ADP Research Institute®

The ADP National Employment Report Methodology

INTRODUCTION

Employment estimates in the Current Employment Statistics (CES) report released by the Bureau of Labor Statistics on the first Friday of each month is one of the first indicators that the U.S. government publishes after the end of each month. As such, it garners considerable attention as an early barometer of the health of the economy, specifically the labor market. The business community keenly follows the CES report, and unexpected movements in the report can move U.S. and global financial markets significantly.

Given ADP's payroll database, which includes about one-fifth of U.S. private payroll employment, it is possible for us to estimate the employment change in the US prior to the release of the by the BLS employment report. To this end, we have developed a methodology for predicting U.S. private-sector employment and publish it in the ADP National Employment Report (NER) two days before the release of the CES report.

This article describes the methodology we use in preparing the NER.

DATA PROCESSING

The data processing involves many steps, including the removal of outliers, aggregating records by industry and company size, the creation of matched pairs, seasonal adjustment, and adjustments to match the industry and size distribution of the official employment data.

Much of the data processing used in preparing the NER mimics the methodology used by the BLS to calculate payroll employment for the CES report. One difference is that the ADP NER includes any active employee within the company, whereas the BLS tracks employees only if they are paid during the applicable month.

We aggregate the data into the 13 NAICS private industry sectors: (1) natural resources and mining, (2) construction, (3) manufacturing, (4) trade, transportation and utilities, (5) information, (6) financial activities, (7) professional, scientific and technical services, (8) management of companies and enterprises, (9) administrative and support services, (10) education, (11) healthcare, (12) leisure and hospitality, and (13) other services.

We then create matched pairs of establishments that have reported employment in two consecutive months. There are more than 460,000 such matched pairs, representing more than 26 million employees in the U.S. Each month's data include only the matched pairs available in that month. Matched pairs are aggregated, and matched-pair growth rates of employment are computed into cells made up of nine size classes and the 13 NAICS super sectors.

The CES measures the number of people on payrolls during the pay period that includes the 12th of the month (the reference period). A pay period can be any length of time; the most common pay frequency is weekly. But a pay period can also cover two weeks; it can be bimonthly, monthly, etc. Since the ADP records provide pay dates rather than pay periods, matched pairs must be constructed using interpolation. By combining the pay date and the frequency of pay, we match the BLS pay period concept as closely as possible.

Most of payroll records are not processed on the 12th of each month - companies often pay workers before or after that date. If there is no recorded employment for a given pay period that includes the 12th, but a record exists for either a later or earlier pay period during the month, we estimate employment for the reference period by linearly interpolating between the level of employment on the prior record and the record for the later pay period. The maximum time range for linear interpolation to capture missing employment on the 12th depends on the payment frequency of an establishment. If the previous record falls out of this range, the recorded employment in the same month immediately after the 12th is used for the employment on the 12th.

After calculating the matched employment growth in every industry by size class, we apply the X-13ARIMA-SEATS method to remove outliers and seasonally adjust these employment growth series. De-seasonalized trends for the industrial cells are recalculated with each new month of data. The X-13ARIMA SEATS is the seasonal adjustment method developed by the U. S. Census Bureau in collaboration with the Bank of Spain that integrates an enhanced version of X-12-ARIMA with an enhanced version of SEATS to provide both X-11 method seasonal adjustments and ARIMA model-based seasonal adjustments and diagnostics. This is the seasonal adjustment software the Census Bureau uses.

Each observation in an industrial cell is then compared with the trend value, and outlier observations are removed. Matched employment growth for the 90 size classes is calculated for the second time using the cleaned data. Each cell's employment growth is seasonally adjusted again using the same X-13ARIMA method.

Complicating this process is the fact that in some months there are five weeks, rather than four, between the survey weeks used by the BLS. The bureau employs an adjustment to account for the longer period. We make a similar adjustment in the NER: regressing the growth rate in each cell on a dummy variable which, if significant, is used to eliminate the long-month effect.

While ADP's raw data represents about one-fifth of all U.S. private employment, the composition of the companies paid by ADP does not match the size and industrial composition of the companies in the general population. Two government sources of data are used to adjust for this. First, the Quarterly Census of Employment and Wages, which is a complete count of payroll employment derived from unemployment insurance tax records, is used to set the correct level of employment by industry. Second, company size data available from the Census Bureau is used to calculate shares of employment by company size. Combining these two sources of data provides a benchmark level of employment by industry and company size.

These benchmark levels are set annually in March, which corresponds to the month in which BLS benchmarks its employment estimates with the Quarterly Census of Employment and Wages (QCEW) data. The seasonally adjusted matched-sample growth rates by industry are computed by taking a weighted average of the matched sample growth rates by size within each industry. The weights are based on monthly interpolations of the March benchmark values. For the time period after the last benchmark, the weights are extrapolated forward to the latest month using the matched-pair growth rates within each size class. The result of this process is the growth rates of size-weighted industries from April 2001 to the current month.

MODEL SPECIFICATION AND ESTIMATION

The ADP model uses historical data beginning in 2001, which covers nearly two full business cycles, enhancing the robustness of the model fit.

A vector autoregressive model (VAR) is used to predict BLS employment. A key benefit of a VAR model is that it captures complex multivariate relationship between the variables included in the model, producing more accurate forecasts. However, VAR models create a tradeoff between oversimplification and over-parameterization. That is, omitting relevant variables may lead to the loss of information, while inclusion of irrelevant variables may create more noise.

The VAR model used by ADP mitigates this tradeoff by using the moving average of the growth rates for a given industry over the previous two months as an explanatory variable. This is a change from our previous model which used three separate lags of industry growth rates. This allows for a reduced number of parameters to be estimated. Instead of three explanatory variables represented by three lags of industry growth rates, our model now uses one explanatory variable represented by a moving average of the first two lags. Moreover, using a moving average of BLS industry growth rates helps reduce the inherent volatility and thus noise in the data.

The ADP model includes:

- Growth rates for each industry derived from the sample data
- Unemployment insurance claims
- Extra cooling/heating degree days, regional snowfall index and the cost of a natural disaster[1]; these variables enter only the equations of such weather sensitive industries as construction, leisure and hospitality, manufacturing, resource and mining, and trade
- A time dummy variable for the month of August
- The growth rate of the leading indicator[2] in the previous month; this variable enters the equation of only two industries: administrative, support and waste management and remediation services (super sector 56) and manufacturing

[1] The historical data (CPI adjusted) for estimation were taken from National Centers for Environmental Information (available online at https://www.ncdc.noaa.gov/billions/events/US/1980-2018). For the forecasts, Moody's Analytics will input advance estimates of dollar cost severity for natural disasters.

[2] Leading indicator is the Conference Board Composite Index of leading indicators

- The growth rate of US retail trade in the previous month; this variable enters the equation for trade
- The vector of the moving averages (or the means) of the employment growth for each of the industries over the previous two months

Natural Disasters

A significant change in enhanced methodology is an explicit accounting for the impact of natural disasters. Hurricanes have had especially large impacts on employment in recent years, but floods, earthquakes and fires have been important at other times.[3] To account for natural disasters, the model now includes a variable which measures the cost of the damage caused by natural disasters. We use an alternating model setup, so that in months when there are no significant natural disasters, the set of explanatory variables consists of the calculated sample growth rate, the growth in unemployment insurance claims, the August dummy variable, weather-related variables, and the moving average of each industry's growth rate over the two previous months. In any month where there is a significant natural disaster, the set of explanatory variables will include only the variables relevant for that particular month and excludes lagged variables such as the moving average of each industry's growth rate over the two previous months.

To capture the post-disaster impact of a natural disaster, the employment growth rates in the post disaster months are adjusted.

Weather Impact

Previous research has documented that temperature, snowfall, and precipitation have a meaningful impact on employment.[4] To account for the effect of temperature, our revised model includes cooling and heating degree days available from the National Centers for Environmental Prediction. More precisely, the variable used in our model is the difference between the 4-week moving average and the 5-year moving average of cooling and heating degree days. These variables are important for weather sensitive industries such as construction, leisure and hospitality, manufacturing, resource and mining, and trade. Snowfalls are captured using the Regional Snow Fall Index from National Centers for Environmental Information.

Labor Strike Impact

BLS employment can be significantly influenced by large employee strikes. Since they do not occur often, in order to account for the impact of strikes which impact at least 10,000 workers

[3] See Stobl, 2011 and Noy, 2009.
[4] See, for example, Burke, Hsiang, and Miguel (2015), Boldin and Wright (2015), Bloesch and Gourio (2015).

in a given month the forecast obtained from the model will be adjusted for the number of employees who took part in the strike, according to BLS.[5] Months with large strikes (i.e., more than 10,000 workers off the job), impact BLS estimates of employment which are used in the model in subsequent months. To prevent a previous strike from materially impacting future months' estimates, lagged BLS data used in the model will be adjusted to remove the strike effect.

August Effect

BLS consistently reports lower levels of employment changes in the month of August, likely in part due to low response rates by larger companies as many of their employees take vacations during the month. This effect is captured in our model by including a dummy variable for the month of August.

References

Bloesch, J. and Gourio, F. (2015). *"The Effect of Winter Weather on U.S. Economic Activity".* Federal Reserve Bank of Chicago Economic Perspectives.

Boldin, M. and Wright, J.H. (September 2015). *"Weather adjusting economic data".* Brookings Papers on Economic Activity.

Burke, M., Hsiang, S. M. and Miguel, E. (2015). *"Global non-linear effect of temperature on economic production".* Nature, 527, pp.235-39.

Strobl, E. (2011). *"The Economic Growth Impact of Hurricanes: Evidence from U.S. Coastal Counties".* Review of Economics and Statistics, 93(2), pp. 575-558.

Noy, I. (2009). *"The Macroeconomic Consequences of Disasters".* Journal of Development Economics, 88(2), pp. 221-231.

[5] https://www.bls.gov/ces/cesstrkhist.htm