Opportunity Insights Economic Tracker
Data Revisions

last updated on 2022-05-18

1 Overview

This document provides a description of major revisions to the data posted by the Opportunity Insights Economic Tracker. The document is organized sequentially by series in the tracker, among series that have had substantive data revisions since June 30th 2021 due to changes in data processing or data sources over time.

This document is updated regularly and the following information is subject to change.

For further information or if you have any questions please feel free to reach out to info@ opportunityinsights.org and someone on our team will be in touch.

2 Data Series

2.1 Consumer Spending

Revisions on March 15th 2022

The consumer spending data was revised with three changes:

1. We have improved our modeling of spending levels over time to better control for growth (or shrinkage) in payment providers' customer bases. Previously we controlled for the sharp discontinuities generated by entries and exits of payment providers, but did not control for steady growth or shrinkage in individual payment providers' customer bases. We have detected an increase in the customer base of existing payment providers during the year 2021, which was biasing our estimates of spending upward. The new processing, described in detail below, causes a downward revision in our estimates of the change in consumer spending—primarily in the second half of 2021.

   Our objective is to simultaneously (1) isolate and remove variation in spending driven by the number of debit/credit cards in our sample changing due to growth or shrinkage in a payment provider's customer base, and (2) incorporate variation in spending driven by the number of debit/credit cards in our sample changing due to changes in the utilization of cards, which can reflect underlying economic conditions. To account for both, we first estimate a two-state switching model of the relationship between Personal Consumption Expenditure as measured by the U.S. Bureau of Economic Analysis and the level of credit and debit spending and number of cards in usage we received from Affinity Solutions. Our estimates imply a state change between February 2020 and August 2020: the number of cards in use is predictive of Personal Consumption Expenditures during this period, but statistically insignificant outside this period. This is consistent with a pattern where

people exhibited extensive margin changes in card spending during the initial waves of COVID-19, but outside of that period changes in the number of cards being used reflect changes in the number of people in-sample as opposed to changes in the propensity to consume of a given set of people.

We therefore construct an estimate of spending that reflects total spending during Feb-Aug 2020 and spending per card outside this period. For each location (l), industry (i) and time (t) we compute:

$$\text{adjusted\_spending}_{l,i,t} = \begin{cases} \frac{spending_{l,i,t}}{cards_{l,\cdot,t}} \times \overline{cards_{l,\cdot,t' \in \text{Jan 2020}}} & \text{if } t \leq \text{January 2020} \\[2em] \frac{spending_{l,i,t}}{cards_{l,\cdot,t}} \times \overline{cards_{l,\cdot,t' \in \text{Jan 2020}}} \times \frac{\overline{cards_{l,\cdot,t' \geq t | t' \in \text{Feb 2020}}}}{\overline{cards_{l,\cdot,t' \in \text{Feb 2020}}}} & \text{if } t \in \text{February 2020} \\[2em] \frac{spending_{l,i,t}}{cards_{l,\cdot,t}} \times \overline{cards_{l,\cdot,t' \in \text{Jan 2020}}} \times \frac{cards_{l,\cdot,t}}{\overline{cards_{l,\cdot,t' \in \text{Feb 2020}}}} & \text{if } t \in [\text{March 2020, July 2020}] \\[2em] \frac{spending_{l,i,t}}{cards_{l,\cdot,t}} \times \overline{cards_{l,\cdot,t' \in \text{Jan 2020}}} \times \frac{\overline{cards_{l,\cdot,t' \leq t | t' \in \text{Aug 2020}}}}{\overline{cards_{l,\cdot,t' \in \text{Feb 2020}}}} & \text{if } t \in \text{August 2020} \\[2em] \frac{spending_{l,i,t}}{cards_{l,\cdot,t}} \times \overline{cards_{l,\cdot,t' \in \text{Jan 2020}}} \times \frac{\overline{cards_{l,t' \in \text{Aug 2020}}}}{\overline{cards_{l,\cdot,t' \in \text{Feb 2020}}}} & \text{if } t \geq \text{September 2020} \end{cases}$$

2. We have improved the internal mapping between Merchant Category Codes (MCCs) to corresponding industry sectors to be more accurate. This causes a noticeable revision in the Transportation series.

3. We have improved the amount of state-level data available by changing the aggregation of masked cells. Previously data in masked county-level cells were included in the national-level data but excluded from the state-level data. Now data in masked cells is included in the state-level data, which increases the set of state-level series available.

2.2 Small Business Revenue & Small Businesses Open

Revisions on March 4th 2022

We now derive our published datasets using a county-level panel of small businesses. In each calendar year, we follow the sample of businesses operating during the first week of the year (i.e. we start following a new panel each calendar year). No new businesses enter our panel during the calendar year. Businesses may exit because they stop operating or because the underlying payment processors ceased providing data.

We detect cases where a payment processor disappears by detecting sharp drops in businesses operating, at the national and the state level. We then adjust the series to remove these drops at the State x Industry level, using the following assumptions:

- In the initial Covid period (March - July 2020) we assume momentum, and impute the value of merchants/sales for the week to continue the rate of change we observe in the 4 weeks prior.
- In the rest of the series (before March 2020 and after July 2020) we assume the series stays constant for the adjusted week.

After performing these adjustments, we aggregate up to the State level, the National x Industry

level, and the National level.

Since the no-entry panel does not cover the ZIP code level, we impute the ZIP code level data to reconcile it with the rest of the geographic levels. We perform an additive adjustment on the ZIP level series so that the weighted sum of the ZIP series aligns with the county level same store series. In doing this, we get the levels from the County no-entry panel, and the within-county variation from the ZIP level data.

Revisions on May 17th 2022

We are now releasing county-level and city-level data derived from the same panel of small businesses described in the March 4th 2022 data revision.

2.3 Job Postings

Revisions on December 15th 2021

At the request of Burning Glass we've suppressed the historical data we publish for the following Job Postings series:

- The state aggregate series, state by education level series, and county aggregate series are now only published for the present and previous 12 months
- The state by supersector series, county by education level series, and county by supersector series are only published for the present and previous 6 months

2.4 Employment

Revisions on June 30th 2021

The Employment data was revised on June 30th 2021 due to three independent changes in methodology.

- **Revisions to address end-of-year "churn" in Paychex client base:** over time, some firms enter and exit Paychex's client base. This is especially concentrated at the end of each calendar year, where there is significant churn as firms renew their payroll processing contracts. This creates two sources of error in the tracker series. First, due to this seasonal pattern, the raw Paychex data displays a downwards trend in employment at the end of each calendar year as some clients leave Paychex, followed by an upward trend in employment at the very beginning of each calendar year as new clients join. Second, we take steps to avoid firm entry and exit that are more responsive to firm entry and exit at finer levels of geography, so that we can minimize the number of discontinuous changes in employment. The firm entry and exit occurring due to end-of-year "churn" in Paychex's client base resulted in a discrepancy between the national employment series and the corresponding series at the state level.

  To avoid these sources of error, we have changed the way in which we process employment data around the end of each calendar year. We now adjust for the end-of-year pattern in the Paychex data using data from the end of 2019. For each date between December 10 2020 and January 10 2021, using Paychex data on employment at the national level, we compute the change in employment relative to December 10 2020 at the two-digit NAICS code x income quartile level. We also compute the change in employment between the corresponding day in the previous year and December 10 2019. We divide the change in employment relative to December 10 2020 by the corresponding change in employment the previous year relative to December 10 2019. At the national level, as of June 2021, this change results in an upwards revision of the employment series by between 1 and

3 percentage points from December 2020 through May 2021. We then apply the same adjustment to each two-digit NAICS code x income quartile cell at the state, county and city levels. For this reason, the state, county and city-level series have also been adjusted upwards by between 1 and 3 percentage points from December 2020 through May 2021.

- **Revisions arising from changes to adjustment for firm entry/exit in Paychex data**: over time, the Paychex sample changes as clients begin to use or stop using Paychex's payroll processing services. We previously adjusted for firm entry and exit separately at the national and state levels; for details on this adjustment, see Appendix D of Chetty, Friedman, Hendren and Stepner (November 2020). This introduced the possibility of discrepancies between the national-level and the (employment-weighted) mean of the state-level employment series. Empirically, these discrepancies were small throughout most of 2020, but began to grow in December 2020 due to increased churn in Paychex's client base at the end of 2020, as described above. Since the firm entry/exit adjustment was applied retrospectively, this led to discrepancies throughout the series.

  We have changed our approach to adjusting for firm entry/exit to avoid these discrepancies. In each county x industry (two-digit NAICS code) x firm size x income quartile cell, we now compute the change in employment relative to January 4-31 2020, and the change in employment relative to July 1-31 2020. For county x industry x firm size x income quartile cells with over 50 employees at any point between January 2020 and the end of the series, we reduce the weight we place on the series if we observe changes in employment that indicate firm entry or exit. In particular, we compute the weight on the cell for county c, industry i, firm size s, and income quartile q as:

$$
\begin{aligned}
\text{Weight}_{c,i,s,q} = \max\Big\{ & 1 - \mathbf{1}\{\text{Min Normed July}_{c,i,s,q} \leq 50\} \times (50 - \text{Min Normed July}_{c,i,s,q}) \times 0.02 \\
& - \mathbf{1}\{\text{Max Normed January}_{c,i,s,q} \geq 150\} \times (\text{Max Normed January}_{c,i,s,q} - 150) \times 0.02, \\
& 0\Big\}
\end{aligned}
$$

  where Min Normed July c, i, s, q is the smallest value of indexed employment we observe at each date relative to its mean level over the period July 1-31 2020, and Max Normed January c, i, s, q is the largest value of indexed employment we observe at each date relative to its mean level over the period January 4-31 2020.

  That is, we reduce the weight we place on the cell by two percentage points for each percentage point of growth we observe above 150 percentage points relative to January 2020. We then further reduce the weight we place on each cell by two percentage points of its January 2020 level for each percentage point of decline we observe below 50 percentage points relative to July 2020.

  At the national level, this change revises the aggregate employment series between -2 (in October 2020) and +2 (in January 2021) percentage points, as of June 2021. This change also substantially revises the state-level employment series. The mean revision (in absolute value) for aggregate employment at the state level is 5.6 percentage points as of June 2021. This revision is largest in March 2021, when the mean revision (in absolute value) is 9.1 percentage points.

- **Revisions to address changes in minimum wages**: we use hourly wage thresholds when constructing employment by income quartile in the Paychex data. The threshold for the bottom quartile of employment is $13; that is, workers who earn below $13 are assigned to the bottom quartile, whereas workers earning above (or exactly) $13 are allocated to

other wage quartiles. On January 1 2021, minimum wage changes came into force in CA, MA, AZ and NY, which caused the minimum wage for some workers to move from below \$13 to above \$13. This resulted in a decline in employment for workers earning below \$13, and a corresponding increase in employment for workers earning above \$13, as firms increased workers' wages in response to the minimum wage change. This was reflected in a decrease in low-income employment and an increase in middle-income employment in the tracker data for states in which the minimum wage had increased above \$13.

Though the tracker data for these states accurately represented trends in employment among workers earning below \$13, the movement of workers around the minimum wage threshold created difficulties when comparing trends in low-wage employment across states. We have taken four steps to address this problem. First, we have created additional series for below-median-income and above-median-income employment, which can be downloaded from this repository. As the national median wage threshold in the Paychex data is \$18.18, the below-median-income series is not affected by the shifting of workers induced by the minimum wage change. Second, we have added annotations to the tracker data indicating minimum wage changes in the relevant states. Third, we have suppressed series cut by income quartile in these states after December 1 2020, to avoid displaying series that are substantially affected by minimum wage changes. Finally, the impacts of the minimum wage on low-income employment in these states also affected trends at the national level: as workers' wages increased above the \$13 threshold, national employment fell. When computing the national-level trend in employment, we now exclude trends in CA, MA, AZ and NY between December 10 2020 and February 10 2021. We continue to use trends in these states when computing national-level employment from February 11 2021 forward.

After making these changes, the (population-weighted) RMSE (root mean square error) of the state-level employment series relative to the CPS is 4.55 percentage points as of April 2021, after removing public sector and furloughed workers and expressing employment in seasonally-unadjusted terms relative to January 2020. Though we will continue to assess our series relative to the CPS, users should note that some amount of noise remains in both our state-level series estimates and in the CPS estimates. Particularly in instances where these two measures of employment differ, users may consider both the Opportunity Insights series and the CPS as helpful inputs in identifying local patterns.

2.5 Unemployment Claims

Revisions on July 29th 2021

The unemployment data was revised on July 29th 2021 to correct a data processing error. Previously we assigned continued PEUC and PUA claims to the end of week date indicated by the Report Date rather than the Reflect Date in the Department of Labor data. Effectively, this meant continued PEUC and PUA claims were offset by one week into the future in our data. We've corrected this, and claims now align to the appropriate week.

2.6 Online Math Participation and Student Progress in Math

Revisions on December 14th 2021

As schools have attritioned from the sample we've revised our methodology to impute missing values for an otherwise active school. Previously school retention in the sample was very high and we imputed across the sample observations missing in a given week for a given school as zeros as the likelihood the omission of data was a sign of no usage but continuity in the sample was quite high. Overtime, and particularly during the 2021 fall semester, a subset of schools

naturally attritioned from the sample of schools using Zearn prior to the onset of the COVID-19 pandemic and we paused our imputation procedure for the newest sample. As such we've now revised our methodology to impute missing values within semester based on semester activity. For this subset of schools that are missing observations for a given week within a semester and are active Zearn users at any point within a semester, we impute these values as zeros within the given semester. As such we avoid imputation of schools observations where a school has dropped from the sample in a given semester and identify weeks missing for schools that remain within the sample for a given semester and impute them as zeros appropriately. This change has resulted in a small downward revision to the estimates post summer 2020.

In addition, previously in the case when correcting for transitory anomalous spikes, week to week changes in the underlying levels of greater than 50 percentage points that immediately revert by an equally sized change in the subsequent week, for a given week for a given school we have changed how both the procedure overall operates and how one sided spikes bounded by a missing value are handled. Overall previously the procedure for spike correction detected and applied the correction sequentially for positive and the negative value spikes. This has been revised to apply these changes simultaneously to a series. Given the size of anomalous change needed for a spike to be identified this has a small impact on the overall series. In the case where a one sided spike was bounded by a missing observation we previously in the 2021 fall semester imputed the transitory anomalous spike as a missing observations as well. As we now impute within semester active schools' missing observations as zeros, we now no longer impute these one sided spikes as missing and instead impute them with the 3 week moving average identically to all other transitory spike corrections in the series. This applies to a small number of cases within certain states and has minimal impact.

2.7 COVID-19 Infections

Revisions on March 17th 2021

Previously we pulled reported cases and deaths from the New York Times' COVID-19 database and reported tests from the COVID Tracking Project. At the conclusion of the COVID Tracking Project's efforts in order to collect testing data we instead began pulling reported cases, deaths, and tests at the county level from the Centers For Disease Control and Prevention's COVID Data Tracker and aggregated to other geographies.

Revisions on August 4th 2021

Previously we pulled reported cases, deaths, and tests at the county level from the Centers For Disease Control and Prevention's COVID Data Tracker and aggregated to other geographies. On July 17th the Centers For Disease Control and Prevention began suppressing reported cases and deaths making aggregations across counties no longer feasible and we began to instead pull reported cases and deaths from the New York Times' COVID-19 database, state level reported tests from the Johns Hopkins Coronavirus Resource Center's U.S. testing database, and county level reported tests from The Centers for Disease Control and Prevention.

2.8 COVID-19 Vaccinations

No substantive data revisions reported.

2.9 Time Outside Home

No substantive data revisions reported.