

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

March 2, 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 Daniel Aaronson Ezra Karger Spencer Krane

March 2, 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 implications of policies implemented during the Covid-19 pandemic. To construct the index, we extract the co-movement in several weekly data series capturing credit & debit card transactions and revenues, mobility, and consumer sentiment as well as monthly retail and food services sales excluding automotive spending (ex. autos) from the MRTS. To ensure that the index remains representative of the sample of firms covered in the MRTS, we use a mixed-frequency dynamic factor model constrained to match the MRTS growth rate at a monthly frequency. We document several interesting features of U.S. retail sales during the 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 the consensus forecasts available at the time or monthly autoregressive models. The gains are substantial, with approximately 70 to 85 percent reductions in mean absolute forecast errors.

JEL Codes: C32, C38, C43, C53, D12

Keywords: dynamic factor model, high-frequency data, retail sales, consumption, Covid-19

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 sources provided private sector analysts and policymakers with valuable real-time information about the effect of the spread of Covid-19 on economic activity. Researchers also used these data to evaluate the impacts of federal, state, and local policies on households and businesses.

Though extremely useful, these data are not typically derived from samples designed to be representative of the population at large. Notably, credit and debit card transactions or business electronic payments 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 time series.[1]

This paper addresses these concerns in the context of high-frequency indicators of retail sales. We create a new *weekly* index that accurately predicts the U.S. Census Bureau's Monthly Retail Trade Survey (MRTS), a carefully designed representative-sample survey that produces the gold standard measure of aggregate U.S. retail spending. The MRTS is released monthly with about a two-week lag. In contrast, our index can be produced weekly, thus providing an early snapshot of consumer activity.[2] Moreover, the weekly frequency of our index also allows analysis of aggregate effects of important changes in policy -- such as mandated lockdowns or the Covid-relief package stimulus payments -- at a more granular level than when using monthly data.

Our index uses alternative data from a number of private vendors: revenue and transaction data from Womply, Facteus, and Consumer Edge; consumer sentiment from Morning Consult; and retail foot traffic from SafeGraph.[3] In addition, we also use publicly available motor fuel consumption data from the U.S. Energy Information Administration (EIA).[4] To construct the index, we estimate a mixed-frequency dynamic factor model that extracts the co-movement in these weekly series and seasonally adjusted monthly retail and food services sales excluding automotive spending (ex. autos) from the MRTS. Furthermore, the model constrains the latent factor in a way that ensures that growth in the resulting Weekly Index of Retail Trade matches the MRTS data at a monthly frequency.

---

[1] See, for example, the case studies in Brave, Butters, and Fogarty (2020).

[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.

[3] All of these alternative data sources have been used extensively during the pandemic to track economic conditions in real time (e.g. Atkinson et al. (2020) and Chetty et al. (2020)).

[4] The EIA data can also be found in the weekly index of Lewis et al. (2020).

We first use our weekly index to document several interesting 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 stay-at-home orders and the federal government disbursed the first round of economic impact payments, the pace of recovery accelerated in April-June. However, progress over the summer of 2020 was uneven, with the variation in weekly sales coinciding with the second wave of Covid-19 cases.
- 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, with the weekly pattern pointing to an important role for the second round of economic impact payments.

We next use the weekly index to "nowcast" the MRTS in real time.[5] Based on out-of-sample monthly predictions, we show that our index more accurately predicted retail and food services sales ex. autos from the MRTS during the March 2020–January 2021 period than either the consensus forecasts available at the time or monthly autoregressive models. The gains are substantial, with approximately 70 to 85 percent reductions in mean absolute forecast error over the other projections.

Our 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). We use several high-frequency data sources of retail sales to inform the index's within-month dynamics:

- Revenue data for more than 400 thousand small businesses, from Womply;
- Credit 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, payroll, government cards, and single-issue gift 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 product supplied, from the U.S. Energy Information Administration.

---

[5] Our use of the word "nowcast" parallels that of Giannone, Reichlin, and Small (2008) in referring to the contemporaneous forecasting of a macroeconomic time series that is published infrequently using information that is available in real time.

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 in the construction of our index. To align ourselves as much as possible with the MRTS, we only include data for firms that are classified with either North American Industry Classification System (NAICS) codes or Merchant Category Codes (MCC) that are included in the retail and food services excluding autos aggregate sales measure.[6]

Turning to the specifics of each data source:

- The Womply data are from payment processors used by more than 400,000primarily brick and mortar establishments in the United States.[7] 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.[8] 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 cards. We use a version of their data in which the number of active cards is scaled to be consistent over time, allowing for changes in spending reflecting 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 about 1.3% of our MRTS benchmark. Data are available at a daily frequency with a five- to seven-day lag.[9]
- 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.[10] The data are available at a daily frequency with a five- to seven-day lag.

---

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

[7] 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.

[8] See https://www.sba.gov/document/support–table-size-standards for the definition of small business thresholds by industry.

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

[10] Facteus' share of retail & 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

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, they represent a small fraction of the spending covered by the MRTS. Second, and likely more important, systematic errors could arise because they reflect nonrandom slices of the MRTS universe.

The MRTS's sample design and data 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, and the Facteus data are weighted towards lower-income and younger consumers. Moreover, by construction, none of these data sources capture cash payments. These discrepancies cannot be eliminated by 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 factor model 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.

Consumer Sentiment, Mobility, and Gasoline Data

To aid in identifying the common factor, our model includes two non-spending series that have also been widely used to examine the impact of the pandemic on economic activity -- consumer sentiment from Morning Consult and retail foot traffic from SafeGraph[11] – 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 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.[12] It thus brings

---

Assistance program, and the $1200 economic impact payments that were all part of the CARES Act passed in March. For more evidence on the impact of the stimulus payments on spending behavior using the Facteus microdata, see Karger and Rajan (2020).

[11] See, for example, the Mobility and Engagement Index described in Atkinson et al. (2020) that is based on the SafeGraph data.

[12] 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/.

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 total visits (or foot traffic) to establishments defined in the MRTS data.[13] 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.[14] These data serve two purposes. First, they complement the SafeGraph data in capturing broad-based mobility. Second, the EIA data provide useful forecasting power for gasoline sales, which are a large and volatile component of the MRTS series.

Figure 1 shows the six high-frequency time series in our seasonally adjusted Weekly Index of Retail Trade from January 2018-January 2021.[15] The figure makes clear the extent of heterogeneity across data sources, even during the depths of the recession that began in March 2020. Our index aims to combine them in a representative way in order to help explain national retail activity at a weekly frequency.

---

[13] 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.

[14] 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.

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

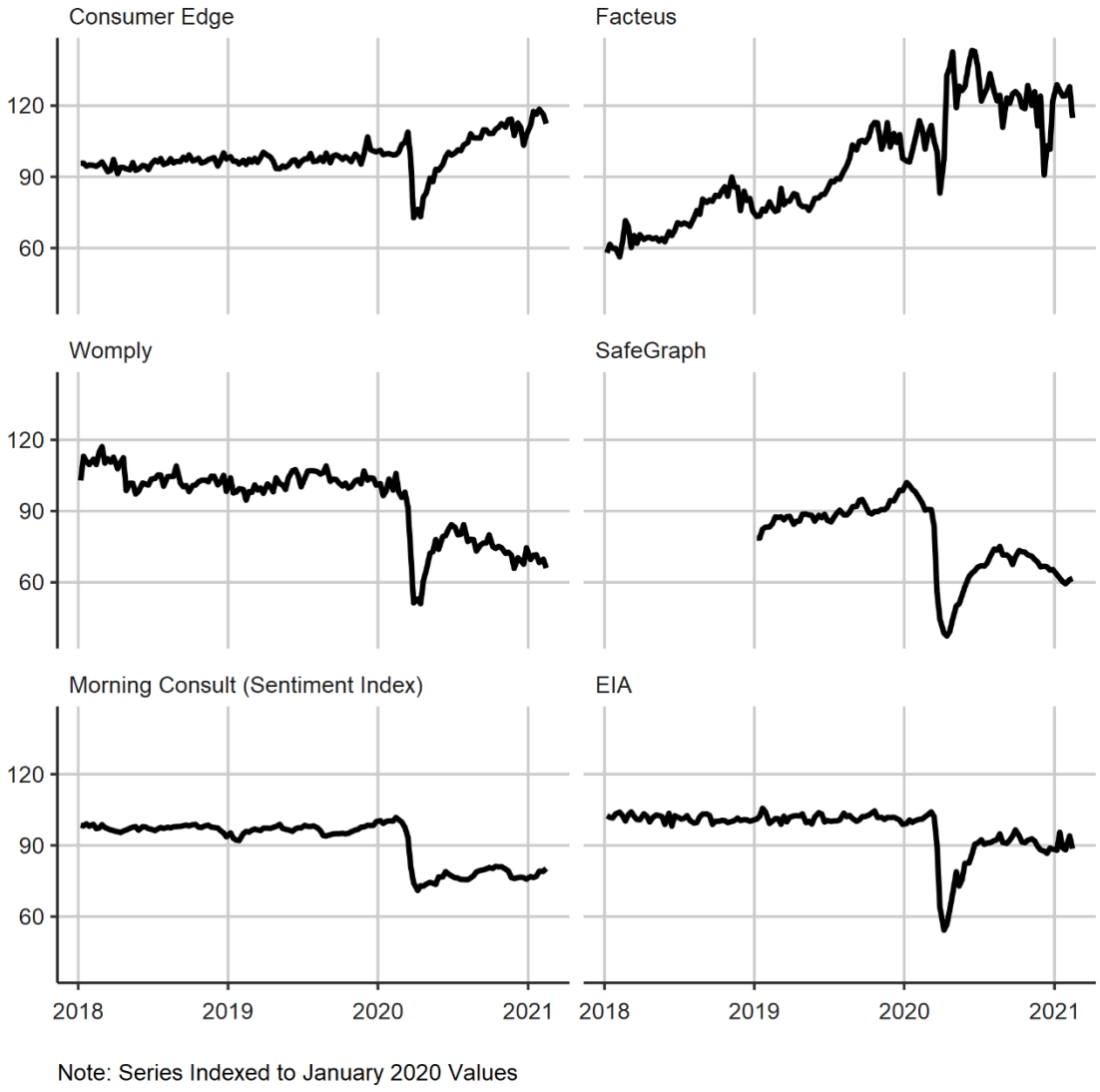


Figure 1: Weekly Data Series, by Source

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.

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.[16] However, unlike a standard

[16] For additional information on dynamic factor models, see Stock and Watson (2011). Similar mixed-frequency models to ours include Aruoba, Diebold, and Scotti (2009) and

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 how this is achieved and the model estimated.

Establishing a Regular Weekly Calendar

Weekly data pose a particular problem when constructing mixed-frequency models. With standard methods of defining weeks, some of the resulting weeks will fall into two months, which is problematic for temporally aggregating a weekly sales measure to the calendar month.[17] Accordingly, we instead define an alternative calendar. To begin, we partition each month into four weeks in the following way: each of the first three seven-day periods within the month are categorized as a week and the remaining seven to ten days are grouped into the fourth week of the month. This process results in 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 data and by averaging across the days of the week for the consumer sentiment data.[18] For the retail foot traffic and motor fuel data, which are reported in calendar weeks, we take weeks that span two months and allocate the visits/fuel supplied to each of the two months proportionally based on the number of days the week has in each month. We then re-allocate visits/supply within the month to match the number of days in each of our four weeks.

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 full seasonal adjustment procedure is described in further detail in the Appendix.

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.

[17] There are at least three common standards for defining weeks from daily data: using the number of seven-day periods from the first day of the year, counting weeks from a consistent Monday starting day, or counting weeks from a Sunday start day.

[18] The Morning Consult data are provided as a 5-day moving average. To create a weekly value we back out the daily index values from the moving average series and then re-average. For details see: https://stats.stackexchange.com/questions/67907/extract-data-points-from-moving-average

A Mixed-Frequency Dynamic Factor Model

In our framework, growth in the weekly data ($W_{n,i,t}$) are "noisy" measures of growth in latent weekly MRTS spending ($F_{i,t}$), with n denoting the indicator series and i the week of month t. Growth is measured in log differences. 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.

The latent factor follows a fourth-order autoregressive process with average growth rate α_F. Additionally, its weekly growth rates are constrained by the "triangle average" formula, which ensures that temporal aggregation of the $F_{i,t}$ generates a monthly series with the same growth rate as observed in the MRTS data. To implement the triangle average, we assume the log difference in monthly MRTS sales, $M_{i=4,t}$, is observed in the fourth week of the month. In calendar time, this occurs near the middle of the following month with the release of the Advance Monthly Retail Trade Survey, or MARTS.

In terms of equations, with L being a weekly lag operator, we have:

$$\begin{aligned} M_{i=4,t} &= \frac{1}{4}F_{i=4,t} + \frac{2}{4}F_{i=3,t} + \frac{3}{4}F_{i=2,t} + F_{i=1,t} + \\ &\quad \frac{3}{4}F_{i=4,t-1} + \frac{2}{4}F_{i=3,t-1} + \frac{1}{4}F_{i=2,t-1} \\ (1-\rho_1 L-\rho_2 L-\rho_3 L-\rho_4 L)F_{i,t} &= \alpha_F + \eta_{i,t} \\ W_{n,i,t} &= \alpha_n + \gamma_n F_{i,t} + \epsilon_{n,i,t} \\ (1-\psi_i L)\epsilon_{n,i,t} &= \upsilon_{n,i,t} \end{aligned}$$

The above model can be put into state-space form using the Harvey (1990) accumulator to handle the mixed-frequency constraint imposed on the latent factor and estimated by maximum likelihood with the Kalman filter assuming that the shocks $\eta_{i,t}$ and $\upsilon_{n,i,t}$ are jointly distributed $N(0,\Sigma)$.[19]

Unlike a standard mixed-frequency dynamic factor model, we do not include a measurement error term for the MRTS data. This means that the model would be over-identified if we assume that Σ is a diagonal variance-covariance matrix, as is typically done. Instead, our preferred specification estimates the elements of Σ associated with the individual covariances between $\eta_{i,t}$ and each of the $\upsilon_{n,i,t}$.[20] Although the covariances turn out to be small, this small modification leads to an exactly identified model as well as modest gains in our ability to nowcast the MARTS data. In essence, these covariances

---

[19] We use the Matlab toolbox MFSS described in Brave, Butters, and Kelley (2020) for this purpose. Furthermore, we restrict $\rho(L)$ and $\psi_i(L)$ to be stationary during estimation.

[20] When Σ is near-diagonal, our model closely resembles that of Aruoba et al. (2016) who create an alternative measurement of U.S. GDP, called GDPplus, that combines information from the product and income sides of the National Income and Product Accounts.

contain useful information about the direction and magnitude of expected deviations of the latent factor from its historical dynamics and, thus, help to improve our estimate of $\rho(L)$.

Once the model has been estimated, to arrive at a weekly level of spending that is benchmarked to the MRTS data, we construct the index $I_{i,t}$ from the Kalman smoothed estimate of the weekly latent factor as follows:

$$I_{i,t} \quad = \exp(M_0 + \sum_{n=1}^{t} F_{i,n})$$

where 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. Given data availability constraints, we are only able to produce an index that begins in the first week of January 2018.[21]

The Estimated Index

Figure 2 shows our Weekly Index of Retail Trade expressed as a seasonally adjusted monthly spending rate, $I_{i,t}$, in comparison to retail and food services sales excluding autos from the MRTS. 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. This is consistent with the 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 stay-at-home orders started to be lifted and the expanded unemployment insurance and first round of economic impact payments from the CARES act began to be disbursed. 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 these rebound effects well in advance of the MRTS.

[21] M_0 corresponds to retail & food services sales ex. auto from the MRTS in December 2017. At the moment, the SafeGraph series only begins in January 2019. 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, Butters, and Kelley (2020).

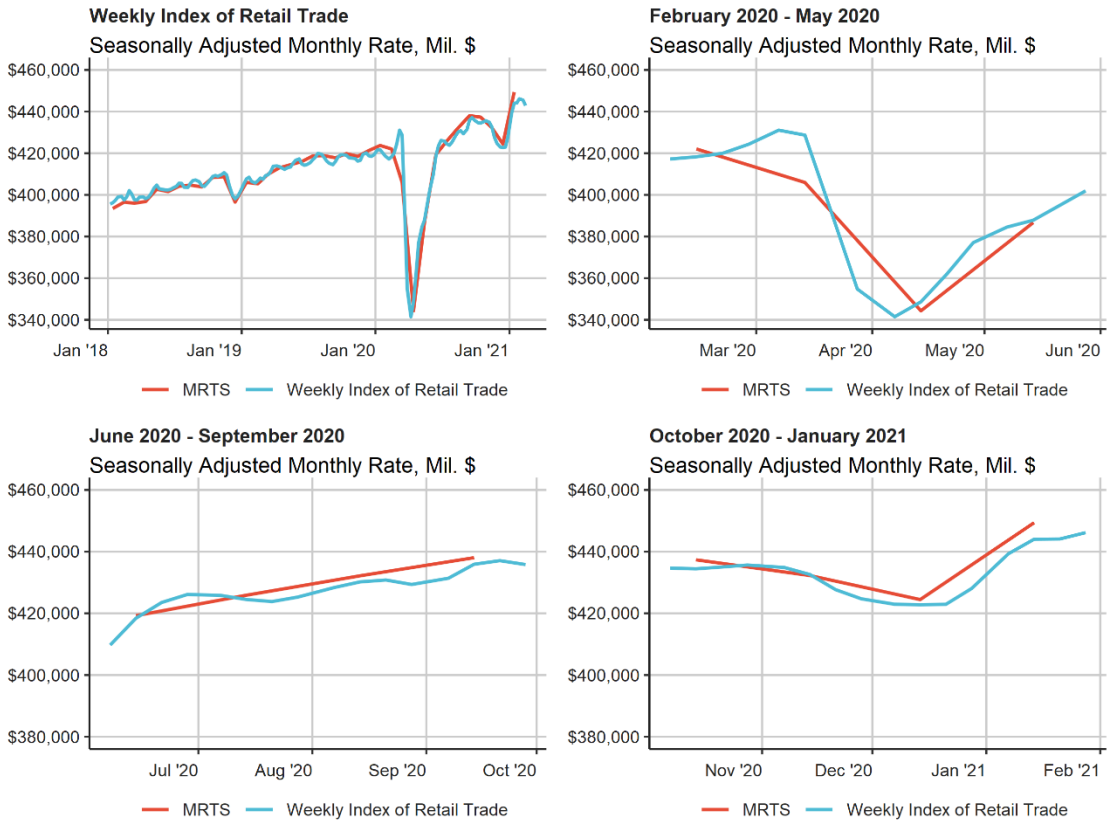


Figure 2: Retail & Food Services Sales ex. Auto

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: 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 approximately equal the MRTS data.

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 cases of Covid-19 began to rise again during the summer, the index shows periods of decline in both July and August; this drop-off is not apparent in the MRTS data because they average in the weeks of substantial positive growth prior to the peaks of each successive summer wave of cases.

The index also highlights the uneven nature of sales during the fourth quarter of 2020 and into the first quarter of 2021. Sales fell sharply starting around Thanksgiving, but were flat for much of December. Going into 2021, they then improved markedly, coinciding with the timing of the second round of economic impact payments in early January 2021. Since that initial burst, however, sales have been roughly flat.

Overall, our weekly index reveals patterns in retail sales that are of interest to those studying the impacts of the pandemic and the policy responses to it that are not always visible in the MRTS data. We capture the quickly changing path of consumer spending

during the pandemic more accurately than has previously been available. Moreover, we do so in a way that preserves the representativeness of our index to the structure of the MRTS.

Our model's combination of time series and benchmarking to the MRTS also 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. [22] These measures also use credit and debit card data, but both come from a single data provider with different coverage and use different statistical methods than our index.

For comparison figure 3 plots the series as a percent deviation from a pre-pandemic January 2020 baseline. [23] All three indexes generally track one another, but ours is substantially less volatile. This suggests that using information from a range of data sources could do a better job at extracting the underlying trend in overall retail activity than any individual data series. For example, the OI measure described in figure 1b of Chetty et al. (2020) has a root mean-squared error of close to 5 percentage points in comparison to the Advance Monthly Retail Trade Survey (MARTS). As we show below, our model results show there is substantial variation over time in the contributions of different indicators to explaining the common factor in overall retail sales, and that such a multi-variate structure can produce superior forecasts of the MARTS, outperforming the consensus of private forecasters as well as simple autoregressive models.

---

[22] See Aladangady et al. (2019) and Dunn, Hood, and Driessen (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/.

[23] In addition, the BEA and Opportunity Insights measures include automotive spending (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 below in section 3.1

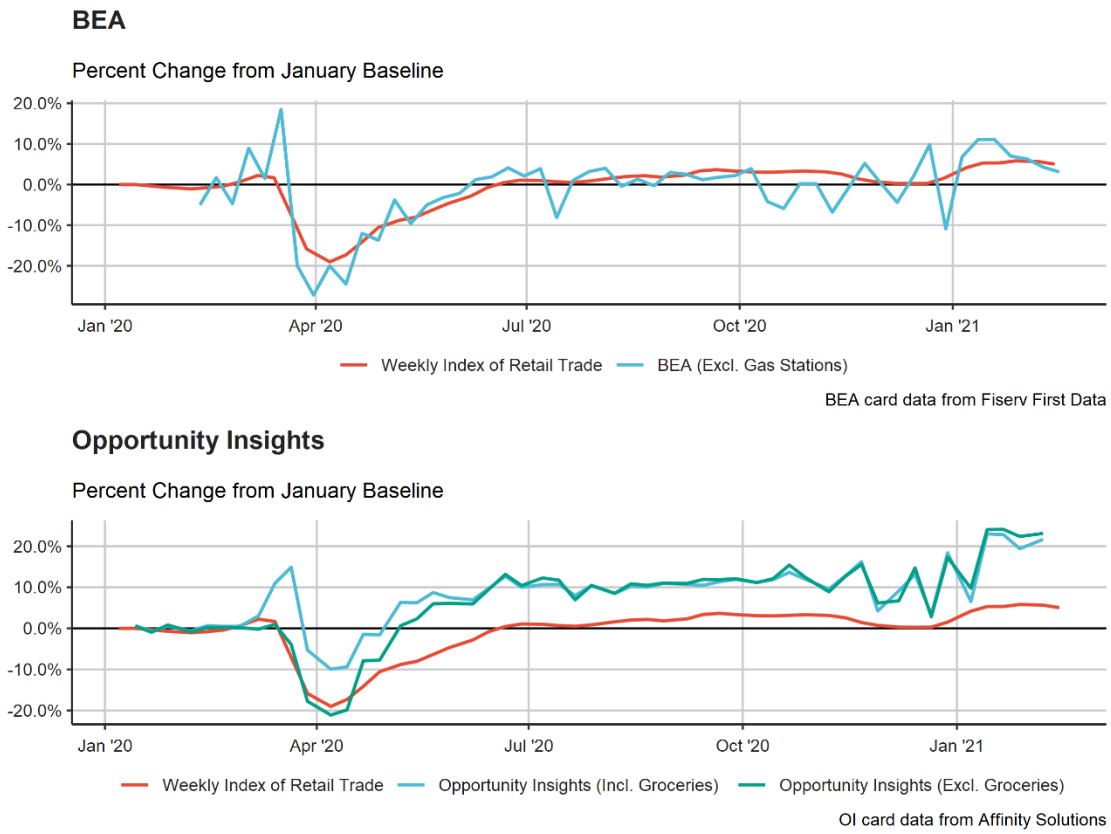


Figure 3: Weekly Measures of Retail & Food Services Sales

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

Contributions of Components to the Weekly Index

To investigate the contributions of the different indicators, figure 4 plots the contribution of each data series to our index for every week in our sample. The variation in the index that cannot be explained by the weekly series is attributed to the MRTS data; this contribution is importantly influenced by 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) often are quite important, they fall short in accounting for the depth of decline of the index in March and contributed little to the rebound in April. At these times, the EIA (purple), and Consumer Edge (orange) data made substantial contributions to the index.

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- January 2021 period. The first column presents the results for the index depicted in figure 3, which is produced by the Kalman smoother. Almost 66 percent of its variation is explained by the MRTS, leaving about 34 percent to be explained by the high-frequency data. Of these data sources, the EIA data explain about 25 percent. While important for a number of specific episodes, together on average the transaction data only

explain roughly 5.5 percent of the variation in the index, with the remaining roughly 3.5 percent accounted for jointly by SafeGraph and Morning Consult.

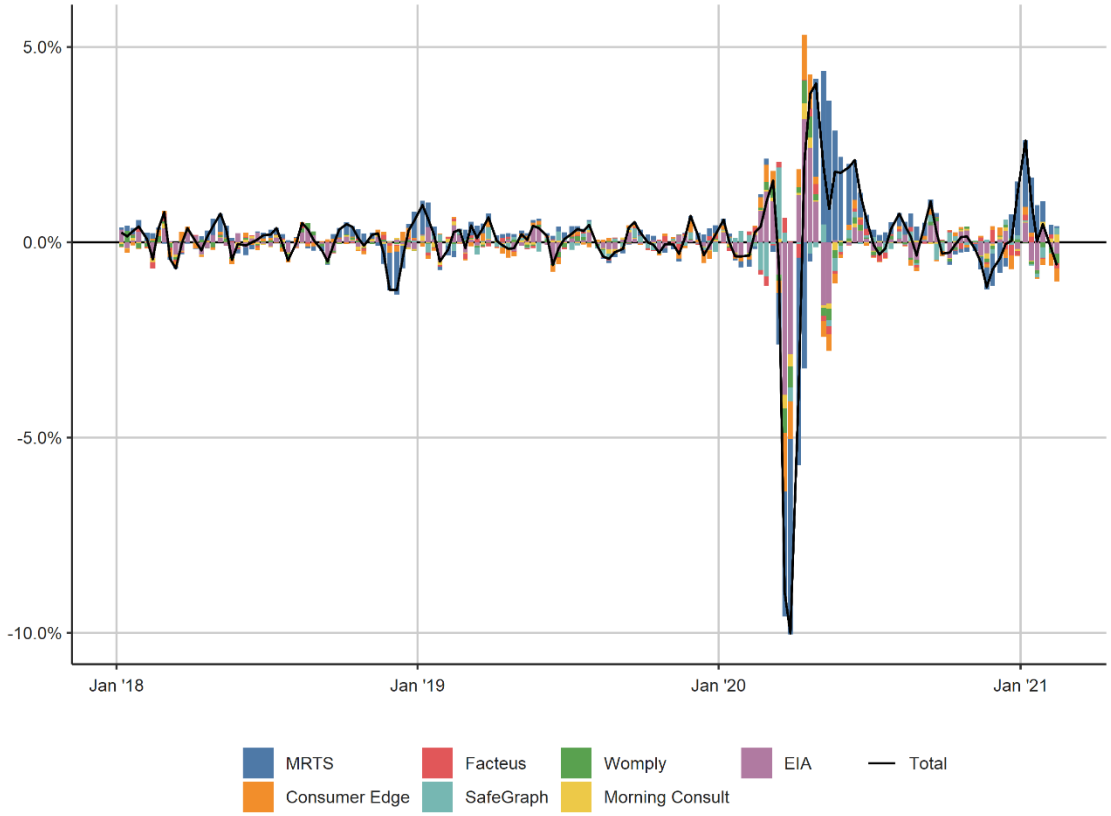


Figure 4: Contributions to Growth for the Weekly Index of Retail Trade

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 the appendix of Brave, Butters, and Kelley (2020).

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 incorporating the impact of omitting each of the weekly data sources one-by-one. In this way, we can capture interdependencies in the estimation of our model. The lower and upper bounds of these ranges are shown in table 1 in brackets. Overall, this exercise suggests that the MRTS data explains no less than 65 and no more than 83 percent of the variation in the growth rate of our weekly index.

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 $(i-1,t)$.[24] This decomposition is substantially different from that for the smoothed index. Here, the Safegraph data explain about 47 percent of the variation, and the EIA, Womply, and Morning Consult data also make important contributions. In contrast, the MRTS data explain very little of the variation, suggesting that much of their impact on the smoothed index comes from the constraint matching it to the MRTS data at the monthly frequency.

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 variance decomposition of nowcasts for the months of March 2020-January 2021 taken from an out-of-sample forecasting exercise with our model. Viewed through this lens, the importance of each high-frequency data source is more balanced, although the SafeGraph data still account for a relatively large share. The contribution of the MRTS is 17.1 percent, similar in size to those of the EIA, Consumer Edge and Womply and about twice that of Morning Consult. We now turn to describing this nowcasting exercise in more detail.

Table 1: Variances Explained by Data Source

Series	Smoothed	Filtered	Nowcast
Census	65.8	2.1	17.1
	[65.8, 82.2]	[1.1, 5.0]	
Consumer Edge	4.1	2.7	14.4
	[4.1, 9.1]	[2.7,19.3]	
Facteus	0.5	4.7	0.6
	[0.1, 0.5]	[0.3, 4.7]	
EIA	24.9	20.3	15.5
	[9.2, 24.9]	[20.3, 53.0]	
Womply	1.0	9.4	14.1
	[1.0, 8.7]	[9.4, 20.8]	
Morning Consult	0.4	13.9	9.2
	[0.4, 3.5]	[13.9, 24.8]	
SafeGraph	3.2	46.9	29.1
	[1.5, 3.6]	[15.3, 46.9]	

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 that always includes the MRTS data. The smoothed and filtered growth rates cover the period from January 2018-January 2021. The out-of-sample nowcasts are for March 2020-January 2021.

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

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.

Constructing Out-of-Sample Predictions

To nowcast the MARTS measure of retail & food services sales excluding autos, 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 $h = 0, 1, 2, 3, or\ 4$.[25]

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).[26] 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.

Figure 5 shows the MRTS data from January 2018 through December 2020 with our January 2021 nowcast and its associated coverage intervals. The MARTS data for January 2021 were released on February 17, 2021. We projected retail & food services sales ex. autos to increase by about 3 percent month-over-month in January 2021, but with a fair amount of uncertainty associated with this forecast, as seen in our coverage intervals in the figure. Consensus forecasts made just prior to February 17 showed an increase of about 1 percent month-over-month, substantially lower than our projection. The February 17 MARTS release instead showed a month-over-month increase of almost 6 percent, almost twice our projected value but within the 70% coverage interval for our projection.

---

[25] 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, Butters, and Kelley (2020) and the documentation for the Matlab toolbox MFSS.

[26] We also account for the impact of data revisions by augmenting the errors in the draws of the estimated end-of-sample variance V 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.

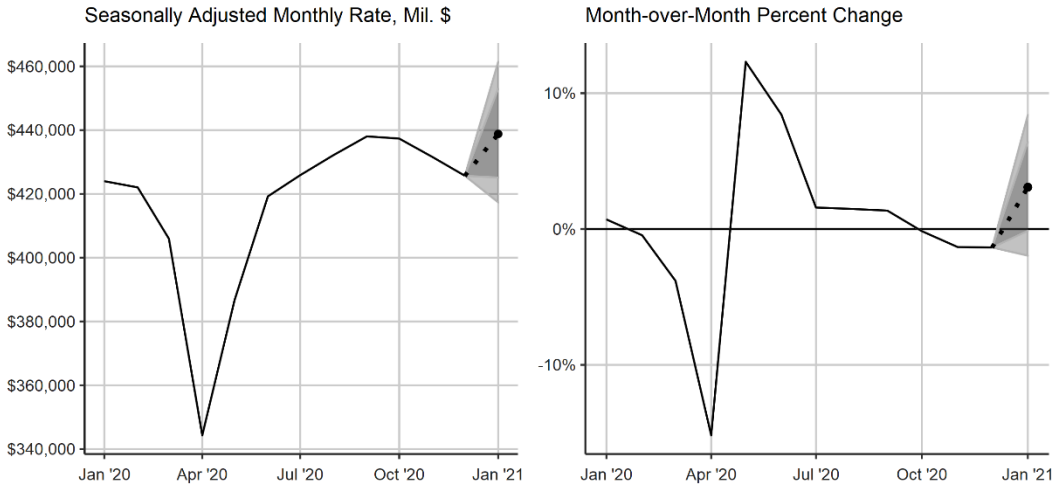


Figure 5: January 2021 Nowcast for Retail & Food Services Sales ex. Auto

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 January 2021 nowcast using weekly data available through the end of the month. The shaded regions represent 70% (dark gray) and 90% (light gray) coverage intervals,

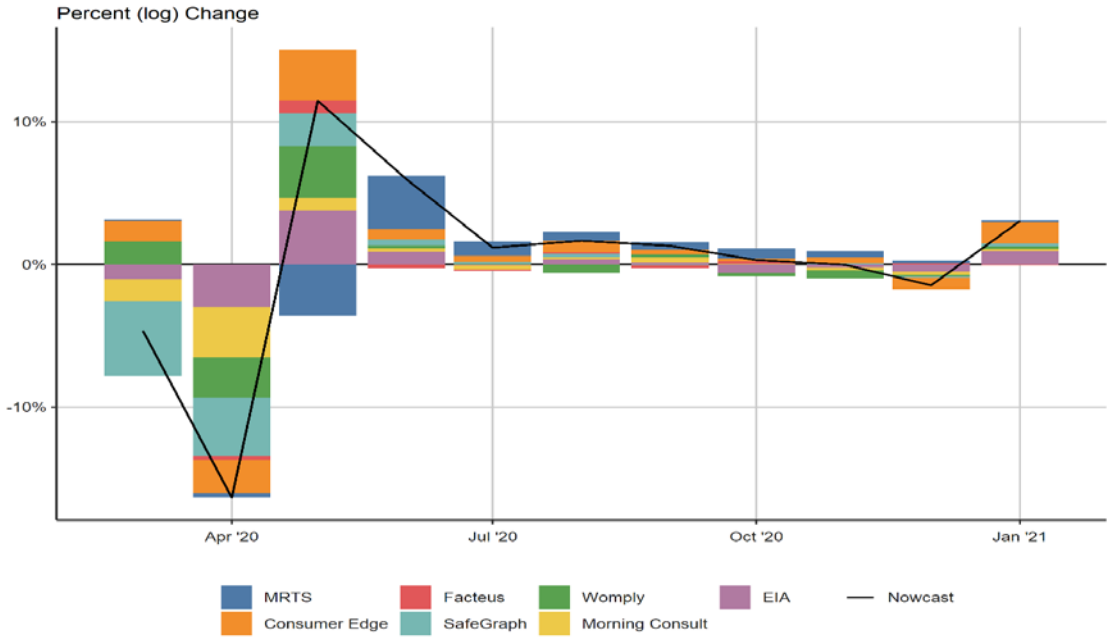


Figure 6: Contributions to Growth for our MARTS Nowcasts

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 January 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.

Out-of-sample Performance

We next evaluate the out-of-sample forecasting performance of our nowcasts by comparing them to consensus forecasts available at the same time and to predictions from a monthly autoregressive forecast 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 the number of parameters in the model.

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 series with the seasonal factors estimated at the time, iteratively re-estimate the model with real-time values for the weekly and lagged MRTS data, and produce nowcasts for the MARTS releases between March and January 2021.

Figure 6 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 only 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 4. For example, the improvement projected for January 2021 retail sales was principally driven by an increase in credit card spending and gasoline consumption. In contrast, last March and April the mobility 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 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 activity.

Table 2 reports mean absolute error ratios comparing the *h=0* nowcasts 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 February 17, 2021. The latter accounts for revisions to the MARTS data, which occur in the two months following the initial release.[27] 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, Leybourne, and Newbold (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 70 percent, for both surveys and releases of the Census data. Moreover, while it is difficult to establish

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

the statistical significance of this result based on only ten observations, the Diebold-Mariano test indicates that these differences are in fact significant at the 95% confidence level.[28] Thus, the nowcasts have a large and economically significant informational advantage relative to the surveys. This is true even abstracting from March-June 2020, where the informational gains are the largest.

Table 2: Mean Absolute Errors Relative to Consensus Forecasts

MRTS Release	First	Current
MCM Research	0.34*	0.31*
Action Economics	0.31*	0.30*

Note: This table shows mean absolute error (MAE) ratios for nowcasts of 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 Service Sales ex. Autos for both the First and Current releases of the Census Bureau's Monthly Retail Trade Survey from March 2020-January 2021. The marker * denotes statistical significance from a Diebold and Mariano (1995) test of equal mean absolute error (i.e. ratios equal to 1) at the 95% confidence level using the Bartlett kernel with lag length equal to sample size (as suggested by Kiefer and Vogelsang (2005)) and the small sample size correction found in Harvey, Leybourne, and Newbold (1997). Both survey forecasts were obtained through a subscription to Haver Analytics. The MCM Research survey is maintained by Informa Global Markets.

We can also characterize the informational advantage provided by the high-frequency data sources 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 monthly AR(1) specifications. The first estimates the model parameters using data from January 2006 through February 2020 and holds them fixed when producing forecasts for March 2020 through January 2021. The second instead recursively updates these parameters by adding the previous month's data to the estimation sample.

To capture the real-time data flow within each month, we produce five nowcasts based on the number of weeks of the high-frequency data available for the month being forecast. 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 importance of each subsequent week's worth of data.

Table 3 demonstrates that our dynamic factor model outperforms a monthly AR(1) specification for the MRTS by about 5-30 percent when no weekly data is used in the model ($h=4$) and an even greater 45-55 percent when only 1 week of high-frequency data is used in the model ($h=3$). The gains steadily increase from there as additional weeks of high-frequency data are available, peaking at a relative performance gain of about 85 percent when all four weeks' worth of high-frequency data are used ($h=0$) in the model.

[28] 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.

Table 3: Mean Absolute Errors Relative to Monthly Autoregressive Models

# of Weeks of Data for Nowcast (h steps ahead)	Fixed	Recursive
0 (h=4)	0.95	0.71*
1 (h=3)	0.57**	0.43**
2 (h=2)	0.43**	0.32**
3 (h=1)	0.21**	0.16**
4 (h=0)	0.17**	0.13**

Note: This table shows mean absolute errors (MAEs) for nowcasts from our preferred 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 January 2021. The recursive window model adds one month of data to the estimation sample iteratively from March 2020 through December 2020 and uses the updated model coefficients to produce one-month ahead forecasts. To produce similar forecasts from our alternative mixed-frequency dynamic factor model we estimate the model separately including from one to four weeks of available weekly data for the month being nowcast. The markers */** denotes statistical significance from a Diebold and Mariano (1995) 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 (Kiefer and Vogelsang, 2005) and the small sample correction found in Harvey, Leybourne, and Newbold (1997). Census retail sales data used to construct mean absolute errors were obtained from Haver Analytics.

Sources of Forecast Error

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

1. *Sampling error*: 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 by including other more representative data in the model and forcing our index to match the MRTS at the monthly frequency. Nonetheless, 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 over this period.) 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 relatively larger misses of our model in the March-June period.

3. *Census seasonal factors and revisions*: The seasonal factors we use to adjust the weekly spending series in our model 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. And in fact, 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 7 decomposes our forecast errors (*h=0* nowcasts minus actual) during the pandemic into each of these three sources. To create this decomposition, we re-ran our nowcasts under alternatives that altered the real-time exercise by using: 1) the model parameters estimated at their latest values, and 2) the latest MRTS data and seasonal factors. Differences between the resulting nowcasts and our original estimates are informative on the roles of parameter instability and data revisions in explaining our errors; the residual can then be attributed to sampling error.[29]

Parameter instability played a large role in explaining forecast errors in the early months of the pandemic. As our parameter estimates have settled down since July 2020, this particular source of error has become less impactful, with the exception of January 2021.

Revisions to the Census data and seasonal factors typically played a minor role with the notable exception of March. The large error then in our forecast for the monthly seasonal factor, which itself derived from a large change in the Census seasonal factor, caused us to *underestimate* the decline in retail sales in March. At the same time, several of the underlying weekly data series *overestimated* the March decline and the two sources of forecast error largely cancelled each other out.

Overall, the errors in our nowcast have instead been largely driven by sampling variability attributable to differences in the coverage of our high-frequency data sources and the MRTS. This was most evident in the early stages of the pandemic in March and June 2020, but was also true at the end of 2020 when retail sales declined more than our model expected in October and November. Similarly, our large miss in January 2021 was also primarily attributable to sampling variability, although parameter instability also factored into our *underestimate* of the rebound in sales.

All together, the mean absolute error for our nowcasts during the pandemic is 0.8 percentage point; this is against an average absolute percent change in MRTS retail & food services ex. autos of 4.8 percent over this period. However, March, June, and November 2020 and January 20201 are the only months in our sample where our absolute errors are greater than 1 percentage point, with the model under-predicting the decline in November 2020, over-predicting the decline in March 2020, and underestimating the strength in sales in June 2020 and January 2021. Excluding those months, the mean absolute error was just 0.2 percentage point, versus an average absolute change in MRTS sales of 4.8 percent.

[29] At the time of writing, the December 2020 and January 2021 MRTS values were still subject to further revisions.

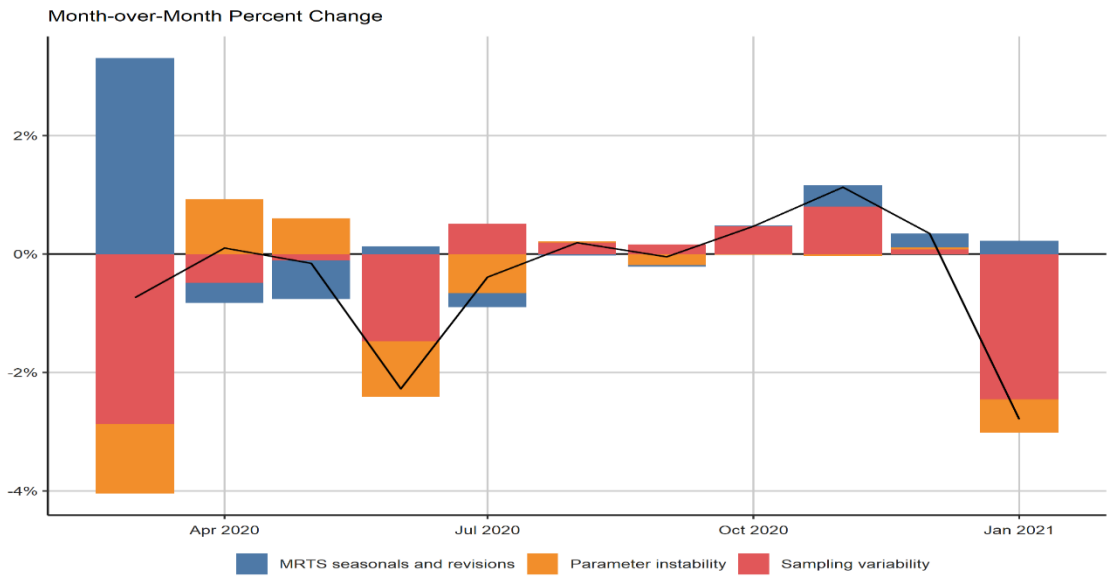


Figure 7: Nowcast Error Decomposition

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 sampling variation, such that all three sources of error sum to the total nowcast error.

Conclusion

We used weekly measures of revenue, credit and debit card transactions, consumer sentiment, mobility, and gasoline shipments to create a seasonally adjusted Weekly Index of Retail Trade that is benchmarked to monthly retail & food services sales (excluding automotive spending) from the Census Bureau's Monthly Retail Trade Survey (MRTS). To do so, we use 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 document 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 could be reliably used to nowcast the Census Bureau's Advance Monthly Retail Trade Survey (MARTS) in real time, besting by about 70 percent both consensus forecasts available at the time and forecasts from monthly autoregressive specifications for the MRTS in terms of lower mean absolute error.

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.

Similarly, we focused on the retail & 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 the effects that we find on consumer spending during the pandemic into GDP.

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." National Bureau of Economic Research.

Alexander, Diane, and Ezra 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 No. WP-2020-12.

Aruoba, S Borağan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song. 2016. "Improving Gdp Measurement: A Measurement-Error Perspective." *Journal of Econometrics* 191 (2). Elsevier: 384–97.

Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti. 2009. "Real-Time Measurement of Business Conditions." *Journal of Business & Economic Statistics* 27 (4): 417–27.

Atkinson, Tyler, Jim Dolmas, K Christoffer, Evan Koenig, Karel Mertens, Anthony Murphy, and Kei-Mu Yi. 2020. "Mobility and Engagement Following the Sars-Cov-2 Outbreak." *Federal Reserve Bank of Dallas Working Paper*, no. 2014.

Brave, Scott A, R. Andrew Butters, and Michael Fogarty. 2020. "The Perils of Working with Big Data and a Small Framework You Can Use to Avoid Them." FRB of Chicago Working Paper No. WP-2020-35.

Brave, Scott A, R. Andrew Butters, and David Kelley. 2020. "A Practitioner's Guide and Matlab Toolbox for Mixed Frequency State Space Models." *Available at SSRN 3532455*.

Brave, Scott, and R. Andrew Butters. 2012. "Diagnosing the Financial System: Financial Conditions and Financial Stress." *International Journal of Central Banking* 8 (2): 191–239.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner. 2020. "Real-Time Economics: A New Platform to Track the Impacts of Covid-19 on People, Businesses, and Communities Using Private Sector Data."

Cleveland, W., and Stuart Scott. 2007. "Seasonal Adjustment of Weekly Time Series with Application to Unemployment Insurance Claims and Steel Production." *Journal of Official Statistics* 23 (2): 209–21.

Consolvo, Victoria, and Kurt G Lunsford. 2019. "Residual Seasonality in Gdp Growth Remains After Latest Bea Improvements." *Economic Commentary*, nos. 2019-05. Federal Reserve Bank of Cleveland.

Diebold, Francis X, and Roberto S Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13 (3): 253–63.

Dunn, Abe, Keyle Hood, and Alexander Driessen. 2020. "Measuring the Effects of the Covid-19 Pandemic on Consumer Spending Using Card Transaction Data." *BEA Working Paper Series*.

Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*. Oxford university press.

Giannone, Domenico, Lucrezia Reichlin, and David Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55 (4): 665–76. https://doi.org/http://dx.doi.org/10.1016/j.jmoneco.2008.05.010.

Harvey, Andrew C. 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge university press.

Harvey, David I, Stephen J Leybourne, and Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13 (2): 281–91.

Karger, Ezra, and Aastha Rajan. 2020. "Heterogeneity in the Marginal Propensity to Consume: Evidence from Covid-19 Stimulus Payments." FRB of Chicago Working Paper No. WP-2020-15.

Kiefer, Nicholas M., and Timothy J. Vogelsang. 2005. "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21 (6). Cambridge University Press: 1130–64. https://doi.org/10.1017/S0266466605050565.

Lewis, Daniel J, Karel Mertens, James H Stock, and Mihir Trivedi. 2020. "Measuring Real Activity Using a Weekly Economic Index." *Federal Reserve Bank of New York Staff Reports*, no. 920.

Mariano, Roberto S., and Yasutomo Murasawa. 2003. "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." *Journal of Applied Econometrics* 18 (4): 427–43.

Stock, James H., and Mark W. Watson. 2011. "Dynamic Factor Models." In. Oxford Handbook of Forecasting.

Appendix

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, 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 to us.

However, we are still able to seasonally adjust our weekly data by making. 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 the MRTS seasonal factors can then be used to seasonally adjust the high-frequency data sources at a first pass. This procedure reliably removes the majority of seasonal patterns in our transaction and foot traffic data (the EIA data are already seasonally adjusted using the Cleveland and Scott (2007) method by Haver Analytics and the Morning Consult data do not exhibit seasonality). To then account for any residual seasonality in the data, we remove prominent week-of-year effects.

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 & food services sales ex. autos series.[30] We assume that the observed monthly Census seasonally adjusted (SA) and non-seasonally adjusted (NSA) time series are the sum of latent (unobserved) weekly series:

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

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.

[30] The seasonal factor for month t is defined as $\frac{SA_t}{NSA_t}$, where SA is the seasonally adjusted Census series and NSA is the unadjusted Census series.

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

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

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:

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

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, $SF_{i,t}^{C}$, is to offset these two calendar sources of weekly sales variation. This suggests setting the weekly seasonal adjustment factors according to the formula:

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

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 (averaging 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 $HF_{i,t}^{NSA}$, the NSA high frequency weekly retail sales data, in order to create a SA series, $HF_{i,t}^{SA}$:

$$
\begin{aligned}
SF_{i,t}^{HF} &= \frac{1}{4} \cdot \frac{1}{\text{avg}_{i,t}\left[\frac{\left(HF_{i,t}^{NSA}/D_{i,t}\right)}{\left(HF_{t}^{NSA}/D_{t}\right)}\right]\frac{D_{i,t}}{D_t}} \cdot SF_t^C \\
HF_{i,t}^{SA} &= HF_{i,t}^{NSA} \cdot SF_{i,t}^{HF} \\
HF_t^{SA} &= \sum_{i=1}^{4} H\,F_{i,t}^{SA}
\end{aligned}
$$

where $SF_{i,t}^{HF}$ and SF_t^{HF} are weekly and monthly seasonal factors and $\text{avg}_{i,t}[\cdot]$ 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, for example whether or not to include holidays in the $D_{i,t}$ and D_t variables. Through testing, we found that the more parsimonious approach - including the holidays - produced better results than only including non-holiday days in the formulas. In addition, SF_t^C, the monthly Census seasonal factor, is not available until the MARTS is published, and so we must estimate it for our nowcasting work. The Census Bureau seasonally adjusts retail sales on a concurrent basis; that is, every month they re-rerun their programs and estimate new factors with all the available data. These programs 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. 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). The residual seasonality in the various weekly data series that we use are most noticeable during the end-of-year holiday period, where there is often a predictable pattern in the weekly growth rates around the first week of the year. This could arise, for example, because the seasonal spending patterns for credit and debit card transactions do not exactly mirror the seasonal pattern in the Census retail sales measure as assumed above.

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 important weeks that we identified for each 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. We do not find evidence of any seasonality in the consumer sentiment data and, therefore, do not seasonally adjust it with any of these procedures.