OPPORTUNITY INSIGHTS
ECONOMIC TRACKER

Data Documentation for the Opportunity Insights Combined Employment Series

We use four data sources to obtain information on employment for the United States workforce: Earnin, Intuit, Paychex, and Kronos.

Descriptions of Each Employment Data Source:
Earnin employment data. Earnin is a financial management application that provides its members with funds in advance of receiving paychecks. Workers sign up for Earnin individually using a cell phone app, which connects to the bank account in which paychecks are deposited. We obtain raw anonymized data from Earnin at the paycheck level with information on home ZIP, workplace ZIP, unemployment status, and earnings and hours worked over the last 28 days.

We construct our analysis sample by restricting the sample to workers who are paid on a weekly or bi-weekly paycycle; these categories account for 92% of paychecks. We also restrict the sample to workers who are active Earnin users, with non-missing earnings and hours worked over the last 28 days. Next, we exclude workers whose reported income over the prior 28 days is greater than $50,000/13 (corresponding to an income of greater than $50,000 annually). Users may continue to use Earnin after they have been laid off; we exclude payments which Earnin classifies as unemployment payments, either based on the user's registration with Earnin as being unemployed, or based on the string description of the transaction. Where a user has previously been unemployed, but stops receiving unemployment checks after a certain data, we treat the user as having been re-employed if they receive a payment amount of $200 within the two weeks following their last unemployment check. Using this approach, we find that 90% of Earnin users are re-employed within fourteen days of receiving their last unemployment check.

We then construct daily employment in a repeated cross-section of the Earnin data. We distribute each individual's paycheck over their paycycle by assuming that individuals are employed for each day in their pay period. We assign workers to locations using their home ZIP codes; we use job ZIP code where home ZIP code is unobserved. To account for the delay in receipt of paychecks, we shift the Earnin series back by one week.

Earnin firm data. We use external data sources to gather further information on firm size and industry. To obtain information on industry, we use a custom-built crosswalk created by Digital Divide Data which contains NAICS codes for each employer in the Earnin data with more than ten Earnin users. To obtain information on firm size, we crosswalk Earnin employers to ReferenceUSA data at the firm location level by spatially matching Earnin employers to ReferenceUSA firms.[1] In total, we match around 70% of Earnin employers to ReferenceUSA firms.

[1] We begin by geocoding Earnin addresses to obtain latitudes and longitudes for each Earnin employer. We then remove common prefixes and suffixes of firm names, such as "inc" and "associated". Next, we compute the trigram similarities between firm names for all Earnin and ReferenceUSA firms within twenty-five miles of another. We then select one "match" for each Earnin firm within the ReferenceUSA data, among the subset of firms within one mile. We first match Earnin employers to ReferenceUSA firms if the firms are within one mile of one another, and

Earnin stimulus data. We also receive transaction-level data on all payments received, which we use to measure the receipt of stimulus checks. We classify a transaction as a stimulus check if the transaction (1) has a string description containing words indicating that it is a stimulus transaction, such as "IRS" or "Economic Impact Payment", (2) is of an amount that could be received as stimulus under the CARES Act, and (3) was received after 10 April 2020.

Intuit employment data. Intuit is a financial and accounting software company that offers payroll services to small businesses as part of its Quickbooks program. We obtain month-on-month and year-on-year changes in employment, and payroll at the state and county level each month. To develop a national series, we take the population-weighted average of state changes in each month. If a county is missing employment or payroll information for any month in the data, it is removed from the entire series. Counties omitted for this reason represent a negligible share of total employees in the data.

Paychex employment data. Paychex is a large, national payroll and human resources company providing services to small and medium-sized businesses. We obtain weekly data on employment, payroll, and hours worked for each county by industry by income quartile by firm size group. A small fraction of the data is not assigned a county; we drop these observations. The Economic Tracker series reflects private-sector employment. Therefore, Paychex establishments whose industries are unclassified, or comprise public administration, are dropped for this series. Firm size is defined by the number of employees at a given establishment in 2019, rather than the total number of employees at a parent, or multi-establishment firm. Observations for which firm size is missing comprise a small fraction of the data, and we drop such observations for any analysis concerning the effects of the Paycheck Protection Program (PPP). We also exclude firms that are new to the data in 2020 from PPP analysis. The data is tabulated over the five weeks ending on the Thursday of the given week. We plot the Paychex data at the beginning of this five-week period.

Kronos employment data. Kronos is a workforce management and human capital management cloud provider used by firms of all sizes. Beginning in March 2020, we obtain weekly data, at the firm level, on "punches" for the past week, where each punch represents an employee clocking into work. For each data point, we observe county, state, and firm size, where firm size is the total number of employees at a parent firm, rather than at a single establishment. The employees in the database average an income of $2,000 - $2,500 a month, or $24,000 - $30,000 per year. We drop any observations that do not record a county, noting that these observations make up a very small fraction of the original data.

---

share the same firm name. Second, where no such match is available, we choose the geographically closest firm (up to a distance of one mile) among all firms with string similarities of over 0.6. Third, where no such match is available, we match an Earnin employer to the ReferenceUSA employer within twenty-five miles with the highest trigram string similarity, provided that the employer has a trigram string similarity of 0.9. We then compute the modal parent-firm match in the ReferenceUSA data for each parent-firm grouping in Earnin. Where at least 80% of locations within a parent-firm grouping in Earnin are matched to a single parent-firm grouping in the ReferenceUSA data, we impute that parent-firm to every Earnin location.

Creating a Combined Employment Series:
In order to more precisely measure dis-aggregated employment declines, we combine employment information from three of our private data providers, Earnin, Intuit, and Paychex. Paychex is broadly representative of the entire United States workforce so we use the Paychex as the base for the combined employment dataset. We then use Earnin and Intuit to refine the series in cells that those datasets best represent. Earnin best represents low-income workers so we use Earnin data to refine declines for Paychex workers in the bottom wage quartile. Intuit best represents higher-income workers so we use Earnin data to refine declines for Paychex workers in the top three wage quartiles. Since we receive each dataset at a different aggregation level, we combine using two different methods.

To combine the Intuit data with the Paychex data, we compute a correction factor to adjust the Paychex data according to trends in Intuit for each geography in each month. To create this scaling factor in a given geography, we reweight the Paychex data at the monthly level to match the national Intuit industry distribution. We then take the weighted average between this reweighted series and the Intuit series to create a combined series (the weights are undisclosed for privacy reasons). The correction factor in each month is then defined as the combined employment decline in that month divided by the reweighted Paychex employment decline in that month. We assume that differences between Intuit and Paychex are constant across different industries and income quartiles within the same geography and month. We therefore adjust each industry by quartile cell in the Paychex data by the same correction factor.

We combine Earnin data with Paychex data in the bottom wage quartile by first converting the Paychex data from weekly to daily by assuming employment is constant within each week. Next, we calculate the employment declines relative to January by day within each two-digit NAICS code in each geography by taking the sum of employment on each day across the two datasets, and then re-weight to match the Paychex distribution of NAICS codes for bottom wage quartile workers in each geography.

At this point we have data at the industry by income quartile level for each geography. We then report combined employment relative to January as a seven-day moving average for each super sector and income quartile both nationally and in each state.

In some cases, Earnin and Intuit data do not provide coverage for a given geographical region or industry. In the interest of data privacy we do not publish disaggregated data for which this is the case, and Paychex alone comprises the series. For similar reasons, we do not publish disaggregated data for which Paychex records less than an average of 100 total employees in the second half of 2019 at the industry by geography, or income quartile by geography level. When aggregating employment series to the geographical level without specifications by industry or income quartile, however, data from all series is used.

Advanced Combined Series. Due to a lag in data acquisition the for combined series, we are not able to observe the combined series for dates more recent than June 27. The Kronos data,

however, is updated frequently, and extends to August 16. Because the Kronos data and the first income quartile of the combined employment data exhibit similar trends in the period for which both series exist, we are able to use Kronos data to forecast the first income quartile of the combined series so that it too extends to August 16. We regress the combined series on the Kronos series for the same date (t), as well as on the Kronos series for up to three observations prior (t-7), (t-14), (t-21). We then use the resulting coefficients on Kronos and its lags to predict the combined series for all dates from June 28 through August 16.