Big Data for Regional Employment and Other Economic Statistics

David Mericle
Chief US Economist
Goldman Sachs
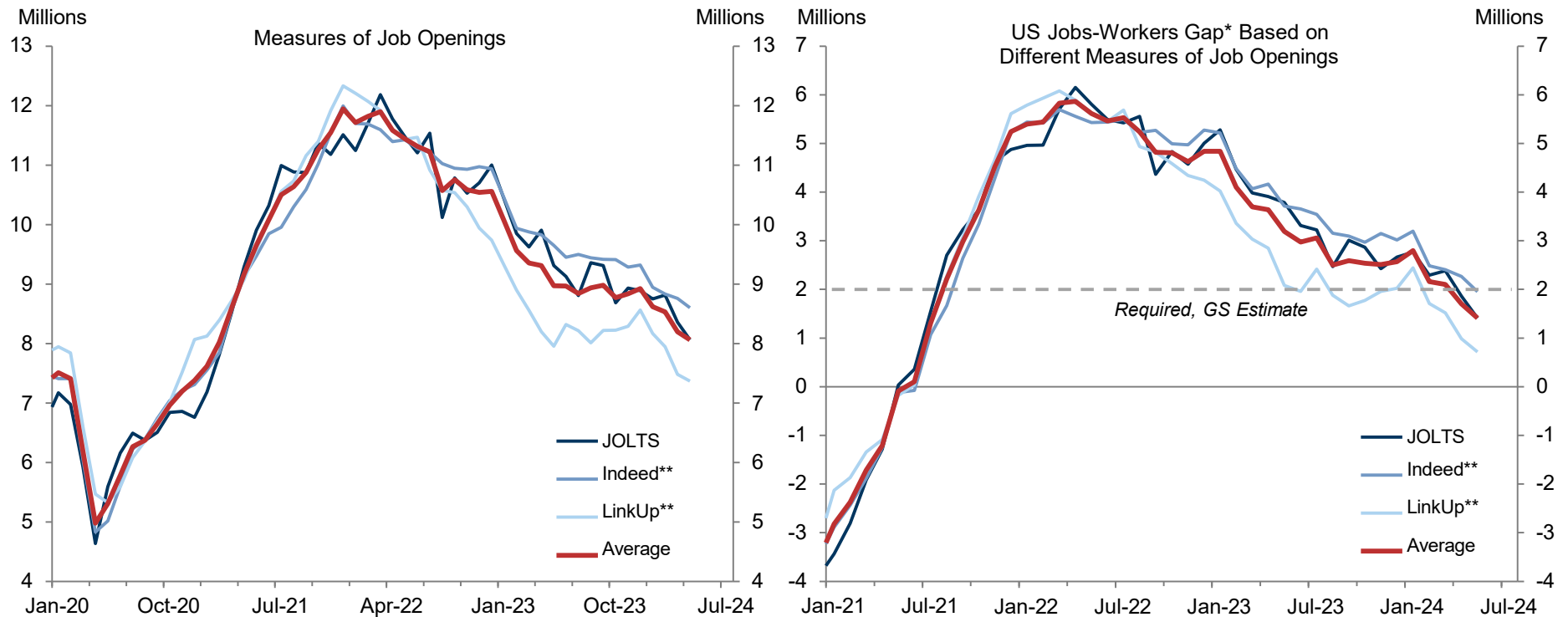
Our Experience Using Alternative Data as Financial Market Economists

- Pre-pandemic, we were generally skeptical of alternative data. Producing high quality economic statistics is an art and a science. The government statistical agencies know how to do it. The private sector simply doesn't. Most of what is available is misleading more often than it is informative.

- Two things changed during the pandemic:
    1. Needed to know the answers immediately.
    2. Macroeconomic fluctuations were so large that the poorer signal-to-noise ratio in alternative data became more forgivable.

- These are not really true anymore. As a result, while we haven't completely stopped using alternative data, we've certainly become more selective.

- We are more inclined to use alternative data at this point when we can get our hands on the raw alternative data and transform it into a credible statistical series ourselves.

Example of a Similar GS Project: the Jobs-Workers Gap

- In 2022 we introduced the jobs-workers gap, initially defined as JOLTS job openings minus the number of people unemployed as a share of the labor force.

- Because the decline in the sample size made JOLTS noisy, we switched the calculation to use an average of official and alternative data measures of openings.

Measures of Job Openings

US Jobs-Workers Gap* Based on Different Measures of Job Openings

Required, GS Estimate

Legend: JOLTS, Indeed**, LinkUp**, Average

* Difference between the number of job openings in the prior month and unemployed workers in the current month.

** Scaled to JOLTS job openings.

Source: Goldman Sachs Global Investment Research.

Feedback 1: Other applications

- Opportunity 1: Real-time distributional data.
  - This is still a weak spot in the US official economic data, even if what is available is better than anywhere else.
  - It's an area where financial company data or private data sets that can be combined with social security numbers could be very useful.

- Opportunity 2: Following workers across jobs.
  - The ability to track workers in payroll data by Social Security number is another potential advantage of private payroll processor or financial data.
  - Potential applications: Are workers who switch jobs getting big raises? Are workers who get laid off finding new jobs? Are workers exiting some struggling industry finding new jobs? Some of this is available, but it's either based on a small sample size (the household survey) or of dubious credibility (private data).

- Opportunity 3: Better real-time spending data by combining official and private data.
  - This is perhaps the most urgent application. The official data are not that useful at a high frequency. But there is a ton of nuance in, say, seasonally adjusting retail sales, so while the alternative data are plentiful, they are unreliable.

Feedback 2: How to incorporate alternative data for timeliness while maintaining accuracy and reliability

- There are potentially huge gains in timeliness. This could eventually become a major advantage of alternative data sources. There is no reason we shouldn't be able to have fairly real-time data on hiring and layoffs or on spending, for example.

- I think the bottom line on accuracy is that the statistical agencies need to have full access to the raw data. You should not be counting on the private sector to have the skill set to turn its raw data into reliable statistical series.   It never will. And even if some company did, people change jobs frequently.

Feedback 3: Issues with alternative data sources like coverage and stability of data providers

In general, the answer to all of these is that again you need access to the full raw data.

- **Coverage.**
    - Most alternative spending data and some labor market data is unrepresentative.
    - *Possible solutions:*
        1. Get a long enough time series to model how to adjust for unrepresentativeness.
        2. Use alternative data to model components (apparel spending, credit card spending, restaurant employment) rather than aggregates.

- **Stability of data providers.**
    - Government statistics can't depend on one private data provider. These series need to be uninterrupted and comparable across decades, and private companies don't always last forever or maintain their data forever. We have experienced many problems like this just in the last four years—many companies that thought it would be cool to provide alternative data during the pandemic no longer provide it.
    - *Possible solutions:*
        1. Diversify data input sources.
        2. Rely on data from too big to fail private companies.
        3. Reach some sort of legal agreement that the data series will be continued.