MARKET INSIGHTS

Fair Weather Trends:

Seasonality and Google Mobility Data

October 2021

D E Shaw & Co

Introduction

Google LLC's Community Mobility Reports offer data that many have found helpful in tracking economic developments since the onset of the COVID-19 pandemic.[1] The raw data from these reports can be misleading, however, as typical seasonal variation obscures the underlying trend.

According to our analysis, in which we account for that seasonal variation in the United States, the overall recovery of U.S. mobility following the massive lockdown-induced decline in early 2020 has been more muted than the raw data suggest, and—notably—the more recent perceived mobility recovery since April 2021 has been essentially nonexistent.

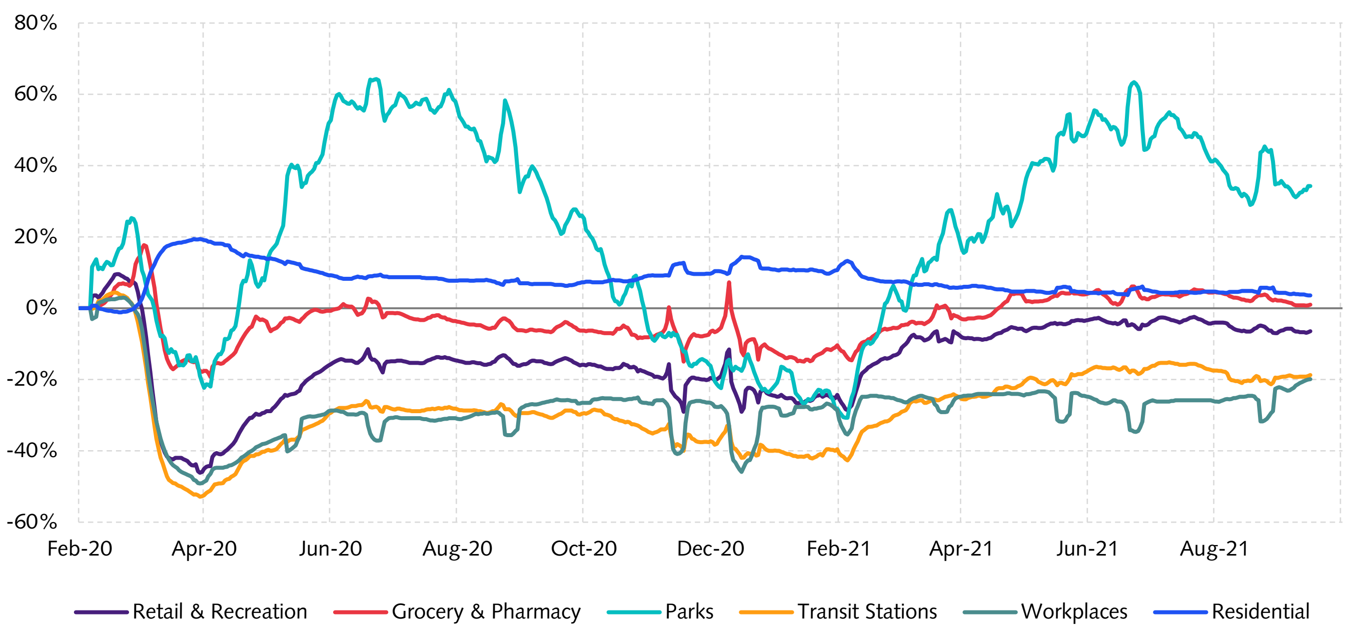
In this paper, we share a novel methodology for adjusting Google mobility data (GMD) to account for the effects of seasonality. This is a particularly difficult problem because of the data series' very short history, which necessitates an adjustment approach that goes beyond traditional methods often applied to macroeconomic time series.

Background on Google Mobility Data

Google began sharing its Community Mobility Reports in early 2020 in response to the onset of the pandemic. GMD is based on anonymized activity data collected from Google users who have turned on their "location history." Google defines six categories for which mobility data are compiled: retail & recreation, grocery & pharmacy, parks, transit stations, workplaces, and residential. Google established "baseline" levels of mobility using median daily values for the five-week period between January 3, 2020 and February 6, 2020. *Figure 1* shows changes in mobility for each category over time relative to that baseline period.

For purposes of this paper, we focus specifically on retail & recreation mobility, as we believe this category maps most directly onto broadly defined economic activity. According to Google, retail & recreation mobility includes locations such as restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters. Other categories, in our view, offer less by way of potential insights. Time spent in workplaces, for example, is likely a poor proxy for

Figure 1 U.S. Google Mobility Data:
Change in 7-Day Moving Average Relative to January 2020 Baseline
(February 15, 2020 to September 30, 2021)



Sources: Google LLC; the D. E. Shaw group.

[1] Available here.

time spent working given the widespread adoption of remote work; parks and residential are not closely tied to market-based economic activity; and each of grocery & pharmacy and transit stations represents a fairly narrow slice of activity.

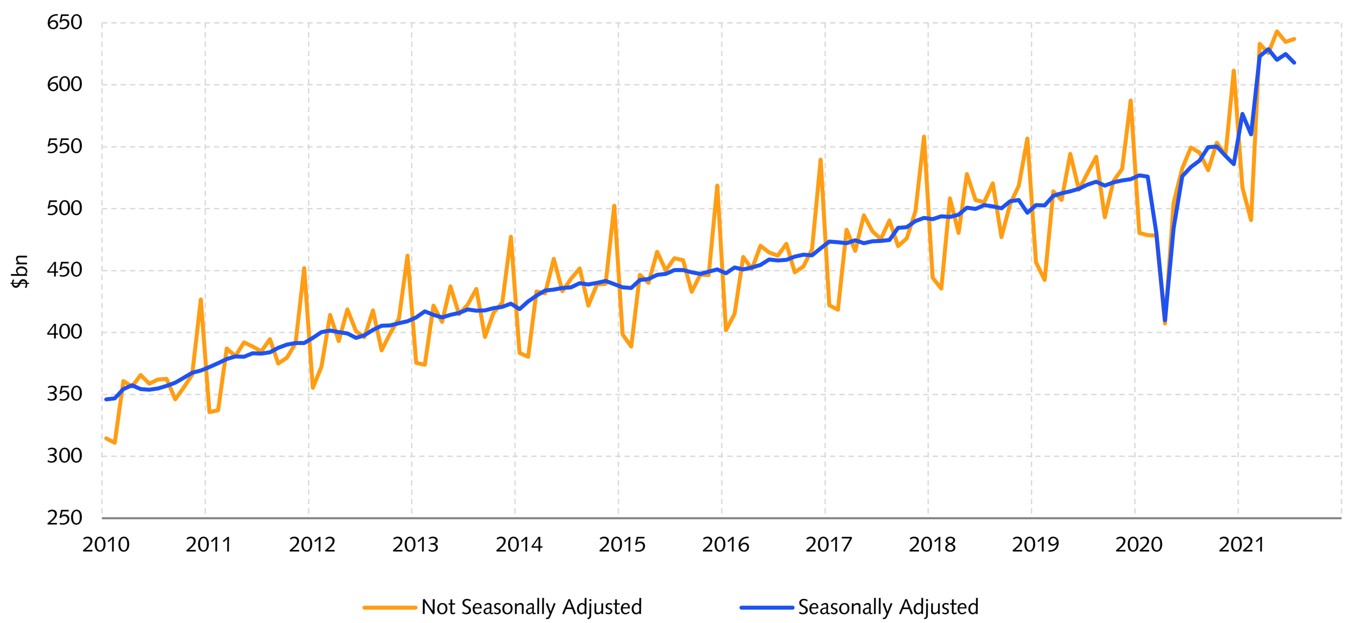
The Importance of Adjusting for Seasonality

Macroeconomic data often show evidence of pronounced seasonality. For example, *Figure 2* shows U.S. retail and food service sales, as measured by the U.S. Census Bureau, on an unadjusted basis (orange line) and a seasonally-adjusted basis (blue line). Given the erratic swings in the unadjusted data, it's much easier to identify important shifts in the economy—such as the onset of the COVID-19 crisis—from the seasonally-adjusted data series. As a result, analysts prefer to work with seasonally-adjusted data when they are available.

One would expect mobility to show a strong seasonal pattern in a manner similar to other economic data; however, GMD are not seasonally adjusted. Importantly, GMD's baseline measurement period falls within a timeframe that is seasonally weak for mobility in the United States, as people generally spend less time outside in the winter and are less apt to travel and shop immediately following the U.S. holiday season. Indeed, for many macroeconomic indicators—such as payroll employment, retail sales, and housing starts—January is consistently close to the low point of the year in non-seasonally-adjusted activity. In other words, in the absence of a pandemic, we would have expected mobility during the rest of 2020 on average to have been substantially higher than during the GMD baseline period.

With all of this in mind, adjusting GMD to account for seasonality allows for a better read on underlying trends in the data. However, doing so is not a straightforward exercise. Traditional methods for seasonal adjustment identify patterns in the data by looking over a long historical time series.[2] For GMD, as with many alternative data sources, its limited history means that use of those methods may not be feasible.

Figure 2 U.S. Retail and Food Services Sales
(January 2010 to July 2021)



Sources: U.S. Census Bureau; the D. E. Shaw group.

[2] For example, X-13ARIMA-SEATS (available here) is the seasonal adjustment technique used for a variety of the U.S. government's macroeconomic indicators, such as the monthly payroll employment estimate from the U.S. Bureau of Labor Statistics. This method requires several years of data in order to uncover seasonal patterns.

Our Alternative Approach

Our new approach to seasonally adjusting GMD addresses this limited history by combining a prior about the underlying drivers of seasonal variation with a richer cross-sectional technique that makes use of U.S.-state-specific data. By contrast, traditional methods of seasonal adjustment are mostly agnostic about why seasonal patterns exist.

We believe that seasonality in economic data is driven mostly by weather *(i.e.*, certain activities are more feasible or more popular depending on the season) and by societal or cultural features of the calendar (*e.g.*, the start of the school year or timing of the U.S. Thanksgiving holiday). We focus specifically on weather, which we think is of primary importance when examining mobility. We use temperature as a proxy for weather because of its regular seasonality and ready availability at the state level.[3]

Our approach to estimating a temperature-based seasonality adjustment then consists of three steps. First, we construct a panel dataset of our quantity of interest (*i.e.*, state-level mobility) that is not seasonally adjusted, as well as corresponding data for average state-level temperatures for each observation period.[4] Next, we compute a simple regression of changes in mobility on changes in temperature to estimate typical sensitivity of mobility to temperature. Finally, we construct an expected average path for temperature over the course of the year at the national level, and we use our sensitivity estimates to calculate expected seasonal change based on that temperature path.

Key to our approach is the availability of state-level data for both independent (temperature) and dependent (mobility) variables, which provides a broad cross-section of data to compensate for a time series history that is too brief to estimate their relationship reliably. Our approach benefits from the fact that temperature often has significant independent variation in the cross-section of states, such as the extreme cold experienced by Texas in February 2021 or the heatwave that descended upon a number of western states in summer 2021 (while weather outcomes in the rest of the country were more typical during these periods).

Validating Our Approach

Before applying our seasonal adjustment approach to GMD, we need to make an initial assessment of its reliability. To do so, we apply the methodology described in the previous section to a data series for which we already have known seasonal factors estimated in the traditional manner. Our aim is to show that we can roughly reproduce these seasonality estimates, by first using a long history of data and then restricting the data history to more closely resemble what is available for GMD.

Because our approach focuses specifically on weather—as compared to holidays, for example—it is most appropriately applied to series for which we believe that seasonal variation is primarily due to weather patterns. Employment in the leisure and hospitality industry is an ideal test case, as it is significantly influenced by weather, is measured and readily available at the state level, and is intuitively quite closely linked to retail & recreation mobility.

Starting first with a panel of ten years of monthly data from the contiguous 48 U.S. states, we run a pooled regression and find with statistical significance that an increase in temperature of 10°F yields, on average, a 2.0% rise in the level of non-seasonally adjusted employment in the leisure and hospitality industry. (Complete results for this regression and the GMD regression described below are available upon request.) Given that the national average temperature usually varies by about 40°F during the year, this implies seasonal swings in leisure and hospitality employment of about 8%, as shown by the orange line in *Figure 3*. Notably, this is almost exactly the amplitude of seasonality reflected in the official seasonal factors from the U.S. Bureau of Labor Statistics (BLS), shown by the blue line in *Figure 3*. As the graph indicates, the pattern of seasonality estimated by BLS is somewhat more nuanced than our technique affords, with its seasonality estimate deviating slightly from our more classic "sine wave" pattern at the start of the school year in August–September and at the peak of the U.S. holiday season in December.

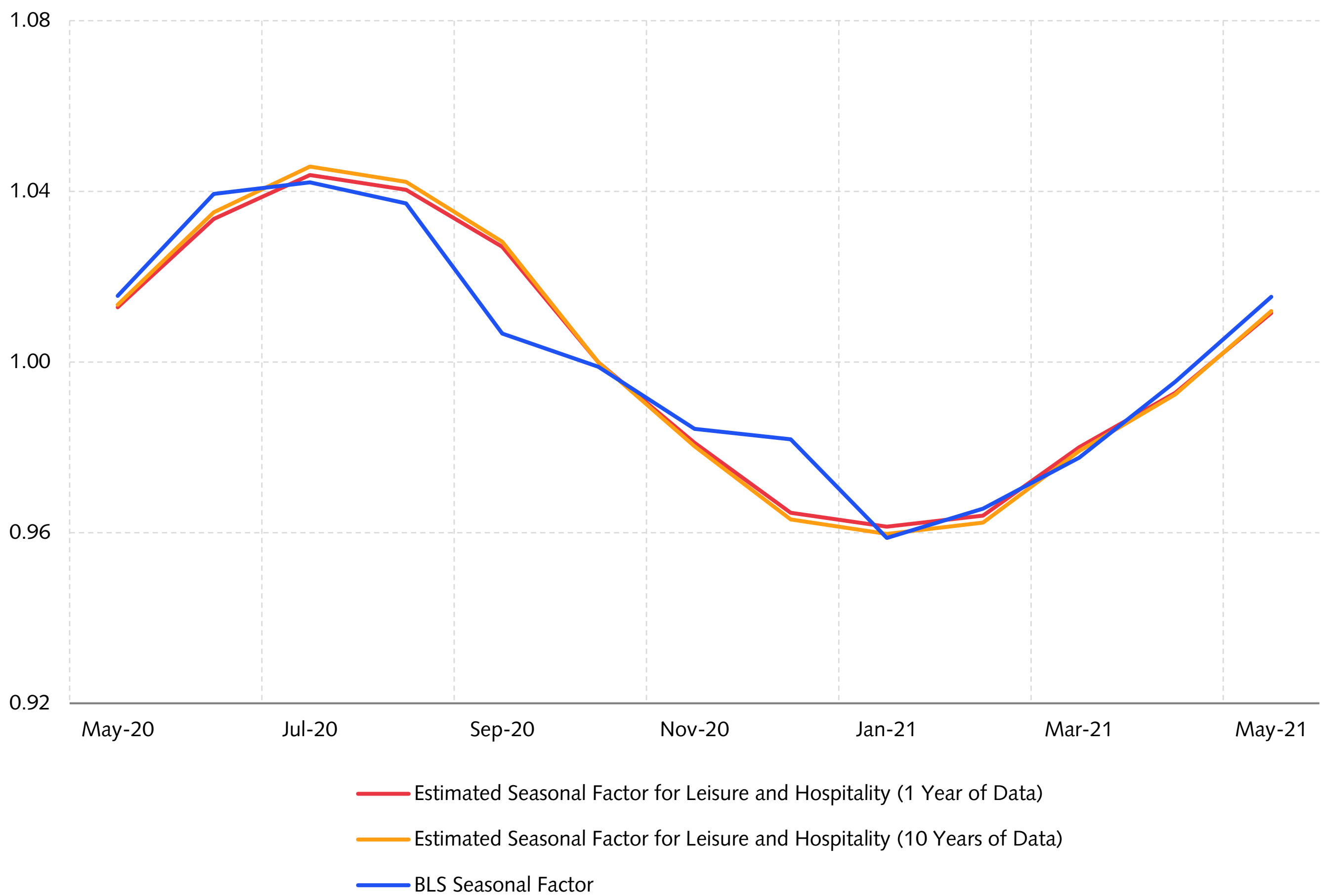
[3] We are aware of past work finding a relationship between mobility and weather, for example Wu, Mooring, and Linz (2021). Although these authors focused on the relationship between temperature and park visitations, their study is primarily focused on public health and so is distinct from our proposed application.

[4] Temperature data are sourced from the National Centers for Environmental Information at the U.S. Department of Commerce's National Oceanographic and Atmospheric Administration.

Next, we confirm whether a similar result can be obtained by using a more limited time series history, akin to what is available for GMD. Using state-level data only since the beginning of 2020,[5] we find that we can replicate almost exactly the same estimated sensitivity of employment to temperature that was established using the ten-year history, as shown by the red line in *Figure 3*. In other words, a state-level approach allowed us to meaningfully reduce the length of the required time series without significantly degrading the fidelity of our estimates in this case.[6]

Because our estimated sensitivities of employment to temperature, derived from both a longer and a shorter time series, line up closely with the official seasonally-adjusted data, we are more confident that we can apply the same methodology to mobility, for which seasonality adjustments don't readily exist.

Figure 3 Estimated Seasonal Factors for Leisure and Hospitality Employment and BLS Seasonal Factor
(May 2020 to May 2021)



Sources: U.S. Bureau of Labor Statistics; the D. E. Shaw group.

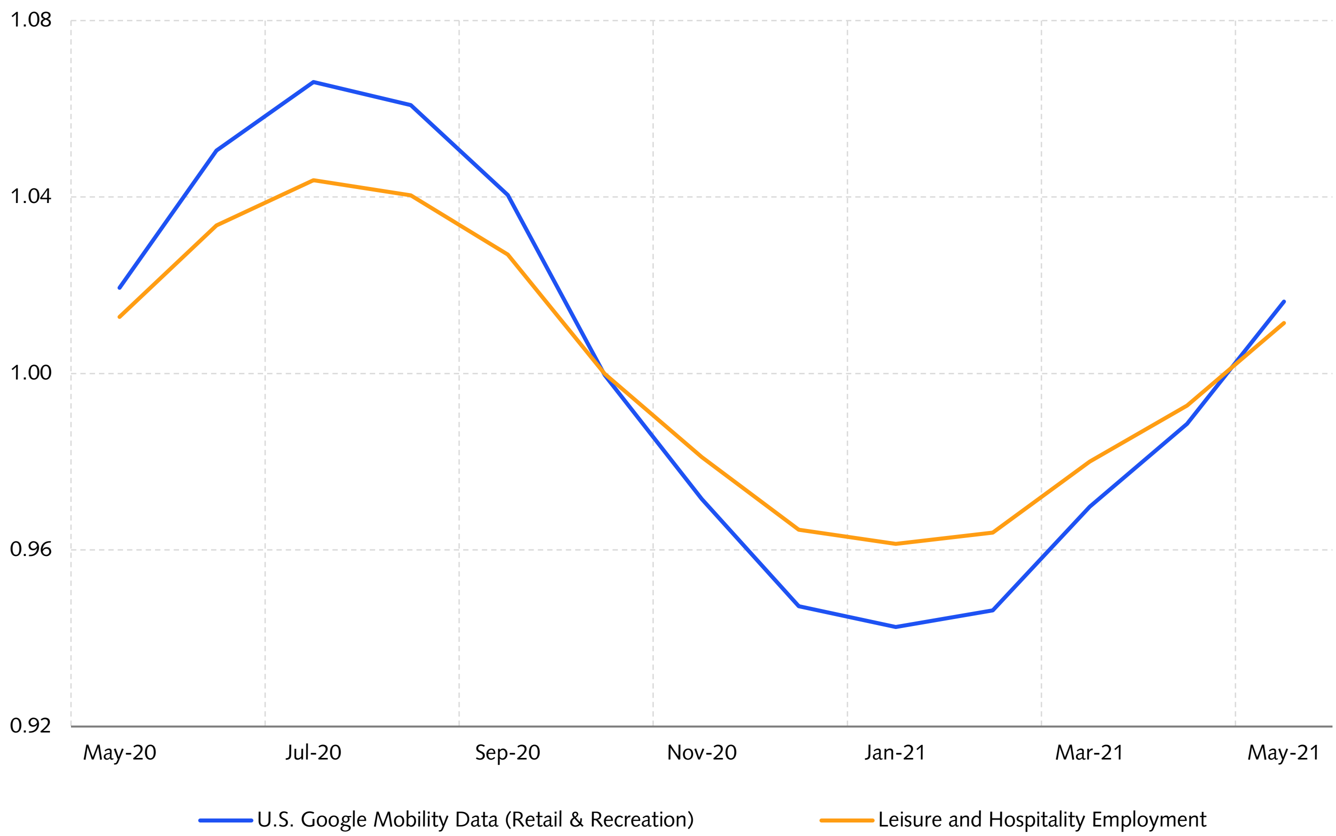
[5] Recognizing that the months of April, May, and June 2020 experienced unusually large employment changes associated with the imposition and subsequent relaxation of COVID-19-related lockdowns, we include binary control variables for those three months to avoid affording that period of time an outsized distorting influence on the estimated temperature sensitivity given the smaller sample.

[6] Some might contend that the causal importance of weather in driving mobility is overstated, reasoning that it is the school vacation schedule in the United States, not temperature, that primarily drives the seasonal pattern. This concern may have some validity, although if the school vacation schedule were the dominant factor, we would not expect our technique using only one year of data to so closely replicate estimated seasonality in leisure and hospitality employment.

Adjusting and Reinterpreting Mobility Data

We now apply our methodology to GMD. To do so, we run a simple regression of changes in retail & recreation mobility on changes in temperature using U.S.-state-level data in order to determine the typical sensitivity of mobility to temperature.[7] We find with statistical significance that a 10°F increase in temperature translates into a roughly 2.9% rise in the level of retail & recreation mobility. As one might expect, *Figure 4* shows that our estimated seasonal variation for mobility (blue line) exhibits a similar pattern to that of leisure and hospitality employment referenced in the previous section (orange line), although it does so with somewhat greater amplitude.

Figure 4 Estimated Seasonal Factors: U.S. Google Mobility Data (Retail & Recreation) and Leisure and Hospitality Employment
(May 2020 to May 2021)



Source: the D. E. Shaw group.

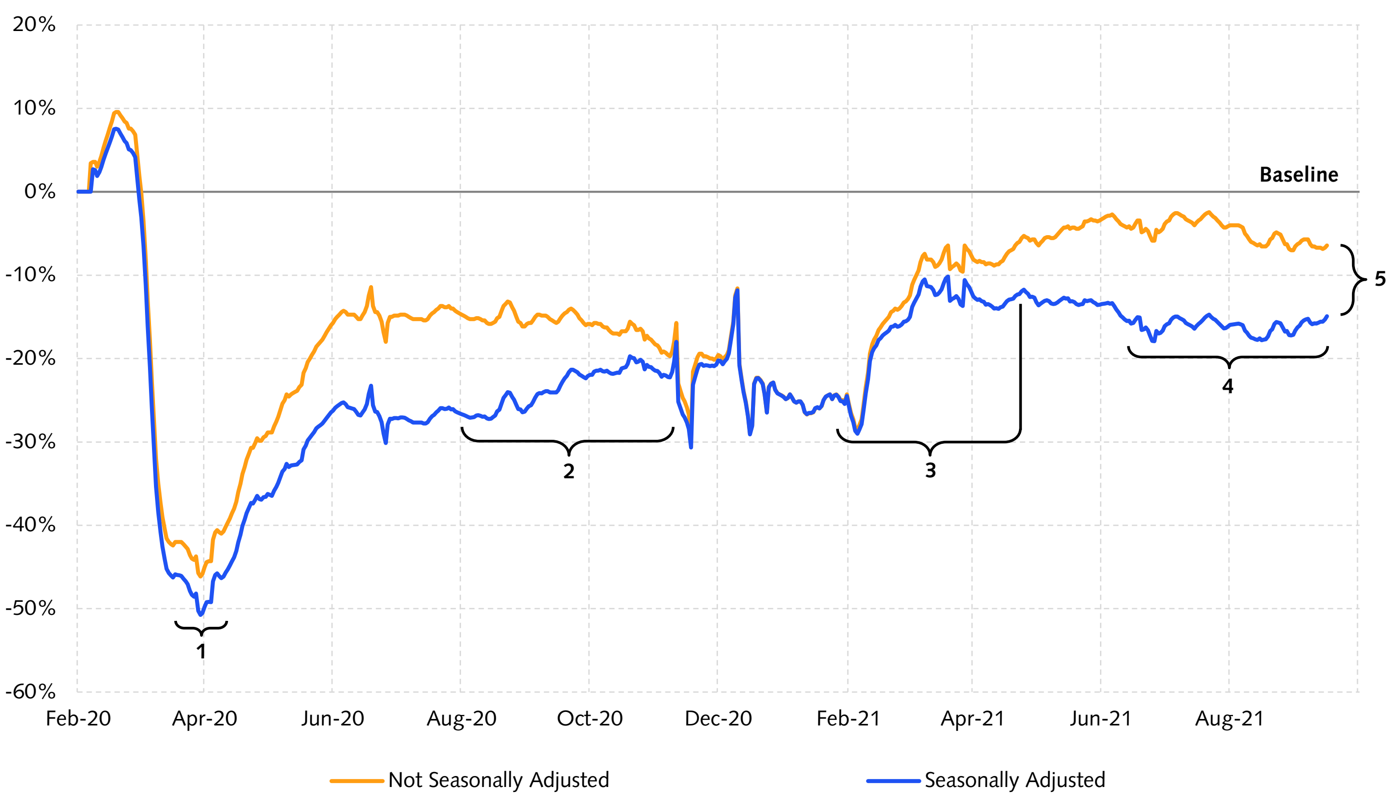
[7] In this case, we use a fixed effects panel regression using both state- and month-fixed effects, although the results are qualitatively similar using either a simple pooled regression or a regression using binary control variables for the COVID-19 shock months.

With these results in hand, we can now adjust the raw GMD series for expected seasonal variation. *Figure* 5 shows the original, non-seasonally-adjusted U.S. retail & recreation mobility data (orange line) and our seasonally-adjusted series (blue line).

Five takeaways stand out from our adjusted results when compared to the unadjusted data in *Figure 5*:

1. The sharp fall in activity evident in GMD after the COVID-19 crisis hit in early 2020 was in fact more dramatic than the raw data suggested because, absent the crisis, mobility would have been seasonally increasing in springtime.
2. Our seasonally-adjusted data indicate an increase in economic activity in late summer and fall 2020, in contrast to the decline in raw mobility that seemed inconsistent with the signal from official economic data at the time, such as nonfarm payroll employment.
3. We find that a modest portion of the upturn in mobility after vaccines became widely available in spring 2021 was seasonally expected, but a robust pace of improvement remained even after seasonal adjustment.
4. Mobility has actually edged down a bit since April 2021 after seasonal adjustment, suggesting that normalization in mobility since the spring has not provided as much continued support to economic recovery as many might believe.
5. Finally, our seasonally-adjusted measure suggests that mobility is still down substantially (about 15%) relative to the pre-COVID baseline, presenting a more pessimistic picture of the overall normalization that has taken place as compared to the unadjusted decrease of only 6.4%. Although not necessarily directly comparable, this seasonally-adjusted number aligns more closely with the level gaps apparent in other key indicators such as leisure and hospitality employment, reported by the BLS, which was down roughly 14% relative to the pre-COVID trend.

Figure 5 U.S. Google Mobility Data Retail & Recreation: Change in 7-Day Moving Average Relative to January 2020 Baseline
(February 15, 2020 to September 30, 2021)



Sources: Google LLC; the D. E. Shaw group.

Conclusion and Directions for Future Work

Google's Community Mobility Reports have been of great interest and utility to many since the early days of the pandemic. As we have seen, the insights derived from GMD are crucial to understanding the economic effects of the pandemic; in turn, we believe it's important to refine those data by adjusting for the effects of seasonality. We view this as just one example of how macroeconomic analysis can be improved by a thoughtful approach to interpreting alternative data sources.

With respect to GMD, we believe our approach is a helpful step, although we have identified a number of promising dimensions along which our work could be extended:

- **Improve model precision:** We have assumed a simple, linear relationship between temperature and mobility. Additional work might relax that assumption and allow for a non-linear relationship between the two quantities, potentially revealing more nuanced patterns of seasonality. Further, we could envision the introduction of two factors to represent climatic variation, such as temperature and precipitation (the typical seasonality for which is slightly out of phase with temperature). Finally, a mobility measure could be constructed to adjust for actual (as opposed to expected) variation in temperature, which might be of interest if national average temperatures are unusual in a given period.
- **Capture additional drivers of variation:** Our approach could be extended to incorporate holiday effects, perhaps by introducing plausible estimates of the magnitude of such effects from an external source and overlaying these estimates on the seasonal pattern derived using our technique. Alternatively, it might be possible to scale off-the-shelf seasonality estimates available for macroeconomic data series such as employment, inclusive of holiday effects, by the peak-to-trough amplitude of seasonality that we estimate for mobility.
- **Expand scope:** In principle, this seasonal adjustment of mobility data could be applied to any country or region that is sufficiently large as to have substantial temperature variation in the geographic cross-section (*e.g.*, Canada or the euro area). The technique could also be applied more broadly than the specific case of GMD, to include other alternative data measures with significant seasonal variation.

Data do not speak for themselves. To make any compilation of raw data more analytically useful, researchers must apply techniques that render that data clean, consistent, and interpretable. Google's Community Mobility Reports typify a range of new alternative data sources that can help inform macroeconomic analysis. We hope this paper will contribute to making mobility data and other innovative data sources increasingly valuable for economic research.

D E Shaw & Co