

Federal Reserve Bank of Chicago

Tracking U.S. Consumers in Real Time with a New Weekly Index of Retail Trade

Scott A. Brave, Michael Fogarty, Daniel Aaronson, Ezra Karger, and Spencer Krane

REVISED
June 18, 2021

WP 2021-05

https://doi.org/10.21033/wp-2021-05

[*] *Working papers are not edited, and all opinions and errors are the responsibility of the author(s). The views expressed do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.*

Tracking U.S. Consumers in Real Time with a New Weekly Index of Retail Trade

Scott A. Brave, Michael Fogarty, Ezra Karger,
Daniel Aaronson, Spencer Krane

Federal Reserve Bank of Chicago

June 18, 2021

Abstract

We create a new weekly index of retail trade that accurately predicts the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS). The index's weekly frequency provides an early snapshot of the MRTS and allows for a more granular analysis of the aggregate consumer response to fast-moving events such as the Covid-19 pandemic. To construct the index, we extract the co-movement in weekly data series capturing credit and debit card transactions, foot traffic, gasoline sales, and consumer sentiment. To ensure that the index is representative of aggregate retail spending, we implement a novel benchmarking method that uses a mixed-frequency dynamic factor model to constrain the weekly index to match the monthly MRTS. We use the index to document several interesting features of U.S. retail sales during the Covid-19 pandemic, many of which are not visible in the MRTS. In addition, we show that our index would have more accurately predicted the MRTS in real time during the pandemic when compared to either consensus forecasts available at the time, monthly autoregressive models, or other commonly-cited high-frequency data that aims to track retail spending. The gains are substantial, with roughly 50 to 75 percent reductions in mean absolute forecast errors.

Keywords: mixed-frequency dynamic factor model, retail sales, consumer spending
JEL: C32, C38, C43, C53, D12

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Corresponding author: Ezra Karger, Federal Reserve Bank of Chicago, Economic Research, 230 S. LaSalle St., Chicago, IL 60604, ezra.karger@chi.frb.org

1 Introduction

With the onset of the Covid-19 pandemic, many researchers turned to new high-frequency data sources to measure the impact of the virus on the U.S. economy. These alternative data provided the private sector and policymakers with valuable real-time information about the spread of Covid-19, its impact on economic activity, and the effects of the resulting federal, state, and local policies on households and businesses.

However, many of these data are not typically derived from samples designed to be representative of the population at large. Notably, credit and debit card transactions or electronic payments to businesses reflect the subset of consumers or businesses that use the particular service gathering the data. Accordingly, it can be problematic to draw inferences about aggregate economic conditions from these non-representative high-frequency time series.[1]

This paper addresses these concerns in the context of high-frequency indicators of U.S. retail sales. We create a new *weekly* index that accurately predicts the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS), a representative-sample survey that produces the gold standard measure of aggregate U.S. retail spending and is used as source data for U.S. GDP. The MRTS is a monthly index released with about a two-week lag. In contrast, our index is weekly, providing an early snapshot of consumer activity.[2] Moreover, the weekly frequency allows for a more granular analysis of fast-moving changes in consumer behavior than can be performed with monthly data, such as the stockpiling of goods immediately preceding the nationwide shutdowns in mid-March 2020 and consumer responses to the Covid-relief stimulus payments.

Our index uses alternative data from a number of private vendors: store revenue from Womply; financial transactions from Facteus and Consumer Edge; consumer sentiment from

[1] See, for example, the case studies in Brave et al. (2020). Other researchers have used carefully designed surveys to produce more representative samples but at the cost of a lower frequency of observation, e.g. Coibion et al. (2020), Bartik et al. (2020), and the U.S. Census Bureau Household and Small Business Pulse surveys described at https://www.census.gov/newsroom/press-kits/2020/pulse-surveys.html.

[2] For example, the March 2020 MRTS was released on April 15, 2020. This means that researchers and policymakers had to wait over a month before understanding how the onset of Covid-19 in early March 2020 affected aggregate consumer spending.

Morning Consult; and retail foot traffic from SafeGraph. We also use publicly available data on gasoline sales from the U.S. Energy Information Administration (EIA). The index is constructed using a mixed-frequency dynamic factor model that extracts the co-movement in these six weekly series and monthly retail and food services sales excluding automotive spending from the MRTS.[3] What is unique about our dynamic factor model is that it constrains the latent factor in a way that ensures that growth in the resulting weekly index matches the MRTS data at a monthly frequency. In doing so, our weekly index retains the representative-sample qualities of the MRTS for aggregate retail spending while still making use of the higher frequency information in our alternative data sources.

We use our weekly index to document several features of U.S. consumer spending during the Covid-19 pandemic, many of which are not visible in the MRTS:

- Retail sales in the U.S. plummeted at the same time as the rapid increase in Covid-19 cases and the closure of nonessential businesses in mid-to-late March 2020; they then began to recover by mid-April 2020. There was also a *stockpiling* effect, where retail sales *increased* substantially in the weeks leading up to the large declines in March. These intra-month patterns are not apparent in the lower frequency MRTS.
- As state and local governments lifted strict stay-at-home orders and the federal government disbursed the first round of economic impact payments, the recovery accelerated in April-June 2020. However, progress over the summer of 2020 was often uneven, with the weekly variation in sales coinciding with developments during the second wave of Covid-19 cases. This pattern is also difficult to see in the lower frequency MRTS, which averages across the ups and downs in weekly sales in July-September 2020.
- Sales fell again in the fourth quarter of 2020 as the third wave of the virus hit but began to recover in late December and early January 2021. The weekly patterns point to the importance of the second round of economic impact payments in boosting spending during the heart of this outbreak. The third round of payments that began in March 2021 further boosted spending and contributed to the substantial volatility in sales observed in our weekly index and the MRTS in the first quarter of 2021.

We also use the weekly index to *nowcast* the MRTS in real time.[4] Based on out-of-sample

[3]We exclude automotive spending from the MRTS in our index since consumer analysts typically track the automobile sector separately using industry reports on vehicle sales instead of the MRTS. Furthermore, vehicle purchases are not likely to be captured to any great degree in our high-frequency data sources.

[4]Our use of the word *nowcast* parallels that of Giannone et al. (2008) in referring to the contemporaneous forecasting of an infrequently published time series using information that is available in real time.

predictions, we show that our index more accurately predicted retail and food services sales ex. autos from the MRTS during the March 2020–April 2021 period than either the consensus forecasts available at the time or monthly autoregressive models. The gains are substantial, with roughly 50-75 percent reductions in mean absolute forecast error.

Our results contribute to several related literatures. A number of recent papers have used credit and debit card transactions data as an alternative high-frequency source to track consumer spending (e.g., Aladangady et al. (2019), Dunn et al. (2020), Chetty et al. (2020), and Carvalho et al. (2020)). Similarly, researchers have used mobility measures (e.g., Atkinson et al. (2020) and Alexander and Karger (2020)) and daily consumer sentiment to track consumer behavior during the Covid-19 pandemic. However, to the best of our knowledge, we are the first to combine such information into a single measure in a way that retains comparability to the U.S. Census Bureau's MRTS data. By using multiple sources of data, we more accurately track the underlying trend in aggregate retail spending relative to existing univariate alternatives, including those used by the BEA and Chetty et al. (2020).

Other work use high-frequency data to forecast consumer spending, retail sales, and GDP, including Ludvigson (2004), Croushore (2005), Galbraith and Tkacz (2018), and Lewis et al. (2020). Our work differs in that we combine weekly data using a multivariate and mixed-frequency data structure akin to the dynamic factor models of Giannone et al. (2008) and Aruoba et al. (2009). Furthermore, we modify a traditional mixed-frequency dynamic factor model to exploit the representative sampling of the MRTS in a way that is similar to what Aruoba et al. (2016) do for U.S. GDP. This allows us to overcome the non-representative sampling of our alternative data sources and contribute in a novel way to the literature on targeted data-dimension reduction using factor models, e.g. Bräuning and Koopman (2014).

2 High-Frequency Data Sources

Our Weekly Index of Retail Trade is targeted to match retail and food services sales excluding automotive spending (ex. autos) from the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS). The MRTS is based on a survey of about 13,000 businesses that sell merchandise and related services to final customers. The data record sales over the entire month and are published with about a two week lag. Sales include both online and in-store purchases.[5]

The MRTS is a vital snapshot of consumer spending on goods in the U.S. As the Census Bureau notes: "These data are widely used throughout government, academic, and business communities. The Bureau of Economic Analysis uses the estimates to calculate Gross Domestic Product. The Bureau of Labor Statistics uses the estimates to develop consumer price indexes and productivity measurements. The Council of Economic Advisers uses the estimates to analyze current economic activity. The Federal Reserve Board uses the estimates to assess recent trends in consumer purchases. The media use the estimates to report news of recent consumer activity. Financial and investment companies use the estimates to measure recent economic trends."[6]

To construct our index, we work with several high-frequency measures of retail activity:

- Revenues for more than 400 thousand small businesses, from Womply;
- Credit and debit card transactions from a sample of over 30 million unique cards, collected by Consumer Edge;
- Transactions from a collection of over 90 million debit, general purpose, single-issue gift, payroll, and government-issued cards from Facteus;
- Consumer sentiment from daily polls, conducted by Morning Consult;
- Retail foot traffic based on cell phone data, collected by SafeGraph; and
- Finished motor gasoline sales, from the U.S. Energy Information Administration.

[5]The sample of firms is chosen approximately every five years (most recently in 2018) and is then updated quarterly to reflect business "births" and "deaths." Added firms are typically represented in the sample with a nine month delay. Estimates are updated annually when the MRTS is benchmarked to the Annual Retail Trade Survey.

[6]See https://www.census.gov/retail/mrts/about_the_surveys.html. Wilcox (1992) describes the process by which the Bureau of Economic Analysis uses the MRTS.

2.1 Revenue and Transactions Data

Although they are correlated, the Womply, Consumer Edge, and Facteus datasets cover different scopes of businesses, consumers, and transactions, and hence each independently provides useful information for our index. To align ourselves as much as possible with the MRTS, we only use data for firms that are classified with either North American Industry Classification System (NAICS) codes or Merchant Category Codes (MCC).[7]

Turning to the specifics of each data source:

- The Womply data are from payment processors used by more than 400,000 primarily small brick and mortar establishments in the United States.[8] At about $400 billion, annual sales at Womply firms amount to a little less than 10% of our MRTS benchmark. Womply filters the data by removing businesses that are inactive or that exceed the Small Business Administration's size threshold.[9] Data are available at a daily frequency with a five- to seven-day lag.

- Consumer Edge tracks consumer spending from more than 30 million unique credit and debit cards. We use a version of the data in which the number of active cards is scaled to be consistent over time; this means that changes in spending will reflect changes in either the number of transactions per card or the size of transactions rather than the size of the underlying sample. The data capture about $60-65 billion in annual sales, or 1.3% of our MRTS benchmark. Data are available at a daily frequency with a five- to seven-day lag.[10]

- Facteus aggregates, anonymizes, and standardizes transaction-level data from a dozen banks, covering a total of more than 90 million debit cards, payroll cards, government benefit cards, and single-issue gift cards. The subset of Facteus transactions that we use amounts to about 1.25% of our MRTS benchmark.[11] The data are available at a daily frequency with a five- to seven-day lag.

[7]NAICS codes are available for Consumer Edge. Facteus and Womply are instead categorized by MCC, which allows us to exclude transactions categorized as automotive.

[8]Chetty et al. (2020) and Alexander and Karger (2020) also use the Womply data in their analysis of the pandemic's effects on economic activity.

[9]See https://www.sba.gov/document/support\OT1\textendashtable-size-standards for the definition of small business thresholds by industry.

[10]Consumer Edge is geared toward providing real-time data for investors. For more information, see https://www.consumer-edge.com/.

[11]Facteus' share of retail and food services sales ex. autos increased to over 2% in the first four to five months after the onset of the pandemic. We view this as consistent with the stabilization of lower-income households' spending by various income support policies, including extended unemployment insurance benefits, the Pandemic Unemployment Assistance program, and the $1200 economic impact payments that were all part of the CARES Act passed in March 2020. For more evidence on the impact of the stimulus payments on spending behavior using the Facteus microdata, see Karger and Rajan (2020).

Two issues suggest that a simple aggregation of these data would have difficulty matching the MRTS. First, sampling error could be large as, even together, the samples represent a small fraction of the spending covered by the MRTS. Second, and likely more important, systematic errors could arise because these data reflect nonrandom slices of the MRTS universe of retail establishments.

The MRTS's sample design and construction methodology are aimed at producing a representative sample of all U.S. retail and food services spending. In contrast, each of our alternative data providers record spending by whomever is covered by their service. For example, Womply measures activity only at small businesses, the Facteus data are weighted towards lower-income and younger consumers, and none of these data sources capture cash payments. Furthermore, these discrepancies cannot be eliminated by simple adjustments such as reweighting the alternative data using the same type-of-store weights that the Census Bureau uses, as these selection issues occur for every type-of-store category.

Instead, we construct a statistical index that treats the individual alternative data as "*noisy*" high-frequency measures of broader spending to isolate the common factor across datasets that is correlated with the MRTS. In this framework, the estimated factor loadings on the component series will capture systematic differences between the spending covered by the component sample and the MRTS universe, while the error term will reflect any remaining random variation idiosyncratic to each data source.

2.2 Consumer Sentiment and Mobility Data

To aid in identifying the common factor, our model includes several non-spending series that have also been widely used to examine the impact of the pandemic on economic activity – a consumer sentiment index from Morning Consult and retail foot traffic from SafeGraph[12] – as well as a publicly available weekly indicator of gasoline consumption from the U.S. Energy Information Administration (EIA). These three additional measures help align our

[12]See, for example, the Mobility and Engagement Index described in Atkinson et al. (2020) produced by the Federal Reserve Bank of Dallas that is based on the SafeGraph data.

high-frequency transaction data with the MRTS universe by capturing drivers of spending related to a broad base of U.S. households.

Morning Consult constructs consumer sentiment indices based on a nationally representative daily email survey of about 6,000 people across the United States. Their survey is based on five questions that are identical to those used in the monthly University of Michigan Survey of Consumers and is available daily, without a lag.[13] It thus brings sample design discipline and timely information on household conditions to our index. We use weekly averages of their overall index of consumer sentiment.

SafeGraph uses cell phone data to measure visits to individual business establishments. They record NAICS code identifiers for these locations, which allows us to measure the total number of visits (or foot traffic) to establishments defined to be in the MRTS universe.[14] We restrict the SafeGraph data to establishments operating in the relevant NAICS codes and normalize the number of visits by the number of devices that SafeGraph tracks in its sample.[15] Since cell phone usage is ubiquitous, this measure may provide a fairly representative measure of the population's visits to retail establishments. They also then would help capture retail sales that are not covered by credit and debit card transactions, such as payments with cash or check. They would not, however, capture online sales, which are included in the MRTS.

We also use EIA's weekly data for finished motor gasoline supplied in our index.[16] These data serve two purposes. First, they complement the SafeGraph data in capturing broad-

[13]For additional details, see https://morningconsult.com/2020/03/10/morning-consult-economic-intelligence-methodology/. More information on the University of Michigan Survey of Consumers can be found at https://data.sca.isr.umich.edu/.

[14]Over 99% of establishments in the SafeGraph data are categorized at the 4-digit NAICS code level. A summary of the foot traffic data available from SafeGraph can be found at https://www.safegraph.com/data-examples/covid19-commerce-patterns.

[15]To do this, we residualize the number of devices tracked in the SafeGraph sample by regressing the number of unique visits to establishments on the number of devices at the daily frequency before aggregating the data to our week definition.

[16]These data are measured in thousands of barrels per day and represent shipments from "*primary*" supply chain facilities (refineries, bulk storage terminals, etc.) to retail outlets and other secondary suppliers. They thus differ from retail sales of gasoline by changes in inventories at retail gasoline stations and the shipment to the other secondary suppliers. These are generally small—indeed, the Bureau of Economic Analysis (BEA) uses these data to help estimate personal consumption expenditures of gasoline in the national accounts.

based mobility. Second, the EIA data provide useful forecasting power for gasoline sales, which are a large and volatile component of the MRTS.

Figure 1 shows the six high-frequency time series in our seasonally adjusted Weekly Index of Retail Trade from January 2018–April 2021.[17] The figure makes clear the extent of heterogeneity across data sources, even during the onset of the Pandemic Recession in March 2020. Our index aims to combine these different sources in a representative way to explain national retail activity at a weekly frequency.

3 Modeling Weekly Retail Sales

The statistical framework underlying our Weekly Index of Retail Trade is a mixed-frequency dynamic factor model with one latent factor.[18] However, unlike a standard dynamic factor model, we impose an additional constraint that forces the latent factor to temporally aggregate to match the MRTS's monthly retail and food services sales excluding autos data. Below, we describe the model and its estimation.

3.1 Establishing a Regular Weekly Calendar

Weekly data pose a particular problem for mixed-frequency models. With standard methods of defining the calendar, some weeks span two months, which is problematic for temporally aggregating a weekly sales measure to a month.[19] Accordingly, we instead define an alternative calendar. To begin, we partition each month into four weeks by categorizing each of the first three seven-day periods within the month as weeks one through three and grouping the remaining seven to ten days into the fourth week of the month. This process results in

[17] Several of these time series exhibit significant seasonal components that have been removed in the figure using the methodology described in the Appendix.

[18] For additional information on dynamic factor models, see Stock and Watson (2011). Similar mixed-frequency models to ours include Aruoba et al. (2009) and Mariano and Murasawa (2003), who both estimate a latent factor describing changes in U.S. economic activity, as well as Brave and Butters (2012), who estimate a latent factor capturing U.S. financial conditions.

[19] Three common standards for defining weeks are: the number of seven-day periods from the first day of the year; a consistent Monday starting day; or a Sunday start day.

a regular calendar in which each week is identified with a unique month.

Within each of our weeks, we aggregate the daily high-frequency data by summing across the days within the week for the transaction and foot traffic data and by averaging across the days of the week for the consumer sentiment data.[20] For the motor fuel data, which are reported in calendar weeks, we disaggregate the weekly data to the daily frequency based on the average daily rate of fuel supplied for each week. We then re-aggregate this data into our four-week partition of each month.

Because the MRTS data are seasonally adjusted, we also seasonally adjust our weekly inputs before estimating our factor model. This seasonal pre-filtering turned out to be critical in extracting the signal from the volatile unadjusted input data. Our adjustment methodology accounts for the regular intra-month patterns in the data and the fact that the fourth week of each month contains a varying number of days. The seasonal adjustment procedure is described in detail in the Appendix.

3.2 Benchmarking to the MRTS

In our framework, growth in the weekly data ($W_{n,i,t}$) are *noisy* measures of growth (in log differences) in latent weekly MRTS spending ($F_{i,t}$), with n denoting the indicator series and i the week of month t. Each series loads onto the latent factor according to loadings γ_n with average growth rates α_n and measurement errors $\epsilon_{n,i,t}$ that follow first-order autoregressive processes,

$$W_{n,i,t} = \alpha_n + \gamma_n F_{i,t} + \epsilon_{n,i,t} \quad (1)$$

$$(1 - \psi_i L)\, \epsilon_{n,i,t} = \upsilon_{n,i,t} \quad (2)$$

[20]The Morning Consult data are a 5-day moving average. We back out daily values from the moving average and then re-average them to our weekly values. For details, see https://stats.stackexchange.com/questions/67907/extract-data-points-from-moving-average.

where L is a weekly lag operator. The factor follows a fourth-order autoregressive process with average growth rate α_F,

$$(1 - \rho_1 L - \rho_2 L - \rho_3 L - \rho_4 L) F_{i,t} = \alpha_F + \eta_{i,t}. \tag{3}$$

The shocks $\upsilon_{n,i,t}$ and $\eta_{i,t}$ are assumed to be jointly distributed $N(0, \Sigma)$.

To maintain the link with the MRTS data, the factor's weekly growth rates, $F_{i,t}$, are constrained by the *triangle average* formula:

$$\begin{aligned} M_{i=4,t} &= \frac{1}{4}F_{i=4,t} + \frac{2}{4}F_{i=3,t} + \frac{3}{4}F_{i=2,t} \\ &+ F_{i=1,t} \\ &+ \frac{3}{4}F_{i=4,t-1} + \frac{2}{4}F_{i=3,t-1} + \frac{1}{4}F_{i=2,t-1}. \end{aligned} \tag{4}$$

where $M_{i=4,t}$ is the log difference in monthly MRTS sales, which for modeling purposes we assume is observed in the fourth week of the month. (Though in calendar time, it is observed near the middle of the following month with the release of the Advance Monthly Retail Trade Survey, or MARTS.) This temporal aggregation constraint ensures that $F_{i,t}$ generates a monthly series with the same growth rate as the published MRTS data.

The above mixed-frequency dynamic factor model is then put into state-space form using the Harvey (1990) accumulator to handle the temporal aggregation constraint imposed on the latent factor. We estimate the model by maximum likelihood using a Kalman filter, where we further restrict $\rho(L)$ and $\psi_i L$ to be stationary processes.[21]

Unlike a standard mixed-frequency dynamic factor model, we do not include a measurement error term for the MRTS data. So, we are in essence recursively benchmarking the latent factor to the representative MRTS data. This restriction, however, also means that

[21] We use the Matlab toolbox MFSS described in Brave et al. (2020) for this purpose.

the model would be over-identified if we make the typical assumption that Σ is diagonal. Instead, we estimate the elements of Σ associated with the covariances between $\eta_{i,t}$ and each of the $\upsilon_{n,i,t}$.[22] This slight modification leads to an exactly identified model, as well as modest gains in our ability to nowcast the MARTS data. This improvement arises because the additional covariances contain useful information about the expected deviations of the MRTS data from their historical dynamics and, thus, helps to improve our high-frequency estimate of $\rho(L)$.[23]

Once the model has been estimated, to arrive at a weekly level of spending that is benchmarked to the MRTS data, we construct the weekly index I_k from the Kalman smoothed estimate of the weekly latent factor as:

$$I_k = \exp\left(M_0 + \sum_{j=1}^{k} F_j\right), \tag{5}$$

where k references the number of weekly observations for the index, $k = 1, ..., K$, and M_0 is the level of Census sales in the base period. Given the triangle average restrictions, the monthly average of the weekly index will approximately equal the monthly MRTS data, with the small difference attributable to the log transformation and the use of a monthly value, M_0, to initialize the index. We correct for this approximation, by rescaling the weekly index such that the average of its weekly values matches the monthly MRTS data when it is observed. Given the availability of our high-frequency data, we are able to produce an index that begins in the first week of January 2018.[24]

[22] The resulting near-diagonal Σ resembles the U.S. GDP measurement error model of Aruoba et al. (2016), who create an alternative measure called GDPplus by combining information from the product and income sides of the National Income and Product Accounts.

[23] For another example of the use of dynamic factor model methods for the purpose of benchmarking to official statistics, see Brave et al. (2021).

[24] M_0 corresponds to retail and food services sales ex. auto from the MRTS in December 2017. Most of our high-frequency time series begin in the first few weeks of 2018, with the exception of the SafeGraph series which begins in the second quarter of 2018. An unbalanced panel for estimation is not a concern for models of this type and just requires that a small modification be made to the Kalman filter as described in Brave et al. (2020).

3.3 The Estimated Index

Figure 2 shows our Weekly Index of Retail Trade as a seasonally adjusted monthly spending rate in comparison to retail and food services sales ex. autos from the MRTS over our entire sample (upper left) and selected time periods. The gray shading plots the number of newly diagnosed Covid-19 cases in the U.S., with the x-axis tick marks showing the first day of each month.

Several interesting features appear in the weekly index that are not readily apparent in the MRTS data. First, weekly sales in March 2020 were very uneven, rising substantially in the first half of the month and then plummeting in the second half (upper right). This is consistent with consumers' increasing awareness of the Covid-19 virus in mid-March (see, e.g. Keane and Neal (2021)) and anecdotal evidence of "*stockpiling*" just prior to the closure of nonessential businesses across much of the U.S. in the second half of March. These closures and the accompanying stay-at-home orders issued by many states and locales then coincides precisely with the sharp decline in the index in the latter half of March.

Subsequently, the weekly index begins to recover around mid-April 2020, as local communities lifted stay-at-home orders and the federal government expanded unemployment insurance and disbursed the first round of economic impact payments. In contrast, in the MRTS the impact of these developments are not apparent until the May data. Note, too, that the advance estimate of the May MRTS was published in mid-June, whereas the first estimate of the index for the second half of April was produced at the beginning of May. Accordingly, the weekly index identified the rebound in sales well in advance of the MRTS.

The recovery continued throughout the summer months of 2020, with the index returning to pre-pandemic levels by July 2020. However, as the figure shows, the recovery was not without some variation at the weekly frequency. In particular, when new cases of Covid-19 began to rise again during the summer, the index exhibited periods of decline in both July and August (lower left). This drop-off is not apparent in the MRTS data, because they average in the weeks of substantial positive growth both before and after the peak of the

summer wave of Covid cases.

The index also highlights the uneven nature of sales during the fourth quarter of 2020 and into the first quarter of 2021 (lower right). Sales fell sharply starting around Thanksgiving, corresponding to the resurgence in Covid-19 cases. Sales remained at a low level through much of December, but then improved markedly around the turn of the year during the week when many Americans received their second round of Covid-19 stimulus payments. After that initial burst, sales fell sharply again, driven in part by severe winter weather across much of the U.S. in February before recovering by the first week of March. They then continued to increase in March, leading up to the third round of Covid-19 stimulus payments—a potential sign of anticipation—before pulling back again in April.

As a policy-relevant application of our index, figure 3 compares the index's week-to-week movements over 2020–21 against the weekly average changes in 2018–19. The vertical dashed lines mark the weeks when each of the initial disbursements of each of the three rounds of economic impact payments took place over the past year—the CARES Act in April 2020, the CRA in January 2021, and the ARP in March 2021. We can see a clear connection between stimulus payments and retail spending, with sales jumping during all three episodes.[25] While the impact of the latter payments on retail sales are apparent in the MRTS as well (see figure 2), the CARES Act payments in April are not because the MRTS data is only available at a monthly frequency.

Our model's combination of time series and benchmarking to the MRTS provides a more stable reading of activity than two alternative measures of weekly retail sales, one from the Bureau of Economic Analysis (BEA) and one from the Opportunity Insights (OI) lab.[26] These measures also use credit and debit card data, but both come from a single data

[25] Although, the first stimulus payments were given out in the midst of a general recovery of consumer spending.

[26] See Aladangady et al. (2019) and Dunn et al. (2020) for more detail on the underlying data and construction of the BEA card spending estimates based on data from Fiserv First Data. Monthly and weekly data are available at https://www.bea.gov/recovery/estimates-from-payment-card-transactions. The OI estimates are based on data from Affinity Solutions and are described in more detail in Chetty et al. (2020). Daily and weekly data are available at https://www.tracktherecovery.org/.

provider with different coverage and use different statistical methods than our index.

Figure 4 compares our index to the BEA and OI series by plotting each series as a percent deviation from a pre-pandemic January 2020 baseline.[27] The indexes track one another and generally increase during weeks with stimulus payments, but ours is substantially less volatile in 2020 and the first half of 2021.[28] This suggests that using information from a range of data sources could do a better job tracking overall retail activity than any individual data series. For example, compared to the MARTS, the OI measure described in figure 1b of Chetty et al. (2020) has a root mean-squared error of close to 5 percentage points, whereas our indicator's out-of-sample RMSE since February 2020 is about half that size. As we show below, there is substantial variation over time in the contributions of different indicators to explaining the common factor in overall retail sales, which our multivariate structure exploits to produce forecasts of the MARTS.

3.4 Contributions of Components to the Weekly Index

Figure 5 plots the weekly contribution of each individual indicator to our overall index. The variation in the index that cannot be explained by the weekly series is attributed to the MRTS data; in turn, this contribution largely reflects the temporal aggregation constraint imposed on the mixed-frequency dynamic factor model. The figure highlights the wide and time-varying range of contributions from these different elements. For example, while the dynamics generated by the MRTS (dark blue) are often important, they fall short in accounting for the depth of decline of the index in March 2020 and contributed little to the rebound in April 2020. At these times, the EIA (purple), and Consumer Edge (orange) data made substantial contributions to the index.

[27]The BEA and Opportunity Insights measures include spending on automobiles and car parts (NAICS code 441) and exclude either gas station sales (BEA) or in some instances groceries (OI). Our Weekly Index of Retail Trade excludes autos. Both the BEA and OI data also use weeks defined as seven-day intervals starting from January 1 of each year. Our Weekly Index of Retail Trade instead uses the week definitions laid out above in section 3.1

[28]One caveat is that our index excludes auto sales, a component of retail sales known to be somewhat more volatile than other components of retail spending. Both the BEA and OI index include whatever auto-related spending is covered in their credit and debit card transaction data sources.

Table 1 decomposes the variance of the weekly growth rate of the index into the percentages explained by each of our six high-frequency data sources and the MRTS data over the January 2018–May 2021 period. The first column presents the results for the index depicted in figure 5, which is produced by the Kalman smoother. Almost 58 percent of its variation is explained by the MRTS, leaving about 42 percent to be explained by the high-frequency data. Of these data sources, the EIA data explain about 27 percent. While important for a number of specific episodes, together the transaction data only explain roughly 12 percent of the variation in the index, with the remaining roughly 2.5 percent accounted for jointly by SafeGraph and Morning Consult.

To assess the degree of uncertainty around the values in table 1, we perform a leave-one-out cross-validation exercise, re-estimating the model each time dropping one of the weekly data series. This produces a range of values showing the impact of omitting each weekly data source one-by-one. In this way, we can capture interdependencies in the estimation of our model when evaluating the relevance of the high-frequency data. The lower and upper bounds of these ranges are shown in table 1 in brackets. Overall, this exercise confirms that each of the high-frequency data sources brings value to the index, with the omission of each weekly series individually resulting in a higher variance explained by the MRTS data.

The Kalman smoother uses data from the entire sample to inform the index in any given period. Of course, in real-time analysis, data for (i, t) and beyond are unknown. To gauge the influence of this future information, the second column of table 1 presents variance decompositions for the Kalman (forward) filtered index, which is the expectation of the index's growth rate based on the data in hand prior to $(i - 1, t)$.[29] This decomposition is substantially different from that for the smoothed index. Here, the high-frequency data explain close to 98 percent of the variation, suggesting that much of the impact of the MRTS data on the smoothed index comes from the constraint matching them to the index at a monthly frequency.

[29]Data from after (i, t) are still used in estimation of the model parameters.

The filtered decompositions also are useful because they give us a sense of how the model may be expected to perform in forecasting the MRTS. The third column of table 1 shows a slightly different variance decomposition of nowcasts for just the months of March 2020–May 2021 taken from an out-of-sample forecasting exercise with our model. Viewed through this lens, the high-frequency data sources explain close to 75 percent of nowcast variance, meaning that the MRTS data, in particular their estimated dynamics, played a larger role during the pandemic than at other times in our sample. At a little more than 25 percent each, the EIA and Consumer Edge data account for the largest shares of nowcast variance explained during the pandemic, with Morning Consult explaining about 15 percent, and Womply and SafeGraph each contributing around 4 percent. We now turn to describing this nowcasting exercise in more detail.

4 Nowcasting the MARTS

Our results so far suggest that the mixed-frequency dynamic factor model could be a useful tool for predicting retail sales. We examine this question in a nowcasting exercise that iteratively re-estimates and forecasts in a fashion meant to mimic the real-time construction of the index during the pandemic period. We then compare its predictions for the Advance Monthly Retail Trade Survey (MARTS) against consensus forecasts available at the time and real-time monthly autoregressive forecasts.

4.1 Constructing Out-of-Sample Predictions

To nowcast the MARTS data, we use recursive estimates of the Kalman filter to project the weekly latent factor forward in time and obtain the resulting nowcast for month T, $M_{i=4,T}$. Forecasts can be made h weeks ahead of each MARTS release, where $h = 0, 1, 2, 3,$ or 4.[30]

[30]Technically, this is done within the Matlab toolbox MFSS by appending missing values for the $W_{n,i,t}$ to extend the estimation period by the number of weeks necessary to reach the end of the month being forecast. $M_{i=4,T}$ is then the end-of-month triangle average of the latent factor tracked in the state-space with an accumulator variable. For further details, see Brave et al. (2020).

Our procedure needs to account for the fact that even when we observe all of the weekly data for a month ($h = 0$), the MARTS data are not yet observed. This means that we cannot fully infer the current month's error structure, or $\eta_{i=1:4,t=T}$ and Σ to be precise. To do so, we use the simulation smoother of Durbin and Koopman (2012).[31] This procedure allows us to simulate multiple paths for the Weekly Index of Retail Trade by taking draws from a multivariate normal distribution of the model's shocks, and so produce coverage intervals as well as point forecasts. To obtain the intervals, we take percentiles of the distribution of $M_{i=4,T}$ from 1.5 million simulated paths. To illustrate, figure 6 shows our May 2021 nowcast of the MRTS data and its associated coverage intervals.

4.2 Out-of-sample Performance

We evaluate the out-of-sample forecasting performance of our nowcasts by comparing them to consensus forecasts available at the time and to predictions from a monthly autoregressive model of the MRTS data. The exercise could help reveal if our index suffers from overfitting, which clearly is a risk given our short sample period and large number of parameters.

For our first out-of-sample exercise, we consider previous forecasts of MARTS when all available weekly data through the end of the month are known, $h = 0$. We seasonally adjust all data with the seasonal factors estimated at the time, iteratively re-estimate the model with real-time values, and produce nowcasts for the MARTS releases between March 2020 and May 2021.

All together, the mean absolute error for our nowcasts since February 2020 is about 1.5 percentage points, with small differences depending on whether you evaluate their accuracy using the simple month-over-month percent changes in real-time or current MARTS values, respectively. This compares with an average absolute percent change in retail and food services sales ex. autos of about 4.5 percent over this period. January-April 2021 are the

[31] We also account for the impact of data revisions by augmenting the errors in the draws of the estimated end-of-sample variance of the model's latent states to reflect historical differences between the advance and final releases of the MRTS using data from the St. Louis Fed's ALFRED database to calculate the variance of revisions in our sample.

only months in our sample where our absolute errors are greater than 2 percentage points in both cases; excluding those four months, the mean absolute error of the other ten months was just a bit above 0.5 percentage point, versus an average absolute change in retail and food services sales ex. autos of still about 4.5 percent.

Figure 7 decomposes our MARTS nowcasts into the contributions from each of the high-frequency data sources and the lagged MRTS data. For ease of viewing, we present the data in terms of month-over-month percent (log) changes. The figure highlights the wide and time-varying range of contributions from the different components that we also saw for the smoothed index in figure 5. For example, the improvements in January and March 2021 retail sales were principally driven by increases in credit card spending and gasoline consumption, which makes sense during a time when stimulus payments were disbursed and vaccines became more accessible. In contrast, in March and April 2020, when lockdowns were instituted and then reversed, the foot traffic and consumer sentiment data made outsized contributions to the forecast. In addition, notably, the lagged values of MRTS played a substantial role only in the May and June 2020 and February and April 2021 nowcasts. This variation over time in contributions to the index highlights the value of using a multi-variate model as opposed to relying on a single indicator to gauge the evolution of aggregate retail spending.

Table 2 reports mean absolute error ratios comparing the $h = 0$ nowcasts (now expressed as simple month-over-month percent changes) against the median projections from two weekly consensus surveys taken on the Fridays prior to the release of the MARTS data during our sample period: Informa Global Markets (MCM Research) and Action Economics. Values less than 1 (greater than 1) in the table denote that our models produced a lower (higher) mean absolute error for the pandemic period.

We report two ratios for each comparison, with the first calculated based on the initial release of the MARTS data being nowcast and the second based on the MRTS data published as of June 15, 2021. The latter accounts for revisions to the MARTS data, which occur in the

two months following the initial release, as well as annual revisions that occur each April.[32] To establish statistical significance for the comparisons, we use the equal mean absolute error test proposed by Diebold and Mariano (1995) with the small sample size correction of Harvey et al. (1997) and standard errors computed with the Bartlett kernel with lag length set equal to sample size as suggested by Kiefer and Vogelsang (2005).

As seen in table 2, the nowcasts of the month-over-month percent change in the MARTS data outperform the consensus forecasts made during the pandemic by about 50 percent, for both surveys and releases of the Census data. Moreover, even with only 15 observations, the Diebold-Mariano test indicates that these differences are significant at standard confidence levels.[33] Thus, the nowcasts have a large and economically significant informational advantage relative to the surveys. While the informational gains are largest for the March–June 2020 period, excluding this period still results in substantial gains in nowcast accuracy.

We can also characterize the informational advantage provided by the high-frequency data over and above the MRTS data. To do so, we compare the accuracy of our nowcasts against one-month ahead forecasts from first-order autoregressive models for the MRTS. We consider two such specifications. The first estimates the model parameters using real-time data from January 2006 through February 2020 and holds them fixed when producing forecasts for March 2020 through April 2021. The second instead recursively updates these parameters by adding the previous month's data to the real-time estimation sample.

To capture the real-time data flow within each month, we produce five nowcasts based on the number of weeks of high-frequency data available. Given the mid-month calendar release of the MARTS and the lags in the availability of the high-frequency data, there is typically at most only one week of high-frequency data available for the current month at the time of release for the previous month's MARTS data. This situation corresponds to either the $h = 3$ or $h = 4$ nowcast. Starting from this baseline, we can then track the relative

[32]The data also undergo an annual revision that is published in April.

[33]One caveat is that we do not account for possible revisions to the high-frequency data. Note too that the results are slightly worse for a version of the model with a diagonal Σ.

importance of each subsequent week's worth of data.

Table 3 demonstrates that based on current MRTS values our dynamic factor model outperforms a monthly AR(1) forecast specification for the MRTS by about 20-30 percent when no weekly data are used in the model ($h = 4$) and an even greater 40-50 percent when only 1 or 2 weeks of high-frequency data are used in the model ($h = 3$ or $h = 2$). The gains steadily increase from there, peaking at about 70-75 percent when all four weeks' worth of high-frequency data are used ($h = 0$) in the model.

4.3 Sources of Forecast Error

While this forecasting performance is impressive, several potential sources of forecast error in our model are worth noting.

1. *Sample composition*: The high-frequency spending data we use collectively account for only a small fraction of the universe of retail sales covered by MRTS and the data are *not* representative of the universe of retail transactions. Our methodology attempts to address this issue, but these corrections are not perfect and a portion of our forecast errors will be caused by these differences in coverage.

2. *Parameter instability*: The wide variation in the data during the first few months of the pandemic mean that period will have an outsized effect on identification. This is an issue for all forecast models estimated using such data. Indeed, between March and June our parameter estimates jumped around considerably before settling down to relatively consistent values in the months since then. This parameter instability is another potential source of forecast error, especially for the March–June 2020 period.

3. *Factor dynamics*: Historically, there has been a moderate degree of persistence in month-to-month growth of the MRTS. The latent factor that represents weekly retail sales in our model follows an autoregressive process that reflects this pattern. These factor dynamics are a possible source of forecast error, especially during periods such as January-April 2021 where there were large month-to-month swings in retail sales. In these instances, the historical persistence in the underlying factor will induce a more stable forecast, even when the underlying weekly spending data point to a larger change in activity.

4. *Census seasonal factors and revisions*: The seasonal factors we use to adjust the weekly spending series are directly proportional to the MRTS seasonal factors (see Appendix). Since the MRTS seasonals are re-estimated concurrently, we must forecast them prior to the MARTS release. Any error in this forecast can propagate into our nowcasts. So, too, can large revisions to past values of the MRTS. At times during the pandemic

both the errors in seasonal factor forecasts and revisions to past MRTS data have been extremely large by historical standards.

Figure 8 decomposes our forecast errors ($h = 0$ nowcasts minus current MRTS data values) during the pandemic into each of these four sources. To create this decomposition, we produced two additional sets of out-of-sample forecasts using: 1) the real-time data but model parameters fixed at estimates made using the entire sample and; 2) the same fixed parameters with the latest MRTS data truncated to match its real-time availability with current seasonal factors. Additionally, we decompose the second set of forecasts into the contributions from the MRTS and the high-frequency data sources. Differences between the resulting forecasts and our original nowcasts are informative about the roles of parameter instability, data revisions, and the factor's autoregressive dynamics in explaining our forecast errors. We then attribute the residual forecast error to the difference between the sample composition of our high-frequency data sources and the universe of retail sales activity.[34]

As seen in the figure, the largest contributions to our nowcast errors is attributable to differences in the sample composition of our high-frequency data sources and the MRTS. This was most evident in the early stages of the pandemic in March, May, and June 2020, although the May and June errors were offset by those emanating from factor dynamics. Composition was also important in explaining large misses in January and March 2021.

Parameter instability played a large role in explaining forecast errors in the early months of the pandemic. As our parameter estimates settled down in the second half of 2020, this source of error became less impactful. However, it has been a more prominent feature of our forecast errors in late 2020 and early 2021.

Factor dynamic induced errors were large in May and June 2020, although as noted above, they largely counteracted the sample composition misses. Dynamics were also important in explaining our over-estimates of sales in February and April 2021 as well as our under-estimate in March 2021. In February and April, retail sales fell month-over-month

[34]At the time of writing, the March and April 2021 MRTS values were still subject to further revisions.

following particularly large stimulus-driven increases in the prior month. Here, the persistent autoregressive factor dynamics led the model astray, as both the January and March increases in sales were much less persistent. This was true as well for the February decline in sales that was driven at least in part by severe weather that substantially impacted the Southeastern U.S.

Revisions to the Census data and seasonal factors typically played a minor role, with the notable exception of March 2020 when there was a large error in our forecast for the monthly seasonal factor due to a large change in the Census seasonal factor. However, while this caused us to *underestimate* the drop in retail sales in March, several of the underlying weekly data series simultaneously *overestimated* the March decline and the two sources of forecast error largely canceled each other out.

5 Conclusion

We used weekly measures of small business revenues, credit and debit card transactions, consumer sentiment, cellphone-based mobility, and gasoline shipments to create a seasonally adjusted Weekly Index of Retail Trade that is benchmarked to monthly retail and food services sales (excluding automotive spending) from the Census Bureau's Monthly Retail Trade Survey (MRTS). To do so, we used a novel mixed-frequency dynamic factor model that combined the valuable real-time information from alternative high-frequency data sources of consumer activity with the representative sampling of the MRTS.

Using our weekly index, we showed that retail sales during the Covid-19 pandemic in the U.S. were highly variable at the weekly frequency, with a considerable amount of interesting variation masked at the monthly frequency by the MRTS. The weekly spending patterns that we documented align well with the timing of the waves of Covid-19 cases, restrictions imposed on businesses and households, and economic impact payments that we have seen during the course of the pandemic.

In addition, we found that the mixed-frequency dynamic factor model used to create our weekly index reliably nowcasts the Census Bureau's Advance Monthly Retail Trade Survey (MARTS) in real time, besting by about 50-70 percent in terms of lower mean absolute error both contemporaneous consensus forecasts and forecasts from monthly autoregressive models of the MRTS.

Going forward, our methodology can be extended to include finer levels of disaggregation, both geographically and by sales categories. For example, the Census Bureau has recently begun to release monthly state-level retail spending measures. Insofar as the high-frequency data sources we used can be disaggregated to the U.S. state level, estimating similar models for state retail sales would be a straightforward extension of our analysis.

Furthermore, we focused on the retail and food services sales ex. autos category of spending, but other categories of consumer spending may also prove to be forecastable. Tailoring our model to the sub-aggregates used by the BEA may be useful for forecasters and others interested in mapping at a high frequency into GDP the effects of policy changes and future business cycles on consumer spending.

References

Aladangady, A., S. Aron-Dine, W. Dunn, L. Feiveson, P. Lengermann, and C. Sahm (2019). From transactions data to economic statistics: Constructing real-time, high-frequency, geographic measures of consumer spending. *NBER Working Paper Series* (26253).

Alexander, D. and E. Karger (2020). Do stay-at-home orders cause people to stay at home? effects of stay-at-home orders on consumer behavior. *FRB of Chicago Working Paper* (WP-2020-12).

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016). Improving gdp measurement: A measurement-error perspective. *Journal of Econometrics 191*(2), 384–397.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Atkinson, T., J. Dolmas, K. Christoffer, E. Koenig, K. Mertens, A. Murphy, and K.-M. Yi (2020). Mobility and engagement following the sars-cov-2 outbreak. *Federal Reserve Bank of Dallas Working Paper* (2014).

Bartik, A. W., M. Bertrand, Z. B. Cullen, E. Glaeser, M. Luca, and C. Stanton (2020). How are small businesses adjusting to covid-19? early evidence from a survey. *NBER Working Paper Series* (26989).

Brave, S. and R. A. Butters (2012). Diagnosing the financial system: Financial conditions and financial stress. *International Journal of Central Banking 8*(2), 191–239.

Brave, S. A., R. A. Butters, and M. Fogarty (2020). The perils of working with big data and a small framework you can use to avoid them. *FRB of Chicago Working Paper* (WP-2020-35).

Brave, S. A., R. A. Butters, and D. Kelley (2020). A practitioner's guide and matlab toolbox for mixed frequency state space models. *Available at SSRN 3532455*.

Brave, S. A., W. Kluender, C. Gascon, and T. Walstrum (2021). Predicting benchmarked u.s. state employment data in real time. *International Journal of Forecasting. forthcoming*.

Bräuning, F. and S. J. Koopman (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis. *International Journal of Forecasting* (3), 572–584.

Carvalho, V. M., S. Hansen, A. Ortiz, J. R. Garcia, T. Rodrigo, S. R. Mora, and J. Ruiz (2020). Tracking the covid-19 crisis with high-resolution transaction data. *CEPR Working Paper Series* (214642).

Chetty, R., J. N. Friedman, N. Hendren, and M. Stepner (2020). The economic impacts of covid-19: Evidence from a new public database built using private sector data. *NBER Working Paper* (27431).

Cleveland, W. and S. Scott (2007). Seasonal adjustment of weekly time series with application to unemployment insurance claims and steel production. *Journal of Official Statistics 23*(2), 209 – 221.

Coibion, O., Y. Gorodnichenko, and M. Weber (2020). The cost of the covid-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending. *NBER Working Paper Series* (27141).

Consolvo, V. and K. G. Lunsford (2019). Residual seasonality in gdp growth remains after latest bea improvements. *Economic Commentary* (2019-05).

Croushore, D. (2005). Do consumer-confidence indexes help forecast consumer spending in real time? *The North American Journal of Economics and Finance 16*(3), 435–450.

Diebold, F. X. and R. S. Mariano (1995, July). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Dunn, A., K. Hood, and A. Driessen (2020). Measuring the effects of the covid-19 pandemic on consumer spending using card transaction data. *BEA Working Paper Series*.

Durbin, J. and S. J. Koopman (2012). *Time series analysis by state space methods*. Oxford university press.

Galbraith, J. W. and G. Tkacz (2018). Nowcasting with payments system data. *International Journal of Forecasting 34*, 366–376.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665 – 676.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Karger, E. and A. Rajan (2020). Heterogeneity in the marginal propensity to consume: Evidence from covid-19 stimulus payments. *FRB of Chicago Working Paper* (WP-2020-15).

Keane, M. and T. Neal (2021, January). Consumer panic in the covid-19 pandemic. *Journal of Econometrics 220*(1), 86–105.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2020). Measuring real activity using a weekly economic index. *Federal Reserve Bank of New York Staff Reports* (920).

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic Perspectives* (2), 29–50.

Mariano, R. S. and Y. Murasawa (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics 18*(4), 427–443.

Stock, J. H. and M. W. Watson (2011). *Dynamic Factor Models*. Oxford Handbook of Forecasting.

U.S. Census Bureau (2017). X-13 arima-seats reference manual. Technical report, Washington DC.

Wilcox, D. W. (1992). The construction of u.s. consumption data: Some facts and their implications for empirical work. *American Economic Review* (4), 922–941.

6 Tables and Figures

Table 1: Variances Explained by Data Source

Series	Smoothed	Filtered	Nowcast
Census	58.3	1.6	24.7
	(58.3, 82.0)	(1.6, 15.2)	
Consumer Edge	10.4	8.6	25.3
	(10.4, 21.9)	(8.6, 34.5)	
Facteus	0.7	0.4	0.4
	(0.1, 0.7)	(0.3, 1.4)	
EIA	27.3	26.2	25.6
	(4.3, 27.3)	(18.7, 32.9)	
Womply	0.9	19.0	4.5
	(0.9, 4.9)	(10.5, 19.0)	
Morning Consult	0.6	10.9	15.5
	(0.6, 3.6)	(10.9, 25.6)	
SafeGraph	1.8	33.4	3.9
	(1.8, 6.7)	(15.2, 33.4)	

Note: This table contains a variance decomposition of the smoothed, filtered, and out-of-sample nowcast for the growth rate of our Weekly Index of Retail Trade by data source. The percentage values in the table sum to 100 (within rounding error) by construction. The values in brackets for the smoothed and filtered variance decompositions are the lower and upper bounds from a leave-one-out cross-validation exercise of the weekly data sources. The smoothed and filtered growth rates cover the period from January 2018-May 2021. The out-of-sample nowcasts are for March 2020-May 2021.

Table 2: Mean Absolute Errors Relative to Consensus Forecasts

	Release	
Forecaster	First	Current
MCM Research	0.55**	0.58**
Action Economics	0.50**	0.53**

Note: This table shows mean absolute error (MAE) ratios for nowcasts from our mixed-frequency dynamic factor model. The ratios are expressed relative to the MCM Research and Action Economics survey median nowcasts of the month-over-month percent change in retail & food services sales ex. auto for both the First and Current releases of the Census Bureau's Monthly Retail Trade Survey from March 2020 through May 2021. The markers */** denotes statistical significance from a Diebold-Mariano test of equal mean absolute error (i.e. ratios equal to 1) at the 90%/95% confidence level using the Bartlett kernel with lag length equal to sample size and a small sample size correction. Both survey forecasts were obtained through a subscription to Haver Analytics. The MCM Research survey is maintained by Informa Global Markets.

Table 3: Mean Absolute Errors Relative to Monthly Autoregressive Models

	AR Type	
Weeks of Data (h Steps Ahead)	Fixed	Recursive
0 ($h = 4$)	0.82^{**}	0.68^{*}
1 ($h = 3$)	0.53^{**}	0.45^{**}
2 ($h = 2$)	0.51^{**}	0.42^{**}
3 ($h = 1$)	0.39^{**}	0.33^{**}
4 ($h = 0$)	0.32^{**}	0.26^{**}

Note: This table shows mean absolute errors (MAEs) for nowcasts of retail & food services sales ex. auto from our mixed-frequency dynamic factor model relative to monthly autoregressive specifications estimated over a fixed and recursive window each with 1 lag. The fixed-window model was estimated over a sample period from January 2006 through February 2020 and its coefficients held fixed to produce one-month ahead forecasts over the period from March 2020 through May 2021. The recursive-window model adds one month of data to the estimation sample iteratively from March 2020 through April 2021 and uses the updated model coefficients to produce one-month ahead forecasts. To produce similar forecasts from our mixed-frequency dynamic factor model we estimate the model separately including from zero to four weeks of available weekly data for the month being nowcasted. The markers */** denotes statistical significance from a Diebold-Mariano test of equal mean absolute error (i.e. ratios equal to 1) at the 90%/95% confidence level using the Bartlett kernel with lag length equal to sample size and a small sample size correction. Retail & food services sales ex. auto data used to construct mean absolute errors were obtained from Haver Analytics.

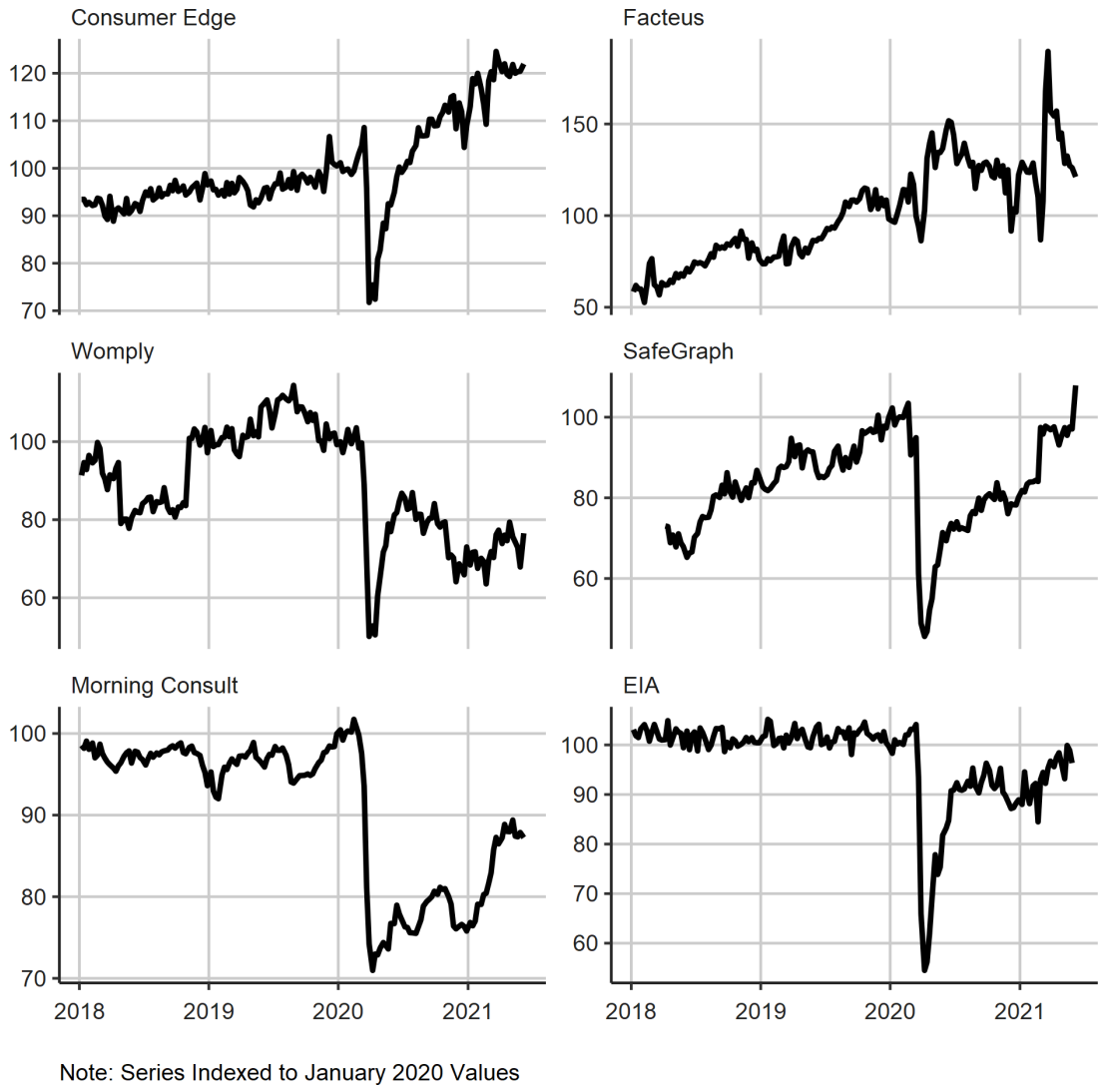


Figure 1: Weekly Data Series, by Source

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: Daily data are aggregated to weekly and seasonally adjusted with the procedure described in the appendix. All of the data series as shown in the figure are set equal to 100 in January 2020.

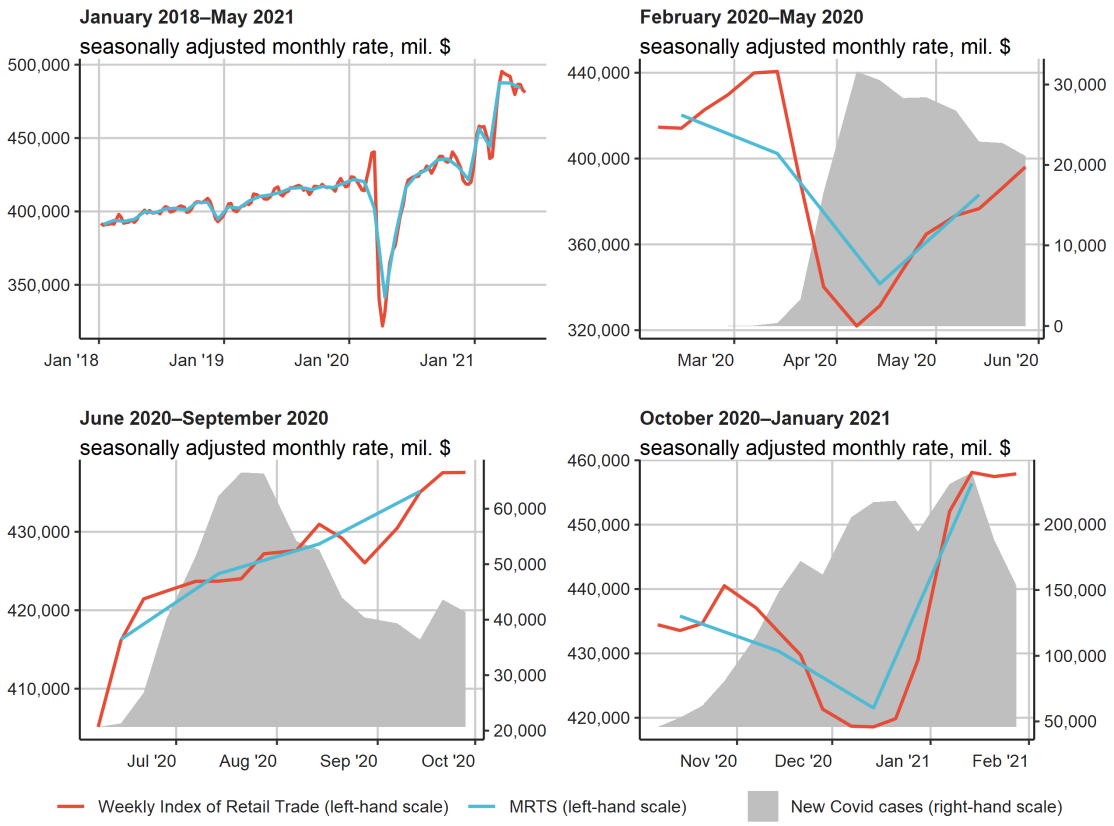


Figure 2: Retail and Food Services Sales ex. Auto

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: X-axis ticks correspond to the first day of the month. We align the monthly MRTS data to the second week of each month to emphasize the temporal aggregation constraint that requires the monthly average of our Weekly Index of Retail Trade to equal the MRTS data. The gray shaded area represents the weekly average of new Covid-19 cases in the United States for various time periods.

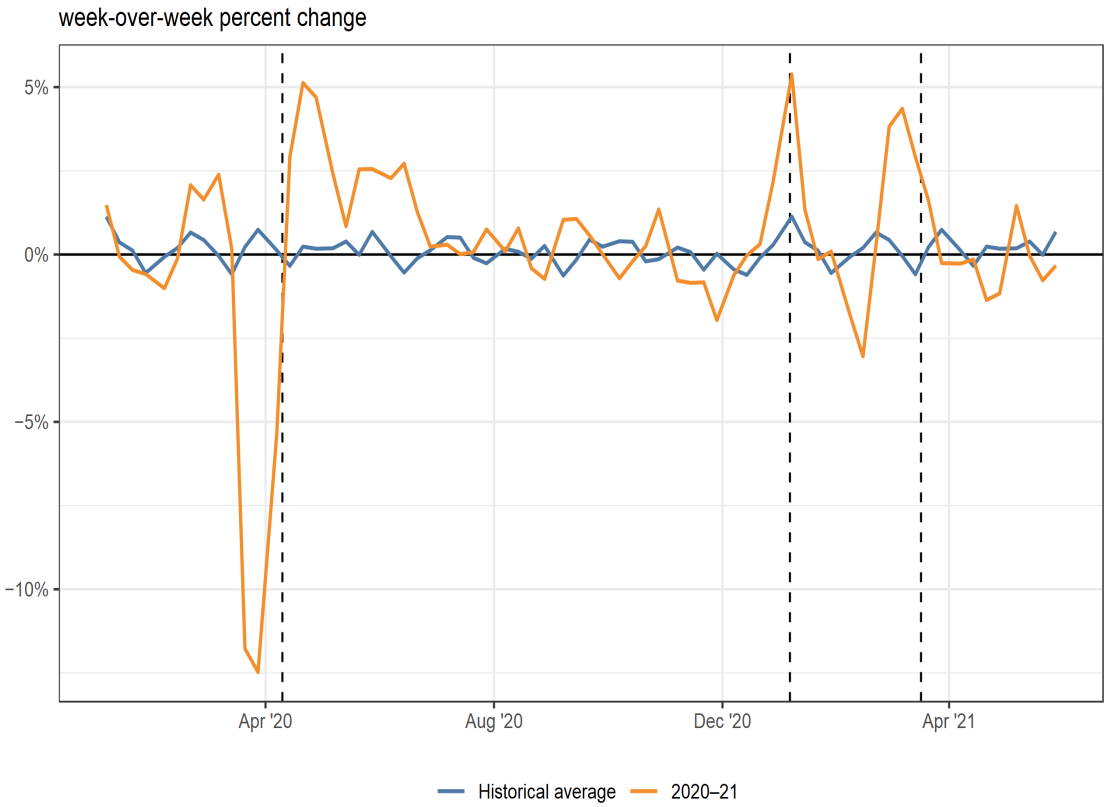


Figure 3: Weekly Retail Sales and the Timing of Economic Impact Payments

Source: Authors' calculations based on data from the U.S. Census Bureau, Energy Information Administration, and Department of the Treasury; Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: The figure compares week-over-week percent changes in our Weekly Index of Retail Trade over the 2020–21 period relative to the average of the same weeks of 2018 and 2019 in order to highlight the impact of the various rounds of Economic Impact Payments on retail and food services sales ex. auto. The first vertical line corresponds to April 10, 2020, when the CARES Act funds were initially disbursed; the second to January 6, 2021, when the second round of payments authorized in the 2020 year-end omnibus federal spending bill were deposited; and the third line to March 17, 2021, the date when the payments authorized in the American Rescue Plan were initially disbursed. All initial payment disbursement dates were determined from the U.S. Treasury's Daily Treasury Statement using data from Haver Analytics.

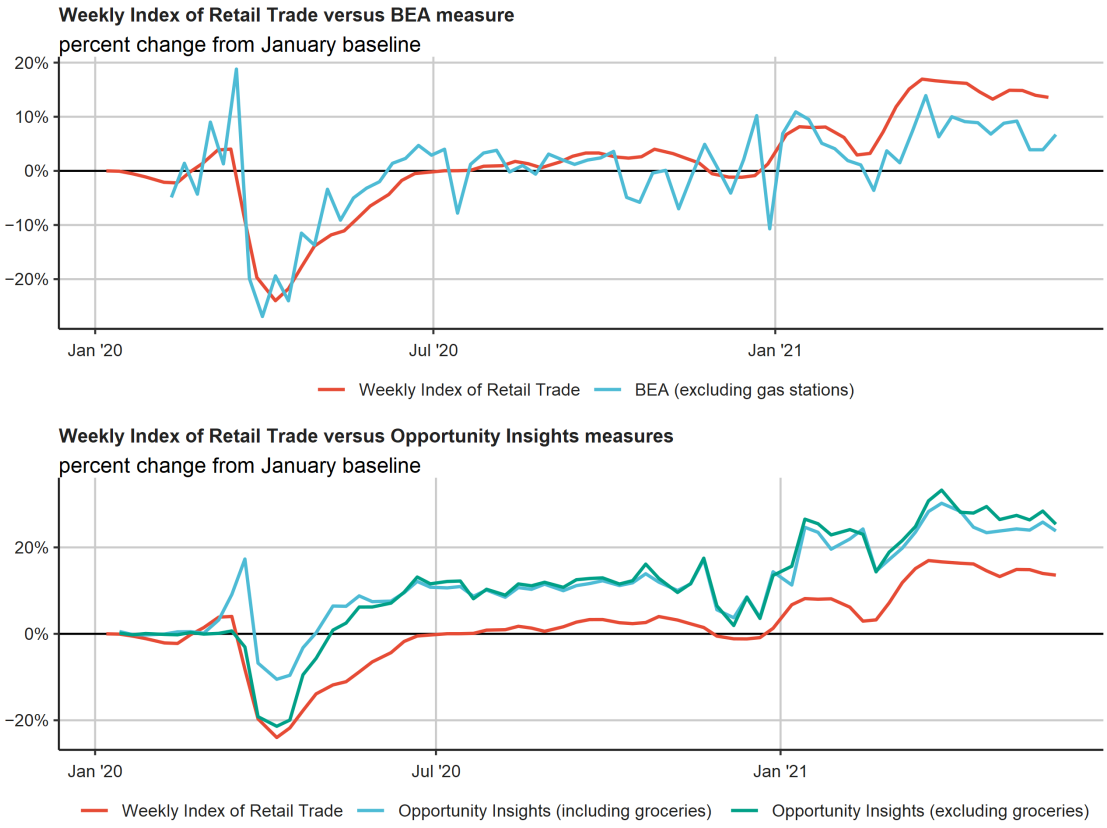


Figure 4: Weekly Measures of Retail and Food Services Sales

Source: Author's calculations. Bureau of Economic Analysis and Opportunity Insights were obtained via Haver Analytics.

Note: The figure compares relative to a pre-pandemic baseline three weekly measures of retail and food services sales: our Weekly Index of Retail Trade and two different measures of credit and debit card spending from the U.S. Bureau of Economic Analysis (BEA) and the Opportunity Insights (OI) lab.

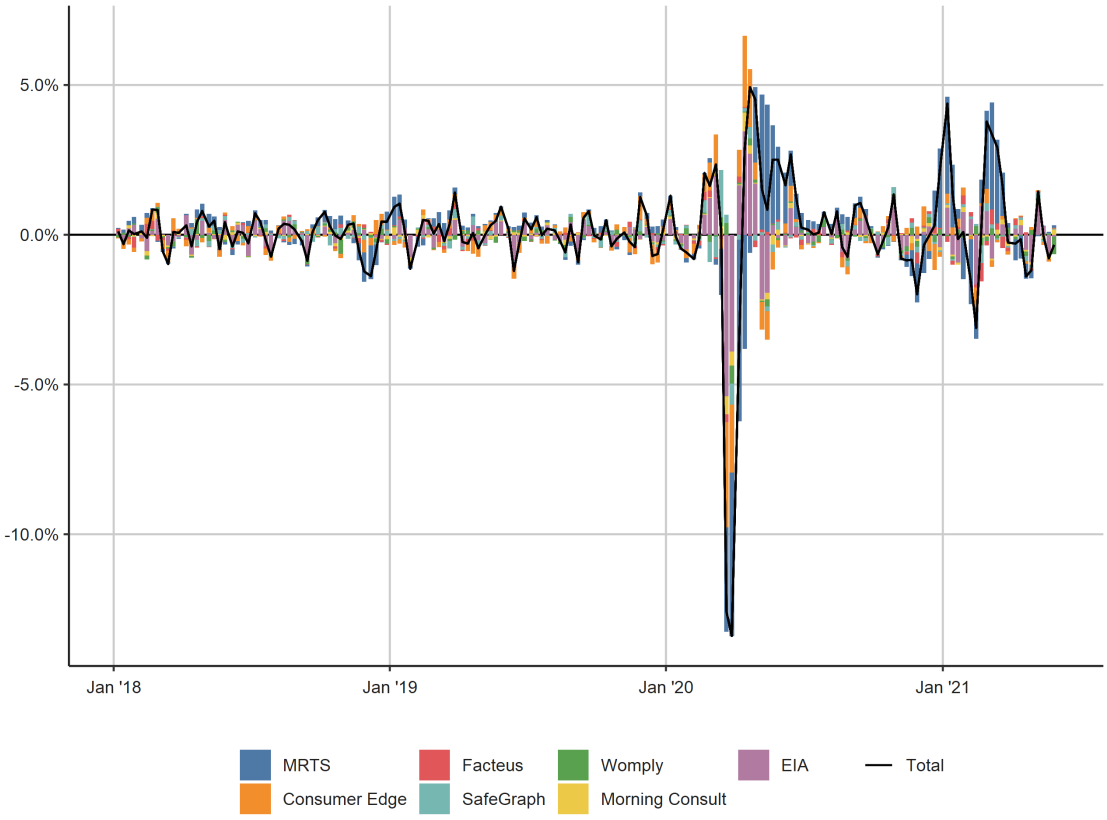


Figure 5: Contributions to Growth for the Weekly Index of Retail Trade

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: The contributions sum to the Kalman smoothed estimate in each week of the percent (log) change of the Weekly Index of Retail Trade. The method of calculation for these contributions is described in detail in Brave, Butters, and Kelley (2020).

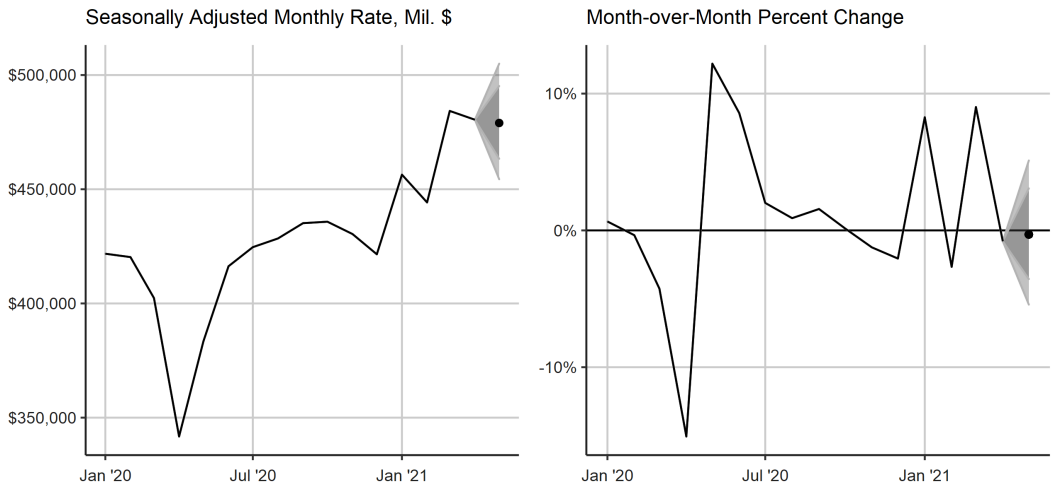


Figure 6: April 2021 Nowcast for Retail and Food Services Sales ex. Auto

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: The figure shows our May 2021 nowcast using weekly data available through end of the month. The shaded regions represent 70% (dark gray) and 90% (light gray) coverage intervals.

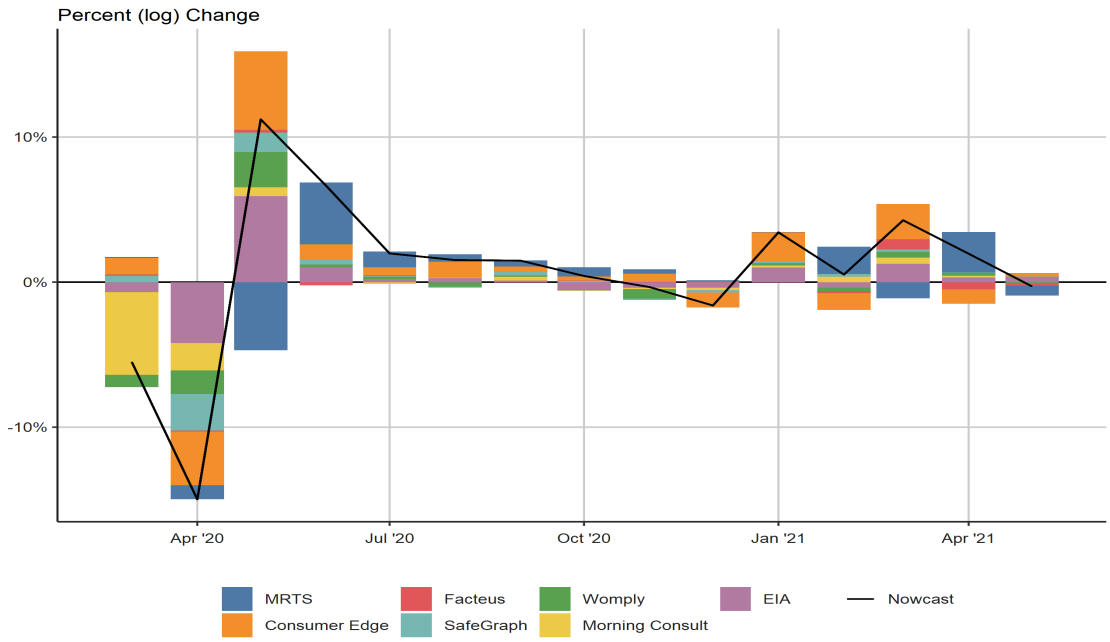


Figure 7: Contributions to Growth for our MARTS Nowcasts

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: The figure displays contributions to our MARTS nowcasts for March 2020 through May 2021 using the weekly data available through the end of each month and previously released MRTS data. The contributions in the figure sum to the nowcast value by construction.

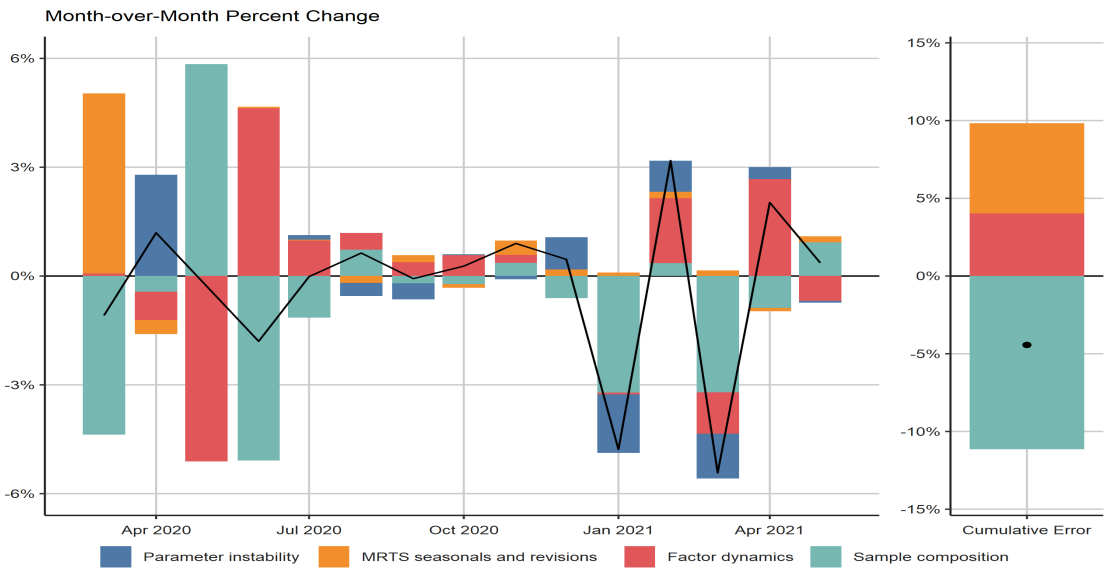


Figure 8: Nowcast Error Decomposition

Source: Authors' calculations based on data from the U.S. Census Bureau and Energy Information Administration, Consumer Edge, Womply, Facteus, Morning Consult, and SafeGraph.

Note: The total nowcast error is defined as the (revised) MRTS value minus our model's nowcast. The portion of the forecast error attributed to parameter instability is defined as the difference between the nowcast and the value from an alternative nowcast where our model parameters have been fixed at their current estimates. The error due to Census revisions is defined as the difference between the alternative nowcast with the fixed model parameters and another alternative nowcast that uses both the fixed parameters and the revised, rather than real-time, MRTS data. We attribute the remaining nowcast error to a combination of sample composition and the factor's dynamics, such that all four sources of error sum to the total nowcast error in each month. The cumulative error shown at the far-right of the figure sums over the forecast error contributions for March 2020 through May 2021.

7 Appendix

7.1 Seasonal Adjustment of Weekly Data

Seasonal variation is large in retail sales, yet none of the high-frequency data are available on a seasonally adjusted basis. Standard seasonal adjustment procedures, such as the X13 ARIMA-SEATS program published by the Census Bureau (U.S. Census Bureau, 2017), are not well suited to the seasonal adjustment of high-frequency data series. In our case, we are also handicapped by the very short history of our time series. For instance, the method described in Cleveland and Scott (2007) that is commonly used to seasonally adjust weekly time series like unemployment insurance claims requires much more than the roughly two years of data that we have available.

However, we are still able to seasonally adjust our weekly data by making two assumptions. First, we assume that the seasonal factors for our weekly data when temporally aggregated to the monthly frequency are the same as the MRTS seasonal factors. Second, we assume that the weekly seasonal factors are a specific function of the monthly seasonal factor, the number of days in the week, and a week-of-the-month effect.

Below we describe how we use the MRTS seasonal factors and week-of-the-year effects to seasonally adjust the high-frequency data sources. This procedure reliably removes the majority of seasonal patterns in our transaction and foot traffic data. We make no adjustments to the EIA data, which are already seasonally adjusted using the Cleveland and Scott (2007) method by Haver Analytics, nor the Morning Consult data, which do not exhibit seasonality.

Derivation of the Weekly Seasonal Factors

To account for seasonality in the high-frequency data, we incorporate the monthly seasonal factors from the Census Bureau's retail and food services sales ex. autos series.[35] We assume that the observed monthly Census seasonally adjusted (SA) and non-seasonally adjusted

[35]The seasonal factor for month is defined as , where is the seasonally adjusted Census series and is the unadjusted Census series.

(NSA) time series are the sum of latent (unobserved) weekly series:

$$\sum_{i=1}^{4} C_{i,t}^{NSA} = C_t^{NSA}$$
$$\sum_{i=1}^{4} C_{i,t}^{SA} = C_t^{SA}$$

where $C_{i,t}^{NSA}$ and $C_{i,t}^{SA}$ are NSA and SA weekly Census retail sales, C_t^{NSA} and C_t^{SA} are NSA and SA monthly Census retail sales, and the subscripts t and i index months and weeks, respectively.

The seasonal factors are multiplicative in levels (additive in logs) at both the weekly and monthly levels

$$C_{i,t}^{SA} = C_{i,t}^{NSA} \cdot SF_{i,t}^{C}$$
$$C_t^{SA} = C_t^{NSA} \cdot SF_t^{C}$$

where $SF_{i,t}^{C}$ is the latent weekly Census seasonal factor and SF_t^{C} is the monthly Census seasonal factor. Together with our assumption that monthly Census retail sales are the sum of latent weekly sales, this yields the following relationship between the monthly and weekly data and seasonal factors:

$$\begin{aligned} C_t^{NSA} \cdot SF_t^{C} &= \sum_{i=1}^{4} C_{i,t}^{NSA} \cdot SF_{i,t}^{C} \\ \Rightarrow \qquad SF_t^{C} &= \sum_{i=1}^{4} \frac{C_{i,t}^{NSA}}{C_t^{NSA}} \cdot SF_{i,t}^{C} \\ \Rightarrow \qquad SF_t^{C} &= \sum_{i=1}^{4} \frac{\left(C_{i,t}^{NSA}/D_{i,t}\right)}{\left(C_t^{NSA}/D_t\right)} \frac{D_{i,t}}{D_t} \cdot SF_{i,t}^{C} \end{aligned}$$

where $D_{i,t}$ is the number of days in week i of month t and D_t is the number of days in month

t . The first term in the sum of the last expression is the ratio of sales per day in week i to sales per day in the month. This will reflect week fixed effects. For example, weeks 2 and 4 may have greater sales because they coincide with pay periods. The second term is the ratio of days in the week to days in the month – in our week dating scheme, the last week of the month has a variable number of days, and this adjusts for the fact that weeks can make up a different share of the total days in the month across months of different length.

The goal of the weekly Census seasonal factor, $SF^C_{i,t}$, is to offset these two calendar sources of weekly sales variation. This suggests setting the weekly seasonal adjustment factors according to the formula:

$$SF^C_{i,t} = \frac{1}{4} \frac{1}{\text{avg}_{i,t}\left[\frac{\left(C^{NSA}_{i,t}/D_{i,t}\right)}{\left(C^{NSA}_{t}/D_{t}\right)}\right]\frac{D_{i,t}}{D_t}} \cdot SF^C_t$$

Here, $\text{avg}_{i,t}$ refers to the average of the ratio of sales per day in week i to sales per day in the month, where the average is taken over some window of months.

We can perform a related exercise to seasonally adjust the high-frequency data. We exploit this structure and the Census monthly seasonal factors to set weekly seasonal factors for $W^{NSA}_{i,t}$, the NSA high frequency weekly retail sales data, in order to create a SA series, $W^{SA}_{i,t}$:

$$SF^W_{i,t} = \frac{1}{4} \frac{1}{avg_{i,t}\left[\frac{\left(W^{NSA}_{i,t}/D_{i,t}\right)}{\left(W^{NSA}_{t}/D_{t}\right)}\right]\frac{D_{i,t}}{D_t}} \cdot SF^C_t$$

$$W^{SA}_{i,t} = HF^{NSA}_{i,t} \cdot SF^W_{i,t}$$

$$W^{SA}_{t} = \sum_{i=1}^{4} W^{SA}_{i,t}$$

where $SF^W_{i,t}$ and SF^W_t are weekly and monthly seasonal factors and $\text{avg}_{i,t}$ is taken over all of

the week i 's of month t in the data sample (for which we have a complete month of data). Averaging the weekly sales share ratios at the week-by-month level allows us to control for varying distributions of sales within the month, which are most noticeable during specific times such as the end-of-year holiday period.

Our approach still has several implementation issues. One, for example, was whether or not to include holidays in the $D_{i,t}$ and D_t variables. Through testing, we found that including the holidays produced better results. A second issue is that SF_t^C , the monthly Census seasonal factor, is not available until the MARTS is published, and so we must estimate it to construct nowcasts. The Census Bureau seasonally adjusts retail sales on a concurrent basis; that is, every month they re-rerun their programs and estimate new factors using all the available data. These programs also produce forecasts for future seasonal factors, which the Census Bureau publishes for the closely related retail and food services sales category. Accordingly, we forecast SF_t^C with a regression using this seasonal factor and 12 lags of the seasonal factor for our sub-aggregate as explanatory variables. In practice, this produces accurate forecasts of SF_t^C with one exception, when Census substantially revised their seasonal factors in March 2020 at the onset of the pandemic. To illustrate, over the nine-month period from January 2020 through September 2020, the mean absolute percent error for our forecasts of the monthly Census seasonal factor was 2.6% . However, excluding the month of March reduces this to just 0.6% .

Residual Seasonality

Even after implementing the seasonal adjustment procedure described above, several of our weekly data series still contain a degree of residual seasonality. Residual seasonality is a phenomenon where there is a predictable seasonal pattern left over in a data series that has already been seasonally adjusted by some standard procedure (Consolvo and Lunsford, 2019). Residual seasonality shows up in the various weekly data series that we use, most noticeably around the first week of the year. This could arise, for example, because the

spending patterns for credit and debit card transactions around the end-of-year holiday period do not exactly mirror the seasonal pattern in the Census retail sales measure.

To account for residual seasonality, we regress each of the spending and foot traffic series, in log growth rates, on a constant and regressors for important retail holidays that move within the year, such as Easter and Thanksgiving, as well as individual week dummies for a small set of weeks that appear associated with outliers for individual series. To get the series back to levels, we take the residuals from these regressions, add back in the mean growth rate taken out by the constant in the regression, and cumulate them forward from the series' initial log-level. We then exponentiate the series to get it back into natural units.