthly.

2.2.3 Baseline Sample and Weighting

Our baseline sample consists of households, which all receive exactly one stimulus payment. Specifically, we begin with all observations in the CE survey and then drop households that do not receive any stimulus payments. This sample restriction avoids the concern that those households that do not receive any stimulus payments are unsuitable as a control group for those that do. Second, the econometric framework, which we use below for estimation, assumes that treatment is an absorbing state, so that there is at most a single treatment for each household. We therefore exclude households that receive multiple stimulus payments

[9]Similar to the definition of personal consumption expenditures (PCE) in the National Income and Product Accounts (NIPA), the CE definition of consumption expenditures does not include property purchases and capital improvements. Such expenditures are instead counted as changes in asset holdings and would be counted as residential investment in the NIPA. While data on property purchases, capital improvements, and related expenditures are collected as part of the CE survey and we have analyzed the expenditure responses on capital improvements to stimulus payments, the estimates were unsurprisingly noisy and largely uninformative. We discuss them in Section 5.2 and report them in Appendix Table B2.

from the sample.[10] Lastly, we drop households who report late stimulus payments. A stimulus payment is defined as late if a paper check was received after August or a direct deposit was received after June (following Parker et al., 2013). This definition builds in a one-month grace period relative to the official disbursement schedule to account for potential delays as well as recall error. As discussed above, the main rationale for excluding late recipients from the sample is that their receipt times are generally not randomized. They could also be more likely to be misreported. All of our estimates use the sampling weights of the CE survey to ensure that the estimates are representative of the relevant population.

3 Empirical Strategy

3.1 Baseline Estimator and Specification

Our empirical strategy adopts an event study approach to analyze the stimulus receipt and estimates dynamic treatment effects using the two-way fixed effect "imputation estimator" by Borusyak, Jaravel, and Spiess (2024) (henceforth BJS) as the baseline. This estimator has several advantages over available alternatives. Relative to implementations of the two-way fixed effect estimator by Ordinary Least Squares (OLS), it avoids multicollinearity problems that can arise in the absence of never-treated units. Further, the BJS imputation estimator avoids the spurious identification of long-run treatment effects and the "negative weighting problem", which can arise when treatment effects are heterogeneous (see Borusyak, Jaravel, and Spiess, 2024). In addition, relative to other estimators that allow for heterogeneity in treatment effects, the BJS imputation estimator is efficient (under spherical errors) – arguably an important property in the context of estimating expenditure responses to stimulus payments. We also perform robustness checks in which we use alternative estimators.

We next discuss the estimator and its intuition. Let i index households and t time in months. An observation is a pair $it \in \Omega$, where Ω is the sample of size N. Each observation is treated exactly once at the event date E_i. The horizon since treatment is denoted by $K_{it} = t - E_i$, so that $K_{it} = 0$ at the time of treatment and $K_{it} > 0$ after treatment. It is also convenient to denote by $D_{it} = 1\{K_{it} \geq 0\}$ current or past treatment of observation it. The set of treated units of size N_1 is then $\Omega_1 = \{it \in \Omega : D_{it} = 1\}$, and the set of untreated units of size N_0 is $\Omega_0 = \{it \in \Omega : D_{it} = 0\}$.

Expenditures are generically denoted by x_{it}. We often distinguish between different types of spending, e.g., on motor vehicles and parts or durables in the entertainment category, but only introduce separate notation when necessary for clarity. Potential expenditures at time t if unit

[10] Households that report multiple stimulus payments *within* the treatment month are included in the sample – unless the disbursement method differs across stimulus payments.

i is never treated is $x_{it}(0)$. For $it \in \Omega_1$, the causal effect of treatment is $\tau_{it} := E[x_{it} - x_{it}(0)]$. Note that the version of the Borusyak, Jaravel, and Spiess (2024) imputation estimator we adopt leaves treatment effect heterogeneity unrestricted.

The implementation of the BJS imputation estimator is as follows. Since all households in the sample receive exactly one stimulus payment at some point in the sample (the treatment), there are no never-treated observations. The counterfactual spending path is therefore constructed from not-yet-treated observations. This is done by first estimating household-specific fixed effects α_i and time fixed effects δ_t on the sample of not-yet-treated observations. These estimates are then used to "impute" the expected counterfactual untreated outcome of a treated observation as $\hat{\alpha}_i + \hat{\delta}_t$. An estimate for the individual treatment effect at time t is then obtained as $\hat{\tau}_{it} = x_{it} - \hat{\alpha}_i - \hat{\delta}_t$, that is, by subtracting from the actual outcome x_{it} with treatment the counterfactual expected outcome without treatment.[11] The estimand of interest is the average treatment effect on the treated at horizon h,

$$\tau_h = \sum_{it \in \Omega_1} w_{it}^h \tau_{it}, \tag{1}$$

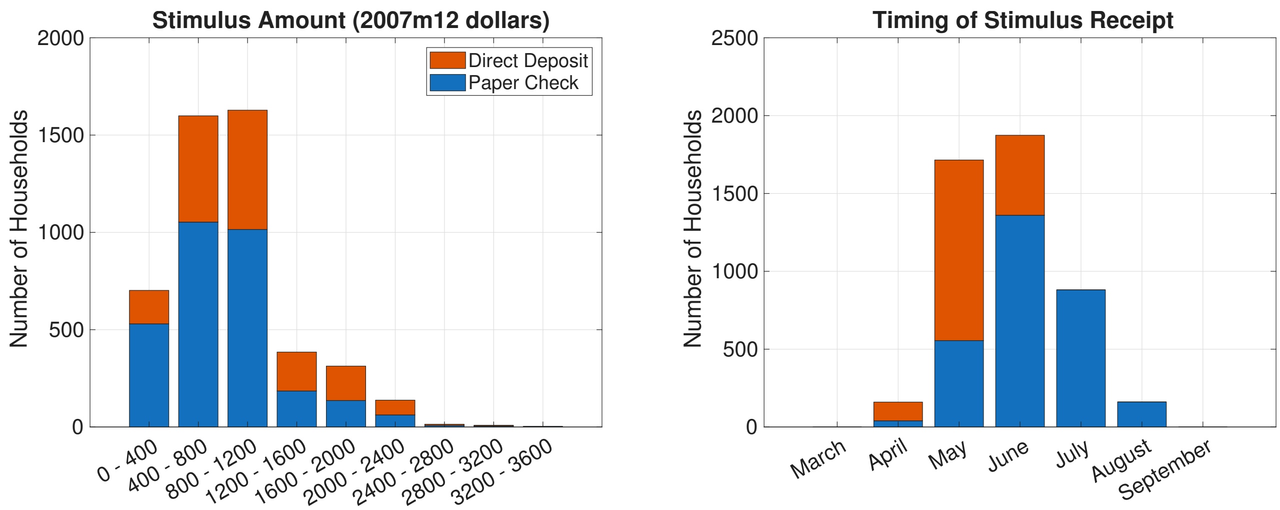
where the weights satisfy $w_{it}^h = 1\{K_{it} = h\} / \sum_{i't'} 1\{K_{i't'} = h\}$. The estimator simply replaces τ_h and τ_{it} with $\hat{\tau}_h$ and $\hat{\tau}_{it}$, respectively. Our key object of interest is the 3-month MPX, constructed as $\mu_{3m} = (\tau_0 + \tau_1 + \tau_2)/905.2$, where 905.2 is the average stimulus payment of treated observations in our baseline sample, measured in December 2007 dollars.[12] For the distribution of stimulus amounts see the left panel of Figure 1.

More precisely, and following Broda and Parker (2014), our baseline specification includes time fixed effects that are specific to the disbursement form by paper check or by direct deposit. That is, instead of imputing the counterfactual expected untreated outcome by $\hat{\alpha}_i + \hat{\delta}_t$, we impute it by $\hat{\alpha}_i + \hat{\delta}_{tm(i)}$, where $m(i)$ denotes the disbursement method. The estimator of the individual treatment effect is then $\hat{\tau}_{it} = x_{it} - \hat{\alpha}_i - \hat{\delta}_{tm(i)}$. Broda and Parker (2014) document that, among other things, both expenditures and household size are greater for households that receive the stimulus payment by direct deposit than for those that receive their stimulus payment as a paper check. Including time fixed effects that are specific to these groups implies that the expenditure responses are identified from within group variation and not the differential behavior of these two groups. The right panel of Figure 1 illustrates the timing of stimulus payments by disbursement method as reported in the CE Survey. Note that unlike

[11] Note that the two-way fixed effects imputation estimator drops observations for which the counterfactual spending path cannot be constructed, either because the household is not in the sample before it is treated (and hence α_i cannot be estimated) or because there are no untreated households in period t to estimate δ_t.

[12] We implement the estimator using the Stata command `did_imputation` written by Borusyak, Jaravel, and Spiess (2024). Standard errors are clustered at the household level throughout.

Figure 1: Reported Amount and Timing by Disbursement Method



Notes: The left figure illustrates the reported stimulus amount and the right figure the reported timing of stimulus receipts in 2008. Both figures are based on our baseline sample (see Section 2.2.3). All data are from the CE survey of the BLS.

Parker et al. (2013) as well as other work that estimates expenditure responses based on the ESA and using the CE survey, the frequency of our expenditure data is monthly.[13]

A useful property of the Borusyak, Jaravel, and Spiess (2024) imputation estimator is that it is linear in the dependent variable, implying that expenditure responses can be additively decomposed – like OLS. For instance, if two expenditure categories a and b satisfy $x_{it} = x_{it}^a + x_{it}^b$, then the estimated treatment effects will also additively decompose, $\hat{\tau}_h = \hat{\tau}_h^a + \hat{\tau}_h^b$ for all h, provided that the sample and the specification are the same. We will make extensive use of this property below. Since the economic decisions to purchase different types of goods and services are likely made jointly, these decompositions should only be viewed as accounting exercises that help understand in which categories expenditures were concentrated in.

3.2 Identification

Consistent estimation requires the assumptions of parallel trends and no anticipation effects.

Assumption 1 (Parallel Trends). *There exist* α_i *and* $\delta_{tm(i)}$ *such that for all* $it \in \Omega$: $E\left[x_{it}\left(0\right)\right] = \alpha_i + \delta_{tm(i)}$.

Roughly, within the paper check and direct deposit groups and in expectation, counterfactual spending of households with treatment would have evolved as in the control group in the absence of treatment. Given that the timing of stimulus receipts was quasi-random among

[13] As do Broda and Parker (2014) we have also calculated the MPX as a weighted average of the MPXs of direct deposit and paper check recipients. The difference from our baseline MPX is very minor.

on-time recipients within the paper check and direct deposit groups, we will assume throughout that the parallel trends assumption holds.[14]

A greater concern for identification is a possible violation of the no-anticipation-effects assumption.

Assumption 2 (No Anticipation Effects)**.** *For all* $it \in \Omega_0 : x_{it} = x_{it}(0)$.

In words, there is no effect of treatment on expenditures before the treatment occurs. Two types of concerns about this assumption are plausible in our context and need to be checked with pre-trend tests.

First, households could have adjusted expenditures before receiving the stimulus payment. As discussed earlier, the disbursement schedule was published before disbursements began and media outlets reported on this schedule. Further, the IRS sent advance notices to households informing them about their individual receipt time (see Section 2.1). Actual changes in expenditures prior to the receipt of the stimulus payment are therefore possible and – we believe – plausible in this empirical environment.

A second possibility is that the timing of the stimulus payment and/or expenditures could be misreported in the CE survey. If, for instance, households increased spending in response to the actual stimulus receipt but incorrectly reported a later receipt month, we would estimate a consumption response prior to the recorded stimulus receipt month – thus leading to a violation of the no-anticipation-effects assumption in the data. Note that interviews in the CE take place every quarter and households need to recall the month in which they received their stimulus payments as well as the month in which expenditures took place. As a result, this violation of the no-anticipation effects assumption could plausibly arise from recall error with regard to the timing of the stimulus receipt, the timing of expenditures, or both.[15]

The event study framework we adopt here is well suited to address concerns about anticipation effects. We will begin our empirical analysis below by analyzing pre-trends – and

[14] We note that the parallel trends assumption is not guaranteed to hold. The reason is that we cannot perfectly identify on-time recipients in the data. We can only restrict the sample based on the reported receipt month. For instance, paper check recipients in June could include households whose on-time receipt date would have been May, but who filed their taxes late and therefore received the stimulus payment in June. In this case the receipt month is not randomized. Relatedly, Thakral and Tô (2022) report balance tests revealing *ex-ante* level differences in several characteristics between cohorts of paper check recipients, but no such differences in the direct deposit sample. It is not clear, however, how to interpret this evidence and its implications—for several reasons. First, they use data from the Nielson Survey, not the CE survey. Second, Kerwin, Rostom, and Sterck (2025) show that common balance tests, which Thakral and Tô (2022) also use, tend to over-reject randomization. Third, the differences across groups appear economically small (see their Appendix Table 5). Lastly, even if the receipt time was not randomized, such differences in levels do not necessarily invalidate the parallel trends assumption – as pointed out by Borusyak, Jaravel, and Spiess (2024). We therefore continue to assume that the parallel trends assumption holds.

[15] Orchard, Ramey, and Wieland (2025) show that households are more likely to report stimulus receipt at the end of the 3-months reference period for their interview. This is inconsistent with randomized receipt times and the correct reporting of receipt times, which should lead to equal probabilities of receipt in all months. Their preferred interpretation is that there is some recall error for receipt times.

find that there is indeed a large positive response one month before households receive their stimulus payment. Since this response plausibly arises because of an early expenditure response or because of misreported timing and is in both cases part of the causal effect we seek to identify, our preferred specifications below redefine the event date E_i such that treatment occurs one month before the stimulus receipt.[16] We provide more discussion of this point in the next section.

3.3 The Semi-elasticity of Expenditures to Transitory Income Shocks

While the MPX answers the question of how much of their stimulus payments households spend, it is not a good measure to compare the responsiveness of spending across expenditure categories. The reason is that one would generally expect that all else equal, categories that receive a larger share of spending unconditionally will also receive a larger share of spending in response to the stimulus payment. To obtain a measure that corrects for this issue, we construct the semi-elasticity of expenditures with respect to the stimulus payment. Specifically, we take the estimate of the category-specific 3-month MPX $\hat{\mu}_{3m}$ and divide it by the category's average quarterly expenditures $\bar{x}_{3m}$ – computed on the same sample of households and before receiving the stimulus payment. This semi-elasticity $\hat{\mu}_{3m}/\bar{x}_{3m}$ answers by how much expenditures change *relative to average quarterly spending* in response to the receipt of a stimulus payment. We report the semi-elasticity in percent per \$1000 of stimulus. A semi-elasticity of, say, 10, would imply that households increased spending by 10 percent of average quarterly spending in response to receiving a \$1000 stimulus payment.

4 Empirical Results

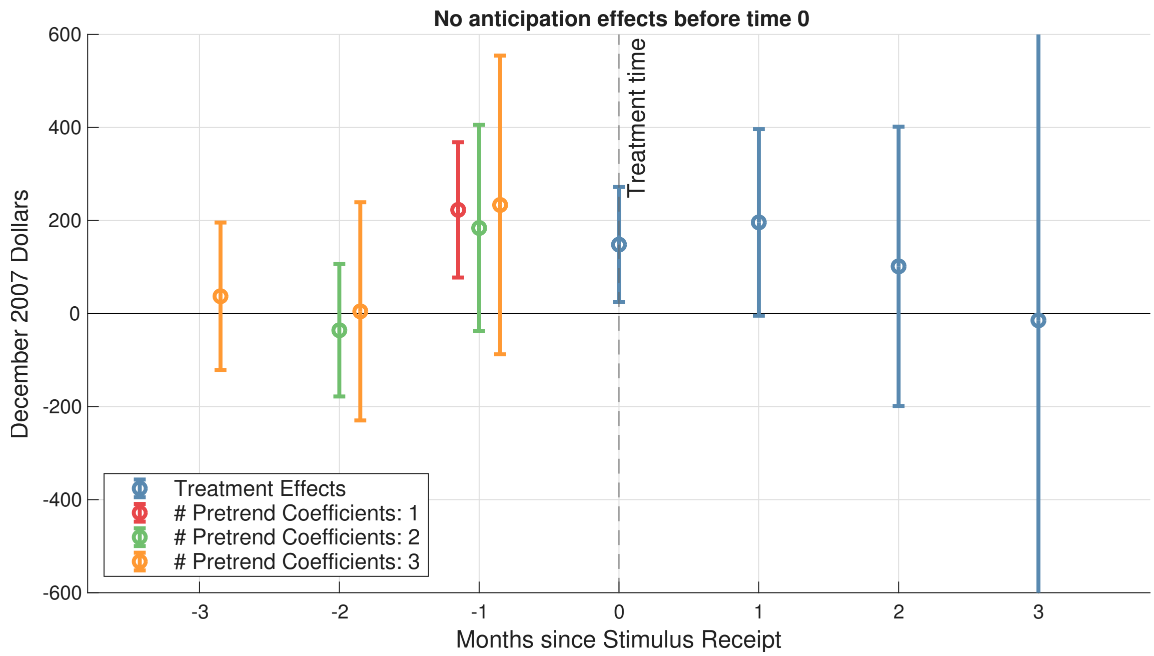
4.1 Pre-trend Tests and Anticipation Effects

To assess the plausibility of our identification assumptions, we next turn to pre-trend tests. Unlike OLS estimation of dynamic treatment effects, the testing for pre-trends in the BJS framework is separate from the estimation of treatment effects.[17] To assess pre-trends, we

[16] While distinguishing these two alternative reasons for anticipation effects is not directly important for our estimates, it is important for comparing our estimates to the literature. If the violation of no anticipation effects arises from anticipation effects, estimates based on specifications in changes, such as Parker et al. (2013) or Orchard, Ramey, and Wieland (2025), would generally be biased downwards. Such a bias is unlikely to arise if the violation is the result of recall error in the timing of expenditures. We have no way of distinguishing between these two possibilities.

[17] This separation has advantages over alternative approaches. First, it often increases statistical power. Second, it avoids size distortions if one conditions on the results of the pre-trend tests. See Borusyak, Jaravel, and Spiess (2024) and Roth (2022) for details.

Figure 2: Pre-trends and Treatment Effects



Notes: The figure shows the pre-trend coefficients $\hat{\gamma}_h$, obtained from specification (2), as well as the horizon h treatment effects $\hat{\tau}_h$, obtained using the BJS imputation estimator as discussed in Section 3.1. The event date E_i is the receipt month and the dependent variable is monthly recorded spending on durable goods. The number of observations is 26,853. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

estimate the specification

$$x_{it} = \alpha_i + \delta_{tm(i)} + \gamma_1 1\{K_{it} = -1\} + \gamma_2 1\{K_{it} = -2\} + \gamma_3 1\{K_{it} = -3\} + \varepsilon_{it} \tag{2}$$

on the sample of untreated observations $it \in \Omega_0$. The coefficients γ_1, γ_2, and γ_3 capture the effects of treatment on spending one, two, and three months in advance, and the null hypothesis is that they are zero, either individually, or jointly. Rejection of these hypotheses indicates a violation of the parallel trends or the no-anticipation effects assumption.

Figure 2 shows the estimated pre-trend coefficients $\hat{\gamma}_h$ along with the treatment effects $\hat{\tau}_h$, which we estimate as discussed above. The dependent variable is expenditures on durables that are recorded at the monthly frequency. Time on the horizontal axis is relative to the stimulus receipt so that 0 denotes the receipt month. The event date E_i is the month in which the stimulus payment is received and the no-anticipation-effects assumption implies that there should be no response before time 0. As the figure shows, the pre-trend coefficients are small three and two months before the stimulus receipt, but become large in the month before the receipt. Spending then remains elevated before falling in months two and three after the receipt.

Even though the estimate $\hat{\gamma}_1$ is economically large, it is not statistically significant. This lack of statistical power arises from the fact that the time series dimension of our panel is short so that neither α_i nor the γ_h can be estimated with high precision. We therefore report two alternative specifications, in which only two and one pre-trend coefficient is included in specification (2). In all cases $\hat{\gamma}_1$ is around 200 implying a response of around \$200 in the month before the stimulus receipt. Further, with only one pre-trend coefficient included in specification (2) the estimate $\hat{\gamma}_1$ is statistically significant ($p = 0.012$).

Since the parallel trends assumption plausibly holds in our empirical environment, we interpret this pre-trend as a violation of the no-anticipation-effects assumption. Households may have increased expenditures in anticipation of receiving the stimulus payment in the subsequent month. Alternatively, this violation could be the result of misreporting in the timing of expenditures or the stimulus receipt. While we cannot distinguish between these two explanations, it is clear that the early expenditure response should be included in the estimates of the MPX, because in both cases it is part of the causal response to the stimulus payment. To do so, we re-define the event date E_i such that treatment occurs one month before the stimulus payment is received. We use this definition of the event date as our baseline. The no-anticipation effects assumption is then imposed on all prior dates, that is, two months before the receipt and earlier. This re-definition of the event date makes our identification assumptions strictly weaker.[18]

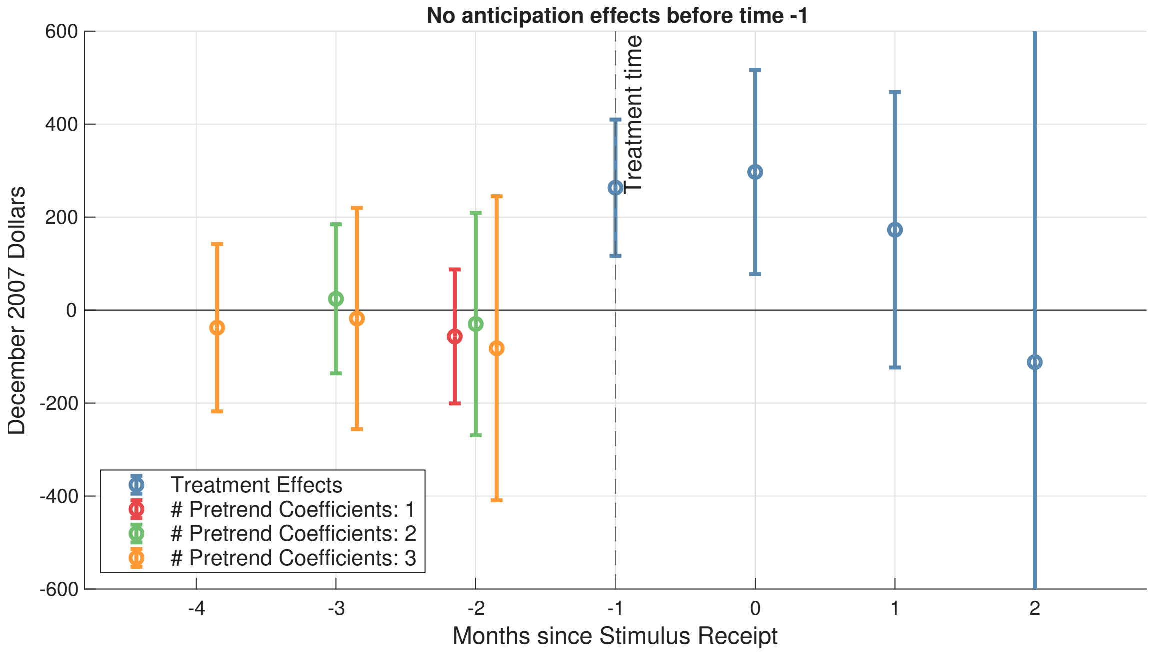
4.2 Baseline Estimates

Figure 3 shows our baseline estimates. There is no evidence for pre-trends after this redefinition of E_i – all pre-trend coefficients are now very close to zero ($p = 0.90$ for the joint test of $\gamma_1 = \gamma_2 = \gamma_3 = 0$). Expenditures then rise to around \$260 one period before receiving the stimulus payment, which now corresponds to the first period of treatment. They continue to rise to almost \$300 before falling to around \$170 a month later. Cumulatively, expenditures on durable goods increase by \$733 over this three months period with a standard error of 338. This corresponds to an MPX of 0.81 (= 733.2/905.2). Since the estimated effect becomes extremely imprecise two months after the receipt, our subsequent analysis will exclusively focus on the three months around the stimulus receipt.

Table 2 shows the estimated three-months spending response and the associated MPX. As discussed earlier, we will use the semi-elasticity of expenditures to the stimulus payment for comparing the responsiveness across categories below. This semi-elasticity is 50.8, which

[18] To see this, note that the parallel trends assumption (Assumption 1) is imposed on all observations and therefore independent of the event date. Further, the no-anticipation effects assumption (Assumption 2) is imposed on untreated observations $it \in \Omega_0$, and this set Ω_0 is strictly smaller after the re-definition.

Figure 3: Baseline Estimates: Pre-trends and Treatment Effects



Notes: The figure shows the pre-trend coefficients $\hat{\gamma}_h$, obtained from specification (2), as well as the horizon h treatment effects $\hat{\tau}_h$, obtained using the BJS imputation estimator as discussed in Section 3.1. The event date E_i is the month prior to the stimulus receipt and the dependent variable is monthly recorded spending on durable goods. The number of observations is 22716. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

implies that expenditures on durables increased by 50.8 percent over average quarterly expenditures for each \$1000 of stimulus, a substantial response. These expenditure responses are estimated for spending we observe at the monthly frequency. Recall from Table 1 that this accounts for 95.1 percent of all spending on durables. Scaling the estimates in column (2) by one over the share of monthly- recorded spending makes therefore little difference, see column (4). Since the scaling is unimportant quantitatively in our environment, we report unscaled estimates for the remainder of the paper.

4.3 Expenditure Decompositions

We next decompose the expenditure responses on durables. We present decompositions along three different dimensions: by expenditure category, into intensive and extensive margin, and by realized expenditure size. Individually, each dimension highlights different properties of the spending responses and jointly they inform models of durable goods consumption. As noted earlier, the BJS imputation estimator is linear in the dependent variable and hence the total durables response exactly decomposes into additive components.

Table 2: Baseline Estimates

	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$ (1)	3-month MPX (2)	Semi-elasticity (3)	Scaled MPX (4)
Monthly-recorded expenditures on durables	733.2	0.81	50.8	0.85
	(338.3)	(0.37)	(23.4)	(0.39)

Notes: The number of observations is 22716. The first column reports the estimates $\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$, which we obtain using the BJS imputation estimator as discussed in Section 3.1. Column (2) reports the MPX, constructed as $(\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2)/905.2$. The semi-elasticity shown in column (3) is expressed in percent per \$1000 of stimulus, see Section 3.3 for details. The scaled MPX, reported in the last column, scales the MPX in column (2) by one over the share of spending that is recorded monthly, see Table 1. Standard errors are clustered at the household level.

4.3.1 Expenditure Categories

We begin with a decomposition by expenditure category. Table 3 shows the estimates for both broad summary categories as well as more detailed subcategories. Among the broad summary categories, spending on motor vehicles and parts accounted for the largest share of expenditures on durables (80.2 percent), followed by entertainment (9.4 percent) and household furnishings and equipment (6.9 percent). Expenditures on miscellaneous durables increased by \$25.2, which is significant at the 10 percent level, but accounts for a small share of the overall response (3.4 percent).

The semi-elasticity of spending on motor vehicles and parts is 73.6, implying that motor vehicle spending increased by 73.6 percent relative to average quarterly spending in response to \$1000 worth of stimulus. This semi-elasticity is greater than those for household furnishings and equipment, entertainment, and miscellaneous durables, suggesting that spending on motor vehicles and parts is more responsive to stimulus payments than spending on other durables. The differences in semi-elasticities, however, are not statistically significant at conventional levels.

The estimates for the more detailed breakdown reveal that the response of vehicle purchases was extremely large. Vehicle purchases alone (net of trade-ins but gross of sales) had an estimated expenditure response of 742.0, corresponding to an MPX of 0.82. The more aggregated response of motor vehicles and parts is smaller than that of vehicle purchases due to a substantial increase in sales. The semi-elasticity of motor vehicle sales is 300, implying that sales increased by 300 percent relative to average quarterly sales for each \$1000 of stimulus. We caveat, however, that both the MPX for sales and the associated semi-elasticity are imprecisely estimated. All other responses were either small (household operations, medical supplies, personal care products and services) or statistically insignificant.

This decomposition suggests that households spent a sizable share of their stimulus payments on vehicles and only a small share on non-vehicle durable goods, if any. One possible

Table 3: Decomposition by Expenditure Category: Durables

	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	3-month MPX	Share (%)	Semi-elasticity
Motor vehicles and parts	588.3	0.65	80.2	73.6
	(259.0)	(0.29)		(32.4)
Vehicle purchases (net outlay)	742.0	0.82	101.2	104.5
	(307.7)	(0.34)		(43.3)
Other vehicle expenses	37.5	0.04	5.1	24.5
	(29.4)	(0.03)		(19.2)
Change in vehicle sales—motor vehicles	-191.3	-0.21	-26.1	300.1
	(150.7)	(0.17)		(236.5)
Household furnishings and equipment	50.5	0.06	6.9	12.7
	(148.2)	(0.16)		(37.1)
Floor coverings	3.4	0.00	0.5	37.9
	(6.8)	(0.01)		(75.6)
Furniture	30.0	0.03	4.1	32.4
	(35.1)	(0.04)		(37.9)
Household textiles	-3.6	0.00	-0.5	-17.1
	(9.6)	(0.01)		(45.3)
Major appliances	-91.0	-0.10	-12.4	-187.3
	(117.2)	(0.13)		(241.1)
Miscellaneous household equipment	0.5	0.00	0.1	0.6
	(22.0)	(0.02)		(25.0)
Small appliances, miscellaneous housewares	7.9	0.01	1.1	62.7
	(7.6)	(0.01)		(59.8)
Household operations	11.9	0.01	1.6	185.4
	(5.4)	(0.01)		(84.0)
Shelter	91.4	0.10	12.5	75.7
	(61.7)	(0.07)		(51.1)
Entertainment	69.2	0.08	9.4	37.5
	(100.2)	(0.11)		(54.2)
Change in vehicle sales—entertainment	12.9	0.01	1.8	-156.6
	(29.5)	(0.03)		(358.6)
Pets, toys, hobbies, and playground equipment	-55.1	-0.06	-7.5	-160.2
	(53.9)	(0.06)		(156.8)
Audio and visual equipment and services	68.4	0.08	9.3	95.6
	(59.4)	(0.07)		(83.0)
Other entertainment supplies, equipment, and services	43.0	0.05	5.9	49.4
	(60.6)	(0.07)		(69.6)
Miscellaneous durables	25.2	0.03	3.4	41.8
	(14.7)	(0.02)		(24.4)
Education	2.4	0.00	0.3	18.2
	(4.6)	(0.01)		(35.9)
Jewely and watches	-1.7	0.00	-0.2	-6.6
	(8.4)	(0.01)		(31.5)
Medical supplies	22.6	0.03	3.1	124.1
	(12.5)	(0.01)		(68.4)
Personal care products and services	2.0	0.00	0.3	72.0
	(0.9)	(0.00)		(33.6)

Notes: The number of observations is 22716. The first column reports the estimates $\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$, which we obtain using the BJS imputation estimator as discussed in Section 3.1. The second column reports the MPX, constructed as $(\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2)/905.2$. The third column reports the share in the total durables response, 733.2 (see Table 2), expressed in percent. The semi-elasticity shown in column four is expressed in percent per \$1000 of stimulus, see Section 3.3 for details. Standard errors are clustered at the household level.

conclusion from these results could be that the responsiveness of durables expenditures to transitory income shocks is generally not large – and that vehicles are an exception. We argue below that this is only partially true. Specifically, we show that expenditures on some non-vehicle durables also increased meaningfully, but that the disaggregation into expenditure categories as shown in Table 3 masks the overall importance of these responses.

4.3.2 Intensive and Extensive Margin

We next explore whether the expenditure responses we estimated above are driven by the extensive margin (i.e., whether to buy) or the intensive margin (expenditures conditional on a purchase). Prior work on plant-level investment and household-level car and home purchases has shown that annual expenditures on durable goods are lumpy (e.g., Doms and Dunne, 1998; Gourio and Kashyap, 2007; Berger and Vavra, 2015). That is, purchases are infrequent and spending conditional on a purchase is large. This literature further emphasizes that aggregate variation in investment is predominantly driven by the frequency of large purchases, and not by time variation of the magnitude of expenditures conditional on a purchase or by time variation in small adjustments. Based on these prior findings one may expect that the extensive margin will be dominant in our environment as well – although it is important to note that our findings below are not directly comparable to those in the literature. Key differences include that our focus is on the household side, that our definition of expenditures does not include home purchases, and that our data is monthly (as opposed to annual).

To assess the roles of both the intensive and the extensive margin, we need to define what constitutes an adjustment. Let $\mathcal{A}_{it}$ denote an indicator that is equal to one if an adjustment occurs. We will set $\mathcal{A}_{it} = 1\{|x_{i,t}| > \kappa\}$ for different values of κ. For instance, if $\kappa = 0$, any nonzero expenditures constitute an adjustment along the extensive margin. For $\kappa = 500$, an adjustment occurs if expenditures exceed \$500 in absolute value, and so forth.[19]

With a given definition of adjustment in hand, we next decompose total expenditures on durable goods into three parts:

$$x_{it} = x_{it}^{ex} + x_{it}^{in} + x_{it}^{0},$$

[19]A common definition of an extensive margin adjustment in the literature is that investment exceeds 20 percent of the capital stock. Clearly, this definition is not appropriate in our context – nor can it be implemented with our dataset since the stock of durables is not measured. Also note that the decomposition of aggregate investment into intensive and extensive margin by Gourio and Kashyap (2007) cannot be implemented on individual-level data. Our approach is similar to theirs, but can be implemented on individual-level data. For the purpose of our application an added advantage of our decomposition is that it is exactly additive.

where

$$
\begin{aligned}
x_{it}^{ex} &= \mathcal{A}_{it} \cdot \bar{x}_{\mathcal{A}}, \\
x_{it}^{in} &= \mathcal{A}_{it} \cdot (x_{it} - \bar{x}_{\mathcal{A}})\,, \\
x_{it}^{0} &= (1 - \mathcal{A}_{it}) \cdot x_{it},
\end{aligned}
$$

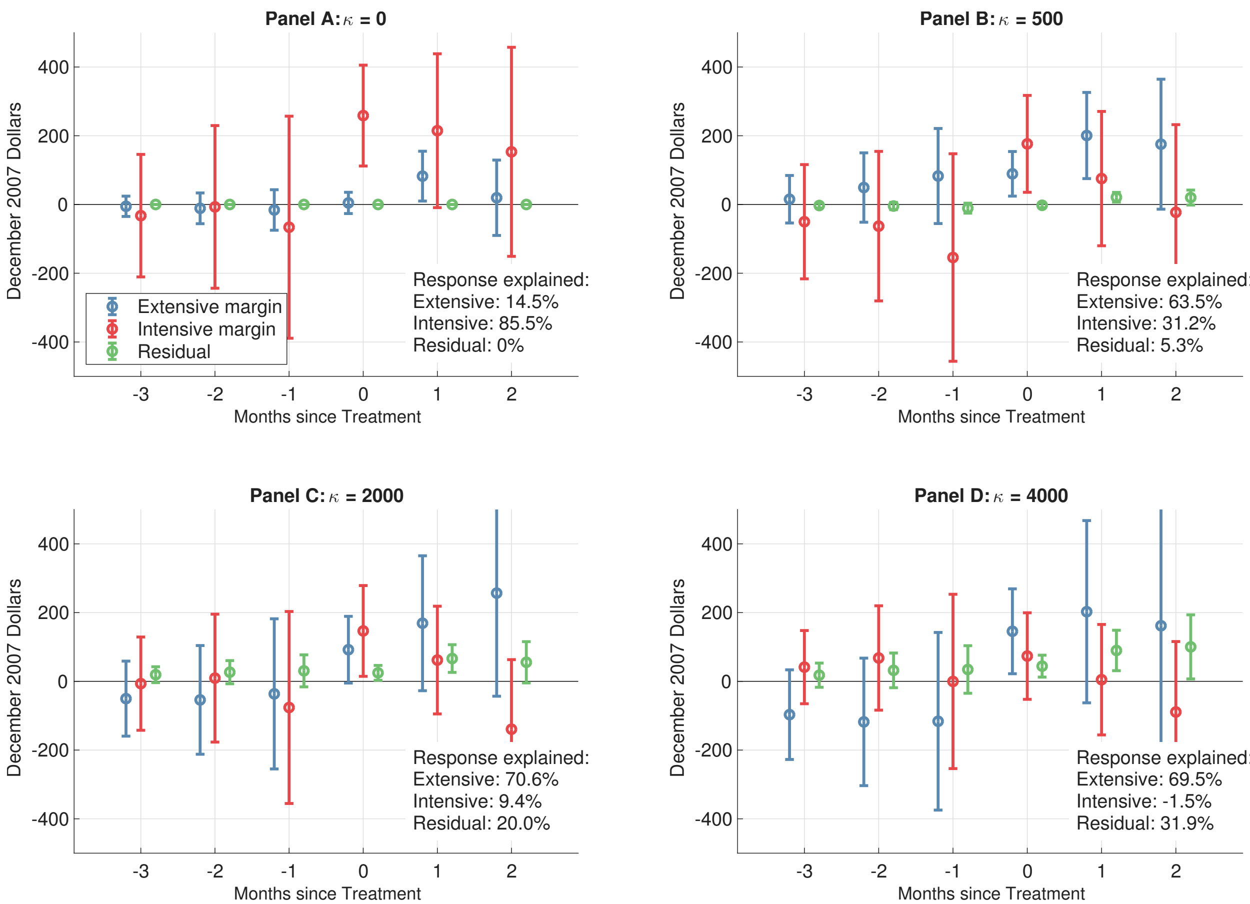
and $\bar{x}_{\mathcal{A}}$ denotes average expenditures conditional on adjustment. We estimate $\bar{x}_{\mathcal{A}}$ on pre-treatment data so that $\bar{x}_{\mathcal{A}} = \frac{1}{|\Omega_0^{\mathcal{A}}|} \sum_{it \in \Omega_0^{\mathcal{A}}} x_{it}$, where $\Omega_0^{\mathcal{A}} = \{it \in \Omega : D_{it} = 0 \text{ and } \mathcal{A}_{it} = 1\}$.

The term x_{it}^{ex} denotes the extensive margin. Since $\bar{x}_{\mathcal{A}}$ is a constant, the extensive margin varies over time only because of changes in adjustment $\mathcal{A}_{it}$. To understand the intensive margin x_{it}^{in}, note that it is different from zero only if an adjustment occurs ($\mathcal{A}_{it} = 1$) *and* expenditures differ from average expenditures conditional on adjustment ($x_{it} \neq \bar{x}^{\mathcal{A}}$). We would estimate a positive effect on the intensive margin if, for instance, the stimulus receipt did not affect adjustment, but if conditional on adjustment expenditures would increase. The term x_{it}^{0} is a residual, which captures small adjustments that are neither included in the intensive nor the extensive margin.

Figure 4 reports the results. The main finding is that the decomposition into the intensive margin, the extensive margin, and the residual category x_{it}^{0} critically depends on the adjustment threshold κ. For $\kappa = 0$, that is, if any nonzero expenditures constitute an adjustment, over 85 percent of the expenditure response is accounted for by the intensive margin. For higher values of κ the extensive margin becomes more important, but so does the residual category x_{it}^{0}. While the extensive margin is important for larger thresholds, for no κ in the figure does it account for the entire response. Jointly, the intensive margin and the residual category of small adjustments explain at least around 30 percent of the response.

The sensitivity of this decomposition to the parameter κ can be rationalized by noting that the frequency of *large* adjustments (i.e., motor vehicles) increased in response to the stimulus payments while the frequency of smaller adjustments did not (or to a lesser extent). Such an increase in large adjustments would drive up the intensive margin for small κ but the extensive margin for large κ. In addition, however, and unlike the decomposition by expenditure category, this cut of the data highlights the importance of smaller adjustments. This insight leads us to our third decomposition by realized expenditure size.

Figure 4: Decomposition by Intensive and Extensive Margin



Notes: The dependent variables are the extensive margin, the intensive margin, and the residual category of monthly recorded spending on durable goods, see text for details. Pre-trends are estimated as discussed in Section 4.1 and treatment effects are estimated as discussed in Section 3.1. The number of observations is 22716. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

4.3.3 Decomposition by Realized Expenditure Size

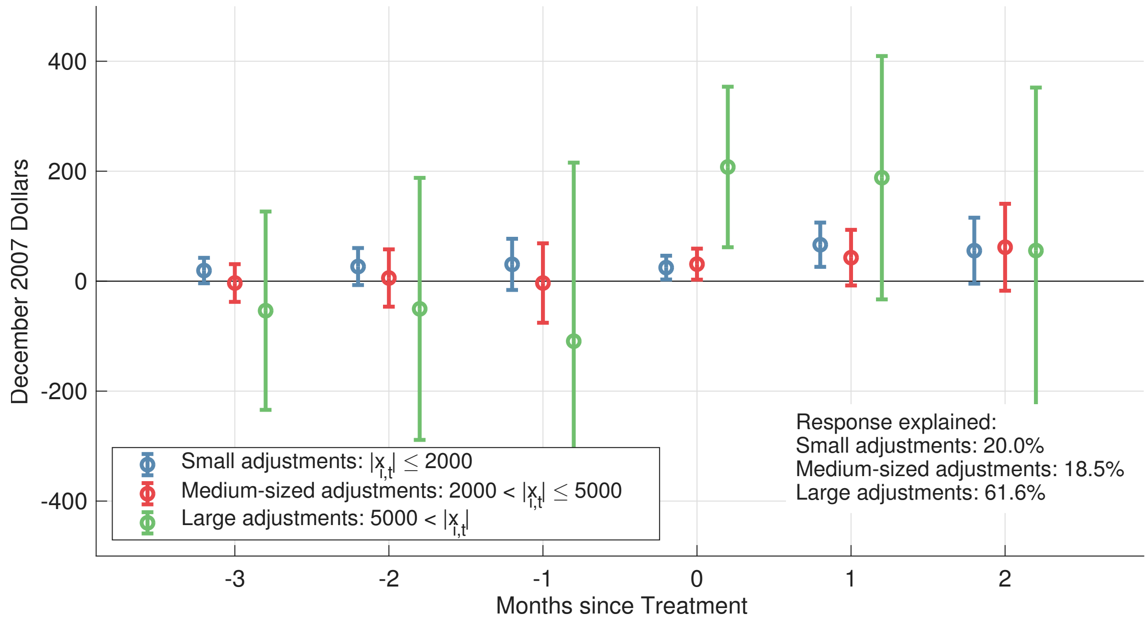
Specifically, we decompose expenditures on durables into small, medium-sized, and large realized adjustments:

$$x_{it} = \underbrace{x_{it}1\{|x_{it}| \leq 2000\}}_{\text{small adjustments}} + \underbrace{x_{it}1\{2000 < |x_{it}| \leq 5000\}}_{\text{medium-sized adjustments}} + \underbrace{x_{it}1\{5000 < |x_{it}|\}}_{\text{large adjustments}}. \tag{3}$$

While the thresholds are, of course, arbitrary they are chosen such that the stimulus payment covers a substantial share of expenditures in the small-adjustments category and that vehicle purchases will predominantly appear in the large-adjustments category (see Table 4 below).

Figure 5 shows the responses. There was a robust and significant response in small adjustments accounting for 20.0 percent of the total response of durables goods. (This response coincides with the residual category in Panel C of Figure 4.) Medium-sized adjustments be-

Figure 5: Decomposition by Realized Expenditure Size



Notes: The dependent variables are the small, medium-sized, and large adjustments as defined in equation (3). Pre-trends are estimated as discussed in Section 4.1 and treatment effects are estimated as discussed in Section 3.1. The number of observations is 22716. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

tween \$2000 and \$5000 did increase, but not significantly so following the stimulus payments. Lastly, large adjustments of over \$5000 increased following the stimulus payment. Of course, these large adjustments are driven by motor vehicle purchases (see Table 3) and are counted as extensive margin adjustments in Panel D of Figure 4.

In summary, these decompositions deliver the following three insights. First, motor vehicles played an outsized role. Second, the extensive margin of large adjustments increased. Taken together, these two points imply that the extensive margin of motor vehicle adjustment was the main driving force of the total durables response, accounting for around 60 to 80 percent. The intensive margin of large adjustments did not play an important role. Third, smaller adjustments also increased significantly while the responses of medium-sized adjustments are imprecisely estimated. The relatively small adjustments account for around 20 percent of the total durables response.

We emphasize the significant response of smaller durables expenditures because without this evidence one could conclude that the MPX for durables is generally small and that motor vehicles are an exception. We instead favor the explanation that expenditures on durables more generally increased if the item was sufficiently small so that the stimulus payment covered a large chunk of the purchase. This interpretation is also supported by the significant response of semi-durables discussed in Section 5.2 below. We will further show below that many vehicle purchases do not require large cash outlays – so that the increase in vehicle purchases does not contradict the hypothesis that the magnitude of cash outlays matters for the response.

4.4 Lumpiness, Resale Markets, and Financing of Motor Vehicles

We next explore whether vehicles are "special". If vehicles differ from other durable goods, it is possible that the MPX on vehicles is large not because they are durable (or at least not only), but because of other properties that they possess. We document that vehicles indeed differ from other durables along several potentially relevant dimensions: First, expenditures on vehicles are typically larger and lumpier than those on other durables. Second, vehicles have a more liquid resale market. Third, households purchasing vehicles often have the option to take out a vehicle loan. After documenting these facts we explore whether the evidence supports the view that these properties affect the expenditure responses to stimulus payments. It is important to note, however, that we do not claim causality in this subsection. The question of whether vehicle spending has a large MPX because vehicles are durable or because of other features is difficult to answer empirically.

4.4.1 Some Facts

Lumpiness We document in Appendix Table B1 that unconditional average quarterly spending on cars and trucks is substantially greater than on any other durable good. In that sense cars and trucks are the most important durables. Cars and trucks are also lumpier: Conditional on a purchase, average quarterly expenditures are often an order or magnitude larger than those on other durables. In addition, the frequency of purchases tends to be lower than those of many other durables. As shown by Beraja and Zorzi (2025), models with lumpy durables adjustment can generate large expenditure responses to transfers.

Resale Markets A second property that sets vehicles apart from other durables is that they have (more) liquid resale markets. Unconditional average expenditures on used cars and trucks are comparable to those on new cars and trucks (see Appendix Table B1), and hence these markets are important. Conditional on a purchase, however, gross expenditures on used motor vehicles are about two fifth of expenditures on new motor vehicles (see Panel A of Table 4). Further, it is common for households to trade in or sell a vehicle when purchasing a (relatively) new one. As a result of these trade-ins and sales, average net expenditures on vehicles conditional on a purchase are about $1,800 lower than gross expenditures, as shown in column (1) Panel A of Table 4. For new motor vehicles trade-ins and sales amount to almost $3,500 (Panel A column (2)). In the case of used vehicles they average around $1,100 (Panel A column (3)). Since lower cash outlays may make it easier for households to purchase

Table 4: Average Cash Outlays for Motor Vehicles

	New and Used (1)	New (2)	Used (3)
	(in 2007 m12 dollars)		
Panel A. All motor vehicles			
Gross expenditures	14,909	25,598	10,174
Trade-ins	1,428	2,972	755
Expenditures net of trade-ins	13,480	22,626	9,419
Sales	388	454	355
Cash outlays net of trade-ins and sales	13,092	22,172	9,064
Panel B. Financed motor vehicles			
Gross expenditures	20,157	26,536	15,116
Trade-ins	2,463	3,598	1,577
Expenditures net of trade-ins	17,694	22,938	13,539
Loans	15,960	20,436	12,406
Cash outlays net of trade-ins and loans (down payment)	1,734	2,502	1,134
Sales	428	494	373
Cash outlays net of trade-ins, loans, and sales	1,306	2,008	760
Panel C. Non-financed motor vehicles			
Gross expenditures	9,273	22,773	6,677
Trade-ins	344	1,201	181
Expenditures net of trade-ins	8,928	21,572	6,496
Sales	341	339	339
Cash outlays net of trade-ins and sales	8,587	21,233	6,157

Notes: This table reports quarterly averages conditional on: purchasing a motor vehicle in Panel A, purchasing a financed motor vehicle in Panel B, and purchasing a non-financed motor vehicle in Panel C. Motor vehicles include autos, trucks, and motorcycles. Sales are included in the quarterly averages if they occur within the same reference period as the vehicle purchase. The data comes from the CE survey of the BLS and the sample includes all households interviewed between 2007q1 and 2009q4.

such items, it is plausible that this feature raises the MPX in comparison to otherwise similar durables.

Financing A third property that distinguishes vehicles from other durables is that a substantial share of vehicle expenditures is financed. This could matter for the expenditure responses for at least two reasons. For one, financing substantially reduces the cash outlay required to purchase the vehicle, as shown in Panel B column (1) of Table 4.[20] It turns out that between

[20] Whether a vehicle was financed or not is measured well in the CE survey. For more details on the data on vehicle financing, see Appendix A.5.

2007q1 and 2009q4 the average cash payment required for a financed vehicle was only $1,306 after taking into account trade-ins and sales, which is quite close to the average stimulus payment of $905.2. The other reason why financing could affect households' expenditure responses is that purchasing a financed vehicle also involves purchasing a financial service, which is bundled together with the vehicle. Since the vehicle serves as collateral for the loan, the interest rate is typically lower than rates for unsecured borrowing (see Appendix Figure B1). This makes buying a financed vehicle one way to borrow cheaply. We view it as plausible that both the low average cash outlay and the interest rate differential relative to unsecured borrowing could affect the responsiveness of financed vehicle expenditures.

As Table 4 makes clear in Panels B and C, the combination of financing and trading in or selling a used vehicle can lead to very small average cash outlays. After trade ins, loans, and sales, new vehicles with financing require an average cash outlay of around $2,000. For used vehicles these outlays are $760, which is less than the average stimulus payment. In addition, cash outlays for used but non-financed vehicles average around $6,000. Though higher, the larger stimulus payments still cover a nontrivial share of such outlays. The only combination that requires large cash outlays is purchasing a new vehicle without financing.

4.4.2 Expenditure Responses

We next establish that from a pure accounting perspective both used vehicle markets and financing played an important role for the estimated expenditure responses. As shown in Table 5, we estimate that households increased spending on used vehicles by $481.1, on average, which accounts for about 65 percent (= $481.1/$742.0) of the total response of motor vehicle spending and for 66 percent (= $481.1/$733.2) of the durables response. Spending on financed vehicles increased by $436.7, which corresponds to 59 percent (= $436.7/$742.0) of the total response of motor vehicle spending and 60 percent (= $436.7/$733.2) of the durables response.

When breaking down the response further into all four combinations of financed or not financed and used or new, the only combination for which the estimated response is small is new vehicle purchases that were not financed. As just discussed, this is precisely the category that requires large cash outlays. For all of the remaining combinations the responses are large: Spending on used vehicles without financing increased by $243.1, spending on new vehicles with financing increased by $198.7, and spending on used and financed vehicles increased by $238.0.

We note that the large semi-elasticity of spending on non-financed and used vehicles stands out. At 207.3 it is more than twice as large as the other three categories, although it is not significantly different at conventional levels. Such a large response suggests that a substantial number of households were unable or unwilling to take out a vehicle loan and used the stimulus

Table 5: Expenditure Responses for New versus Used and Financed versus Non-financed Vehicles

	Total		New vehicles		Used vehicles	
	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	Semi-elasticity	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	Semi-elasticity	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	Semi-elasticity
Motor vehicle purchases	742.0	104.5	260.9	80.2	481.1	125.1
	(307.7)	(43.3)	(242.7)	(74.6)	(173.6)	(45.1)
Not financed	305.3	156.8	62.2	80.3	243.1	207.3
	(131.9)	(67.7)	(56.2)	(72.5)	(119.0)	(101.4)
Financed	436.7	84.7	198.7	80.2	238.0	89.0
	(268.0)	(52.0)	(235.5)	(95.0)	(136.8)	(51.1)
Down payment	77.0	162.5	38.5	137.3	38.5	198.8
	(42.6)	(89.8)	(36.5)	(130.2)	(22.9)	(118.3)
Loan	359.7	76.9	160.2	72.9	199.5	80.4
	(234.5)	(50.1)	(206.2)	(93.8)	(119.4)	(48.1)

Notes: The number of observations is 22716. Treatment effects are estimated as discussed in Section 3.1. Due to survey error and missing data, the loan and down payment variables are imputed in approximately 40 percent of cases. They are constructed such that the loan and down payment sum to vehicle expenditures. For details, see Appendix A.5. Standard errors that are clustered at the household level.

payment to cover a part of the purchase price.

We next ask how much of financed vehicle spending was paid down at the time of purchase and how much was financed. To answer this question we use data from the CE survey on loans and down payments. Unlike expenditure data and the financing flag we used for the analysis above, these data are measured with greater error. For instance, the reported down payment plus loan often do not equal reported vehicle expenditures. In addition, some records are missing. We discuss how we address these issues in Appendix A.5. In short, we scale incorrect records so that the loan and down payment sum to vehicle spending when the reporting error is small. For large errors and when records are missing we impute the down payment and loan with their unconditional shares times the purchase price of the vehicle.

With this caveat in mind we turn to the estimates. Of the spending on financed motor vehicles the large majority was financed. New vehicle loans increase by \$359.7, on average, accounting for 82 percent (=\$359.7/\$436.7) of the rise in spending on financed vehicles. The remaining 18 percent were paid down at the time of purchase. These financing shares were similar for new and used vehicle purchases, although they are more precisely estimated in the case of used vehicles.

Households therefore substantially lever up after receiving stimulus payments. Specifically, our estimates imply that the average household borrows an additional 40 cents of vehicle loans for every dollar they receive (=\$359.7/\$905.2), see Table 5. We call this object the MPLU (for levering up). Since we do not observe credit card and other forms of borrowing following the stimulus payment, this estimate of the MPLU is likely a lower bound.

That households lever up following stimulus payments is surprisingly absent from much of the literature concerned with the consumption and saving responses to transfers. Typically, the assumption is that the household can only spend or save the income received. This includes most theoretical work on HANK models and some empirical work. For instance, when mapping the survey responses from the categorical answers of mostly spending, mostly saving, or mostly paying off debt to the MPC, Shapiro and Slemrod (2003b) assume that individual MPCs are distributed between zero and one. If households lever up, however, they have more cash available than the government transfer alone. Our estimated MPLU of 0.4 implies that the average household spends or saves \$1.4 instead of just the \$1 from the government. Note that some evidence on MPC heterogeneity (e.g., Misra and Surico, 2014) is consistent with this result.

5 Robustness and Extensions

In this section we turn to robustness checks and some extensions.

5.1 Robustness

We first ask whether our results are robust to excluding certain observations potentially recorded with error. For instance, Parker et al. (2013) exclude observations with nondurables expenditures below the first percentile, households reporting impossible age changes, and households reporting implausible changes in the number of children, etc. (see Parker et al., 2011, p. 9 and their Appendix C). Orchard, Ramey, and Wieland (2025) exclude additional observations. When we follow their definitions of what constitutes likely erroneous cases and exclude such households from the analysis, the estimates stay quite close to the baseline, as shown in Panel A of Table 6.

We next estimate expenditure responses on a sample with paper check recipients only. Since paper check recipients made up the majority of the sample and there was more variation in their disbursement time, see the right Panel of Figure 1, there is sufficient identifying variation for such an analysis. For direct deposit recipients on the other hand, both the sample size and the time variation in disbursement are too limited. As Panel A of Table 6 shows, the expenditure response of paper check recipients is somewhat smaller than our baseline estimate. Since the average size of their stimulus payment, however, is also smaller than that of direct deposit recipients, the estimate of their 3-month MPX, 0.74, remains large and close to the baseline estimate.

Table 6: Alternative Samples, Dependent Variables, and Specifications

Difference from Baseline	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	3-month MPX	N
Reference point: baseline estimate	733.2	0.81	22716
	(338.3)	(0.37)	
Panel A. Alternative Samples			
Intersection with PSJM sample	800.2	0.87	21224
	(364.9)	(0.40)	
Intersection with ORW sample	668.7	0.73	20241
	(364.0)	(0.40)	
Paper check sample	619.0	0.74	15264
	(360.5)	(0.43)	
Strict on-time recipients	1056.0	1.25	17792
	(795.2)	(0.94)	
Panel B. Alternative Dependent Variables			
Nominal	725.3	0.77	22716
	(336.3)	(0.36)	
No winsorization	681.9	0.75	22716
	(369.2)	(0.41)	
Winsorization at 95th percentile	765.5	0.85	22716
	(304.1)	(0.34)	
Panel C. Alternative Specifications			
With PSJM controls	731.1	0.81	22716
	(340.2)	(0.38)	
Cohort-disbursement method instead of household FEs	760.1	0.84	22716
baseline sample	(340.1)	(0.38)	
Cohort-disbursement method instead of household FEs	869.2	0.96	23651
extended sample	(340.8)	(0.38)	

Notes: The first column reports the estimates $\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$, which we obtain using the BJS imputation estimator as discussed in Section 3.1. The second column reports the MPX, constructed as $\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$ divided by the average stimulus payment in the sample of treated observations. *Intersection with PSJM sample* begins with our baseline sample but only retains households that are also in Parker et al.'s (2013) sample of on-time recipients. *Intersection with ORW sample* begins with our baseline sample but only retains households that are also in Orchard, Ramey, and Wieland's (2025) baseline sample of SP recipients. *Paper check sample* drops direct deposit recipients from our baseline sample. *Strict on-time recipients* drops direct deposit recipients after May and paper check recipients after July 2008. *Nominal* uses undeflated expenditure and SP data with expenditure data still winsorized at the 98th percentile. *No winsorization* uses real but unwinsorized data. *Winsorization at 95th percentile* uses real data where the absolute value of expenditures at the UCC level have been winsorized at the 95th percentile. *With PSJM controls* includes the age of the reference person, the number of adults, and the number of children as control variables. *Cohort-disbursement method instead of household FEs* drops the individual fixed effects but adds cohort-disbursement method fixed effects, where *baseline sample* uses the baseline sample and *extended sample* uses a larger sample that can be used because individual fixed effects no longer need to be estimated from pre-treatment data. Standard errors are clustered at the household level.

In our next robustness check we adopt a stricter definition of what constitutes on-time receipt of the stimulus payment. Specifically, we define a paper check as late if it arrived after July and a direct deposit as late if it was received after May of 2008 – in line with the official disbursement schedule for on-time tax filers (see Internal Revenue Service, 2008b). We implement this robustness test because the receipt time of late recipients was generally not randomized, so that there is no guarantee that the parallel trends assumption holds. On the other hand, this definition may be too strict. Specifically, it does not allow for recall error of one month or genuine delays that may have occurred, for instance, with the delivery of paper checks. The estimate, reported in Panel A of Table 6, is larger than the baseline, but also noisy. Nonetheless this robustness check demonstrates that too lax a definition of on-time receipt is not the reason why our baseline estimates are large.

We next turn to several robustness exercises, which examine whether our results are robust to measuring the dependent variable, monthly-recorded spending on durable goods, differently. Recall that our baseline estimates are based on expenditure data which were deflated and winsorized at the 98th percentile. Correspondingly, the stimulus payment, used to construct the MPX, was also deflated. Panel B of Table 6 shows estimates for alternative choices. The estimates are very similar when (i) we use nominal (but still winsorized) data, (ii) we use expenditure data that is not winsorized but still deflated, or (iii) we winsorize real expenditures at the 95th percentile. As is intuitive, the standard errors fall for more tightly winsorized data.

Panel C of Table 6 considers two additional checks. In one specification we add the age of the household's reference person, the number of adults, and the number of children as controls (as in Parker et al., 2013). Including these controls has essentially no effect on the estimates. Second, we consider a specification in which we include cohort-disbursement method fixed effects instead of individual fixed effects. Since the treatment time within the groups of paper check and direct deposit recipients should be unrelated to household characteristics due to randomization we view this specification as econometrically defensible. The main advantage of this specification is that the individual fixed effects need not be estimated from the relatively short period of untreated observations of each household. When we estimate this specification on the baseline sample the MPX remains very similar to our baseline estimate. When extending the sample, which is possible because untreated counterfactual outcomes are now imputed from the cohort-disbursement method (and time-disbursement method) fixed effects, but not individual fixed effects, the estimate of the MPX rises.

In Table 7 we consider a range of alternative estimators. After the "negative weighting problem" was discovered for regression-based estimators in event study environments with staggered treatment timing and heterogeneous treatment effects (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021, among others), a number of alternative esti-

Table 7: Alternative Estimators

Estimator	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	3-month MPX
Reference point: baseline estimator	733.2	0.81
(Borusyak, Jaravel, and Spiess, 2024)	(338.3)	(0.37)
Sun and Abraham (2021)	693.8	0.77
	(335.3)	(0.37)
Callaway and Sant'Anna (2021)	698.8	0.77
	(375.9)	(0.42)
de Chaisemartin and D'Haultfœuille (2024)	698.7	0.77
	(362.6)	(0.40)
Dube et al. (2025)	926.4	1.02
	(337.2)	(0.37)
Ordinary Least Squares	746.2	0.82
	(346.5)	(0.38)

Notes: All estimates are based on the baseline sample with $N = 22716$ and use the sampling weights of the CE survey. They are also based on comparable specifications, which include individual fixed effects (whenever the estimator is not in differences) and time-disbursement method fixed effects. The estimators are implemented with the following Stata commands: the Sun and Abraham (2021) estimator is implemented with `eventstudyinteract`, the Callaway and Sant'Anna (2021) estimator with `csdid`, the de Chaisemartin and D'Haultfœuille (2024) estimator is implemented with `did_multiplegt_dyn`, and the Dube et al. (2025) estimator is implemented with the command `lpdid`. Standard errors are clustered at the household level.

mates have been developed. Our baseline estimator, proposed by Borusyak, Jaravel, and Spiess (2024), offers one solution to the negative weighting problem, but there are others. When we alternatively use the estimators by Sun and Abraham (2021), Callaway and Sant'Anna (2021), or de Chaisemartin and D'Haultfœuille (2024), the estimates are very similar. The estimator by Dube et al. (2025), which is a version of Jordà's (2005) local projections that avoids "forbidden comparisons" by adding a "clean control condition", delivers a somewhat larger estimate. When we use Ordinary Least Squares, which can be subject to the negative weighting problem if treatment effects are sufficiently heterogenous, the estimates are also similar to our baseline.

In summary, our results are robust to applying additional sample restrictions and using certain subsamples, they are robust to using alternative choices when constructing the dependent variable, they are robust to adding controls and replacing the individual fixed effects with cohort-disbursement method fixed effects, and they are robust to using other commonly-used estimators for event studies with staggered treatment timing.

5.2 Extensions

We study several additional outcomes in Appendix B. Two categories of spending related to our main outcome of interest, consumer durables, are spending on semi-durables and capital improvements. Semi-durables consist primarily of clothing and footwear. Capital improvements mostly include expenditures on labor and materials for various home improvements. These are not counted in consumption expenditures as defeined by either the BLS or the BEA, but as changes in assets and residential investment, respectively. When we estimate the response of semi-durables to the stimulus payment, we find that spending on semi-durables rises significantly by 54 dollars over three months, corresponding to an MPX of 0.06. Note that the relatively strong response is consistent with findings by Browning and Crossley (2009). The estimate for home improvements is small and imprecise. Both estimates are available in Panel A of Appendix Table B2.

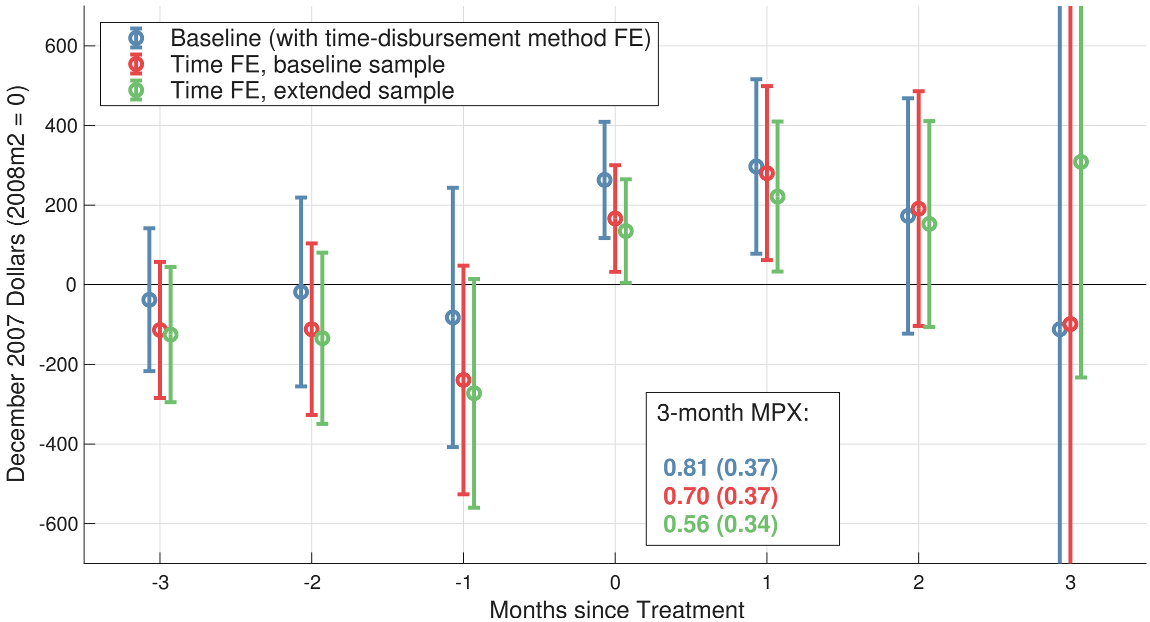
We also study the effect on leasing expenditures. While leasing services are classified as nondurable, this category is of interest to us because it is possible that households cut back on leasing when purchasing a car or truck after receiving the stimulus payment – a type of income- or cash-induced substitution. When we estimate the response of leasing expenditures to the stimulus payment, we obtain a sharp zero: There is essentially no change on regular leasing expenditures, cash down payments for new car and truck leases, or termination fees (see Panel B of Appendix Table B2). Additionally, these estimates are precise. It is therefore not the case that households cut back on leasing expenditures when purchasing a car or truck in response to receiving the stimulus payment.

5.3 Stronger Identification Assumptions

Lastly, we discuss two alternative identification assumptions as well as their implications for the durables MPX estimates. Both of these assumptions are stronger and the estimates therefore less plausible. We provide this discussion, because it is helpful for comparing our estimates to the literature.

First, we consider an alternative parallel trends assumption, which assumes that for all $it \in \Omega : E\left[x_{it}\left(0\right)\right] = \alpha_i + \delta_t$. This assumption differs from Assumption 1 because the expenditure trend, as captured by δ_t, is now common across disbursement methods. Under this assumption direct deposit recipients and paper check recipients are no longer allowed to be on different expenditure trajectories. As noted previously, whether a household received the stimulus payment by direct deposit or paper check was not random. In addition, Broda and Parker (2014) showed that paper check recipients differ from direct deposit recipients in a number of characteristics. This assumption is therefore strong. If true, however, imposing it raises the

Figure 6: Time Fixed Effects instead of Time-Disbursement Method Fixed Effects



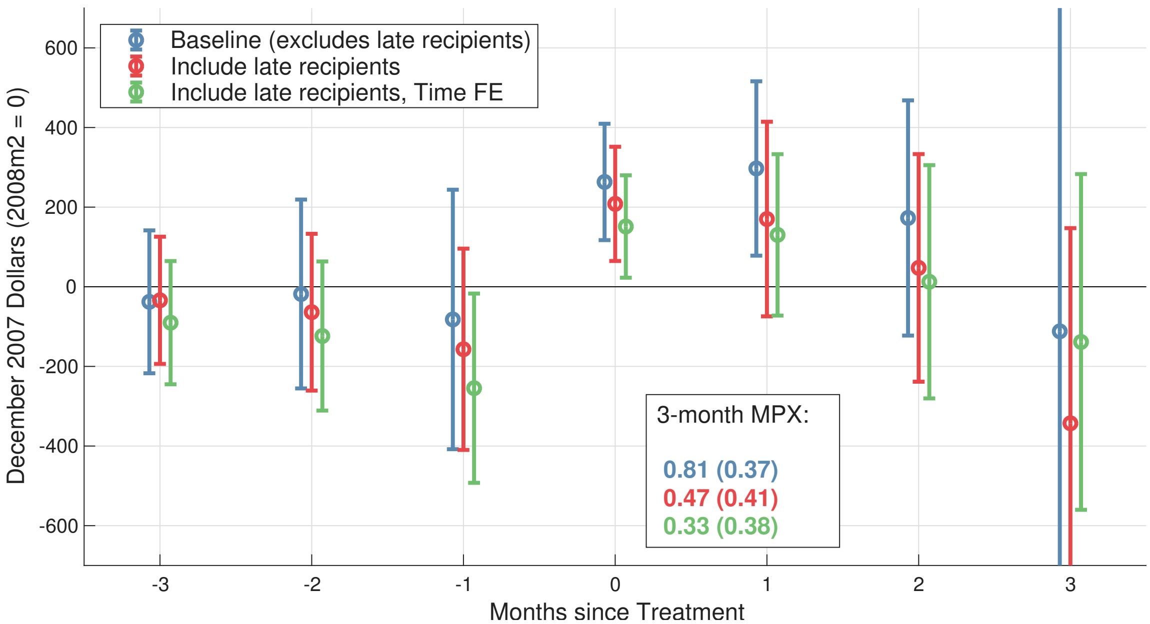
Notes: The figure shows three different sets of estimates for the pre-trend coefficients $\hat{\gamma}_h$, obtained from specification (2), as well as the horizon h treatment effects $\hat{\tau}_h$, obtained using the BJS imputation estimator as discussed in Section 3.1. In all cases the event date E_i is the month prior to the stimulus receipt and the dependent variable is monthly recorded spending on durable goods. *Baseline* is the baseline specification with time-disbursement method fixed effects estimated on the baseline sample with 22716 observations. *Time FE, original sample* replaces the time-disbursement method fixed effects with time fixed effects and is estimated on the baseline sample of on-time recipients with 22716 observations. *Time FE, extended sample* also replaces the time-disbursement method fixed effects with time fixed effects and is estimated on the baseline sample of on-time recipients, but uses the maximum number of observations, 25494. This number is greater than in the baseline sample because with this specification not-yet treated paper check recipients can be used to construct counterfactuals for direct deposit recipients and vice versa. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

sample size and therefore statistical power, because not-yet-treated paper check recipients can be used as control observations for direct deposit recipients, and vice versa.

Figure 6 shows the estimates and compares them to the baseline with time-disbursement method fixed effects (in blue). The red circles represent the estimates when the sample is left unchanged. Relative to the baseline estimates the estimated MPX falls slightly to 0.70. When estimating the expenditure response on the extended sample, the estimated MPX falls further to 0.56 (estimates in green). In addition to delivering smaller estimates of the MPX, this stronger parallel trends assumption also leads to the emergence of large pre-trends. One month prior to treatment expenditures fall to -272 (with a p-value of 0.12) in the extended sample, which is larger in absolute value than any of the monthly treatment effects in this specification. We interpret this pre-trend as a warning sign, suggesting that this version of the parallel trends assumption is potentially violated.

Second, we include late recipients in the sample. Since households received their stimulus payment late if they filed their taxes late, there is no guarantee that on-time and late tax

Figure 7: Including Late Recipients in the Sample



Notes: The figure shows three different sets of estimates for the pre-trend coefficients $\hat{\gamma}_h$, obtained from specification (2), as well as the horizon h treatment effects $\hat{\tau}_h$, obtained using the BJS imputation estimator as discussed in Section 3.1. In all cases the event date E_i is the month prior to the stimulus receipt and the dependent variable is monthly recorded spending on durable goods. *Baseline* is the baseline specification, which excludes late recipients, and is estimated on the baseline sample with 22716 observations. *Include late recipients* adds late recipients to the sample but leaves the specification unchanged. The number of observations is 31556. *Include late recipients, Time FE* also adds late recipients to the sample but additionally replaces the time-disbursement method fixed effects with time fixed effects only. The number of observations is 32277. The vertical bars denote 90 percent confidence bands based on standard errors that are clustered at the household level.

filers are comparable to one another. This variation was not randomized. An additional concern is that late recipients had some control over the receipt time of the stimulus payment, as discussed above. We therefore prefer to exclude late recipients from the sample in our baseline analysis.

As Figure 7 illustrates the estimates fall substantially when late recipients are included in the sample. This is true for both the estimated treatment effects and the pre-trend coefficients. Relative to the baseline estimates with a 3-month MPX estimate of 0.81 (in blue), the estimate falls to 0.47 when late recipients are included while the specification remains unchanged (in red). When, in addition, the specification is modified to only include time fixed effects, and not time-disbursement method fixed effects as in our baseline specification, the 3-month MPX drops further to 0.33 (in green). The inclusion of late recipients in the sample therefore has drastic effects on the estimates.

Including late recipients in the sample also leads to the emergence of large pre-trends. In the specification with time fixed effects only, expenditures drop to -255 one month before treatment and this effect is significant at the 10 percent level. There are therefore not only *ex-*

ante reasons to exclude late recipients from the sample. The pre-trend tests provide evidence against the identification assumptions made for this larger sample as well.

We note that the estimates from the sample that includes late recipients and uses a specification with time fixed effects only (green in Figure 7) is the closest analogue to the estimates of Orchard, Ramey, and Wieland (2025). Specifically, Orchard, Ramey, and Wieland (2025) also include late recipients in their sample and use time fixed effects as opposed to time-disbursement method fixed effects in their specifications. Although additional differences in the empirical setup exist, we show in Appendix C that the inclusion of late recipients in the sample explains the main discrepancy between theirs and our estimates. Importantly, this finding is independent of which estimator and specification is used: it holds both for our estimation setup and for Orchard, Ramey, and Wieland's (2025). We also show in Appendix C that late direct deposit recipients are on a dramatically different durables spending path than on-time direct deposit recipients. This likely makes them an invalid control group – at least for our estimation with durables expenditures as the outcome of interest. Our conclusion is therefore that it is more prudent to exclude late recipients from the sample.

6 Conclusion

We estimate expenditure responses on durables after transitory income shocks based on the ESA of 2008. Our main finding is that the durables MPX is large. Households spent around 80 cents on durables for every dollar of stimulus. This response was predominantly driven by purchases of financed motor vehicles, used motor vehicles, or both, and adjustments largely occurred along the extensive margin. To purchase these vehicles, the average household levered up by 40 cents on the dollar using vehicle loans. Smaller purchases of other durables also increased significantly, and accounted for about 20 percent of the total response.

In conjunction with other recent work by Borusyak, Jaravel, and Spiess (2024), among others, our findings imply that the link between current income and spending is much stronger for durables than it is for nondurables. This suggests that HANK models should explicitly feature consumer durables and their financing, similar to the work of Beraja and Zorzi (2025). Since the spending dynamics of durables are different from those of nondurables—for instance, because durables spending can be subject to the reversal property—such models have different implications for the business cycle and policy (e.g., McKay and Wieland, 2021; Mian and Sufi, 2012), the very topics for which these models were developed. We note that the state of the economy in 2008 was unusual in many ways, which may imply that our estimates are not externally valid.

A question we do not resolve in this paper is what the aggregate implications of stimulus programs such as the ESA of 2008 are. Based on a two-agent New Keynesian model with a durable goods sector, Orchard, Ramey, and Wieland (2025) argue that large expenditure responses at the micro level generate implausible counterfactuals in the aggregate. In order for such models to generate plausible counterfactuals they require (i) smaller MPXs and (ii) dampening general equilibrium forces such as steep supply curves.

Although our estimates of the MPX are large, we speculate that they do not necessarily imply implausibly large aggregate effects. One reason is that there is evidence that supply curves are convex so that they can be steep at times of sudden high demand (Boehm and Pandalai-Nayar, 2022).[21] A second potential reason is that the role of the used durables market for the general equilibrium effects of stimulus programs is – to our knowledge – not fully understood. Studying markets of used capital goods, Lanteri (2018) documents that used vehicle prices are highly volatile and procyclical (see also Gavazza and Lanteri, 2021, for related facts). If fiscal stimulus leads to a sharp rise in the relative price of used vehicles or congestion, substantial crowding out effects are possible.

References

Aaronson, Daniel, Sumit Agarwal, and Eric French. 2012. "The Spending and Debt Response to Minimum Wage Hikes." *American Economic Review* 102 (7):3111–39.

Abel, Andrew B. 1990. "Chapter 14 Consumption and investment." In *Handbook of Monetary Economics*, vol. 2, edited by Benjamin M. Friedman and Frank H. Hahn. Elsevier, 725–778.

Auclert, Adrien. 2019. "Monetary Policy and the Redistribution Channel." *American Economic Review* 109 (6):2333–67.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "The Intertemporal Keynesian Cross." *Journal of Political Economy* 132 (12):4068–4121.

Beraja, Martin and Nathan Zorzi. 2025. "Durables and Size-Dependence in the Marginal Propensity to Spend." *Mimeo* .

Berger, David and Joseph Vavra. 2015. "Consumption Dynamics During Recessions." *Econometrica* 83:101–154.

Boehm, Christoph E. 2020. "Government consumption and investment: Does the composition of purchases affect the multiplier?" *Journal of Monetary Economics* 115:80–93.

Boehm, Christoph E. and Nitya Pandalai-Nayar. 2022. "Convex Supply Curves." *American Economic Review* 112 (12):3941–69.

Boehm, Johannes, Etienne Fize, and Xavier Jaravel. 2025. "Five Facts about MPCs: Evidence from a Randomized Experiment." *American Economic Review* 115 (1):1–42.

[21] For estimates for durable goods industries see Boehm and Pandalai-Nayar (2022, Appendix D.5).

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* 91 (6):3253–3285.

Broda, Christian and Jonathan A. Parker. 2014. "The Economic Stimulus Payments of 2008 and the aggregate demand for consumption." *Journal of Monetary Economics* 68 (S):20–36.

Browning, Martin and Thomas F. Crossley. 2009. "Shocks, Stocks, and Socks: Smoothing Consumption Over a Temporary Income Loss." *Journal of the European Economic Association* 7 (6):1169–1192.

Bureau of Economic Analysis. 2022. "Table 2.4.5U. Personal Consumption Expenditures by Type of Product." Tech. rep., Bureau of Economic Analysis. https://apps.bea.gov/iTable/iTable.cfm?ReqID=19&step=4&isuri=1&0=flatfiles#reqid=19&step=4&isuri=1&1921=flatfiles, accessed 07/27/2022.

Bureau of Labor Statistics. 2017. "Concordance Mapping of CE UCC Codes to PCE Series from Table 2.4.5U: Expenditures by Type of Product - 9-21-2018." Tech. rep., Bureau of Labor Statistics. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwicqpTk_pn5AhXVgGoFHQyMAp0QFnoECAkQAQ&url=https%3A%2F%2Fwww.bls.gov%2Fcex%2Fpce-concordance-2017.xlsx&usg=AOvVaw3vb6P30h8BijJA4YRao-Dr, accessed 07/27/2022.

———. 2022a. "Appendix 5. Consumer Expenditure survey item name (universal classification codes-UCC) to Consumer Price Index item titles (entry level item-ELI) concordance." Tech. rep., Bureau of Labor Statistics. https://www.bls.gov/cpi/additional-resources/ce-cpi-concordance.htm, accessed 07/27/2022.

———. 2022b. "Item Aggregation Tree." Tech. rep., Bureau of Labor Statistics. Obtained via email from Jonathan Church, email address cpi_info@bls.gov, on 09/22/2020.

Callaway, Brantly and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics* 225 (2):200–230. Themed Issue: Treatment Effect 1.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9):2964–96.

———. 2024. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *The Review of Economics and Statistics* :1–45.

Doms, Mark E. and Timothy Dunne. 1998. "Capital Adjustment Patterns in Manufacturing Plants." *Review of Economic Dynamics* 1 (2):409–429.

Dube, Arindrajit, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. 2025. "A Local Projections Approach to Difference-in-Differences." *Journal of Applied Econometrics* 40 (7):741–758.

Fagereng, Andreas, Martin B. Holm, and Gisle J. Natvik. 2021. "MPC Heterogeneity and Household Balance Sheets." *American Economic Journal: Macroeconomics* 13 (4):1–54.

Fuster, Andreas, Greg Kaplan, and Basit Zafar. 2020. "What Would You Do with $500? Spending Responses to Gains, Losses, News, and Loans." *The Review of Economic Studies* 88 (4):1760–1795.

Ganong, Peter, Fiona Greig, Pascal Noel, Daniel M. Sullivan, and Joseph Vavra. 2024. "Spending and Job-Finding Impacts of Expanded Unemployment Benefits: Evidence from Administrative Micro Data." *American Economic Review* 114 (9):2898–2939.

Gavazza, Alessandro and Andrea Lanteri. 2021. "Credit Shocks and Equilibrium Dynamics in Consumer Durable Goods Markets." *The Review of Economic Studies* 88 (6):2935–2969.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225 (2):254–277. Themed Issue: Treatment Effect 1.

Gourio, Francois and Anil K Kashyap. 2007. "Investment spikes: New facts and a general equilibrium exploration." *Journal of Monetary Economics* 54 (Supplement):1–22.

Hausman, Joshua K. 2016. "Fiscal Policy and Economic Recovery: The Case of the 1936 Veterans' Bonus." *American Economic Review* 106 (4):1100–1143.

Havranek, Tomas and Anna Sokolova. 2020. "Do consumers really follow a rule of thumb? Three thousand estimates from 144 studies say "probably not"." *Review of Economic Dynamics* 35:97–122.

Hsieh, Chang-Tai. 2003. "Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund." *American Economic Review* 93 (1):397–405.

Internal Revenue Service. 2008a. "Facts about the 2008 Stimulus Payments." FS-2008-15, https://www.irs.gov/pub/irs-news/fs-08-15.pdf, last accessed 07/03/2023.

———. 2008b. "IRS Announces Economic Stimulus Payment Schedules, Provides Online Payment Calculator, and How to Fix It." IR-2008-44, https://www.irs.gov/pub/irs-news/ir-08-044.pdf, last accessed 07/03/2023.

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

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

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1):161–182.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review* 108 (3):697–743.

Kaplan, Greg and Giovanni L. Violante. 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica* 82 (4):1199–1239.

Kerwin, Jason T., Nada Rostom, and Olivier Sterck. 2025. "Striking the Right Balance: Why Standard Balance Tests Over-Reject the Null, and How to Fix It." *Mimeo* .

Kueng, Lorenz. 2018. "Excess Sensitivity of High-Income Consumers*." *The Quarterly Journal of Economics* 133 (4):1693–1751.

Laibson, David, Peter Maxted, and Benjamin Moll. forthcoming. "A Simple Mapping from MPCs to MPXs." *Journal of Finance: Insights and Perspectives* .

Lanteri, Andrea. 2018. “The Market for Used Capital: Endogenous Irreversibility and Reallocation over the Business Cycle.” *American Economic Review* 108 (9):2383–2419.

Lewis, Daniel, Davide Melcangi, and Laura Pilossoph. forthcoming. “Latent Heteogeneity in the Marginal Propensity to Consume.” *Review of Economic Studies* .

McKay, Alisdair and Johannes F. Wieland. 2021. “Lumpy Durable Consumption Demand and the Limited Ammunition of Monetary Policy.” *Econometrica* 89 (6):2717–2749.

Mian, Atif and Amir Sufi. 2012. “The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program*.” *The Quarterly Journal of Economics* 127 (3):1107–1142.

Misra, Kanishka and Paolo Surico. 2014. “Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs.” *American Economic Journal: Macroeconomics* 6 (4):84–106.

Orchard, Jacob D, Valerie A Ramey, and Johannes F Wieland. 2025. “Micro MPCs and Macro Counterfactuals: The Case of the 2008 Rebates.” *The Quarterly Journal of Economics* 140 (3):2001–2052.

Parker, Jonathan A., Jake Schild, Laura Erhard, and David S. Johnson. 2022. “Economic Impact Payments and Household Spending during the Pandemic.” *Brookings Papers on Economic Activity* :81–130.

Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland. 2011. “Consumer Spending and the Economic Stimulus Payments of 2008.” Working Paper 16684, National Bureau of Economic Research.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. 2013. “Consumer Spending and the Economic Stimulus Payments of 2008.” *American Economic Review* 103 (6):2530–2553.

Roth, Jonathan. 2022. “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends.” *American Economic Review: Insights* 4 (3):305–322.

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

Sahm, Claudia R., Matthew D. Shapiro, and Joel Slemrod. 2010. “Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications.” *Tax Policy and the Economy* 24 (1):69–110.

Shapiro, Matthew D. and Joel Slemrod. 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.

Souleles, Nicholas S. 1999. "The Response of Household Consumption to Income Tax Refunds." *American Economic Review* 89 (4):947–958.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2):175–199. Themed Issue: Treatment Effect 1.

Thakral, Neil and Linh T Tô. 2022. "Anticipation and Consumption." *Mimeo* .

Wilcox, David W. 1989. "Social Security Benefits, Consumption Expenditure, and the Life Cycle Hypothesis." *Journal of Political Economy* 97 (2):288–304.

For Online Publication

A Appendix: Data

A.1 Data from the CE Survey

Data on household expenditures come from the Interview Module of the CE survey. These data are available for download from the website https://www.bls.gov/cex/pumd_data.htm (last accessed 9/22/2025). We predominantly use data from the `fmli`YY files and the `mtbi`YY files (where YY is the two-digit interview year). The `fmli`YY files contain data on consumer unit characteristics and sampling weights, among other things, and the `mtbi`YY files contain most household-level expenditures at the UCC level. In addition, we use some information from the `itbi`YY files and the `ovb`YY files. We use data from 1997 to 2019. Over this period the CE's definitions of consumption expenditures are relatively consistent over time. Data on the stimulus payments comes from the `rbt08` file.

To process and clean the data, we also use auxiliary files available from the BLS website. These include in particular, the hierarchical groupings of expenditures, available at https://www.bls.gov/cex/pumd-getting-started-guide.htm#section3.3.2 (last accessed 9/22/2025), which define consumption expenditures for all years in the CE survey.

A.2 Other Data Sources

In addition to the data from the CE survey, we use several other data sources. These are

- Personal Consumption Expenditures by Type of Product, BEA Table Table 2.4.5U
- Price Indexes for Private Fixed Investment in Structures by Type, BEA Table 5.4.4
- Item-level data from the Consumer Price Index, BLS (https://download.bls.gov/pub/time.series/cu, last accessed 9/4/2023)
- Survey of Consumer Finances, for the variables used see the note of Figure B1.

A.3 Definition of Durability

We define durables as follows. The main criterion for durability is that the item or service purchased does not have a nontrivially short service life. The second criterion is that it not a rental or a lease. We include repairs and some maintenance expenditures in our definition of durables. Lastly, we err on the side of inclusion. For instance, if a UCC description reads "rental and repair" of a particular item, we classify it as durable (because we classify repairs as durable).

We cross-check our definition of durables with similar definitions from the BEA (for personal consumption expenditures (or PCE)) and the BLS (for the consumption definition underlying the CPI). When available, these definitions mostly align with ours. We note, however, that due to different definitions of narrow expenditure categories across these consumption definitions, many UCCs do not match uniquely to categories from the PCE or CPI.

Semidurables are defined as clothing and footwear as well as related repair services.

Table A1: Expenditures by CE Expenditure Categories and Durability

CE Expenditure Category	Nondurables		Semidurables		Durables	
	# of UCCs	Quarterly expenditures	# of UCCs	Quarterly expenditures	# of UCCs	Quarterly expenditures
Alcoholic beverages at home	2	34	0	0	0	0
Alcoholic beverages away from home	3	44	0	0	0	0
Cash contribution	9	352	0	0	0	0
Change in vehicle sales	0	0	0	0	8	-88
Drugs	1	84	0	0	0	0
Education	6	198	0	0	5	19
Other entertainment supplies, equipment, and services	12	15	0	0	16	80
Fees and admissions	10	136	0	0	0	0
Food away from home	6	549	0	0	0	0
Food at home	4	1122	0	0	0	0
Apparel and services: Footwear	0	0	4	35	0	0
Gasoline and motor oil	5	630	0	0	0	0
Household furnishings and equipment	2	1	0	0	72	289
Household operations	19	221	0	0	5	8
Health insurance	13	394	0	0	0	0
Apparel and services: Children under 2	1	8	4	9	0	0
Life and other personal insurance	2	68	0	0	0	0
Medical services	8	173	0	0	0	0
Medical supplies	2	1	0	0	4	20
Apparel and services: Men and boy	0	0	29	67	0	0
Miscellaneous	15	137	0	0	0	0
Other apparel products and services	4	24	4	3	3	25
Pensions and Social Security	5	1262	0	0	0	0
Personal care products and services	1	69	0	0	2	2
Pets, toys, and playground equipment	3	65	0	0	3	28
Public transportation	9	115	0	0	0	0
Reading	2	14	0	0	3	13
Shelter	23	2272	0	0	36	167
Tobacco products and smoking supplies	2	75	0	0	0	0
Television, radios, sound equipment	9	153	0	0	18	83
Utilities, fuels, and public services	36	851	0	0	0	0
Other vehicle expenses	28	401	0	0	22	155
Vehicle purchases (net outlay)	0	0	0	0	6	739
Apparel and services: Women and girls	0	0	29	107	0	0
Total	242	9469	70	221	203	1540

Notes: This table reports the number of UCCs and average quarterly expenditures of households in 2007m12 dollars. The data are from the BLS and the sample includes all households interviewed between 2007q1 and 2009q4.

Of the 514 UCCs that are counted as consumption expenditures over the period from 2007q1-2009q4, we classify 203 as durable and 70 as semidurable. The remaining 242 UCCs are classified as nondurable goods or services. Table A1 provides an overview by expenditure category.

A.4 Monthly-recorded Spending

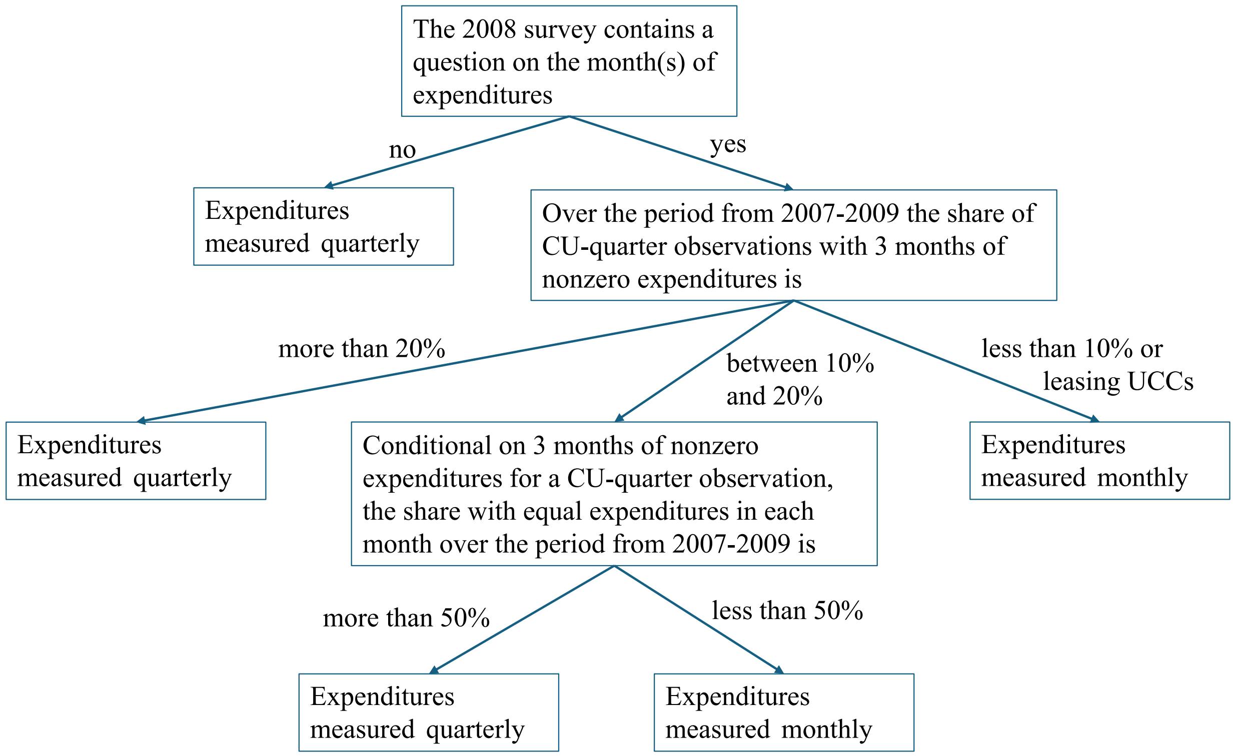
To assess whether a narrow expenditure category (UCC) is recorded at the monthly as opposed to quarterly frequency we proceed as follows. We first check whether information on the month(s) of expenditures is collected in the 2008 survey. To do so, we manually link each UCC to one or more survey questions from the 2008 questionnaire (available at https://www.bls.gov/cex/capi/2008/, last accessed 10/16/2025) and check whether the survey asks about the timing of a purchase. It often does. For instance, if a household purchased a major appliance such as a stove or a dishwasher, the survey asks "When did you purchase it?" (see https://www.bls.gov/cex/capi/2008/csxsection6a.htm#MAJTYPE, last accessed 16/10/2025).

A question in the 2008 survey on the precise month of expenditures is a necessary condition for spending on a UCC to be recorded monthly. It is not clear, however, that such a question is sufficient. The reason is that there could be a bias towards reporting smoothed expenditures for all three months of the reference quarter. An example is spending on school or college tuition. In this case, the survey asks "In what month was the payment made?", but also suggests "Enter 13 for same amount each month of the reference period" (see https://www.bls.gov/cex/capi/2008/csxsection16.htm#EDMONTHA, last accessed 16/10/2025).

To ensure that such smoothing across months does not interfere with our objective of accurately measuring monthly expenditures, we require additionally that a large share of household-quarter observations with nonzero expenditures during the reference quarter does not report positive expenditures in all three months of the quarter. By requiring at least one month of zero expenditures, this filter removes UCCs that report smoothed consumption across months in the reference quarter. For some UCCs we check additionally if the reported monthly expenditures appear excessively smooth, i.e. equal, to accurately reflect true expenditures.

We deviate from this classification scheme only for UCCs on the leasing of cars and trucks. The CE survey asks explicitly about the first and the last month of the lease, as well as extra charges such as cash down payments at the beginning of the leasing period and termination fees (see https://www.bls.gov/cex/capi/2008/csxsection10302.htm, last accessed 21/10/2025). Based on these questions, we expect that expenditures on leased cars and trucks are accurately recorded at the monthly frequency even though many households will report regular monthly payments. Figure A1 visualizes our classification into monthly- and quarterly-recorded spending. As Table 1 shows 95.1 percent of durables expenditures and all expenditures on semidurables are recorded monthly.

Figure A1: Monthly- versus Quarterly-Recorded Spending



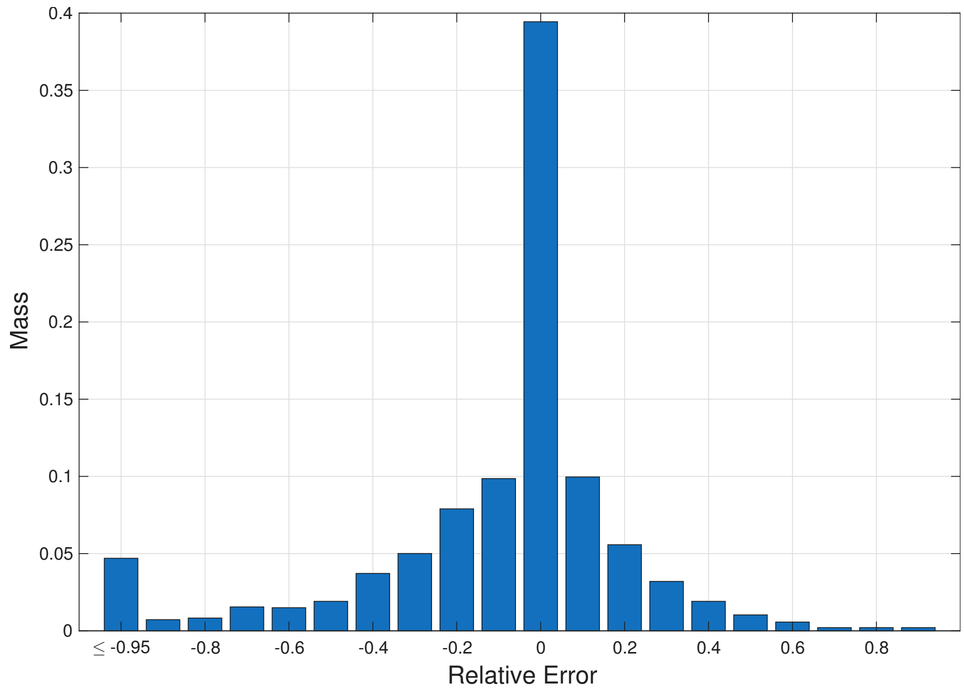
Notes: The figure illustrates how we classify UCCs into those that provide information at the monthly frequency and those that provide information only at the quarterly frequency.

A.5 Data on Vehicle Financing

In addition to expenditure data we use data related to the financing of vehicles in Section 4.4. These data come from the `ovb`YY files, which we combine with information from the `mtbi`YY files. Both sets of files are populated with data from the Interview Module of the CE survey, and are available for download from the website https://www.bls.gov/cex/pumd_data.htm (last accessed 08/29/2023). These datasets contain information on the vehicle principal balance, the initial down payment, and whether the vehicle was financed or not, among other variables. Each of these variables is specific to whether the motor vehicle purchased was a car, a truck, or a motorcycle and whether it was used or new. For what follows we define the *type* of each vehicle expenditure as one of the following: new auto, new truck, new motorcycle, used auto, used truck, or used motorcycle expenditures. We denote the vehicle expenditure type by θ.

One issue with these data is that the vehicle principal balance (henceforth loan) plus down payment in the month of purchase does not always equal expenditures net of trade-ins. We confirmed in an email exchange with the CE staff that this identity is not enforced in the survey and that failure of the identity arises from survey response error. To assess the

Figure A2: Histogram of the Relative Error



Notes: Observations vary by household, vehicle expenditure type, and month. The sample includes all cases interviewed between 2007q1 and 2009q4 in which households had strictly positive vehicle spending on a particular vehicle type, the vehicle was financed, and financial information was not missing. The share of observations with missing data on loan principal and down payment is 15.4 percent. The relative error is computed from equation (A1) using unwinsorized and undeflated data.

magnitude of this problem we compute the relative error as

$$\mathrm{RE}_{i\theta t} = \frac{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})} - \mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})} - \mathrm{Down\ Payment}_{i\theta t}^{(\mathrm{n,uw})}}{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})}}, \tag{A1}$$

where the superscript $(\mathrm{n}, \mathrm{uw})$ indicates that the data is undeflated and unwinsorized.

Figure A2 shows that the error is often sizable. Only in approximately 39 percent of cases is the error zero or small. Note that below negative -0.95, the error is censored in the histogram. In addition to the survey error illustrated in Figure A2, a discrepancy arises from the fact that we deflate and winsorize vehicle expenditures prior to using the data in our empirical analysis.

For clarity in Section 4.4 we prefer to enforce the identity that the loan principal plus down payment equals expenditures net of trade-ins. To do so, we proceed as follows. First, if the relative error (A1) is zero, we scale the loan and down payment by the ratio of deflated and winsorized (r, w) to undeflated and unwinsorized $(\mathrm{n}, \mathrm{uw})$ vehicle expenditures (see top row of Table A2). In the second case, if the relative error is between zero and 25 percent in absolute value, we re-scale the loan and down payment variables by the ratio of real and winsorized vehicle expenditures to the sum of the nominal and unwinsorized down payment and principal. The rationale here is that errors of this magnitude could plausibly arise from survey error.

Despite this error, we deem this data informative about the relative sizes of the loan and down payment and hence we scale the responses only to preserve the identity of interest. Lastly, if the relative error exceeds 25 percent in absolute value, we discard the individual responses and impute the loan and down payment variables using the average ratio of loan size to vehicle expenditures and down payment to vehicle expenditures (see the third row in Appendix Table A2). We also use this imputation when information on the loan and down payment is missing. This imputation occurs in about 40 percent of cases. Conditioning on a relative error of zero, the average down payment is about 8 percent of total vehicle expenditures.

Table A2: Constructing the Loan and Down Payment Variables

Relative Error (A1)	Loan	Down Payment
$\mathrm{RE}_{i\theta t} = 0$	$\frac{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}}{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})}} \cdot \mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})}$	$\frac{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}}{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})}} \cdot \mathrm{Down\ Payment}_{i\theta t}^{(\mathrm{n,uw})}$
$0 < \lvert\mathrm{RE}_{i\theta t}\rvert \leq 25\%$	$\frac{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}}{\mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})} + \mathrm{Down\ Payment}_{i\theta t}^{(\mathrm{n,uw})}} \cdot \mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})}$	$\frac{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}}{\mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})} + \mathrm{Down\ Payment}_{i\theta t}^{(\mathrm{n,uw})}} \cdot \mathrm{Down\ Payment}_{i\theta t}^{(\mathrm{n,uw})}$
$25\% < \lvert\mathrm{RE}_{i\theta t}\rvert$	$\left.\overline{\frac{\mathrm{Loan}^{(\mathrm{n,uw})}}{\mathrm{Vehicle}^{(\mathrm{n,uw})}}}\right\rvert_{\theta} \cdot \mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}$	$\left.\overline{\frac{\mathrm{Down\ Payment}^{(\mathrm{n,uw})}}{\mathrm{Vehicle}^{(\mathrm{n,uw})}}}\right\rvert_{\theta} \cdot \mathrm{Vehicle}_{i\theta t}^{(\mathrm{r,w})}$

Notes: This table summarizes how we construct the loan and down payment variables used in our empirical analysis. Throughout variables superscripted $(\mathrm{n,uw})$ indicate that the original data has not been deflated or winsorized, whereas variables with the superscript $(\mathrm{r,w})$ have been deflated and winsorized at the 98th percentile as discussed in the text. The average ratio of the loan size to vehicle expenditures is constructed as

$$\left.\overline{\frac{\mathrm{Loan}^{(\mathrm{n,uw})}}{\mathrm{Vehicle}^{(\mathrm{n,uw})}}}\right\rvert_{\theta} = \frac{1}{\lvert\Omega_{\theta}^{\mathrm{noRE}}\rvert} \sum_{it \in \Omega_{\theta}^{\mathrm{noRE}}} \frac{\mathrm{Loan}_{i\theta t}^{(\mathrm{n,uw})}}{\mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})}},$$

where $\Omega_{\theta}^{\mathrm{noRE}} = \left\{ it \in \Omega : \mathrm{Vehicle}_{i\theta t}^{(\mathrm{n,uw})} > 0 \text{ and } \mathrm{RE}_{i\theta t} = 0 \right\}$. The average ratio of down payments to vehicle expenditures is constructed analogously. In order to accommodate the possibility that financing patterns may vary for different vehicles and new versus used purchases, we construct these ratios separately for each of the six types of vehicle purchases: new auto purchases, new truck purchases, new motorcycle purchases, used auto purchases, used truck purchases, used motorcycle purchases.

B Appendix: Additional Results

Table B1: Expenditures and Purchase Fequencies of Top 15 Durables UCCs

	Quarterly		
	Average expenditures *(in 2007m12 dollars)*		Frequency of purchases *(%)*
	Unconditional	Conditional on purchase	
Item description			
	Top 15 items		
Used cars	186	8,804	2.1
New cars	185	21,232	0.9
New trucks	174	25,943	0.7
Used trucks	169	10,082	1.7
Other repair and maintenance services	46	1,019	4.5
Televisions	41	920	4.5
Computers and computer hardware for nonbusiness use	38	616	6.2
Tires – purchased, replaced, installed	28	339	8.3
Toys, games, arts and crafts, and tricycles	26	135	18.9
Roofing and gutters	25	1,915	1.3
Heat, A/C, electrical work	24	546	4.3
Sofas	23	973	2.4
Jewelry	21	254	8.4
Painting and papering	18	1,444	1.2
Motor repair, replacement	17	567	2.9

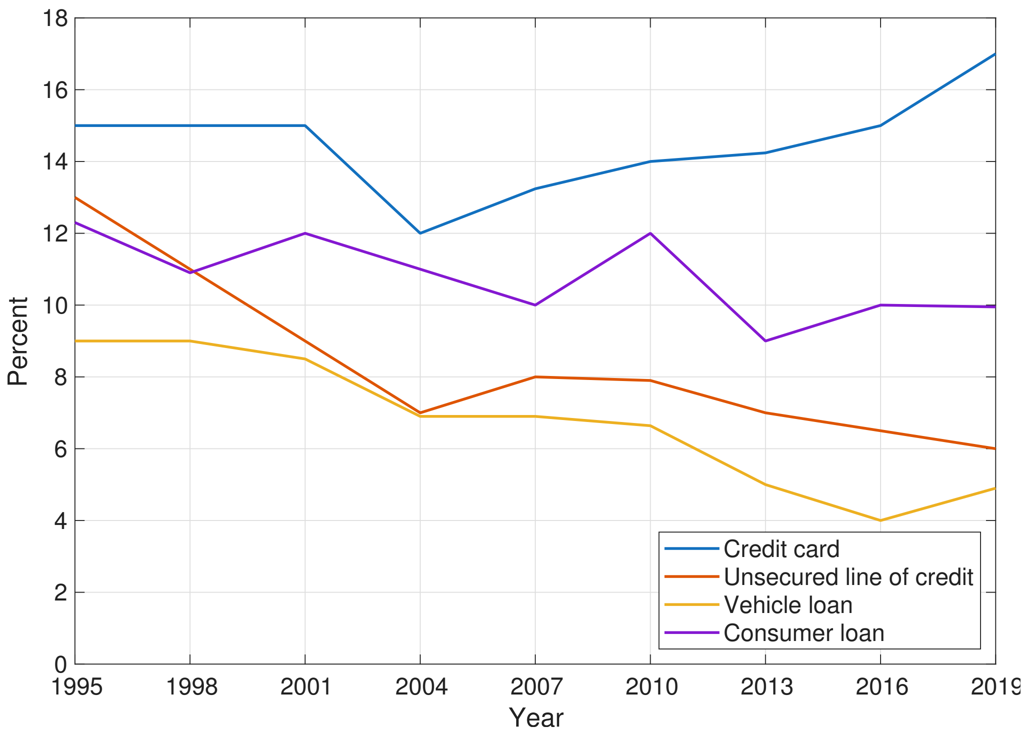
Notes: This table reports average quarterly expenditures and the average quarterly frequency of purchases using data from the CE survey of the BLS. The sample covers the period from 2007q1 to 2009q4. Expenditures are deflated using item-level CPI as discussed in Section 2.2.1 and aggregated to the reference quarter of the interview. Unconditional average spending is calculated over all observations. Conditional average spending is calculated over a sample of household purchasing the item in question. Frequency is the share of observations with non-zero expenditures per quarter. Note that vehicle spending is net of trade-ins but gross of sales, which is how the CE survey defines these UCCs.

Table B2: Extensions

	$\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$	3-month MPX
Panel A. Semidurables and Capital Improvements		
Semidurables	54.0	0.06
	(20.7)	(0.02)
Capital Improvements	-77.3	-0.09
	(91.9)	(0.10)
Panel B. Car and Truck Leasing		
Car and Truck Lease Payments	-4.5	0.00
	(4.0)	(0.00)
Cash down payment for Car and Truck Lease	1.0	0.00
	(1.6)	(0.00)
Termination Fee for Car and Truck Lease	-1.8	0.00
	(1.2)	(0.00)

Notes: The number of observations is 22716. The first column reports the estimates $\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2$, which we obtain using the BJS imputation estimator as discussed in Section 3.1. The second column reports the MPX, constructed as $(\hat{\tau}_0 + \hat{\tau}_1 + \hat{\tau}_2)/905.2$. *Semidurables* includes mostly clothing and footwear, see Table A1 for details. *Capital Improvements* includes labor and materials for various capital improvements (UCCs 220512, 220513, 220611, 220612, 220615, 220616, 790610, 790611, 790620, 990950). Standard errors are clustered at the household level.

Figure B1: Interest Rates on Vehicle and Other Loans



Notes: The data are from the Survey of Consumer Finances (SCR). The figure reports medians of annual percentage rates for each type of loan, where all medians are computed using the survey weights. The SCF variables we use are: x7132 for credit card rates; x1111, x1122, and x1133 for interest rates of lines of credit where we condition on these being unsecured (x1103, x1114, x1125); x2219, x2319, x2419, and x7170 for vehicle loan rates; and x2724, x2741, x2824, x2841, x2924, and x2941 for consumer loan rates.

C Appendix: Comparison to Orchard, Ramey, and Wieland (2025)

The magnitudes of our estimated expenditure responses differ from those of Orchard, Ramey, and Wieland (2025) (henceforth ORW). Since our focus and estimation setup is different from ORW's, one would not expect the estimates to align exactly. What is surprising, however, is that we find greater expenditure responses than do ORW although we only study a subset of spending. In this appendix we explain why our estimates differ from those of ORW.

C.1 Differences in the Estimation Setup

There are five differences between ORW's and our estimation setup. First, the dependent variable differs. While ORW study total expenditures, we study expenditures on durable goods only. In addition, ORW use the BEA's definition of personal consumption expenditures, while our definition is close to the BLS'. We also winsorized expenditures at the UCC level prior to aggregating them to total expenditures. Second, the estimating equation and the estimator differ. ORW's baseline specification is an adaptation from Parker et al.'s (2013) specification, which, in turn, is motivated by the consumption Euler equation. Relative to Parker et al. (2013), ORW added a lagged stimulus indicator, allowed for cohort-specific treatment effects, and added several controls. By contrast, we estimate dynamic treatment effects using the BJS imputation estimator. Third, the time aggregation differs. While ORW use quarterly expenditure data, our data is monthly. Fourth, we redefined treatment as occurring one month prior to receiving the stimulus payment to address the rise in households' expenditures in that month. As discussed in Section 4.1, this redefinition strictly weakens our identification assumptions. Taken together, the previous two points imply that the three-months MPX we estimate differs from ORW's estimate with regard to the timing of expenditures relative to the time of the stimulus receipt. Lastly, the sample differs.

It turns out that the main difference between ORW's and our estimates comes from a difference in the sample, and specifically, the inclusion of late recipients. If late recipients are included in the sample, the estimated expenditure responses are small. If they are excluded from the sample, the estimates are large. While ORW's sample includes late recipients, our baseline sample does not, because the treatment time of late recipients is generally not randomized (see discussion in Section 2.1). Recall that we identified recipients as late (following Parker et al., 2013) if they received direct deposits after June of 2008 or paper checks after August of 2008. Table C1 shows the timing of stimulus receipts as reported in the CE survey. Among paper check recipients 5.7% report dates we classify as late. At 14.8% the share of late recipients is substantially greater for households receiving direct deposits.

Table C2 illustrates how the inclusion of late recipients in the sample affects the estimated expenditure responses for three different specifications. Panel A shows the estimates from ORW's baseline analysis. Specification (1) shows the estimate when late recipients are included in the sample. This estimated quarterly expenditure response of 320.72 coincides with the estimate from their Table 3 Panel B specification (4). When we exclude late recipients from the sample in specification (2) of Panel A, the estimate rises to a value of 832.8.[22] Panel B

[22]Note that when late recipients are dropped from the sample, the treatment effects of cohorts interviewed after

Table C1: Timing of Stimulus Receipts

	Paper Check			Direct Deposit		
	Count	Share (%)	Late	Count	Share (%)	Late
2008 April	40	1.3	No	120	5.7	No
2008 May	555	17.5	No	1,160	55.1	No
2008 June	1,360	42.8	No	514	24.4	No
2008 July	881	27.7	No	170	8.1	Yes
2008 August	161	5.1	No	71	3.4	Yes
2008 September	86	2.7	Yes	29	1.4	Yes
2008 October	42	1.3	Yes	21	1.0	Yes
2008 November	38	1.2	Yes	15	0.7	Yes
2008 December	8	0.3	Yes	0	0.0	Yes
2009 January	4	0.1	Yes	3	0.1	Yes
2009 February	3	0.1	Yes	2	0.1	Yes
Late	181	5.7		311	14.8	
Total	3,178			2,105		

Notes: This table presents the timing of stimulus receipt dates as reported in the CE survey. The sample includes all households that receive exactly one stimulus payment over their time in the sample, as well as households that receive multiple stimulus payments within the same receipt month with the same disbursement method.

shows the estimates for ORW's implementation of the BJS imputation estimator. When late recipients are included in the sample, the expenditure response is estimated to be 351.0 (see specification (1) of Panel B or their Appendix Table C.4., specification (4)). When they are dropped from the sample in specification (2) of Panel B, the estimate rises to a value of 1012.3. Lastly, Panel C presents estimates following our preferred estimation strategy. Specification (1) of Panel C reports our baseline estimate, which is based on a sample that excludes late recipients. In this case the estimated expenditure response is 733.2. This estimate falls to 425.7 when late recipients are added to the sample (see specification (2) of Panel C). Hence, in all three cases are the estimates large when late recipients are excluded from the sample and small if not.

C.2 Why does the Inclusion of Late Recipients Affect the Estimates?

We next discuss why the inclusion of late recipients affects the estimated expenditure responses. For concreteness, we focus on the implications for the BJS imputation estimator. The implications for other estimators are similar.

August 2008 ($\beta_{2,3}$, $\beta_{2,4}$, $\beta_{2,5}$, etc. in ORWs equation (27)) are no longer identified. The reason is that the treatment indicator of these cohorts plus the lagged treatment indicators of previously treated cohorts sum to the time fixed effect, creating a perfect multicollinearity problem. Intuitively, there are no untreated control observations left after August 2008 when the sample is limited to on-time recipients, and hence the statistical model cannot construct treatment effects for these cohorts. Following the literature on event studies with heterogeneous treatment effects (see Section 5.1), we therefore construct the average treatment effect reported in specification (2) of Panel A in Table C2 only from the first three cohorts for which the counterfactual without treatment is constructed from the not-yet treated observations. As in ORW, the weights are constructed from Sun and Abraham's (2021) routine `eventstudyweights`, but re-scaled so that they sum to one for the first three cohorts. Note that this point only applies to specification (2) of Panel A in Table C2.

Table C2: The Role of Late Recipients in the Sample

Panel A: ORW baseline estimates

Dependent variable: total expenditures	(1)	(2)	(3)
Quarterly expenditure response	320.7	832.8	907.5
	(496.7)	(344.7)	(370.7)
Include late recipients	Y	N	Y
Interview month FE	Y	Y	N
Interview month × Disbursement method × On-time/late FE	N	N	Y
Observations	10,076	9,473	9,775

Panel B: ORW estimates (BJS specification)

Dependent variable: total expenditures	(1)	(2)	(3)
Quarterly expenditure response	351.0	1012.3	818.0
	(557.4)	(297.6)	(410.6)
Include late recipients	Y	N	Y
Interview month FE	Y	Y	N
Interview month × Disbursement method × On-time/late FE	N	N	Y
Observations	5,511	4,578	4,117

Panel C: Our baseline specification

Dependent variable: monthly-recorded durables expenditures	(1)	(2)	(3)
3-month expenditure response	733.2	425.7	751.9
	(338.3)	(370.0)	(316.7)
3-month MPX	0.81	0.47	0.84
	(0.37)	(0.41)	(0.35)
Include late recipients	N	Y	Y
Household FE	Y	Y	Y
Month × Disbursement method FE	Y	Y	N
Month × Disbursement method × On-time/late FE	N	N	Y
Observations	22,716	31,556	24,615

Notes: The estimates in Panels A and B are produced from ORW's replication code and therefore adopt their variable definitions, use their samples, and use their estimating equations and estimators. Specifications (1) in each panel replicate ORW's estimates (see their Table 3 Panel B specification (4) and Appendix Table C.4. specification (4)). Specifications (2) and (3) in this table modify the sample and the set of fixed effects as indicated in the table. Panel C is based on our own analysis. Standard errors that are clustered at the household level and reported in parentheses.

The inclusion of late recipients affects the estimates in two ways. First, late recipients are part of the control group prior to being treated. Specifically, data from late recipients prior to treatment are used to estimate the month-disbursement method fixed effects, which are necessary to impute the counterfactual spending path of treated observations (see Section 3.1). Second, the effect of treatment on the expenditures of late recipients will affect the estimated average treatment effects if their individual treatment effects differ from those of on-time recipients.

It is possible to disentangle these two effects by including late recipients in the sample, while estimating separate time fixed effects for on-time and late recipients. More precisely, there are now four sets of time fixed effects: all combinations of on-time versus late recipients and paper check versus direct deposit recipients. Including these sets of time fixed effects in the estimation allows on-time and late recipients to be on different time trends within the paper check and direct deposit categories and implies that data from late recipients cannot be used to estimate the time-disbursement method fixed effects of on-time recipients. However, the inclusion of late recipients in the sample continues to affect the estimated average treatment effects to the extent that their individual treatment effects differ from those of on-time recipients.

When we include late recipients in the sample but separately estimate time fixed effects for all four combinations of on-time versus late and paper check versus direct deposit recipients, the estimated expenditure responses are large and very similar to our baseline estimates. This is shown in Panel C of Table C2 column (3). The estimates are also large when analogous fixed effects are included in ORW's baseline specification (column (3) of Panel A) or in their implementation of the BJS imputation estimator (column (3) of Panel B). Hence, the inclusion of late recipients in the control group drives the main difference in our estimates relative to ORW. Specifically, late recipients appear to be on a different time trend than on-time recipients.

C.3 (Pre-)Trend Tests

We next implement a variety of tests to check whether a different expenditure trend of late recipients is statistically detectable in the data. These tests are based on regressions estimated on the sample of untreated observations. Assumptions 1 (parallel trends) and 2 (no anticipation effects) imply that for all observations it in the set of untreated observations Ω_0, the (conditional) mean of expenditures is

$$E\left[x_{it}\right] = \alpha_i + \delta_{tm(i)}. \tag{C1}$$

We therefore estimate specifications of the form

$$x_{it} = \alpha_i + \delta_{tm(i)} + W_{it}'\gamma + \varepsilon_{it}, \tag{C2}$$

where W_{it} is a vector of variables specifying departures from the conditional mean (C1). For instance, W_{it} could include a set of indicator functions that switch on $1, ..., k$ periods before treatment. Alternatively, W_{it} could be a linear time trend specific to observations that receive their stimulus payment late, and so forth. Regardless of the specification of W_{it} the null hypothesis holds that $\gamma = 0$. For more details on pre-trend tests in this estimation framework, see Borusyak, Jaravel, and Spiess (2024).

Table C3 shows the results for different choices of W_{it}. Specifications (1) and (2) report estimates of pre-trends – coefficients on indicators that switch on one, two, and three periods prior to treatment. As a benchmark, specification (1) reports estimates from our baseline sample, which excludes late recipients. (These estimates coincide with those plotted in Figure 3). Specification (2) is estimated on a sample that includes late recipients. The estimated pre-trend coefficients one and two periods prior to treatment become more negative, but remain statistically insignificant. Note that this negative pre-trend is consistent with the "rebate reporting bias" identified by ORW. Specification (3) estimates pre-trend coefficients that are specific to late recipients only. According to these estimates late recipients have 539.8$ greater expenditures than on-time recipients one month prior to treatment. Further, they have 345.8$ and 596.7$ greater expenditures two and three months before treatment. While none of the estimates is statistically significant, these estimates are large – roughly twice the size of the estimated monthly treatment effects – and raise questions about whether it is appropriate to include late recipients in the control group. In specification (4) we estimate the coefficient on a linear time trend that is specific to late recipients. The estimate is 90.6, implying that expenditures of late recipients grow by 90.6$ per month relative to on-time recipients, which is substantial. This estimate is also significant at the 10 percent level. In specification (5) we estimate pre-trend coefficients separately for late direct deposit and late paper check recipients. Relative to on-time direct deposit recipients late direct deposit recipients have much higher expenditures one, two, and three periods before treatment. These estimates are so large that they are also statistically significant at conventional levels ($p = 0.021$, $p = 0.029$, and $p = 0.055$ for one, two, and three months before treatment), despite the small number of late recipients. Late paper check recipients have lower expenditures than their on-time counterparts. While also large in absolute value, the number of late paper check recipients is too small to estimate these coefficients precisely. Lastly, specification (6) estimates separate linear time trends for late direct deposit and late paper check recipients. The estimate for late direct deposit recipients is large and statistically significant at the five percent level. It implies that late direct deposit recipients increase consumption by 136.8$ per month when compared to on-time direct deposit recipients. By contrast, the estimated trend for late paper check recipients is small in absolute value and statistically insignificant.

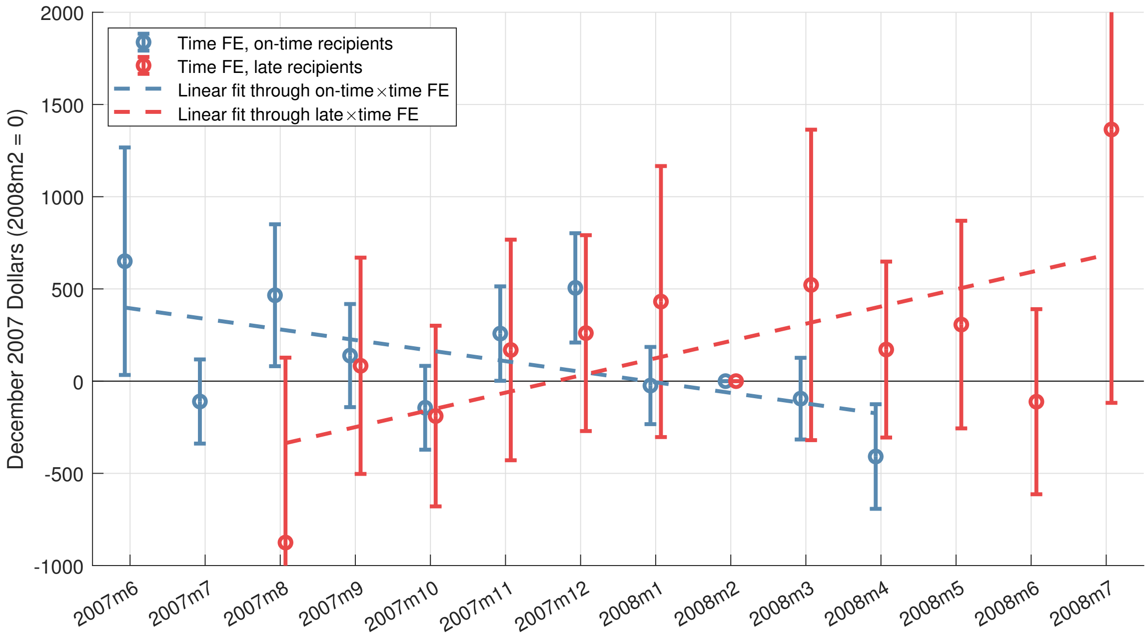
Table C3: (Pre-)Trend Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$1\{K_{it} = -1\}$	-82.1	-157.0					-36.9
	(198.7)	(154.2)					(190.6)
$1\{K_{it} = -2\}$	-18.2	-63.9					13.0
	(144.6)	(120.1)					(139.4)
$1\{K_{it} = -3\}$	-37.8	-34.0					2.4
	(109.4)	(97.4)					(106.3)
$1\{K_{it} = -1\} \cdot 1\{\text{Late SP}_i\}$			554.2				
			(495.4)				
$1\{K_{it} = -2\} \cdot 1\{\text{Late SP}_i\}$			370.6				
			(339.5)				
$1\{K_{it} = -3\} \cdot 1\{\text{Late SP}_i\}$			599.8				
			(395.8)				
Linear monthly trend $\cdot\, 1\{\text{Late SP}_i\}$				90.6			
				(50.5)			
$1\{K_{it} = -1\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Direct Deposit}_i\}$					1155.3		
					(500.3)		
$1\{K_{it} = -2\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Direct Deposit}_i\}$					705.1		
					(323.0)		
$1\{K_{it} = -3\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Direct Deposit}_i\}$					963.5		
					(502.3)		
$1\{K_{it} = -1\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Paper Check}_i\}$					-855.6		
					(1193.6)		
$1\{K_{it} = -2\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Paper Check}_i\}$					-573.8		
					(892.9)		
$1\{K_{it} = -3\} \cdot 1\{\text{Late SP}_i\} \cdot 1\{\text{Paper Check}_i\}$					-320.6		
					(646.4)		
Linear monthly trend $\cdot\, 1\{\text{Late SP}_i\} \cdot 1\{\text{Direct Deposit}_i\}$						136.8	
						(64.3)	
Linear monthly trend $\cdot\, 1\{\text{Late SP}_i\} \cdot 1\{\text{Paper Check}_i\}$						-17.6	
						(73.7)	
Include late recipients	N	Y	Y	Y	Y	Y	Y
Household FE	Y	Y	Y	Y	Y	Y	Y
Month × Disbursement method FE	Y	Y	Y	Y	Y	Y	N
Month × Disbursement method × On-time/late FE	N	N	N	N	N	N	Y
Observations	17,881	19,027	19,027	19,027	19,027	19,027	19,027

Notes: This table presents the results of several (pre-) trend tests, which are implemented by estimating versions of equation (C2) in the sample of untreated observations. The dependent variable is expenditures on durable goods that are measured at the monthly frequency and expressed in 2007m12 dollars. Standard errors that are clustered at the household level and reported in parentheses.

Our interpretation of these test results is that on-time and late direct deposit recipients are on dramatically different expenditure trends. Specifically, durables expenditures of late direct deposit recipients are increasing over time relative to those of on-time direct deposit recipients. This point is visualized in Figure C1. Including late direct deposit recipients in the sample without allowing these sets of households to be on a different time trend violates the parallel trends assumption (Assumption 1). We therefore prefer to either (i) exclude late

Figure C1: Trends in Expenditures of on-time versus late Direct Deposit Recipients



Notes: The figure visualizes trends in expenditures of direct deposit recipients until 2008m7, the last month for which the time fixed effects are used to impute the counterfactual spending path of on-time recipients. The circles are time fixed effects specific to on-time and late recipients and estimated in the sample of untreated direct deposit recipients. The excluded base month is 2008m2 (the last month before the first treatment begins). The specification also includes household fixed effects. The vertical bars indicate 90 percent confidence bands based on standard errors that are clustered at the household level. The dashed lines are fitted through the time fixed effects of on-time or late recipients.

recipients from the sample as we do in our baseline analysis or (ii) to include them but to also include separate sets of time fixed effects for late direct deposit and late paper check recipients. A pre-trend test for this latter case is shown in specification (7) of Table C3. All pre-trend coefficients are estimated to be very close to zero. Further, the estimated 3-month expenditure response is 751.9 as shown in specification (3) of Panel C in Table C2, which is very similar to our baseline estimates.

We note that (pre-)trends may differ in ORW's sample because the dependent variable differs. When we estimate their specifications on their original samples but include separate sets of time fixed effects for on-time direct deposit recipients, on-time paper check recipients, late direct deposit recipients, and late paper check recipients, the estimated expenditure responses remain large (see specifications (3) in Panels A and B of Table C2). This suggests that differing trends between these groups may also lead to violations of the identification assumption in ORW's analysis.