When Stimulus Shows Up:

Measurement and the Micro-Macro Consumption Puzzle

Edmund Crawley, William L. Gamber, and William Goodwin*

Federal Reserve Board

November 18, 2026

Abstract

We assess whether aggregate spending following major U.S. fiscal stimulus events is consistent with micro-based MPC estimates. While existing work interprets 2001 and 2008 data as implying that micro MPCs are too large, we argue that the associated fiscal stimulus was too small relative to ordinary spending volatility to clearly distinguish large from small MPCs. However, the 2021 stimulus was substantially larger, and the response shows up clearly in the aggregate data. High-frequency data confirm the timing and incidence of the spending response, and the MPC path implied by the aggregate dynamics sits comfortably within the range of existing micro estimates: in short, there is no micro-macro consumption puzzle.

Keywords: marginal propensity to consume, fiscal stimulus, retail sales, high-frequency spending data

JEL Codes: E21, E62, H31

*Contact information: edmund.s.crawley@frb.gov, will.gamber@frb.gov, and liam.m.goodwin@frb.gov. We thank Jacob Orchard for helpful comments and discussions. The views expressed in this paper are solely those of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System.

1. Introduction

The marginal propensity to consume (MPC) out of a one-time income shock is among the most intensively measured quantities in macroeconomics. The best micro evidence indicates that the spending response is sizable and front-loaded: Graham and McDowall (2026) estimate that households spend roughly 10 percent of a payment on nondurable goods within a week of receipt, with the cumulative MPC reaching 19 percent after one month and 25 percent after three.[1] Micro-based consumption estimates now anchor a substantial body of quantitative macroeconomics: heterogeneous-agent models are calibrated to reproduce them (Kaplan and Violante, 2014, 2022), and the intertemporal MPC paths they imply serve as sufficient statistics for the output effects of fiscal policy (Auclert, Rognlie and Straub, 2024).

However, existing analysis of the aggregate data in the United States following the 2001 and 2008 federal stimulus programs casts doubt on these measurements. Neither program left a clear footprint in monthly spending: the disbursement months look much like the months around them, and aggregate time-series regressions find no detectable effect of the rebates on consumption (Taylor, 2009). More recently, Orchard, Ramey and Wieland (2025*a*) and Orchard, Ramey and Wieland (2025*b*) have reinforced the skeptical case, arguing that if the micro MPCs were correct, aggregate consumption in the absence of the 2001 and 2008 rebates would have followed implausible paths. They conclude that the true MPCs out of the rebates must have been considerably smaller than the household studies suggest. We label this apparent tension between the household evidence and the aggregate data the "micro-macro consumption puzzle."

This paper argues that the micro-macro consumption puzzle largely reflects measurement and statistical power: the transfers in 2001 and 2008 were small relative to the volatility of aggregate spending, which renders the 2001 episode uninformative and the 2008 episode only weakly informative. The 2021 episode does not face the same limitations, for three reasons. First, the third round of Economic Impact Payments (EIP3) under the American Rescue Plan Act, totaling roughly $410 billion, was far larger: more than four times the 2001 rebate and more than twice the 2008 rebate as a share of disposable personal income (DPI). Second, it was disbursed much faster, with the majority of the total arriving within the first few days of the program. Third, new high-frequency spending datasets let us measure the timing and incidence of the response with a precision not available for the earlier episodes.

Following the logic of Orchard, Ramey and Wieland (2025*a*), we ask for each stimulus episode whether realized spending departed from the counterfactual path it would have followed absent the transfer. The answer is informative only if the induced spending is large relative to the ordinary uncertainty around that

[1] We view the Graham and McDowall (2026) estimates as representative of a long literature estimating the MPC. See Crawley and Theloudis (2024) for a recent review of estimates of the MPC.

counterfactual path. We show that the 2001 rebate does not meet this informativeness condition in monthly retail sales data: using an MPC path derived from the Graham and McDowall (2026) micro-estimates, the induced spending is comparable in size to the ordinary month-to-month variation in the series. The absence of a visible response thus does not provide meaningful evidence against a substantial consumption response to the stimulus. The 2008 rebate, roughly twice as large as a share of DPI, offers somewhat more power, and in contrast to previous estimates we find directional evidence of a response in nominal retail sales. The aggregate data are consistent with the micro evidence in suggesting that the 2008 rebate stimulated consumer spending, implying a three-month cumulative MPC of approximately 25 percent, albeit with uncertainty that limits inference. Following EIP3 in 2021, the aggregate data instead speak with clarity in favor of the micro estimates: retail sales jump far outside the range of paths consistent with their ordinary behavior in the month the payments arrive, and the implied cumulative MPC is close to 20 percent at the three-month horizon. Because retail sales cover only goods and food services, that figure is likely a lower bound on the MPC out of total consumption. When the stimulus is large enough to show up in aggregate data, it shows up about where and when the micro estimates say it should, if somewhat less front-loaded.

We begin by applying a consistent methodology across each of the 2001, 2008, and 2021 episodes. Our benchmark spending series is Census retail sales excluding motor vehicles, a monthly survey measure available since 1992 that records sales directly. For each episode we first construct a "macro counterfactual": we model aggregate spending with an ARIMA process estimated on data outside the stimulus windows, then use the estimated model and a Kalman filter/smoother to interpolate through each window. The macro counterfactual is the expected path of retail sales over the stimulus months given the typical behavior of the series, and its confidence bands capture the plausible distribution of sales around that expectation. The divergence of realized retail sales from these confidence bands is indicative of whether the fiscal transfers converted into aggregate spending over the window. In 2001, realized sales never leave the macro confidence bands; given the rebate's small size, we would not expect them to. In the summer of 2008, realized retail sales rise modestly above the confidence bands, and in March 2021 they jump far beyond them.

We next construct a "micro counterfactual," subtracting from realized sales the induced spending implied by the Graham and McDowall (2026) profile convolved with each program's disbursement schedule. Comparing the pair of counterfactuals enables us to assess their consistency: if the micro counterfactual falls visibly below the range of plausible paths for the macro counterfactual, then the data suggest that the micro MPC estimates are too large. If instead the counterfactuals are statistically indistinguishable, then we conclude that the micro-based estimates of the MPC are consistent with the macro data. Across all three stimulus episodes, the micro counterfactual stays largely within the confidence bands, resembling the macro counterfactual. The first exception is September 2001, when the micro counterfactual falls below the bands;

we argue that this reflects the economic response to the September 11 attacks and is uninformative about the effect of the transfers. The second exception is at the start of the 2021 window, where we interpret the low micro counterfactual as indicating a slower initial spend-out than implied by the Graham and McDowall (2026) profile, perhaps reflecting changes in the spending patterns of households that entered the spring of 2021 with unusually large liquid savings (Parker et al., 2022). Section 3.8 and Appendix H show that the same general conclusions emerge using the data and forecasting framework of Orchard, Ramey and Wieland (2025*a*) and Orchard, Ramey and Wieland (2025*b*) once forecast uncertainty is accounted for.

While spending clearly jumped in March 2021, its attribution to EIP3 is not a given: the stimulus was disbursed into an economy that was already in the process of reopening in the wake of the COVID-19 pandemic. Vaccinations accelerated through the spring of 2021, and federal guidance on public activity loosened over the same months in which the payments went out: in principle, the surge in spending could merely reflect accelerated recovery. In addressing this identification problem, we leverage daily data from Numerator's receipt-based household panel and Fiserv's aggregate card transactions, joining the growing body of work that measures spending with private transaction data (Aladangady et al., 2019; Chetty et al., 2024). These data play two roles in our analysis. First, they verify our measurement of retail sales: neither dataset shares a sampling frame with the Census survey, yet both closely reproduce the March 2021 spending jump in timing and in magnitude. Second, they sharpen the timing: we re-aggregate daily spending into "shifted months" that run from the 12th of one month to the 11th of the next, so that the first shifted month after EIP3 disbursement began on March 12 (March 12–April 11) excludes spending from before the payments. Measured this way, spending peaks precisely in the disbursement window and steps down over the following months. As discussed in Appendix D, reopening proceeded gradually through the spring and summer. It can explain the gradual rise in spending that our macro counterfactual captures, but not the sharp spike and subsequent step-down around the payment dates.

The household-level detail in the Numerator panel lets us examine the reopening confounder in the cross-section through two placebo tests. Under the assumption that the reopening drove the spending surge, payment eligibility would not correlate with increases in individual spending: reopening presumably affected payment-eligible and ineligible households similarly, so spending across the two groups would have risen together. Vaccination status, by contrast, should predict spending: because public-health guidance loosened first for vaccinated individuals, vaccinated households would have led the reopening. On both counts, the data show the opposite. Spending growth among payment-eligible households diverged from that of ineligible households by about 10 percentage points in the month with the largest payment batches for EIP3. A gap of that size is unmatched elsewhere in six years of monthly data, excepting April 2020, during which the vast majority of the payments (EIP1) associated with the Coronavirus Aid, Relief, and

Economic Security (CARES) Act were distributed. At the same time, a split by vaccination status exhibits no significant divergence in the payment months. The Numerator cross-section thus strongly supports the stimulus payments, not the reopening, as the primary factor determining the timing and incidence of the aggregate spending response.

Having found strong evidence that 2021 contains the necessary signal, we conclude by asking what MPC path is most consistent with the macro data. Inverting our convolution procedure, we estimate by generalized least squares the daily cumulative MPC profile that best fits excess realized spending relative to the macro counterfactual. In the 2021 episode, we acquire fairly precise measures of the cumulative MPC at short horizons, with the Census, Numerator, and Fiserv data yielding estimates ranging from 11 to 12 percent at one month, 12 to 23 percent at three months, and 12 to 28 percent at six months. The 2008 episode delivers a similar path with higher uncertainty, reflecting the smaller stimulus volume relative to retail sales volatility. Across episodes, the Census and Fiserv data estimate paths that are somewhat less front-loaded but otherwise largely consistent with the Graham and McDowall (2026) profile, while the Numerator data struggle to pick up a response outside of the first shifted month. These macro-based estimates fall within the wide range of household-level estimates of spending out of the pandemic-era payments (Parker et al., 2022; Karger and Rajan, 2021; Baker et al., 2023; Coibion, Gorodnichenko and Weber, 2020).

The remainder of the paper proceeds as follows. Section 2 describes the three datasets and our benchmarking procedure. Section 3 presents our counterfactual analysis on Census retail sales for all three episodes, and discusses the relation of our work to Orchard, Ramey and Wieland (2025*a*) and Orchard, Ramey and Wieland (2025*b*). Section 4 turns to the daily data and the timing of the 2021 response, while Section 5 presents the cross-sectional evidence on transfer eligibility and vaccination. Section 6 presents our macro-data-implied MPC paths, and Section 7 concludes.

2. Data

We measure household spending in three datasets. Our baseline analysis is carried out using aggregate retail sales data from the United States Census Bureau. These data start in 1992, covering the three stimulus periods we study. For the spring 2021 stimulus period, we also use Numerator's Total Commerce Panel and credit and debit card transaction data from Fiserv. As described below, these daily datasets allow us to more cleanly identify the timing of spending, and thus distinguish the stimulus response from other concurrent events, in particular the reopening of the economy after the pandemic. Numerator's panel also contains cross-sectional variation, which we use in Section 5 to test whether the spending response we observe in March 2021 was driven by stimulus or the reopening economy.

2.1. Census retail sales

Our primary aggregate series is nominal retail sales, excluding motor vehicles and parts, plus food services from the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS). The MRTS is a stratified survey of retail and food services firms that reports monthly retailer revenue by NAICS sector. The benchmarking exercise of Section 2.4 uses the non-seasonally adjusted sector-level series; our counterfactual exercises in Section 3 use the published seasonally adjusted aggregate retail sales series excluding motor vehicle and parts dealers. The aggregate series begins in January 1992, when Census transitioned to NAICS-based industry reporting, providing roughly three decades of data for estimating the ARIMA model that underlies our counterfactual procedure.

We exclude motor vehicle and parts dealers throughout the analysis. Doing so aligns the Census aggregate with the Numerator and Fiserv datasets described below, neither of which captures motor vehicle dealer transactions. Our retail sales aggregate therefore consists of eleven non-motor-vehicle NAICS subsectors of retail trade (NAICS 442–454) together with food services and drinking places (NAICS 722).

2.2. Numerator Total Commerce Panel

Numerator's Total Commerce Panel is a longitudinal transaction-level household panel covering retail spending.[2] Panelists install a mobile application (Receipt Hog) through which they log purchases by photographing paper receipts or by linking their loyalty and online shopping accounts. They can also grant Receipt Hog access to their email to automatically import receipts. Panelists are occasionally asked demographic questions in the app. They are incentivized to participate with gift cards (Numerator, 2025). We use Numerator's "static panel," a subset of 100,000–200,000 households with continuous and sufficiently complete transaction history over a 12-month period. Numerator selects these users from a larger sample of over 1 million users to be demographically and nationally representative.

Each transaction is associated with a banner-level retailer identifier and a basket-level dollar amount. We aggregate transactions to national totals using Numerator's projection, which multiplies each basket's pre-tax dollar value by a household-level demographic weight and a national projection factor equal to the ratio of total U.S. households to static panel size (Hacıoğlu Hoke, Feler and Chylak, 2024). We use a crosswalk to assign each retailer identifier to a 3-digit NAICS code. Appendix A reports the precise variable construction, our data quality filters, and the NAICS crosswalk. In addition to transaction-level spending, the dataset includes household characteristics, updated intermittently, including income, marital status, and responses to topical questionnaires (e.g., vaccination status and preferences in early 2021).[3]

[2] We use the January 16, 2026 vintage.

[3] Household-level demographic attributes, including income bracket and marital status, are not consistently refreshed on a monthly

2.3. Fiserv card transactions

The Fiserv card transactions dataset is a daily aggregate measure of retail spending constructed from anonymized credit and debit card transactions processed by Fiserv (formerly First Data Merchant Services), the largest payment processor in the United States by transaction volume. We use the methodology and data infrastructure documented in Aladangady et al. (2019), which describes the construction of daily retail-spending indexes from the underlying transactions, mapped to 3-digit NAICS subsectors via a Merchant Category Code crosswalk developed jointly by staff at the Census Bureau and the Bureau of Economic Analysis. We do not have access to transaction-level microdata; our analysis uses data aggregated to 3-digit NAICS subsectors. The data run from 2010 through January 2025.

The data capture card spending only and so miss cash transactions. Cards account for roughly 30 percent of consumer payments by dollar value in aggregate, with a higher share for retail goods and restaurant spending (Aladangady et al., 2019). The data are constructed to isolate underlying retail activity from movements driven by Fiserv's own business decisions, such as merchants joining or leaving the processing network, or acquisitions of competing platforms. To do this, Aladangady et al. (2019) apply a constant-merchant filter that retains only merchants with stable attachment to the network over rolling fourteen-month windows, and construct daily indexes by rescaling and averaging across these overlapping samples.

2.4. Benchmarking and aggregation

We construct retail-sales-benchmarked monthly series from each of Numerator and Fiserv through three steps: sector-level seasonal adjustment, scaling to Census retail sales, and aggregation across sectors.

For each alternative data source, we compute the average level of expenditures in each 3-digit NAICS subsector over November 2020–October 2021, a 12-month benchmark period centered on the EIP3 disbursement. We then define a scaling factor as the ratio of Census expenditures to the dataset-specific expenditures over this period. These scaling factors are applied at the industry level, allowing the aggregate scaling to vary with the composition of spending. Scaling by sector addresses potential differences in sectoral coverage across these three datasets.

After constructing industry-level series, we aggregate daily data to the monthly frequency and seasonally adjust each series using X-13ARIMA-SEATS (hereafter X-13; Findley et al., 1998), as detailed in Appendix B. Sector-level seasonal adjustment is consistent with the methodology used by the Census. For each source and sector, we apply the sector-specific scaling factor to the seasonally adjusted source series.

basis, which introduces some classification noise into the eligibility-based exercise of Section 5.

The scaled, seasonally adjusted sector-level series are then summed to produce a single source-specific monthly aggregate, directly comparable to Census retail sales.

In Section 4, we introduce an alternative shifted-month aggregation to better align measured spending with the disbursement of the EIP3 stimulus checks, which were rolled out starting on March 12, 2021. For this alternative, we aggregate spending starting on the 12th day of the month and ending on the 11th of the following calendar month. To maintain a consistent seasonal adjustment strategy, we design an alternative X-13 procedure (AltX-13; Appendix B) where the seasonal adjustment process is a function of the number of trading days in the shifted months, rather than in calendar months.

3. Census Retail Sales Counterfactual Analysis

In this section, we test whether Census retail sales during the 2001, 2008, and 2021 stimulus episodes are consistent with MPCs estimated from household-level data. For each episode, we calculate two counterfactuals. The macro counterfactual is the path that we would expect retail sales to follow through the stimulus window, assuming sales behave in the same way as they do in normal times. This macro counterfactual comes with confidence bands that show how far realized retail sales would normally deviate from this path. The micro counterfactual is calculated as realized retail sales minus the spending implied by combining the MPC profile from Graham and McDowall (2026) with the size and timing of the payments. If the micro MPC estimates are consistent with the aggregate data, we would expect the micro counterfactual to lie within the confidence bands around the macro counterfactual.

We first lay out the methodology for the macro and micro counterfactuals, then describe the episodes and report the results for each. At the end of the section, we relate our approach and findings to those of Orchard, Ramey and Wieland (2025*a*) and Orchard, Ramey and Wieland (2025*b*).

3.1. Macro counterfactual

The key idea is to treat data in the periods around large stimulus episodes as missing and fill them in with a model estimated on how retail sales behave in normal times. We call these periods the stimulus windows and denote them by $\mathcal{T}$. Outside these periods, we assume that log nominal retail sales, denoted by y_t, follows an ARIMA(1,1,0) process with drift.[4] Specifically, with y_t observed without measurement error,

$$\Delta y_t = \mu + \phi \Delta y_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2). \tag{1}$$

[4]We model the counterfactual using an ARIMA(1,1,0) process with drift for several reasons. We find that first-differencing achieves stationarity, and that together the autoregressive and drift terms are sufficiently flexible to account for both the secular growth trend in retail sales and short-run dynamics, substantially reducing innovation variance and standard errors relative to a no-drift baseline (see Appendix E.1).

We first estimate μ, ϕ, and σ_η^2 using maximum likelihood on data excluding $\mathcal{T}$ and the pandemic months of March–July 2020.[5] We then ask: for $t \in \mathcal{T}$, what is the distribution of y_t, given what we know about y_t for $t \notin \mathcal{T}$? Given the linear specification for y_t and that the innovations are Gaussian, y_t is conditionally Gaussian, and the conditional means and variances summarize these distributions:

$$y_t \mid \{y_s\}_{s \notin \mathcal{T}} \sim \mathcal{N}\left(y_{t|s\notin\mathcal{T}}, P_{t|s\notin\mathcal{T}}\right).$$

Because the model is Markov in the pair $(y_t, \Delta y_t)$, we use a (slightly non-standard) Kalman filter and smoother to efficiently compute $y_{t|s\notin\mathcal{T}}$ and $P_{t|s\notin\mathcal{T}}$. The algorithm is described in Appendix E.

The path of conditional means is the path we would expect under the assumption that the data generating process did not change in the stimulus window, given the level and growth rate of retail sales just before and after the window. The conditional variances give the confidence bands.[6] The bands also allow for the fact that μ, ϕ, and σ_η^2 are estimated: we draw 500 sets of parameters from their estimated sampling distribution, compute the conditional mean and variance for each, and combine them using the law of total variance. We chose to condition the counterfactual on the data both before *and after* the stimulus window under an assumption that the effect of stimulus programs on the level of household spending was temporary and had largely subsided by the end of the stimulus window. This decision is particularly relevant for the 2021 episode, which coincided with the reopening of the economy and the associated permanent shift up in the level of retail sales.[7]

Ideally, the stimulus windows we choose would not include other unusually large shocks to the economy. However, the 2001 stimulus came just months before 9/11, which led retail sales to decline sharply, and the 2008 stimulus came a few months before Lehman Brothers failed in September. These events contribute to making it harder to distinguish the effect of stimulus payments in the aggregate data.

3.2. Micro counterfactual

The micro counterfactual is realized retail sales minus the spending out of stimulus that the micro MPCs imply, given the size and timing of the payments. We use the nondurable goods MPCs estimated by Graham and McDowall (2026). They identify a daily MPC out of expected tax refunds using household-level transactions data from a large U.S. bank. We use these estimates because i) the nondurable goods expenditure category in that paper lines up fairly well with our chosen retail sales categories, ii) the large

[5] For ease of notation, we omit hats on the estimated ARIMA parameters here and in Appendix E; they are used only in the specification tests of Appendix E.1.

[6] The bands are pointwise. Section 3.7 reports joint tests over each window that account for the correlation of counterfactual errors across months.

[7] Appendix H shows an alternative forecast for 2021 that only conditions on the level of retail spending *before* the stimulus episode, and we find that this alternative likely overstates the size and persistence of the impact of the 2021 stimulus program.

sample size means that sampling error is negligible (Graham and McDowall, 2026), such that we do not need confidence bands around the micro counterfactual, and iii) the daily frequency allows us to match spending to the exact dates that payments arrive.

We construct the micro counterfactual in four steps.

1. **Daily MPC profile**. We define MPC_k to be the MPC on day k after a stimulus payment arrives and fit a double-exponential form, the sum of two exponential decay functions:

$$MPC_k = a_1 \exp(-b_1 k) + a_2 \exp(-b_2 k). \tag{2}$$

   We choose the four parameters in this function to match four targets: an MPC of 4 percent on day 0 and cumulative MPCs of 10, 19, and 25 percent by days 7, 30, and 90, respectively. After day 150, we depart from the form in Equation (2) and assume no further spending.

2. **Payment schedule**. We use data retrieved primarily from U.S. Treasury press releases on the timing, amount, and number of stimulus payments distributed during the 2001, 2008, and 2021 stimulus programs. Our estimates are detailed in Appendix C.

3. **Convolution**. We calculate implied spending on day d as the sum, over arrival days s up to and including d, of payments arriving on day s times MPC_{d-s}:

$$S_d = \sum_{s \leq d} C_s \, MPC_{d-s}, \tag{3}$$

   where C_s is the total value of payments arriving on day s. We aggregate these daily spending effects to a monthly frequency by summing all spending that falls in each month.

4. **Subtract**. The micro counterfactual is then calculated as realized retail sales minus the implied monthly spending from the previous step.

The high frequency of the MPC estimates is important because the timing of the transfer disbursement and spend-out patterns within each month affect the monthly spending response. For example, if the payments arrive only in the second half of a month, a monthly MPC allocates all the first-month spending to that month. This overstates micro-implied spending in that month and could make the micro estimates look inconsistent with the aggregate data.[8]

[8]See Appendix G for a comparison to results using monthly MPC estimates.

3.3. Stimulus episodes

We study three stimulus episodes:

1. **The Economic Growth and Tax Relief Reconciliation Act of 2001**, which provided advance refund tax credits during the recession that began in March 2001. Checks were mailed in batches starting in late July and continued in volume through September, with payments totaling $36.4 billion, about $35 billion of it in ten weekly batches.

2. **The Economic Stimulus Act of 2008**, which similarly offered tax rebates following the onset of the recession at the end of 2007. Payments were disbursed beginning in late April and continued in volume into July, totaling $108 billion.

3. **The American Rescue Plan Act of 2021**, which funded EIP3 to support recovery from the COVID-19 pandemic. These payments began on Friday, March 12, 2021, continued in large volume through early April, and then tapered off through the summer. EIP3 payments totaled $410 billion.

We do not analyze the first round of Economic Impact Payments (EIP1), paid in April 2020. It arrived as spending was collapsing for reasons unrelated to the transfers, at the onset of the pandemic, so its aggregate effect cannot be separated from the lockdown; indeed, we exclude March through July 2020 when estimating the ARIMA process.

Table 1 reports the size of each program. As a share of DPI, EIP3 was more than four times the 2001 rebate and more than twice the 2008 rebate.

Table 1: Stimulus program size relative to income and retail spending

	Total	Percent of		
Program	($ billions)	DPI	Retail sales	GDP
2001 rebates	36	0.5	1.4	0.3
2008 rebates	108	1.0	3.1	0.7
2021 EIP3	410	2.2	7.2	1.7

Notes: Stimulus amounts represent total federal stimulus check disbursements, including payments made outside the weekly batches of our disbursement schedules (Appendix C), about 3, 15, and 4 percent of the 2001, 2008, and 2021 totals. Gross Domestic Product (GDP) and Disposable Personal Income (DPI) are from Federal Reserve Economic Data (FRED), averaged over the respective calendar year. Retail spending (MRTSSM44Y72USS) is the annual sum of monthly values from FRED.

The choice of stimulus window affects the macro counterfactual. Given that the estimates in Graham and McDowall (2026) imply that most of the response of consumption to a transfer occurs in the first few

months, the stimulus window covers each month when payments were distributed in volume and the following three months. In 2008, that window would run through October and include the collapse of Lehman Brothers in September. To avoid conflating the stimulus with the onset of the financial crisis, we end the 2008 window in August.[9] The resulting windows are July–December 2001, April–August 2008, and March–July 2021.

3.4. 2021 episode

Figure 1 shows the results for the 2021 stimulus episode. The black line shows monthly nominal retail spending, excluding motor vehicles, from the U.S. Census. As the figure shows, spending jumped up in March 2021, the month in which EIP3 began, and stayed elevated thereafter. The dashed blue line and shaded region show the macro counterfactual for March through July 2021. Note that the confidence bands widen toward the middle of the stimulus window, where the interpolation is furthest from the observed endpoints.

The stimulus payments are clearly visible beyond the usual noise in the aggregate data. Realized spending in March 2021 lies far above the upper edge of the bands and the jump coincides with the rapid distribution of stimulus checks from mid-March. The macro counterfactual suggests that without EIP3, spending would have risen more gradually over spring 2021, consistent with a gradual reopening of the economy as households received their COVID-19 vaccines.

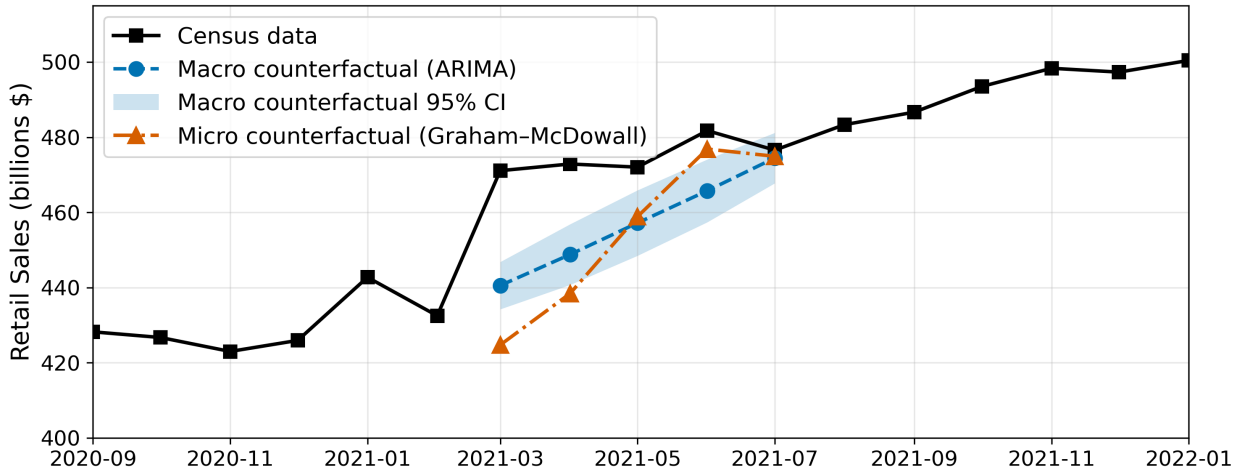
The micro counterfactual also shows a more gradual increase in spending over the spring. This suggests that the MPCs from Graham and McDowall (2026), estimated on microdata, are broadly consistent with the macroeconomic dynamics over the period. That said, it does appear that the Graham and McDowall (2026) MPCs imply a front-loaded response to stimulus income that is not present in the aggregate data: spending in March and April was not quite as elevated as expected using the MPC estimated from microdata, while spending in June was a little higher than expected. This may have been because households at this time held an unusually large quantity of liquid assets (Parker et al., 2022). We return to this discrepancy in Section 6.

3.5. 2008 episode

Turning to 2008: stimulus checks began to go out at the end of April and were gradually disbursed through July. As shown in Figure 2, over this period, nominal retail sales rose persistently, before beginning to decline in August and continuing to fall after the failure of Lehman Brothers in September. On its face this pattern suggests that the stimulus checks may have raised aggregate retail spending.

[9] Ending the window in September instead lowers the counterfactual path of spending and raises the implied MPC (see Appendix F).

Figure 1: Retail sales around EIP3 in 2021



Notes: Monthly nominal retail sales excluding motor vehicles and parts dealers from the U.S. Census Bureau's Monthly Retail Trade Survey, accessed via FRED (series: MRTSSM44Y72USS). The solid black line is realized sales; the dashed blue line and shaded region show the ARIMA(1,1,0) counterfactual and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. Billions of dollars per month, seasonally adjusted.

The macro counterfactual and its confidence bands provide some evidence that the stimulus did increase aggregate spending.[10] Realized spending rises modestly above the 95 percent confidence bands in June and July, months during and just after the checks arrived.

As the dash-dotted orange line shows, the micro counterfactual is close to the macro counterfactual, and it falls within the confidence bands in every month, coming closest to the lower edge in May.[11] The aggregate data are thus consistent with the Graham and McDowall (2026) MPC profile, which reaches a cumulative 25 percent at three months. However, the confidence bands are large and the aggregate data are also consistent with a range of MPCs, including smaller ones than the profile we use.

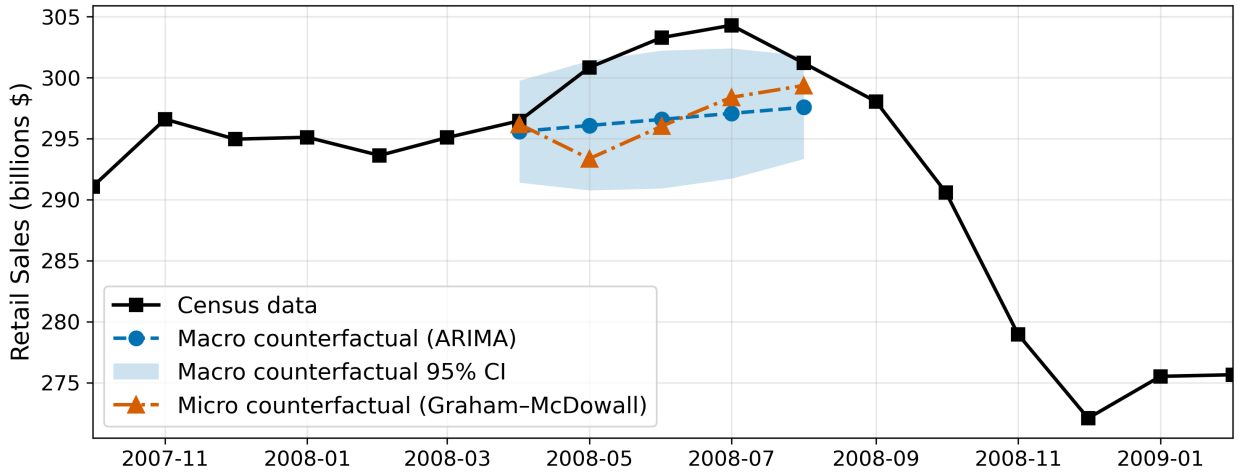
3.6. 2001 episode

The final episode that we analyze is the 2001 Economic Growth and Tax Relief Reconciliation Act, shown in Figure 3. Stimulus checks began to go out in July 2001, and while retail spending rose in August, it fell in September before largely resuming its prior trajectory in October. Realized spending never leaves the 95 percent confidence bands: even though these months included the events of 9/11, the movements in the data were broadly consistent with historical patterns in retail spending.

[10] Our baseline analysis uses a stimulus window from April to August 2008. Appendix F presents sensitivity analysis with alternative stimulus window end dates (April–July and April–September). Ending in July leaves the micro counterfactual inside the confidence bands in every month, but realized spending then exceeds the upper edge of its band only marginally, in July; including September, the month of the Lehman Brothers collapse, implies a larger stimulus effect.

[11] We emphasize that not only the level, but the shape of the micro counterfactual—a fall into the spring then recovery into August—is consistent with macroeconomic shocks over this period. The near-collapse of Bear Stearns in March, falling house prices, the comovement of consumer sentiment and stock prices, and a surge in oil prices that peaked in June and July before falling sharply can all help to explain the counterfactual dynamics (see Appendix Table 7).

Figure 2: Retail sales around the 2008 stimulus program



Notes: Monthly nominal retail sales excluding motor vehicles and parts dealers from the U.S. Census Bureau's Monthly Retail Trade Survey, accessed via FRED (series: MRTSSM44Y72USS). The solid black line is realized sales; the dashed blue line and shaded region show the ARIMA(1,1,0) counterfactual and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. Billions of dollars per month, seasonally adjusted.

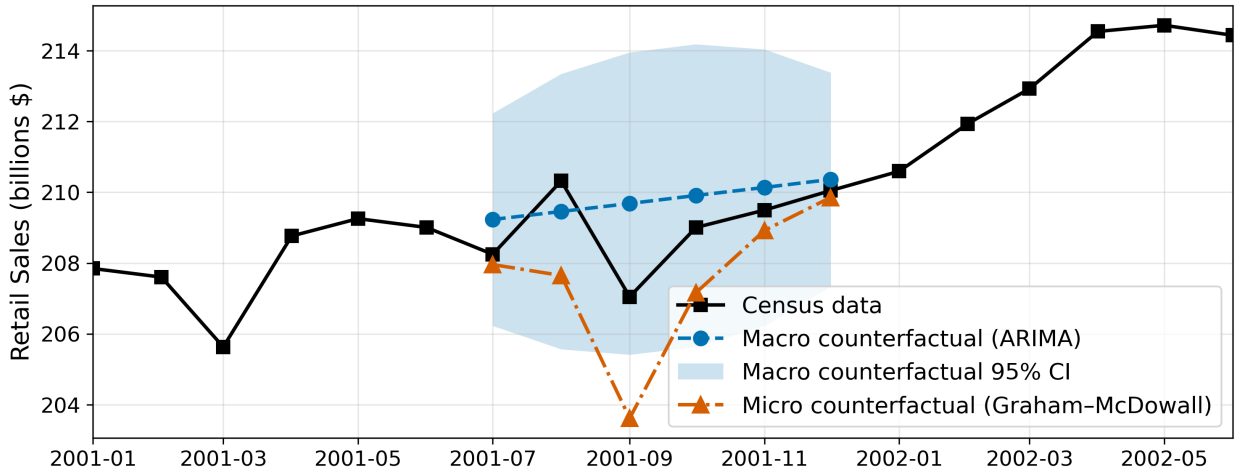
We would not expect a visible response given the relatively small size and slower disbursement of the payments during this episode (total payments were 1.4 percent of annual retail sales; see Table 1). The Graham and McDowall (2026) MPC profile implies monthly spending from the payments peaks at around $3.4 billion, relative to confidence band widths of $6–9 billion, so even under the micro MPCs retail sales would not leave the bands, and the absence of a visible response says little about the size of the MPC.

The micro counterfactual lies inside the confidence bands for every month except September. In September, it is reasonable to think that 9/11 would have caused spending to drop below the confidence bands in the absence of any stimulus payments, so our read is that this data point reflects a unique month in American history rather than evidence against the micro MPCs. Overall, the 2001 episode does not allow us to distinguish between large and small MPCs in the macro data.

3.7. Joint tests over the stimulus windows

The confidence bands in Figures 1–3 are pointwise: each describes the plausible distribution of the counterfactual in a single month. However, if the counterfactual errors are strongly correlated within a given window, then several months above the bands may reflect a single deviation in the level of spending rather than several independent pieces of evidence. We therefore also test, for each episode, whether realized spending over the whole window is consistent with no stimulus effect given the macro counterfactual and taking the joint distribution of counterfactual errors across months into account. We estimate the covariance matrix Σ of the counterfactual across the n months of the window from 5,000 counterfactual paths. For each

Figure 3: Retail sales around the 2001 stimulus program



Notes: Monthly nominal retail sales excluding motor vehicles and parts dealers from the U.S. Census Bureau's Monthly Retail Trade Survey, accessed via FRED (series: MRTSSM44Y72USS). The solid black line is realized sales; the dashed blue line and shaded region show the ARIMA(1,1,0) counterfactual and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. Billions of dollars per month, seasonally adjusted.

path we draw the ARIMA parameters from their estimated sampling distribution and then the innovations over the window from their distribution conditional on the data outside it, as in the MPC bootstrap of Section 6. With X the vector of realized sales minus the macro counterfactual over the window, the test statistic $X'\Sigma^{-1}X$ is approximately distributed χ^2_n under the null that the stimulus had no effect.[12]

Table 2: Joint tests of the macro counterfactual

Episode	Months in window	Months outside 95% bands	Joint test p-value
2001	6	0	0.29
2008	5	2	0.15
2021	5	4	<0.001

Notes: The null hypothesis is that realized retail sales over the stimulus window are a draw from the macro counterfactual, i.e. that the stimulus had no effect. The joint test uses the full covariance of the counterfactual across the months of the window, including innovation and parameter uncertainty. Months outside the 95% bands counts the months in which realized sales lie outside the pointwise bands in Figures 1–3.

Table 2 reports the results. In 2001 and 2008 the tests fail to reject the null, with p-values of 0.29 and 0.15 respectively, implying that the observed path of deviations from the macro counterfactual does not provide meaningful evidence of a spending response to the stimulus. In 2008, this comes despite the fact that realized spending left the confidence bands in both June and July, highlighting that the stimulus event lacks the statistical power to speak conclusively about the consumption response. In 2021 however, the p-value is below 0.001 and the null hypothesis is rejected decisively, confirming a significant spending response.

[12] We find that our p-values are essentially unchanged when computed from the distribution of the statistic across the counterfactual draws rather than from the χ^2_n approximation, which allows for the non-normality introduced by parameter uncertainty.

3.8. Relationship to Orchard, Ramey and Wieland (2025*a*,*b*)

Our paper is closely related to Orchard, Ramey and Wieland (2025*a*) and Orchard, Ramey and Wieland (2025*b*) (henceforth "ORW"), who study the 2008 and 2001 rebates. They argue that the large micro MPCs estimated from these episodes imply implausible paths for aggregate consumption. Like us, they construct a micro counterfactual by subtracting spending implied by a micro-estimated MPC (Parker et al., 2013 for 2008, Johnson, Parker and Souleles, 2006 for 2001) from realized spending, measured as real personal consumption expenditures (PCE) in 2008 and real nondurable spending in 2001. They then judge the plausibility of this micro counterfactual against forecasts from a variety of professional forecasters and a conditional forecast from a monthly vector autoregression with exogenous variables (VARX), which they call the "pessimistic forecast." This pessimistic forecast plays the same role as our macro counterfactual. One difference is that it is solved forward from the months before the rebate, while ours is also anchored to realized spending after the stimulus window.

In 2008, ORW use the gap between the pessimistic forecast and realized spending to bound the stimulus effect of the rebate. We measure the uncertainty around the forecast, as realized spending will ordinarily deviate from the forecast and the size of ordinary deviations is needed to interpret any gap.

Appendix H reproduces their VARX exercises, using their replication code. We add 95 percent confidence bands that include both parameter and shock uncertainty, as do the bands around our own macro counterfactuals. In 2001, both of their counterfactuals lie within these confidence bands, or at their lower edge. In 2008, the counterfactual with an MPC of 30 percent does, the one with an MPC of 50 percent dips below the lower edge, and the one with an MPC of 90 percent lies well below. When we repeat the exercise with retail sales and the Graham and McDowall (2026) MPC profile, the micro counterfactual lies within the bands in 2001 and in every month of 2008 except May, when it is just below the lower edge.

Orchard, Ramey and Wieland (2025*a*) lay out a story of the 2008 episode, motivated by the results of Parker et al. (2013), that is driven by motor vehicles. They estimate that all of the micro MPC comes from motor vehicles and write down a model in which prices for these goods rise in general equilibrium following a stimulus policy because the short-run supply curve is upward-sloping. This effect dampens the aggregate response. Our retail sales aggregate excludes motor vehicles, removing most of this channel, and we read the wider MPC literature as evidence that high MPCs do not only reflect durables (Graham and McDowall, 2026).

We agree with ORW that spending left little footprint in aggregate data in 2001, but, as shown above, a response of the micro-implied size would not be detectable. For 2008, realized retail sales rise modestly above the bands in June and July, but the data are also consistent with a range of MPCs. The 95 percent

confidence bands around the results using the method and data in Orchard, Ramey and Wieland (2025*a*) imply the data cannot rule out an MPC of 30 percent in this period. The 2021 episode, which ORW do not study, was large enough that the response we measure there is strong evidence for an aggregate MPC in line with the micro estimates.

While we find there is no micro-macro consumption puzzle for retail spending, summing our estimated MPC of around 25 percent and the estimated MPC on motor vehicle spending from Orchard, Ramey and Wieland (2025*a*) of about 30 percent implies a total MPC of above 50 percent. Such an MPC would be inconsistent with the dynamics of total PCE, as shown in Orchard, Ramey and Wieland (2025*a*). That said, we restrict our exercise to retail spending, rather than overall PCE, which includes services that are less well measured at a monthly frequency.

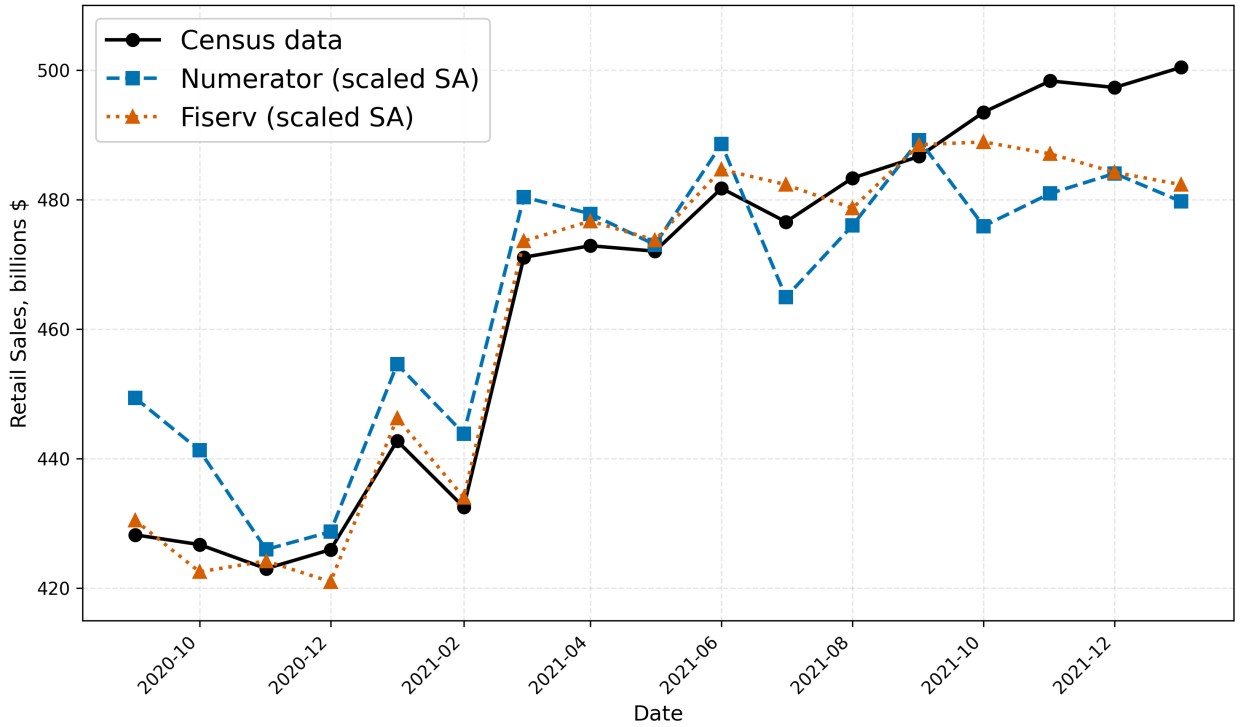
4. Evidence from Daily Spending Data

The analysis up to this point has relied on the Census retail sales series, providing a consistent measure of spending across our stimulus events of interest. We now turn to the Numerator and Fiserv data for further analysis of the 2021 stimulus episode. Together, these datasets offer three things the Census series cannot. First, they provide independent verification: constructed from consumer-reported spending receipts (Numerator) and card transactions (Fiserv), their sampling frames are independent of the Census. Agreement between these series and the Census results is therefore evidence that the measured response reflects actual spending rather than a feature of the Census survey. Second, because both series are at a daily frequency, we can re-aggregate to shifted months before seasonally adjusting, better aligning our data to the timing of the stimulus payments and sharpening our identification. Third, the Numerator panel carries household-level demographic and survey information, which allows us to evaluate the post-COVID economic reopening as a confounder. We present our discussion of this last point in Section 5.

4.1. Cross-source validation

We begin by verifying that the scaled, seasonally adjusted Numerator and Fiserv aggregates constructed in Section 2.4 reproduce the dynamics of Census retail sales. Figure 4 plots the three series at calendar-month frequency over the period around the 2021 stimulus. By construction, the levels of the three series are anchored to one another: industry-level scaling factors equate average spending from each source to that of the Census over the November 2020–October 2021 benchmark period. However, the comovement of the series is not imposed. The relative monthly distribution of spending throughout the period surrounding the stimulus event is determined independently by each source's sample. In addition, because each industry

Figure 4: Census, Numerator, and Fiserv retail sales, scaled and seasonally adjusted

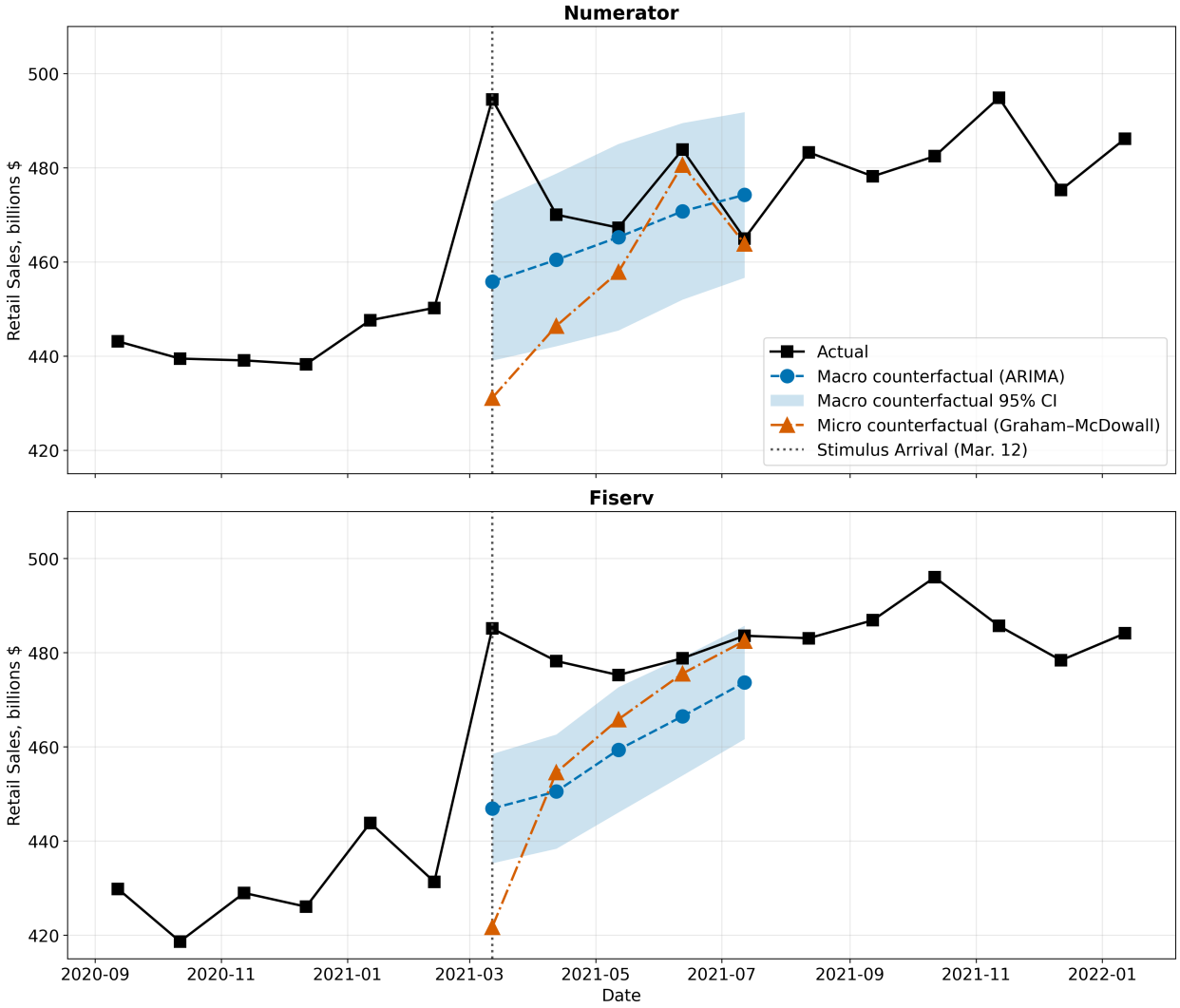


Notes: All three series are scaled to Census retail sales over the November 2020–October 2021 benchmark period; their month-to-month comovement is not imposed by this scaling. Billions of dollars per month, seasonally adjusted.

scaling factor is a constant, changes in the measured composition of spending are reflected in the month-to-month movements of the aggregate series. The close alignment of the three series, and in particular their synchronized jump in March 2021, therefore reflects common movements in the underlying data rather than the benchmarking.

After the recovery from the spring 2020 trough, spending moved roughly sideways through the latter half of 2020, rose in January 2021 as the second round of Economic Impact Payments (EIP2) was disbursed, fell back in February, and then jumped sharply in March with the arrival of EIP3. All three series register this movement as occurring at the same time and with a similar magnitude. The most visible discrepancy is in the amplitude of the month-to-month fluctuations: the Numerator series is noticeably more volatile than either Fiserv or Census, a feature that bears directly on the statistical power of the counterfactual exercise below. Following the jump in March 2021, Census sales continue to drift up gradually. Fiserv tracks Census closely through September; from October the Census series keeps rising while the Fiserv and Numerator series remain roughly flat.

Figure 5: Shifted-month retail spending around the 2021 stimulus: Numerator and Fiserv



Notes: Each panel aggregates daily spending into shifted months running from the 12th of one calendar month to the 11th of the next (e.g., the March observation spans March 12–April 11). Series are scaled to Census retail sales over the November 2020–October 2021 benchmark period and seasonally adjusted with AltX-13 (Appendix B). The solid black line is realized spending; the dashed blue line and shaded region are the ARIMA(1,1,0) counterfactual and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. The vertical dotted line marks the onset of EIP3 disbursement. Top panel: Numerator; bottom panel: Fiserv. Billions of dollars per month, seasonally adjusted.

4.2. The timing of the response

The daily structure of the Numerator and Fiserv data allows us to better align our monthly unit of aggregation with the timing of the EIP3 stimulus. The first and largest batch of the program, worth almost $250 billion, was disbursed during the week of March 12, 2021. As described in Section 2.4, we re-aggregate the Fiserv and Numerator data into shifted months running from the 12th of one calendar month to the 11th of the next, isolating the impact response of the stimulus in our shifted-March observation. This realignment concentrates the impact response in a single observation, raising power, and keeps the tail of the EIP2 spending response out of the shifted-February observation that anchors the start of the counterfactual. Re-estimating our ARIMA process and applying the same Kalman filter/smoother and Graham and McDowall (2026) counterfactual described in Section 3 to the shifted-month series, we interpolate over five shifted months beginning with shifted-March. Figure 5 presents the results of this exercise. In both panels, realized shifted-March spending rises sharply on impact, landing well beyond the upper edge of the confidence band. The response is very large and clearly identified in both sources.

The path of the response is similarly informative. Diverging from the calendar-month series, spending steps noticeably down over the following shifted months after peaking. This pattern is difficult to reconcile with post-COVID reopening as an alternative explanation for the elevated spending in the spring of 2021. Whereas reopening proceeded gradually through the spring and summer as vaccinations increased and public guidance loosened, on its own predicting a fairly smooth upward drift in spending, realized spending spiked as most checks arrived and stepped down in the following months.

Even so, the micro counterfactual derived from Graham and McDowall (2026) initially lies below the 95 percent confidence band of the macro counterfactual. By late spring the ordering for Fiserv reverses, with the micro counterfactual lying near the upper edge of the band. Thus, to the extent the two counterfactuals are in tension, the tension concerns primarily the timing of the response rather than its cumulative magnitude, consistent with the Census evidence of Section 3. The Graham and McDowall (2026) profile appears somewhat too front-loaded and less persistent than necessary to adequately describe the spending response to EIP3. These patterns help to motivate our flexible estimation procedure of Section 6, where we recover the intertemporal MPC profile implied directly by aggregate spending and our macro counterfactual.

5. Demographic Heterogeneity

So far, we have shown that there is evidence in aggregated data that the 2021 stimulus checks led household spending to rise. We now exploit the cross-sectional heterogeneity in the Numerator panel as another test of whether the observed increase in household spending can be attributed to the stimulus program.

The primary alternative hypothesis that we want to consider is that the surge in spending in March reflected reopening and the accelerating vaccination campaign rather than the transfers themselves. If that were the case, we would expect spending to increase in all households, regardless of whether they were eligible for a stimulus check. Similarly, we would expect vaccination status to predict the size of the increase in spending.

We now present a comparison of per-household spending for different groups in the Numerator data. For each group we form a macro counterfactual, as in Section 4, and ask whether realized spending departs from it. Each figure in this section pairs this counterfactual (top panel) with the difference between the two groups in month-over-month spending growth over the full sample (bottom panel). In the bottom panel we highlight the EIP disbursement months and compare them against the mean and standard deviation of the difference in all other months.

5.1. Stimulus check eligibility

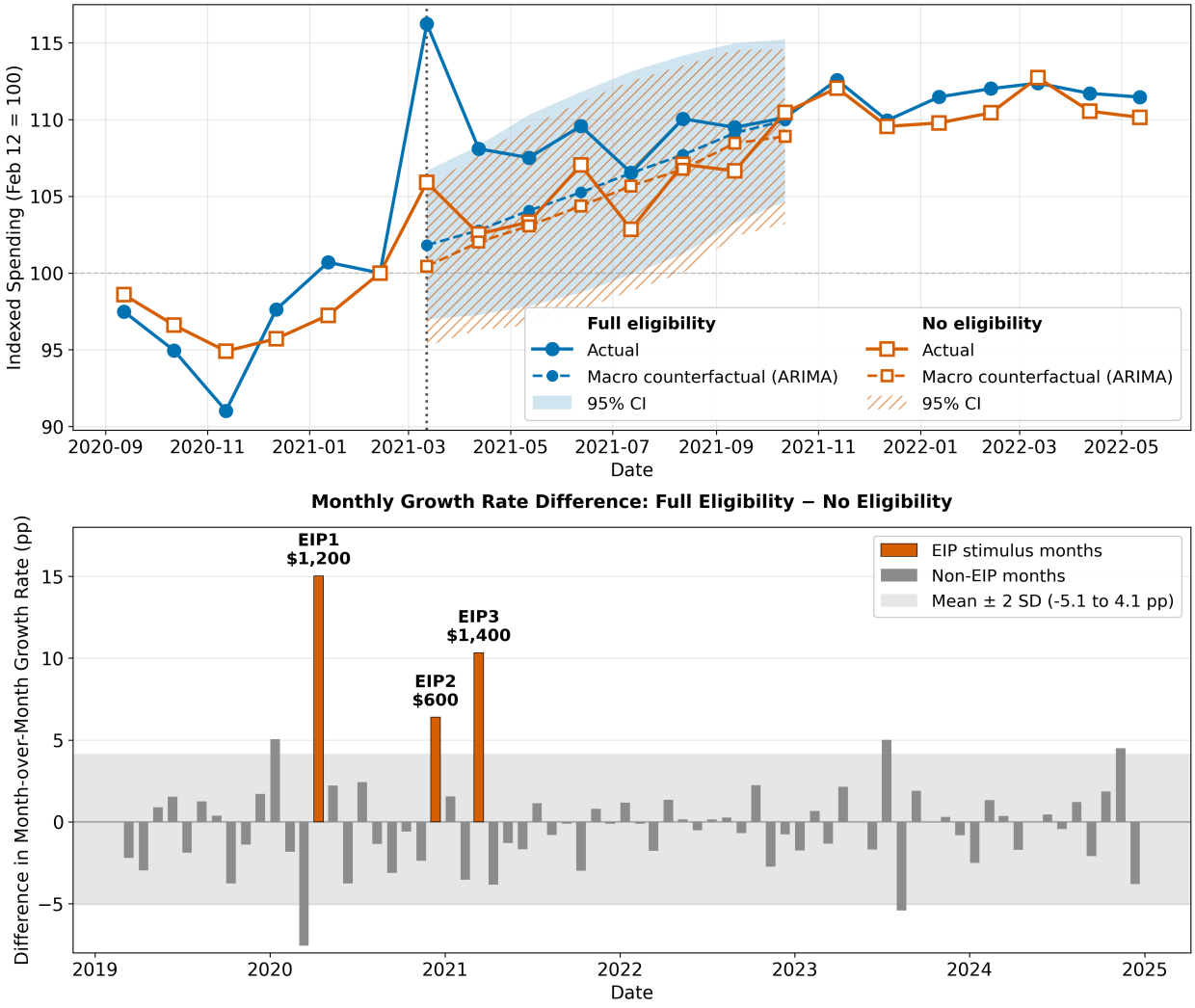
We classify panelists by their eligibility for the third round of Economic Impact Payments using the household income bracket and marital status reported in December 2020.[13] We define a *Full eligibility* group of households entitled to the full payment, corresponding to household income below $70,000, and a *No eligibility* group of ineligible households, corresponding to married households with income of $175,000 or more together with unmarried individuals with income of $125,000 or more. These cut-offs are edges of Numerator's income brackets. We set aside households in the phase-out region and those whose filing status cannot be determined in our data.[14]

Figure 6 reports the result. The top panel shows that spending in the *Full eligibility* group jumped sharply in March 2021, rising far above the upper edge of its counterfactual band. Spending then fell back, staying above the counterfactual path through the summer but inside its confidence band. The *No eligibility* group also rose in March, but by considerably less, rising to about the level of the upper edge of its confidence band and staying inside it afterward. That said, the increase in spending among the ineligible group could suggest that some portion of the rise in aggregate spending in March was due to factors unrelated to the transfer payments, such as reopening. Alternatively, because Numerator uses self-reported income and marital status that are not refreshed on a monthly basis, some households that in fact received a payment may be misclassified as ineligible. Either way, this comparison rejects the hypothesis that all of the March increase was due to reopening.

[13] EIP3 eligibility was determined by adjusted gross income (AGI) on the 2020 tax return, or on the 2019 return if a 2020 return had not been processed. For citizens and resident aliens, these two variables are effectively sufficient to determine eligibility under the rules summarized in Appendix C.

[14] The thresholds of the first two rounds differed but reached broadly the same households, and we apply the EIP3-based classification throughout the sample.

Figure 6: Per-household spending by EIP3 eligibility



Notes: The top panel plots per-household retail spending for the *Full* and *No* eligibility groups, each indexed to shifted-February 2021 (beginning February 12) = 100. Solid lines are realized spending; dashed lines are the group-specific ARIMA(1,1,0) counterfactuals, and the shaded (Full) and hatched (No) regions are their 95 percent confidence bands, interpolated over the eight shifted months from March through October 2021. The vertical dotted line marks the onset of EIP3 disbursement on March 12, 2021. The bottom panel plots the difference between the two groups in month-over-month spending growth over the full sample; bars for the three EIP disbursement months are highlighted and labeled, and the shaded band is the mean ± two standard deviations of the difference across all other months. Growth differences in percentage points.

The lower panel plots the difference in month-over-month spending growth between these two groups from 2019 through 2024. The largest differences coincide with the disbursement of stimulus checks: in the months of EIP1 and EIP3, in which households received $1,200 and $1,400 payments, respectively. The month of EIP2, the smallest of the three payments at $600, shows a more modest positive difference that remains within the two-standard-deviation band.

5.2. Vaccination

We next analyze whether vaccination status, as reported in a topical questionnaire fielded in early 2021, can account for the rise in spending in March. We compare individuals vaccinated before February 2021 with those not yet vaccinated at that date but who reported intending to be. If vaccination and a reopening of the economy drove the spring surge, we should expect spending among households who were already vaccinated to pull away from those still waiting to be vaccinated.

The top panel of Figure 7 shows that spending rose similarly in both groups. The lower panel shows that the difference in spending growth between the vaccinated and the not-yet-vaccinated was small and well inside the two-standard-deviation band in March 2021. Vaccination status thus does not appear to predict spending in the months of the third stimulus check.[15]

Taken together, these results show that the March surge was concentrated among eligible households and unrelated to vaccination status, pointing to the payments rather than reopening as its source.

6. MPC Estimation

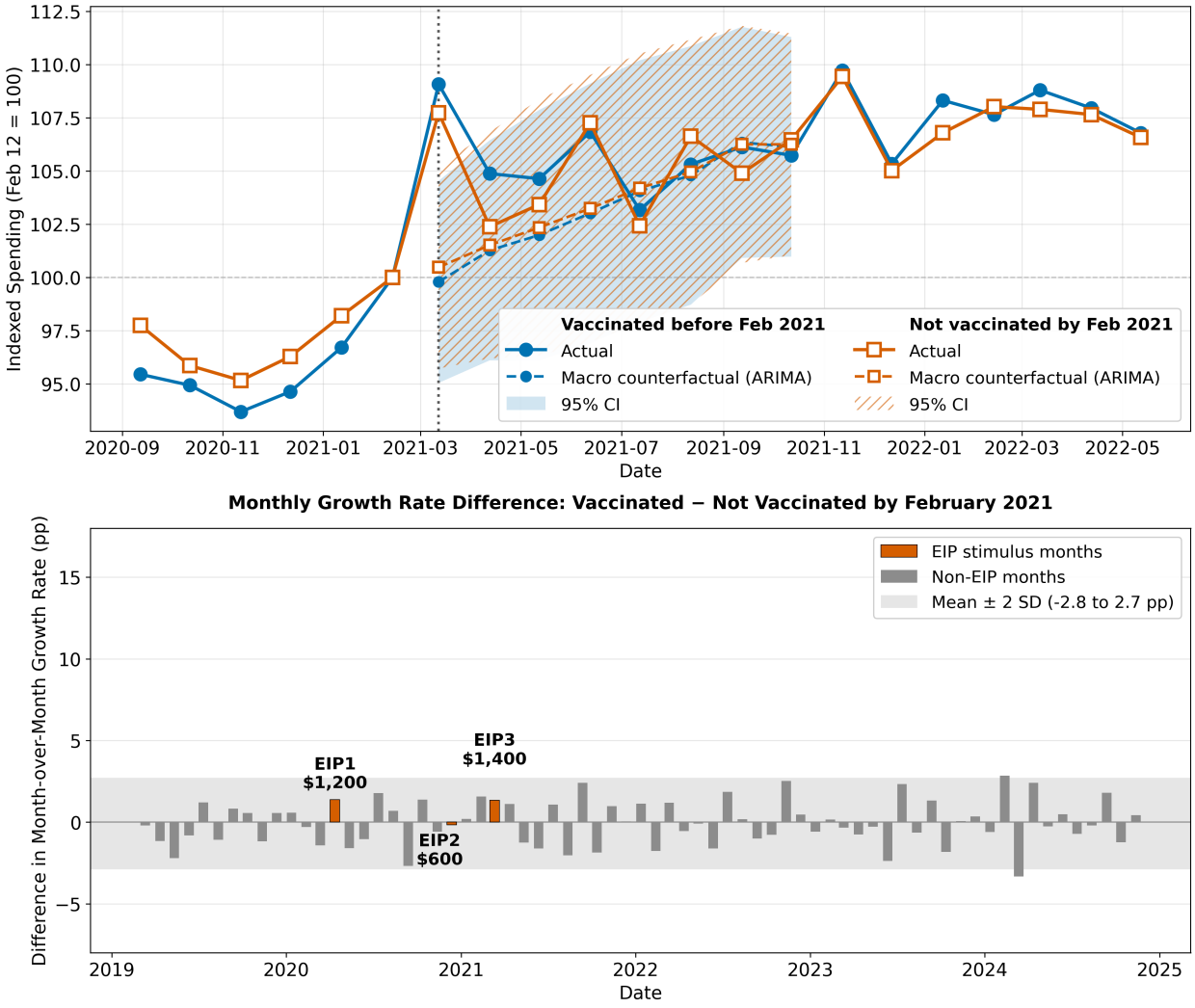
In Section 3 we tested whether the aggregate data are consistent with a particular MPC profile. In this section we instead determine the MPC profile that best fits the aggregate data. We estimate it for the 2008 and 2021 episodes, the two episodes that provide directional evidence consistent with a spending response.

6.1. Methodology

We estimate the MPC profile in two steps. First, we compute excess spending in each month t of the stimulus window as $X_t = Y_t - \hat{Y}_t$, where Y_t is observed retail sales and $\hat{Y}_t$ is the macro counterfactual from Section 3. Second, we fit the four parameters of the double-exponential form for the daily MPC that we defined in Equation (2) to match this excess spending. Specifically, we choose the parameters by generalized least squares: they minimize $(X - S)'\Sigma^{-1}(X - S)$, where X stacks excess spending over the months of the stimulus window, S stacks the monthly spending implied by the parameters and the payment schedule

[15] The March increase was in fact largest among respondents who reported no plans to be vaccinated (not shown).

Figure 7: Per-household spending by vaccination status



Notes: As in Figure 6, for panelists grouped by vaccination status as reported in an early-2021 questionnaire: those vaccinated before February 2021 and those not yet vaccinated at that date who reported intending to be. The sample is restricted to questionnaire respondents and is smaller than the full panel. Top panel: per-household spending indexed to shifted-February 2021 = 100, with group-specific macro counterfactuals and 95 percent confidence bands (shaded for the vaccinated group, hatched for the not-yet-vaccinated group) interpolated over the eight shifted months from March through October 2021. Bottom panel: difference in month-over-month spending growth between the two groups, with EIP disbursement months highlighted against the mean ± two standard deviations of all other months. Growth differences in percentage points.

through Equation (3), and Σ is the covariance of the macro counterfactual across those months (Section 3.7). Weighting by Σ^{-1} rather than by each month's own variance accounts for the correlation of counterfactual errors across months. It gives less weight to smooth, sustained deviations of the kind the counterfactual itself often produces, and more weight to the month-to-month shape of excess spending.

We compute standard errors by bootstrap. In each of 5,000 draws, we first draw the parameters of the ARIMA process from their estimated sampling distribution (Section 3.1). We then draw a path for $\hat{Y}_t$ across the months of the stimulus window from its distribution given those parameters and the data outside the window.[16] Counterfactual errors are strongly correlated across months, which widens the uncertainty around cumulative MPCs at longer horizons. Because each draw is a full correlated path, the bootstrap standard errors account for this correlation. For each draw we repeat the second step, holding Σ fixed, and report the standard deviation of the cumulative MPC at one, three, and six months.

6.2. Results

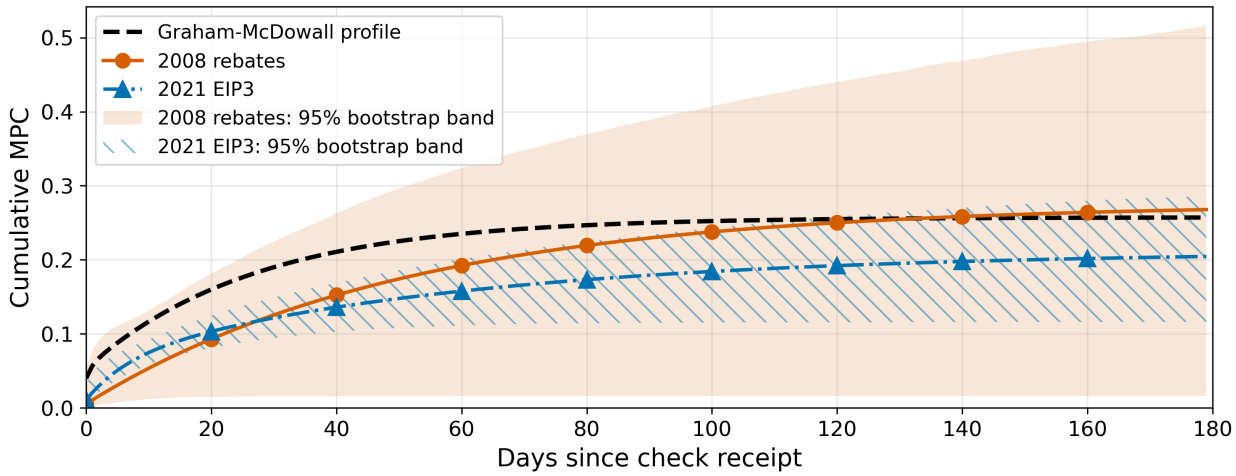
Figure 8 shows the daily cumulative MPC estimates for the 2008 and 2021 periods, along with those from Graham and McDowall (2026).[17] In both stimulus episodes, the six-month cumulative MPC is close to the 26 percent of the Graham and McDowall (2026) micro estimate: 27 percent in 2008 and 20 percent in 2021. However, the timing differs noticeably. In both episodes the estimated MPC rises more gradually than the Graham and McDowall (2026) profile: after one month it is 12 percent in both 2008 and 2021, compared with 19 percent for the micro profile.

The MPC estimates from the 2008 and 2021 counterfactual periods are statistically indistinguishable. The standard errors around the 2008 episode MPC are much larger than those around the 2021 MPC, reflecting the smaller size of the 2008 stimulus. The six-month cumulative MPC for 2008 is 27 percent, with a wide 95 percent band that runs from close to zero to about twice the Graham and McDowall (2026) value.

Table 3 compares the estimated MPCs across Census, Fiserv, and Numerator data. Looking just at the 2021 stimulus episode, the results are consistent across all three datasets at short horizons: MPCs range from 11 to 12 percent at one month, 12 to 23 percent at three months, and 12 to 28 percent at six months. The Census and Fiserv estimates are broadly consistent with Graham and McDowall (2026), but imply a slower spend-out of stimulus checks than in that micro-based estimate, while the Numerator data struggle to pick up a response outside of the first shifted month. However, the uncertainty in the Numerator estimates at longer horizons is large, and the six-month Graham and McDowall (2026) value lies within its 95 percent confidence band.

[16] We draw the path with the simulation smoother of Durbin and Koopman (2012).
[17] We do not estimate the MPC for 2001, which lacks the power to distinguish large from small MPCs (Section 3).

Figure 8: Cumulative marginal propensity to consume in Census retail sales data



Notes: Cumulative MPC by days since payment receipt, estimated from monthly nominal retail sales excluding motor vehicles and parts dealers (FRED series MRTSSM44Y72USS). The solid line with circles (2008) and the dash-dotted line with triangles (2021) are the double-exponential MPC profiles that best fit, by generalized least squares, the difference between realized sales and the macro counterfactual, given the payment schedule. The shaded (2008) and hatched (2021) regions show pointwise 95 percent bands, the 2.5th to 97.5th percentiles of the cumulative MPC across bootstrap draws. The dashed line is the Graham and McDowall (2026) MPC profile.

Table 3: Estimated marginal propensity to consume

	Cumulative MPC		
Dataset	1 month	3 months	6 months
Panel A: 2021 Stimulus			
Census	0.12	0.18	0.20
	(0.01)	(0.03)	(0.04)
Fiserv	0.12	0.23	0.28
	(0.02)	(0.04)	(0.07)
Numerator	0.11	0.12	0.12
	(0.02)	(0.05)	(0.06)
Panel B: 2008 Stimulus			
Census	0.12	0.23	0.27
	(0.05)	(0.09)	(0.12)
Graham and McDowall (2026)	0.19	0.25	0.26

Notes: Point estimates are from the fit to the observed data; bootstrap standard errors in parentheses. Cumulative MPC estimates based on the macro counterfactual and generalized least squares fitting of a double-exponential spending profile, weighting by the inverse of the full covariance of the counterfactual across the months of each window. Bootstrap draws of the counterfactual allow for uncertainty about the ARIMA parameters and for correlation across months (see Section 6). Numerator and Fiserv are aggregated into shifted months running from the 12th of one month to the 11th of the next (e.g., March spans March 12–April 11) and seasonally adjusted with AltX-13 (Appendix B); Census uses calendar months.

7. Conclusion

The effectiveness of fiscal stimulus is among the most studied and widely debated issues in macroeconomics and public policy. However, the fundamental question of to what extent aggregate spending actually reflects government transfers to households has thus far remained largely unresolved. Household-level micro studies estimate large marginal propensities to consume out of stimulus payments, but the aggregate data during historical transfer programs, in particular the 2001 and 2008 U.S. stimulus episodes, have shown no clear and obvious spending response, despite marked increases in disposable income.

Our paper aims to resolve this apparent micro-macro consumption puzzle. We argue that the 2001 rebate was too small, relative to ordinary volatility in aggregate spending, to detect a response even if the micro MPCs are correct. The 2008 rebate, about twice as large as a share of DPI, gives the data somewhat more power. Retail sales rise modestly above the macro confidence bands in the months the checks arrived, and the implied three-month cumulative MPC is approximately 25 percent, in line with the micro estimates, though with wide uncertainty that limits inference. The 2021 Economic Impact Payments, by contrast, were more than four times the 2001 rebate and more than twice the 2008 rebate as a share of DPI. For the first time, the aggregate data speak clearly: retail sales in March 2021 jumped far outside the range of ordinary behavior precisely when the payments arrived. The micro counterfactual built from the Graham and McDowall (2026) MPCs broadly aligns with the macro counterfactual in all three episodes. In 2021 the macro data point to a somewhat slower spend-out than the micro profile implies. Once we account for the uncertainty around the counterfactual, we find no conflict between the micro MPCs and aggregate retail sales.

In addressing this puzzle, we view our work as offering several additional contributions. Most importantly, we thoroughly analyze the consumption response to EIP3 in the spring of 2021 in the context of prior estimates. In particular, we find that our 2021 results are unlikely to be driven by the reopening of the economy after COVID-19. Daily data from Numerator and Fiserv, which do not share a sampling frame with the Census survey, reproduce the March 2021 jump in both timing and magnitude. When we re-aggregate them into months starting on March 12, the day disbursement began, spending peaks as the payments arrive and steps down over the following months. Gradual reopening would not produce that pattern. We leverage information on income, marital status, and vaccination status to compare consumption responses across eligible recipients and vaccinated and not-yet-vaccinated households. In both cases the results point to the payments as the main driver of the March increase. Spending among ineligible households also rose, so reopening may have contributed, but it cannot account for the whole increase. With the importance of the stimulus in driving the aggregate data established, we additionally estimate the aggregate MPC path

directly from the macro data in 2021 rather than merely testing the plausibility of our micro counterfactual. Across the Census and Fiserv data, our point estimates for the cumulative MPC reach between 20 and 28 percent at six months, confirming the applicability of prior micro estimates to the 2021 episode, though perhaps suggesting a somewhat less front-loaded response. The Numerator estimates are lower at longer horizons, but imprecise enough to remain consistent with the micro estimates.

Our analysis has several limits. First, we measure the response only in *nominal retail sales*. We therefore do not address how much of the induced spending passed through to prices, or how it affected spending on services and motor vehicles. Orchard, Ramey and Wieland (2025*a*) highlight the importance of these factors for general equilibrium effects: they model durables as subject to upward-sloping short-run supply, implying that they respond to stimulus spending with increases in relative prices that crowd out other purchases and attenuate the aggregate stimulus response. Second, we do not assess the effectiveness of transfers *for improving welfare*: we abstract entirely from measuring the welfare gains from consumption smoothing, the distortionary costs of deficit financing, and distributional consequences. Our paper is limited to asking whether and to what extent fiscal stimulus in the U.S. has been effective at *inducing consumption spending on retail sales*.

Within those bounds, the aggregate data are consistent with the household evidence. While the 2001 and 2008 episodes lacked statistical power, and as such were not able to speak conclusively for or against the micro estimates, the 2021 episode shows a clear and highly significant consumption response to EIP3. When the stimulus is large enough to show up in aggregate data, it shows up where and when the micro estimates say it should.

References

Aladangady, Aditya, Shifrah Aron-Dine, Wendy Dunn, Laura Feiveson, Paul Lengermann, and Claudia Sahm. 2019. "From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2019-057.

American Hospital Association. 2021. "CDC Ends Indoor Mask Requirements for Fully Vaccinated People." *AHA News*.

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

Baker, Scott R., Robert A. Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis. 2023. "Income, Liquidity, and the Consumption Response to the 2020 Economic Stimulus Payments." *Review of Finance*, 27(6): 2271–2304.

Burbio. 2021. "Week of 3/1: The Long Phase-Ins."

Centers for Disease Control and Prevention. 2021*a*. "CDC Issues First Set of Guidelines on How Fully Vaccinated People Can Visit Safely with Others." *CDC Newsroom*.

Centers for Disease Control and Prevention. 2021*b*. "CDC Updates Operational Strategy for K-12 Schools to Reflect New Evidence on Physical Distance in Classrooms." *CDC Newsroom*.

Centers for Disease Control and Prevention. 2021*c*. "Interim Public Health Recommendations for Fully Vaccinated People." Updated April 2021; includes guidance relaxing outdoor masking for fully vaccinated individuals.

Chetty, Raj, John N. Friedman, Michael Stepner, and Opportunity Insights Team. 2024. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data." *Quarterly Journal of Economics*, 139(2): 829–889.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2020. "How Did U.S. Consumers Use Their Stimulus Payments?" National Bureau of Economic Research Working Paper 27693.

Crawley, Edmund, and Alexandros Theloudis. 2024. "Income Shocks and Their Transmission into Consumption." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 2024-038.

Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods.* . 2nd ed., Oxford University Press.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen. 1998. "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program." *Journal of Business & Economic Statistics*, 16(2): 127–152.

Graham, James, and Robert A. McDowall. 2026. "Mental Accounts and Consumption Sensitivity across the Distribution of Liquid Assets." *American Economic Journal: Macroeconomics*, 18(1): 1–33.

Hacıoğlu Hoke, Sinem, Leo Feler, and Jack Chylak. 2024. "A Better Way of Understanding the US Consumer: Decomposing Retail Sales by Household Income." Board of Governors of the Federal Reserve System FEDS Notes, Washington. October 11, 2024.

Internal Revenue Service. 2008. "Facts about the 2008 Stimulus Payments."

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.

Kaiser Family Foundation. 2021. "KFF COVID-19 Vaccine Monitor: What We've Learned."

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

Kaplan, Greg, and Giovanni L. Violante. 2022. "The Marginal Propensity to Consume in Heterogeneous Agent Models." *Annual Review of Economics*, 14: 747–775.

Karger, Ezra, and Aastha Rajan. 2021. "Heterogeneity in the Marginal Propensity to Consume: Evidence from Covid-19 Stimulus Payments." Federal Reserve Bank of Chicago Working Paper 2020-15.

Morning Consult. 2021. "Good News for U.S. Economy: Comfort Spikes Among Vaccinated Americans Following Big Shift in CDC's COVID-19 Guidance."

Numerator. 2025. "Numerator Data Foundations." Numerator, Inc. Panel methodology documentation.

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

Orchard, Jacob D., Valerie A. Ramey, and Johannes F. Wieland. 2025*b*. "Using Macro Counterfactuals to Assess Plausibility: An Illustration Using the 2001 Rebate MPCs." *Economic Journal*. Forthcoming.

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*, Fall: 81–130.

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.

Restaurant Dive. 2021. "Study: 60% of Consumers Are Comfortable Dining Out."

Robert E. Wenzel. 2001. "Testimony of Deputy Commissioner of Internal Revenue Robert E. Wenzel Before the House Ways and Means Subcommittee on Oversight: Mailings on Advance Payment Checks." *Testimony before the House Committee on Ways and Means, Subcommittee on Oversight*.

SafeGraph. 2021. "Fitness Deep Dive." White paper on gym foot traffic and visitation patterns.

SafeGraph. 2022. "Coffee Deep Dive: Waking Up to 2022." White paper on coffee shop foot traffic and visitation patterns.

Taylor, John B. 2009. "The Lack of an Empirical Rationale for a Revival of Discretionary Fiscal Policy." *American Economic Review*, 99(2): 550–555.

U.S. Bureau of Labor Statistics. 2009. "Consumer Expenditure Survey Results on the 2008 Economic Stimulus Payments (Tax Rebates)."

U.S. Congress. 2001. "Economic Growth and Tax Relief Reconciliation Act of 2001." H.R. 1836, Public Law 107-16, 107th Congress.

U.S. Congress. 2008. "Economic Stimulus Act of 2008." H.R. 5140, Public Law 110-185, 110th Congress.

U.S. Department of the Treasury. 2001. "Week Ten: Treasury to Mail Out 8.607 Million Checks on Friday." Advance Tax Refund Program Weekly Update.

U.S. Department of the Treasury. 2008. "Week 11 Wrap-Up: Treasury Sent 7.530 Million Stimulus Payments This Week." Economic Stimulus Payment Program Weekly Update.

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*a*. "More Than 1.1 Million Additional Economic Impact Payments Disbursed Under the American Rescue Plan; Payments Total Approximately 164 Million."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*b*. "More Than 1.8 Million Additional Economic Impact Payments Disbursed Under the American Rescue Plan; Total Payments Reach Nearly 167 Million."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*c*. "More Than 2.3 Million Additional Economic Impact Payments Disbursed Under the American Rescue Plan; Total Payments Top 169 Million."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*d*. "Treasury and IRS Disburse Another 37 Million Economic Impact Payments from the American Rescue Plan."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*e*. "Treasury and IRS Disburse More Economic Impact Payments Under the American Rescue Plan."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*f*. "Treasury and IRS Disbursed 90 Million Economic Impact Payments from the American Rescue Plan."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*g*. "Treasury, IRS Disburse 2 Million More Economic Impact Payments Under the American Rescue Plan; VA Beneficiaries Bring Total to Approximately 159 Million as Payments Continue."

U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service. 2021*h*. "Treasury, IRS Disburse 25 Million More Economic Impact Payments Under the American Rescue Plan."

U.S. General Accounting Office. 2001. "Tax Administration: Interim Report on Advance Tax Refunds." U.S. General Accounting Office GAO-02-257.

U.S. General Accounting Office. 2002. "Tax Administration: Advance Tax Refund Program Was a Major Accomplishment, but Not Problem Free." U.S. General Accounting Office GAO-02-827.

U.S. Government Accountability Office. 2022. "Stimulus Checks: Direct Payments to Individuals During the COVID-19 Pandemic." U.S. Government Accountability Office GAO-22-106044.

World Health Organization and Our World in Data. 2021. "Coronavirus (COVID-19) Vaccinations."

A. Data Processing Details

A.1. Numerator projection variables

Each transaction in the Numerator Total Commerce Panel is projected to national spending using household-level demographic weights and a national scaling factor. For transaction i at time t, projected national spending is:

$$\text{Spend}_{i,t} = \texttt{BASKET_SUB_TOTAL}_{i,t} \times \texttt{DEMO_WEIGHT}_{i,t} \times \texttt{NATIONAL_FACTOR}_{t} \tag{4}$$

where:

- **BASKET_SUB_TOTAL**: Pre-tax dollar value of the transaction. For transactions with missing pre-tax values (approximately 13 percent of records), we compute this as BASKET_TOTAL minus BASKET_SALESTAX, or use BASKET_TOTAL directly if sales tax information is unavailable.
- **DEMO_WEIGHT**: Household-level demographic weight adjusting for panel representativeness relative to the U.S. population across eight Census demographic dimensions: age, gender, household size, presence of children under eighteen, ethnicity, household income, Census Division, and urbanicity.
- **NATIONAL_FACTOR**: The ratio of total U.S. households to static panel size in month t. This factor scales household spending to national totals and varies as the panel grows (from approximately 100,000 households in 2019 to nearly 200,000 by 2025).

A.2. Data quality filters

We apply three filters to ensure data quality and prevent outlier amplification:

1. **Receipt amount bounds**: BASKET_SUB_TOTAL $\in$ [0, \$10,000]. This removes extreme outliers likely representing data errors rather than consumer transactions. The filter affects fewer than 0.0002 percent of transactions but prevents massive projection errors when extreme receipts are amplified through weighting factors.
2. **Non-null weighting factors**: We require non-missing values for DEMO_WEIGHT and NATIONAL_FACTOR. Transactions with missing weights cannot be reliably projected to national spending.
3. **Valid user demographics**: We restrict to transactions from users that can be matched to valid demographic information in the static user file, ensuring household characteristics are available for cross-sectional analysis.

A.3. NAICS sector classification

We map Numerator's proprietary retailer classification (`BANNER_ID`) to NAICS retail and food services subsectors using a manually curated crosswalk containing approximately 15,000 unique retailer identifiers. The twelve NAICS subsectors used throughout the analysis are:

- **NAICS 442**: Furniture and Home Furnishings Stores
- **NAICS 443**: Electronics and Appliance Stores
- **NAICS 444**: Building Material and Garden Equipment Dealers
- **NAICS 445**: Food and Beverage Stores
- **NAICS 446**: Health and Personal Care Stores
- **NAICS 447**: Gasoline Stations
- **NAICS 448**: Clothing and Clothing Accessories Stores
- **NAICS 451**: Sporting Goods, Hobby, Musical Instrument, and Book Stores
- **NAICS 452**: General Merchandise Stores
- **NAICS 453**: Miscellaneous Store Retailers
- **NAICS 454**: Nonstore Retailers
- **NAICS 722**: Food Services and Drinking Places

Numerator banners without a standard retail classification are pooled into an "Unclassified" category, which we scale to Census gasoline stations (NAICS 447), reflecting the fact that gas stations and convenience stores account for most of the spending in the category. Multi-banner retailers operating multiple formats (e.g., Walmart operating both general merchandise and grocery) are disaggregated to the banner level to preserve sector-specific spending patterns. The Fiserv data use an analogous NAICS mapping via Merchant Category Codes as documented in Aladangady et al. (2019), which assigns gasoline stations directly to NAICS 447.

B. Seasonal Adjustment

We apply sector-level X-13ARIMA-SEATS seasonal adjustment to Numerator and Fiserv data, following the methodology used by the U.S. Census Bureau for the Monthly Retail Trade Survey (Findley et al., 1998). As with the published Census figures, we seasonally adjust each NAICS sector separately before aggregating.

B.1. Standard X-13 adjustment (calendar months)

Our seasonal adjustment procedure closely mirrors Census practice. X-13 performs a two-stage decomposition. First, a regARIMA model removes calendar effects and outliers from the raw series. The preadjusted series then passes through the X-11 filter-based decomposition, which separates it into trend-cycle, seasonal, and irregular components. X-13 automatically selects the ARIMA specification by minimizing the corrected Akaike information criterion (AICc), chooses between additive and multiplicative decomposition based on series characteristics, and extends the series with ARIMA forecasts and backcasts so that the X-11 moving-average filters can be applied symmetrically near the sample edges.

We include trading day adjustments in the regARIMA stage. These regressors account for the fact that months with more weekends or more weekdays systematically differ in retail activity, and they allow the seasonal adjustment to distinguish calendar composition effects from true seasonal patterns. Trading day adjustments are a standard component of Census seasonal adjustment and constitute the primary calendar effect in monthly retail sales data.

For simplicity, we do not include holiday regressors (Easter, Thanksgiving, Christmas, etc.) in the regARIMA preadjustment, whereas Census does. This omission represents a minor departure from official methodology. Holiday effects in monthly data are small relative to trading day effects because most major holidays fall on predictable dates or weekends, and their impact is largely absorbed by the seasonal component itself.

B.2. Shifted-month adjustment (AltX-13)

In Section 4, we re-aggregate daily spending into shifted months that begin on the 12th rather than the 1st of each month, aligning the observation window with the EIP3 disbursement schedule. Each shifted month runs from the 12th day of calendar month M through the 11th day of month $M + 1$.

Standard X-13 trading day regressors count weekdays within calendar month boundaries. When data are aggregated from March 12 to April 11, the weekday composition differs from that of March 1–31. For instance, March 1–31, 2021, contained five Wednesdays, while March 12–April 11 only contained four. Applying standard X-13 to shifted-month data would use incorrect weekday counts and produce misspecified trading day adjustments.

Our solution, which we call AltX-13, computes custom trading day regressors that reflect the actual weekday counts in each shifted-month period. For each period, we count the occurrences of Monday through Sunday between the shifted boundaries and construct regressors as the difference between each weekday count and the Sunday count, following X-13 convention. We then pass these regressors to X-13 as exoge-

nous variables and disable the built-in trading day adjustment. All other aspects of the procedure—ARIMA model selection, log transformation, outlier detection, and X-11 decomposition—remain unchanged. The seasonal patterns themselves depend only on the sequence of period numbers (1 through 12), not on calendar dates, so the shifted boundaries do not affect the seasonal filter.

For clarity in figures, we label each shifted-month observation with the date on which its period begins. The observation covering March 12–April 11, 2021 is labeled March 12, 2021, making the alignment with payment timing visually apparent.

C. Stimulus Payment Distribution Paths

C.1. 2021: American Rescue Plan Act (EIP3)

The American Rescue Plan Act, signed March 11, 2021, authorized the third round of Economic Impact Payments. Eligible individuals received $1,400 per person ($2,800 for married couples filing jointly) plus $1,400 per qualifying dependent with a valid Social Security number. Payments phased out for single filers with adjusted gross income (AGI) between $75,000 and $80,000, married couples filing jointly between $150,000 and $160,000, and heads of household between $112,500 and $120,000.

The Treasury Department distributed approximately 176 million payments totaling $410 billion between March 12 and the end of 2021. Table 4 summarizes the disbursement schedule used in our analysis, which includes the first 11 batches distributed through early June (U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service, 2021*f*,*d*,*e*,*h*,*g*,*a*,*b*,*c*; U.S. Government Accountability Office, 2022).

C.2. 2008: Economic Stimulus Act

The Economic Stimulus Act of 2008 was signed into law on February 13, 2008. Eligible taxpayers received payments based on tax year 2007 filings: $300–$600 for individuals and up to $1,200 for married couples filing jointly, plus $300 per qualifying child. Eligibility required at least $3,000 in qualifying income (wages, Social Security benefits, or certain veterans' benefits). Payments phased out for AGI above $75,000 (individuals) or $150,000 (married filing jointly).

Payments were distributed in 11 weekly batches from late April through mid-July 2008 based on the last two digits of taxpayers' Social Security numbers. Direct deposits began April 28, 2008, with paper checks following in early May. Approximately 112 million payments totaling $92 billion were distributed in the 11 weekly batches; including later payments and recovery rebate credits claimed through mid-2009, the total reached 130 million payments and $108 billion. Table 5 shows the distribution schedule with midpoint

Table 4: EIP3 distribution schedule, 2021

Batch	Interpolated Arrival	Payments (millions)	Value ($B)
March 12–17	March 14	90.0	242.0
March 19–24	March 21	37.0	83.0
March 26–31	March 28	4.0	10.0
April 2–7	April 4	25.0	36.0
April 9–14	April 11	2.0	3.4
April 16–21	April 18	2.0	3.4
April 23–28	April 25	2.0	4.3
April 30–May 5	May 2	1.1	2.0
May 7–12	May 9	1.0	1.8
May 14–26	May 16	0.9	1.8
May 14–26	May 23	0.9	1.8
May 28–June 9	May 30	1.2	2.1
May 28–June 9	June 6	1.2	2.1
Other	N/A	7.6	16.0
Total		175.8	409.6

Notes: The vast majority of early payments were made by direct deposit, including all $242 billion in the first batch. Interpolated arrival dates are offset by 2 days from batch start dates to account for the front-loaded MPC profile. Split batches are divided evenly for weekly modeling: May 14–26 is $3.5 billion (1.8 million payments) and May 28–June 9 is $4.2 billion (2.3 million payments). Entries are rounded to one decimal, so rows may not sum exactly to the totals. "Other" denotes additional payments distributed June through December 2021, which are excluded from the convolution method establishing our micro counterfactual. Sources: U.S. Department of the Treasury, Internal Revenue Service, and Bureau of the Fiscal Service, 2021*f*,*d*,*e*,*h*,*g*,*a*,*b*,*c*; U.S. Government Accountability Office, 2022.

dates used for modeling the spending response (U.S. Congress, 2008; Internal Revenue Service, 2008; Parker et al., 2013; U.S. Department of the Treasury, 2008; U.S. Bureau of Labor Statistics, 2009).

Table 5: Economic Stimulus Act of 2008: payment distribution schedule

Batch	Interpolated Arrival	Payments (millions)	Value ($B)
April 28–May 2	April 30	7.7	7.1
May 5–9	May 7	22.2	20.1
May 12–16	May 14	15.6	13.6
May 19–23	May 21	6.2	4.9
May 26–30	May 28	5.8	4.3
June 2–6	June 4	9.1	6.8
June 9–13	June 11	9.5	7.0
June 16–20	June 18	9.1	6.9
June 23–27	June 25	9.7	7.5
June 30–July 4	July 2	10.0	7.8
July 7–11	July 9	7.5	5.8
Other	N/A	17.6	16.6
Total		130	108.4

Notes: Midpoint dates calculated as the center of each week's distribution period. Direct deposits began April 28; paper checks followed in early May. "Other" denotes additional payments and recovery rebate credits claimed through mid-2009, which are excluded from the convolution method establishing our micro counterfactual. Sources: Parker et al., 2013; Internal Revenue Service, 2008; U.S. Department of the Treasury, 2008; U.S. Bureau of Labor Statistics, 2009.

C.3. 2001: Economic Growth and Tax Relief Reconciliation Act

The Economic Growth and Tax Relief Reconciliation Act of 2001 was signed into law on June 7, 2001. The act provided advance refund tax credits for eligible filers with tax liabilities on their tax year 2000 return: up to $300 for individuals, $500 for heads of household, and $600 for married couples filing jointly. Eligibility required that filers were not claimed as dependents on another return and were not classified as non-resident aliens. Unlike later programs, direct deposits were not available; all payments were mailed as paper checks.

The Treasury mailed checks in 10 weekly batches from July 20 through September 21, 2001, distributed according to Social Security number. Each batch contained roughly 8–9 million checks worth $3–4 billion. Approximately 83 million checks totaling $35 billion were distributed during the initial ten-week period, with the total reaching $36.4 billion across 86 million checks by December 2001 due to extensions and late-filed returns. Table 6 shows the distribution schedule with midpoint dates used for modeling the spending response (U.S. Congress, 2001; U.S. General Accounting Office, 2001; Robert E. Wenzel, 2001; U.S. Department of the Treasury, 2001; U.S. General Accounting Office, 2002).

Table 6: Economic Growth and Tax Relief Reconciliation Act of 2001: payment distribution schedule

Batch	Interpolated Arrival	Payments (millions)	Value ($B)
July 20	July 26	7.9	3.3
July 27	August 2	8.1	3.4
August 3	August 9	8.2	3.5
August 10	August 16	8.2	3.5
August 17	August 23	8.2	3.5
August 24	August 30	8.3	3.6
August 31	September 6	8.3	3.5
September 7	September 13	8.4	3.6
September 14	September 20	8.6	3.7
September 21	September 27	8.6	3.7
Other	N/A	3.2	1.1
Total		86	36.4

Notes: All payments were mailed as paper checks; no direct deposits were available. Interpolated arrival dates are therefore offset by 6 days from mailing dates to account for mail delivery and check processing. "Other" denotes additional payments from extensions and late-filed returns through December 2001, which are excluded from the convolution method establishing our micro counterfactual. Sources: U.S. Congress, 2001; U.S. General Accounting Office, 2001; Robert E. Wenzel, 2001; U.S. Department of the Treasury, 2001; U.S. General Accounting Office, 2002.

D. Economic Reopening in Spring 2021

EIP3, the largest of the three pandemic-era payment rounds, likely had a major effect on spending in spring 2021, but identifying that effect is complicated by a concurrent factor: the reopening of the economy and its recovery from the pandemic. By constraining the ability of consumers to work and shop in public spaces, the pandemic collapsed demand and ushered in high unemployment through the middle of 2020. Only in the spring of 2021 did restrictive guidelines begin to relax and vaccines begin to be distributed to the broader public. This timeline implies that consumer spending over March and April of 2021, the periods during which we are interested in measuring the effects of EIP3, was plausibly encouraged by recovery from the pandemic and the initial release of pent-up consumer demand. The counterfactual methods employed throughout this paper are designed with this potential confounder in mind. Importantly, the macro counterfactual methods predict a steady increase in consumption throughout the spring, consistent with gradual reopening, rather than the sharp March spike we observe in the data. We therefore argue that our excess-spending measures mainly reflect the stimulus rather than reopening. Below we present a brief summary of the relevant information for the pandemic-induced economic reopening during the period of interest.

On March 8, 2021, less than a week before the release of the first batch of EIP3 payments, the Centers for Disease Control and Prevention (CDC) released its first set of guidelines on the activities that people who

were fully vaccinated against COVID-19 could safely resume (Centers for Disease Control and Prevention, 2021*a*). Most importantly, the guidance indicated that fully vaccinated individuals could safely interact with other vaccinated or unvaccinated individuals in-person, without needing to wear a mask or practice social distancing, representing "a first step toward returning to everyday activities." The CDC continued to loosen guidance throughout the spring, lowering requirements for physical distancing in schools (Centers for Disease Control and Prevention, 2021*b*), eliminating quarantine, testing, and masking recommendations in outdoor public spaces for vaccinated individuals (Centers for Disease Control and Prevention, 2021*c*), and eventually retracting its guidance on indoor masking (American Hospital Association, 2021).

Vaccination rates in the United States peaked during spring 2021 (World Health Organization and Our World in Data, 2021). On January 1, 2021, the daily percent of the U.S. population receiving vaccination doses was just over 0.1 percent. By March 11, the day before EIP3 disbursement began, that daily measure had jumped to over 0.7 percent. The vaccination rate hit its peak on April 12 at just over 1 percent, and rates remained fairly elevated through the rest of the spring. As of March 2021, 32 percent of U.S. adults reported having received at least one dose of a COVID-19 vaccine (Kaiser Family Foundation, 2021).

Survey evidence indicates that consumer comfort with in-person activities was rising through spring 2021 (Morning Consult, 2021). Restaurant dining provides a representative example: as of late April, 60 percent of consumers reported comfort dining out, up significantly from the 42 percent peak in 2020 (Restaurant Dive, 2021). Data on virtual school attendance suggest that by the beginning of March, the proportion of K-12 students attending virtual-only schools fell from 55 percent at the start of the year to 27.5 percent, a factor which would encourage a return to normalcy in the work schedules of parents as well (Burbio, 2021). Foot traffic reports also highlight increasing returns to public spaces like gyms and coffee shops (SafeGraph, 2021, 2022).

Together, this evidence confirms that multiple factors contributed to economic reopening throughout spring 2021—precisely the gradual recovery our macro counterfactual methods predict. Crucially, however, none of these factors exhibits a discrete shift coinciding with the March 12 disbursement date. Vaccination rates, CDC guidance relaxations, and consumer comfort all evolved continuously through the spring rather than spiking in mid-March. This temporal pattern, combined with the demographic evidence in Section 5 showing that spending increases were concentrated among eligible households and invariant to vaccination status, strongly supports our conclusion that the sharp March surge in retail sales reflects spending out of stimulus payments rather than pandemic reopening.

E. Kalman Filter Algorithm

We summarize the algorithm used to compute the counterfactual from the ARIMA(1,1,0) model in Equation (1), where y_t is log nominal retail sales. Because the data are in y_t rather than in its growth rate, we write the model in state-space form with a state vector that contains both the log level and its growth rate, $x_t = (y_t, \Delta y_t)'$:

$$x_t = c + Fx_{t-1} + g\,\eta_t, \qquad c = \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \quad F = \begin{pmatrix} 1 & \phi \\ 0 & \phi \end{pmatrix}, \quad g = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

which restates Equation (1) together with the identity $y_t = y_{t-1} + \Delta y_t$. Tracking the level is what lets the counterfactual use the level of retail sales on both sides of the window: the growth rates inside the window must add up to the observed change in log sales across it.

Consider one stimulus window, $\mathcal{T} = \{\underline{t}, \ldots, \bar{t}\}$; the windows are far apart, and we treat each separately. Because there is no measurement error, the state is known exactly in month $\underline{t} - 1$, from $y_{\underline{t}-1}$ and $y_{\underline{t}-2}$, and again in month $\bar{t} + 2$, from $y_{\bar{t}+2}$ and $y_{\bar{t}+1}$. Month $\bar{t} + 2$, rather than $\bar{t} + 1$, is the first month after the window in which the whole state is known: $y_{\bar{t}+1}$ is observed, but its growth rate depends on the missing $y_{\bar{t}}$. Because the model is Markov, these two states summarize everything the data outside the window say about the months inside it. We write $x_{t|t'}$ and $\Sigma_{t|t'}$ for the mean and covariance matrix of x_t given the data through month t', and $x_{t|s\notin\mathcal{T}}$ and $\Sigma_{t|s\notin\mathcal{T}}$ for the mean and covariance matrix given all data outside the window. The key departures from the standard Kalman filter/smoother routines are: (1) there is no measurement error, so outside the window the update step sets the state equal to the data, and (2) we condition only on data outside the stimulus window $\mathcal{T}$, so inside the window the update step leaves the prediction unchanged.

1. **Kalman filter**. We start from the known state in month $\underline{t} - 1$, so that $x_{\underline{t}-1|\underline{t}-1} = x_{\underline{t}-1}$ and $\Sigma_{\underline{t}-1|\underline{t}-1} = 0$, and repeat the following two steps for each month $t = \underline{t}, \ldots, \bar{t} + 2$.

   (a) The prediction step:

$$x_{t|t-1} = c + Fx_{t-1|t-1} \tag{5}$$

$$\Sigma_{t|t-1} = F\Sigma_{t-1|t-1}F' + \sigma_\eta^2\, gg' \tag{6}$$

   (b) The update step:

$$x_{t|t} = x_{t|t-1} \tag{7}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} \tag{8}$$

Note that in this case, the updated estimate follows the ARIMA(1,1,0) process, Equation (5), and the uncertainty around the estimate follows Equation (6).

2. **Kalman smoother.** After the forward pass is complete, we then use the backward smoother to incorporate future information into our estimates of x and Σ. The smoothing gain L_t determines how much weight to place on future information when revising the filtered estimates. Starting from the known state, $x_{\bar{t}+2|s\notin\mathcal{T}} = x_{\bar{t}+2}$ and $\Sigma_{\bar{t}+2|s\notin\mathcal{T}} = 0$, for $t = \bar{t}+1, \bar{t}, \ldots, \underline{t}$:

$$L_t = \Sigma_{t|t} F' \Sigma^{-1}_{t+1|t} \tag{9}$$

$$x_{t|s\notin\mathcal{T}} = x_{t|t} + L_t(x_{t+1|s\notin\mathcal{T}} - x_{t+1|t}) \tag{10}$$

$$\Sigma_{t|s\notin\mathcal{T}} = \Sigma_{t|t} + L_t(\Sigma_{t+1|s\notin\mathcal{T}} - \Sigma_{t+1|t})L_t' \tag{11}$$

The conditional mean $y_{t|s\notin\mathcal{T}}$ is the first element of $x_{t|s\notin\mathcal{T}}$, and its variance $P_{t|s\notin\mathcal{T}}$ is the top-left element of $\Sigma_{t|s\notin\mathcal{T}}$. Because the level is known on both sides of the window, $P_{t|s\notin\mathcal{T}}$ is largest in the middle of the window. Because there is no measurement error, $\Sigma_{t+1|t}$ can be singular (for example, $\Sigma_{\underline{t}|\underline{t}-1} = \sigma^2_\eta\, gg'$), in which case the inverse is replaced by the Moore–Penrose pseudo-inverse. In practice, we compute these moments with the missing-data Kalman smoother in the `statsmodels` SARIMAX routine, run over the full sample with y_t set to missing for $t \in \mathcal{T}$; this gives the same conditional means and variances.

3. **Parameter uncertainty.** Steps 1 and 2 take $\theta = (\mu, \phi, \sigma^2_\eta)$ as known. We repeat them for 500 draws θ^* from a normal distribution centered at the maximum likelihood estimate with its estimated covariance matrix, discarding draws with $|\phi| \geq 1$ or $\sigma^2_\eta \leq 0$. The variance of y_t is then the average of $P_{t|s\notin\mathcal{T}}$ across draws plus the variance of $y_{t|s\notin\mathcal{T}}$ across draws. The point estimates use the maximum likelihood estimate.

4. **Transform back to levels.** We reconstruct sales by exponentiating $y_{t|s\notin\mathcal{T}}$ and then adjust standard errors using the delta method.

Note that this procedure answers the question: what is our best estimate of how retail sales would have evolved during the stimulus windows $\mathcal{T}$ if they had followed the same statistical process as in normal times? The procedure excludes information from the stimulus windows, and therefore also abstracts from any other shocks that occurred during them.

E.1. Model specification

We use a single specification, ARIMA(1,1,0) with drift, for all three spending series: Census retail sales, and the shifted-month Fiserv and Numerator series of Section 4. The specification has three components (first-differencing, a drift term, and an autoregressive term), and we test each on every series. All estimates use exact maximum likelihood on the same sample as the corresponding counterfactual, which masks the pandemic months and the stimulus windows.

The order of differencing comes from the augmented Dickey–Fuller (ADF) test, whose null is a unit root, and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test, whose null is stationarity. These tests require an uninterrupted sample, so for Census we apply them to log retail sales from January 1992 through February 2020 rather than masking the stimulus windows. In levels, with a constant and linear trend, the ADF test does not reject a unit root (p-value = 0.85) and the KPSS test rejects trend stationarity (p-value < 0.01). In first differences, with a constant, the ADF test rejects (p-value < 0.01) and the KPSS test does not (p-value > 0.10). Log retail sales are therefore integrated of order one. The same tests on the longest uninterrupted stretch of the Fiserv series (April 2010 to January 2020) again point to first-differencing. For Numerator, that stretch covers only 41 months after the reopening, too short for the tests to be informative, and we difference it for consistency with the other two series.

The drift term captures the secular growth of nominal spending. Comparing ARIMA(p,1,q) models with $p, q \in \{0, 1\}$ with and without drift, likelihood ratio tests reject a zero drift in every ARMA order for Census and Fiserv, and in every order except the random walk for Numerator. Omitting the drift forces mean growth to zero and pushes the trend into the short-run dynamics: for Census, $\hat{\phi} = 0.20$ without drift, compared with $\hat{\phi} = 0.003$ (s.e. 0.03) with it.

The autoregressive term captures short-run dynamics, and here the series differ. Fiserv and Numerator both show significant negative autocorrelation in monthly growth: $\hat{\phi} = -0.43$ (s.e. 0.07) for Fiserv and $\hat{\phi} = -0.39$ (s.e. 0.12) for Numerator, and a likelihood ratio test rejects $\phi = 0$ in both (p-values < 0.001 and = 0.004). The Bayesian information criterion (BIC) selects ARIMA(1,1,0) with drift for Fiserv. For Numerator it marginally prefers an MA(1) term to an AR(1) term, by a BIC difference of 1.3, too small to discriminate between the two. Census growth, by contrast, shows no short-run dynamics once drift is included. We nonetheless apply the same model to Census for consistency, and the estimated autoregressive coefficient is very close to zero, so the model effectively collapses to a random walk with drift, the specification BIC selects for Census. Finally, note that the counterfactuals are not sensitive to these choices: replacing the AR(1) term with an MA(1) term, including both, or dropping both leaves the interpolated paths and their confidence bands essentially unchanged for all three series.

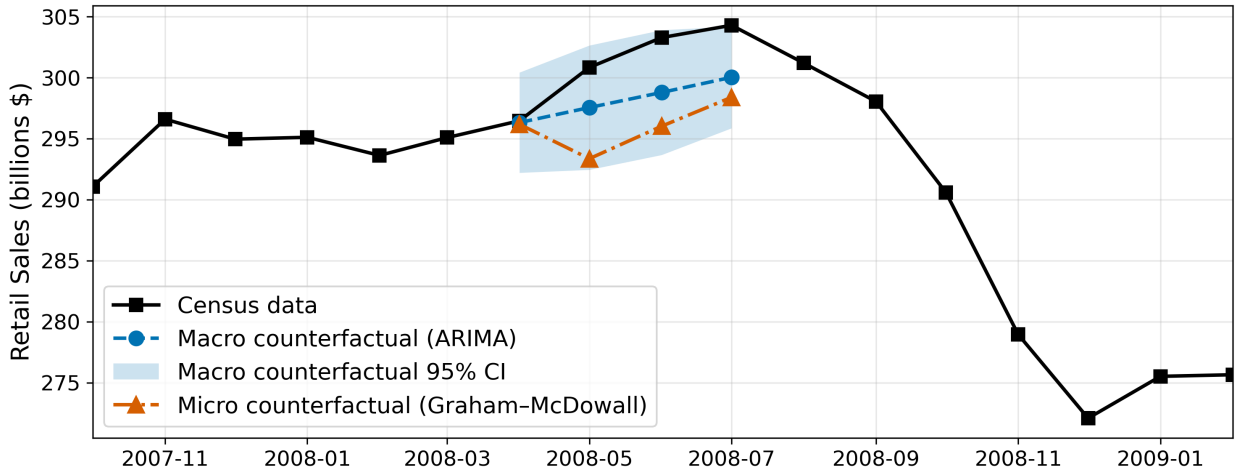
F. Sensitivity Analysis: 2008 Stimulus Window

The 2008 counterfactual analysis in the main text uses an stimulus window from April to August 2008. This choice reflects the timing of the stimulus check disbursements, which began in late April and continued through July. To assess the robustness of our findings to this choice, we conduct a sensitivity analysis with alternative window end dates.

Figure 9 shows the counterfactual when the stimulus window ends one month earlier (April to July). Figure 10 shows the counterfactual when the stimulus window ends one month later (April to September). In both cases, actual retail sales rise above the macro counterfactual during the stimulus disbursement period. Ending the window in July leaves the micro counterfactual inside the confidence bands in every month, as in the baseline. Realized sales, however, are then inside the bands in June and exceed the upper edge only marginally in July, so the evidence of a response is weaker than in the baseline.

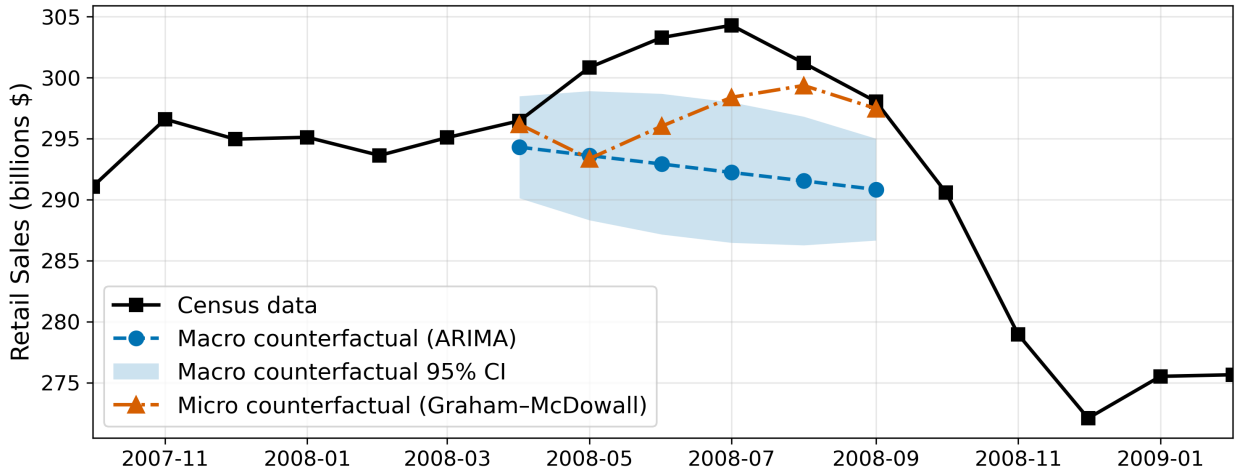
The estimated stimulus effect is sensitive to whether September is included in the stimulus window. September 2008 saw the collapse of Lehman Brothers and a sharp decline in retail sales, which pulls the macro counterfactual down across the whole window. Including September (Figure 10) roughly doubles excess spending over the window, from about $23 billion to about $49 billion, and the micro counterfactual rises above the confidence bands from July through September. With this window, the data imply a larger MPC than the Graham and McDowall (2026) profile. Table 7 reports monthly macroeconomic indicators for 2008.

Figure 9: 2008 counterfactual: one month earlier (April–July)



Notes: As in Figure 2, but with the stimulus window ending in July. Monthly nominal retail sales excluding motor vehicles and parts dealers from the U.S. Census Bureau's Monthly Retail Trade Survey, accessed via FRED (series: MRTSSM44Y72USS). The solid black line is realized sales; the dashed blue line and shaded region show the ARIMA(1,1,0) counterfactual for April–July 2008 and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. Billions of dollars per month, seasonally adjusted.

Figure 10: 2008 counterfactual: one month later (April–September)



Notes: As in Figure 2, but with the stimulus window ending in September. Monthly nominal retail sales excluding motor vehicles and parts dealers from the U.S. Census Bureau's Monthly Retail Trade Survey, accessed via FRED (series: MRTSSM44Y72USS). The solid black line is realized sales; the dashed blue line and shaded region show the ARIMA(1,1,0) counterfactual for April–September 2008 and its 95 percent confidence band; the dash-dotted orange line is the micro counterfactual implied by the Graham and McDowall (2026) MPC profile. September 2008 includes the Lehman Brothers collapse. Billions of dollars per month, seasonally adjusted.

Table 7: Macroeconomic indicators during the 2008 financial crisis

Month	Consumer Sentiment	DJIA Change	Case-Shiller Index	WTI Oil Price
2007 Average	85.6	66.8	179.9	72.4
January	78.4	−614.5	171.1	93.0
February	70.8	−384.0	169.2	95.3
March	69.5	−3.5	167.9	105.6
April	62.6	557.2	167.3	112.6
May	59.8	−181.8	167.0	125.4
June	56.4	−1,288.3	166.5	133.9
July	61.2	28.0	165.7	133.4
August	63.0	165.9	164.3	116.6
September	70.3	−693.3	161.9	103.9
October	57.6	−1,525.6	159.2	76.7
November	55.3	−496.0	156.1	57.4
December	60.1	−52.7	152.5	41.0

Notes: Consumer Sentiment is the University of Michigan Consumer Sentiment Index. DJIA Change shows month-over-month change in the Dow Jones Industrial Average at close. Historical DJIA data from Yahoo Finance. Case-Shiller Index is the S&P/Case-Shiller U.S. National Home Price Index (not seasonally adjusted). WTI Oil Price is the monthly average West Texas Intermediate spot crude oil price, dollars per barrel. Consumer Sentiment, Case-Shiller, and WTI data from FRED (series UMCSENT, CSUSHPINSA, and WTISPLC).

G. Monthly MPC Convolution Method

This appendix describes an alternative computational method for constructing the Graham and McDowall (2026) counterfactual spending paths used in Section 3.

G.1. Motivation

The baseline approach distributes each stimulus check's spending effect across 151 days using the daily MPC profile from Graham and McDowall (2026), then aggregates these daily flows into monthly totals. Past studies have used monthly measures of the MPC to compute such counterfactuals. Here we show that taking the monthly approach could lead one to draw incorrect conclusions about the plausibility of micro-based estimates of the MPC.

G.2. Method

Let $m_d(k)$ denote the daily (non-cumulative) MPC for day k after check receipt, as estimated by Graham and McDowall (2026). The daily MPC satisfies $\sum_{k=0}^{150} m_d(k) = 0.257$, representing a 25.7 percent cumulative MPC over 151 days.

We construct a monthly (non-cumulative) MPC profile by summing over 30-day windows:

$$m_m(j) = \sum_{k=30j}^{30(j+1)-1} m_d(k), \quad j = 0, 1, \ldots, 5 \tag{12}$$

This yields a six-month profile: $m_m = [0.187, 0.047, 0.015, 0.005, 0.002, 0.000]$, which preserves the total MPC of 0.257 (the entries are rounded to three decimals, so they sum to 0.256). Because the windows are 30 days long but are applied to calendar months of 28 to 31 days, the monthly profile is a small approximation to the daily one.

Let C_t denote total stimulus checks arriving in calendar month t. Monthly spending from the convolution is:

$$S_t = \sum_{s \leq t} C_s \cdot m_m(t - s) \tag{13}$$

The micro counterfactual Y_t^{micro} subtracts this estimated spending from realized retail sales Y_t:

$$Y_t^{\text{micro}} = Y_t - S_t \tag{14}$$

G.3. Comparison with the baseline method

Figure 11 compares the monthly convolution method (bottom panel) with the baseline daily-aggregation approach (top panel) for the 2021 stimulus episode. The two approaches allocate first-month spending differently. In our payment schedule (Table 4), the $335 billion that arrived in March comes in three batches, with interpolated arrival dates of March 14, 21, and 28, two days after each batch began. The daily method counts in March only the spending that occurs before the month ends: 18 days of spending for the March 14 batch, 11 days for the March 21 batch, and 4 days for the March 28 batch. The monthly method instead places each batch's entire 30-day first-month MPC in March. As a result, it moves about $16 billion of spending into March ($63 billion rather than $46 billion), most of it from April. This timing change moves the March micro counterfactual further below the confidence band surrounding the macro counterfactual path, from $425 billion to $408 billion, while the April value moves from just below the band to inside it.

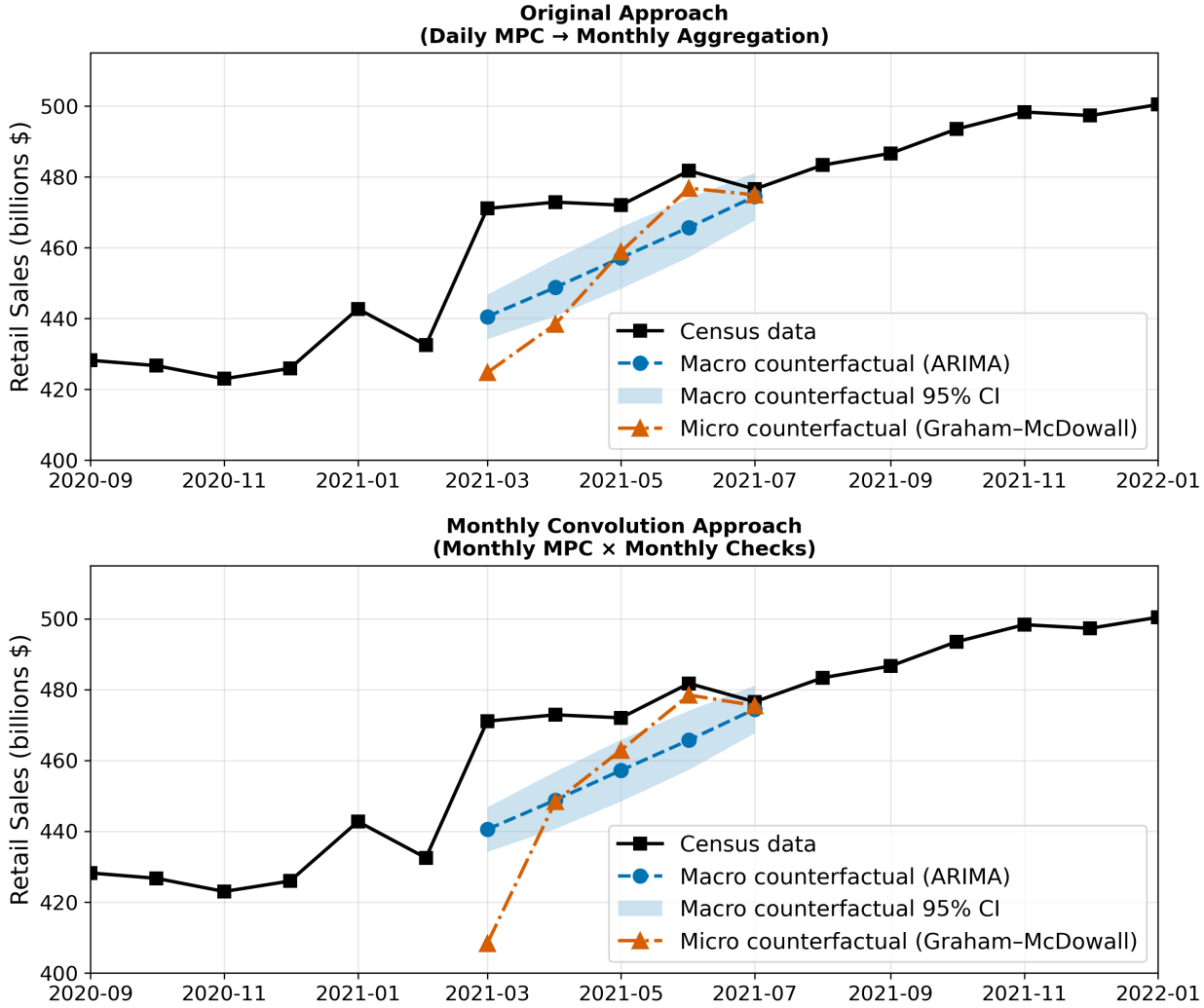
H. Orchard–Ramey–Wieland Counterfactuals with Forecast Uncertainty

In this appendix, we reproduce the time-series forecasting exercises behind Figure 4, Panel A, of Orchard, Ramey and Wieland (2025*b*) and the top-left panel of Figure IV of Orchard, Ramey and Wieland (2025*a*), leaving the construction of the forecasts and counterfactuals unchanged. In addition, we add 95 percent confidence bands that account for both parameter and shock uncertainty in order to better evaluate the plausibility of counterfactual spending paths.

We then repeat each exercise with nominal retail sales, the series used in the main text. Unlike monthly PCE, retail sales are measured directly each month rather than partly interpolated, and they align more closely with the spending categories measured in many micro MPC studies.[18] We also apply the same machinery to the 2021 episode. The 2001 episode is consistent with the Graham and McDowall (2026) counterfactual but provides little evidence for or against a large spending response. In 2008, ORW's method applied to retail sales suggests that the spending response may have been slower than the Graham and McDowall (2026) profile implies. As with our own methodology, the March 2021 episode speaks more clearly than either 2001 or 2008. Realized spending escapes the confidence bands immediately and does not return, although the spending appears to start at a slower pace than in the counterfactual.

[18] Orchard, Ramey and Wieland (2025*a*) show in their online appendix A.3 that retail sales and PCE on goods move together in 2008, but they do not repeat their forecast comparison with retail sales.

Figure 11: Comparison of MPC calculation methods: 2021 stimulus



Notes: Top panel reproduces the baseline approach from Figure 1. Bottom panel applies the monthly convolution method. Both use ARIMA(1,1,0) with drift (dashed blue line and 95 percent confidence band). Dash-dotted orange lines show micro counterfactuals. Y-axes are identical (400–515 billion dollars) to facilitate comparison.

H.1. Forecasts and confidence bands

The forecasting models follow the replication code of the two papers. Both are recursive monthly VARX models in log levels, estimated equation by equation by ordinary least squares with six lags of each variable. The 2001 system contains Michigan consumer sentiment, ordered first and entering the remaining equations contemporaneously, together with log real consumption, log real disposable income, and the log PCE deflator; the National Bureau of Economic Research (NBER) recession indicator and a September 2001 dummy enter exogenously. The 2008 system replaces sentiment with the Gilchrist–Zakrajšek excess bond premium and adds the log real oil price and a Lehman Brothers dummy to the exogenous block. Each system is solved dynamically over the relevant forecast window, anchored at the last pre-forecast month, conditioning on the realized paths of the exogenous variables. Both systems are estimated from 1984 (1992 for retail sales, which begins then) through 2019, as in the original code. We refer to the resulting path as the *conditional forecast*.[19]

The confidence bands are computed by Monte Carlo. Each of 2,000 replications draws one coefficient vector for the whole system from its estimated joint sampling distribution, which allows for correlation across equations, and a sequence of residual shocks, each obtained by resampling a whole vector of in-sample residuals with replacement. The system is re-solved under those draws; the band is the empirical 2.5th to 97.5th percentile of the simulated paths.

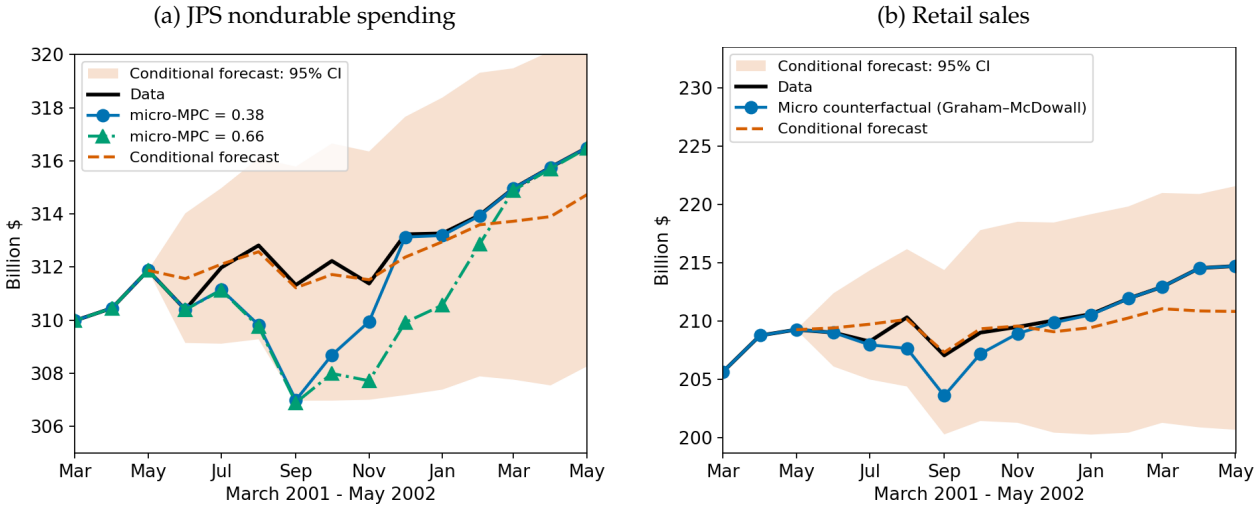
H.2. The 2001 rebate

Panel (a) of Figure 12 reproduces Figure 4, Panel A, of Orchard, Ramey and Wieland (2025*b*), adding the confidence band. The black line is real spending on the nondurable aggregate of Johnson, Parker and Souleles (2006) (henceforth JPS), re-based so that real and nominal spending coincide in May 2001, the month at which the forecast is anchored. The dashed line is the conditional forecast, solved from June 2001 onward, and the two remaining lines are the published counterfactuals: realized spending minus the rebate consumption responses implied by Johnson, Parker and Souleles (2006), with micro MPCs of 38 and 66 percent corresponding to three-month and six-month responses. The counterfactuals fall well below the conditional forecast during and after the disbursement months, but they remain just inside the 95 percent band, only touching its lower edge in September 2001. September 2001 is also the month in which the comparison is least informative about the rebate, since the level of spending there is dominated by the effects of 9/11.

[19] Orchard, Ramey and Wieland (2025*b*) and Orchard, Ramey and Wieland (2025*a*) label this path the "pessimistic forecast." We prefer *conditional forecast* because it describes what the object is: the model's conditional-mean path given the realized values of the exogenous variables. Conditioning on the adverse realizations of 2001 and 2008 lowers the forecast of consumption in those episodes relative to the unconditional forecast, but the construction itself is not inherently pessimistic.

Panel (b) repeats the exercise with nominal retail sales, the Census aggregate used throughout the main text, with the counterfactual constructed as in Section 3: realized sales less the induced spending implied by the Graham and McDowall (2026) daily MPC profile convolved with the path of rebate disbursements. This counterfactual remains inside the band in every month. Measured with the monthly source data and judged with forecast uncertainty, the 2001 episode carries no evidence against an MPC consistent with Graham and McDowall (2026).

Figure 12: 2001 rebate: ORW conditional forecast with full confidence bands



Notes: Panel (a) reproduces Figure 4, Panel A, of Orchard, Ramey and Wieland (2025*b*), with a 95 percent confidence band added. The data line is real spending on the Johnson, Parker and Souleles (2006) nondurable aggregate, re-based so that real equals nominal spending in May 2001; the solid line with circles and the dash-dotted line with triangles are the Orchard, Ramey and Wieland (2025*b*) implied rebate counterfactuals at the Johnson, Parker and Souleles (2006) micro MPCs of 38 and 66 percent, respectively. Panel (b) repeats the exercise with nominal retail sales excluding motor vehicles and parts dealers (FRED series MRTSSM44Y72USS), the series used in the main text, with the forecasting system estimated from 1992, when the series begins; the solid line with circles subtracts from realized sales the induced spending implied by the Graham and McDowall (2026) daily MPC profile convolved with the rebate disbursement path. In both panels the dashed line is the conditional forecast, solved dynamically from June 2001 and conditioning on realized NBER recession dates and a 9/11 dummy, and the shaded region is the 95 percent band from 2,000 Monte Carlo replications incorporating both parameter and shock uncertainty. Billions of dollars per month, seasonally adjusted.

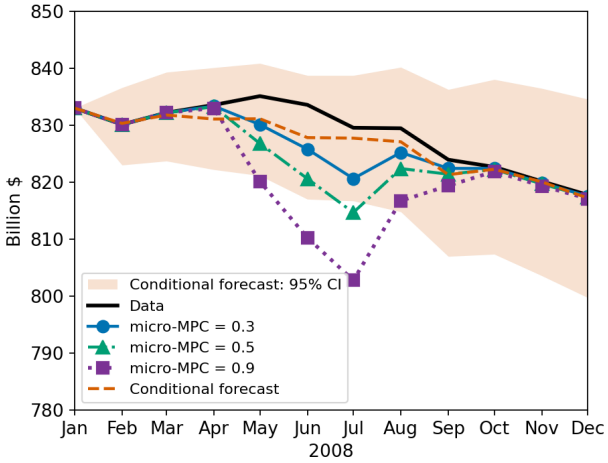
H.3. The 2008 rebate

Panel (a) of Figure 13 reproduces the top-left panel of Figure IV of Orchard, Ramey and Wieland (2025*a*). The black line is total real PCE at a monthly rate, re-based so that real equals nominal spending in January 2008, the anchor month; the conditional forecast is solved from February 2008, conditioning on realized recession dates, oil prices, and the Lehman dummy. The counterfactuals apply the rebate consumption responses excluding general equilibrium effects at micro MPCs of 30, 50, and 90 percent from the baseline model in Orchard, Ramey and Wieland (2025*a*). For the MPC of 30 percent, the counterfactual remains inside the band in every month, sitting roughly four billion dollars above the lower edge at its trough in July. The larger MPC of 50 percent pushes just past the lower edge of the confidence band at its trough, while the

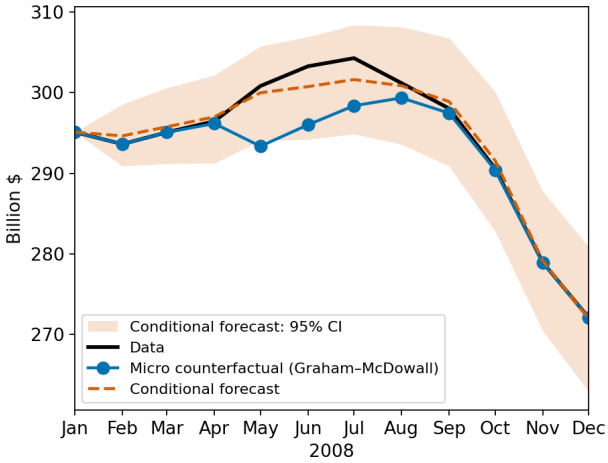
MPC of 90 percent sits well outside the realm of plausibility. Panel (b) substitutes nominal retail sales and the Graham and McDowall (2026) profile convolved with weekly check distributions. In this retail sales exercise, the micro counterfactual extends below the confidence bands in May, providing further evidence that the spending response to stimulus payments may be slower than the spending response to tax refunds measured in Graham and McDowall (2026).

Figure 13: 2008 rebate: ORW conditional forecast with full confidence bands

(a) Total real PCE



(b) Retail sales



Notes: Panel (a) reproduces the top-left panel of Figure IV of Orchard, Ramey and Wieland (2025*a*), with a 95 percent confidence band added. The data line is total real PCE at a monthly rate, re-based so that real equals nominal spending in January 2008; the lines with markers follow Orchard, Ramey and Wieland (2025*a*) in subtracting model-implied partial equilibrium rebate consumption responses at micro MPCs of 30 percent (solid, circles), 50 percent (dash-dotted, triangles), and 90 percent (dotted, squares) from the total. Panel (b) repeats the exercise with nominal retail sales excluding motor vehicles and parts dealers, the series used in the main text, with the forecasting system estimated from 1992; the solid line with circles subtracts from realized sales the induced spending implied by the Graham and McDowall (2026) daily MPC profile convolved with the rebate disbursement path. In both panels the dashed line is the conditional forecast, solved dynamically from February 2008 and conditioning on realized NBER recession dates, real oil prices, and a Lehman Brothers dummy, and the shaded region is the 95 percent band from 2,000 Monte Carlo replications incorporating both parameter and shock uncertainty. Billions of dollars per month, seasonally adjusted.

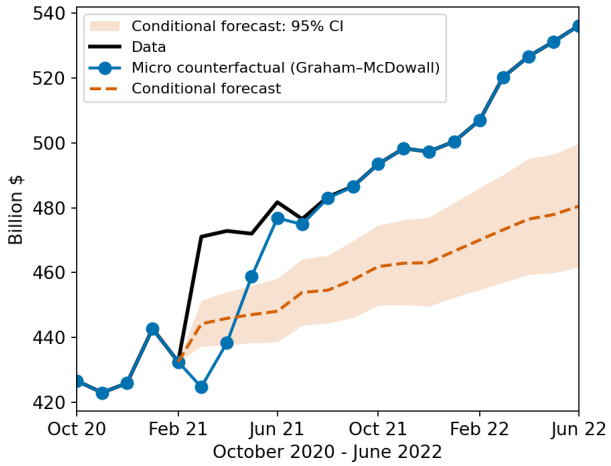
H.4. The 2021 payments

Figure 14 applies the 2008 forecasting system, unchanged, to nominal retail sales around the third round of Economic Impact Payments. The system is estimated from 1992 through December 2019, so that the collapse and rebound of 2020 do not contaminate the coefficients, and solved from March 2021, anchored at February 2021. Realized retail sales jump in March 2021 to well above the upper edge of the band and remain outside it in every month through June 2022. This is the same forecasting machinery that, in 2001 and 2008, could not distinguish realized spending from its conditional forecast.

The micro counterfactual reflects the findings documented in Sections 3 and 4: the front-loaded Graham and McDowall (2026) profile removes more spending in March than the gap between the data and the forecast, possibly because during this period households had accumulated significant liquid wealth and

therefore had a lower MPC than in normal times. Spending remains well above the confidence band around the conditional forecast throughout the period shown, likely reflecting the reopening of the economy. This is one motivation for the two-sided approach in the main text, which judges plausibility from the timing of month-to-month changes rather than the medium-term trend.

Figure 14: 2021 payments: retail sales, ORW 2008 forecasting system



Notes: The 2008 forecasting system of Figure 13, estimated on nominal retail sales from 1992 through December 2019 and solved dynamically from March 2021, anchored at February 2021. The solid line with circles subtracts from realized sales the induced spending implied by the Graham and McDowall (2026) daily MPC profile convolved with the March 2021 (EIP3) disbursement path; the January 2021 (EIP2) payments are left in the data, so the counterfactual coincides with realized sales before March 2021. The shaded region is the 95 percent band from 2,000 Monte Carlo replications incorporating both parameter and shock uncertainty. Billions of dollars per month, seasonally adjusted.