Always Designing for People™

The ADP National Employment Report

Introduction

Payroll employment 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 report, and unexpected movements in the report can move U.S. and global financial markets significantly.

Given Automatic Data Processing Inc.'s large payroll database, which includes about one-sixth of U.S. private payroll employment, it is possible to estimate the employment change in the US prior to the official release by the BLS payroll employment report (Current Employment Statistics). To this end, we have developed a method of predicting the private-sector employment two days before the release of the CES report.

Although ADP's database does not precisely match the distribution of employment as reported by the BLS, we adjust the ADP data to match the CES distribution. The methodology is described below.

Data processing

The data processing involves many steps, including the removal of outliers, identifying payroll records by industry and company size in order to aggregate the individual payroll record data into industries 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 mimics the methodology used by the BLS to calculate payroll employment for the CES report.

The data files used in the model start from 2001. We use the following data fields from weekly data files: number of employees, frequency of payment, processing date of payment, industry and payroll processing service center information.

Unlike the BLS, which tracks employees paid during the month, the ADP methodology includes employees who are active. These are chosen over paid because employees may receive more than one paycheck for incentive pay or bonuses. This would result in multiple records and, therefore, the inclusion of such records would overstate the actual number of employees on a company's payroll.

Then the raw data is checked for outliers, anomalies and inconsistencies. This may mean adjusting records for which there are missing or incorrect data fields. If the correct data field (such as the industry or location) cannot be determined, the record is excluded.

Always Designing for People™

As a final step before creating the matched pairs, companies are classified 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.

Matched pairs

We then create matched pairs of establishments that have reported employment in two consecutive months. There are more than 400,000 such matched pairs, representing more than 23 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 the payroll records are not processed on the 12th of the month. Rather, they may pay workers before or after that date. If there is no recorded employment for the 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.

Seasonal adjustment

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.

Always Designing for People™

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. A similar adjustment is made using the ADP payroll data: regressing the growth rate in each cell on a dummy variable which, if significant, is used to eliminate the long-month effect.

Adjusting the ADP data to align with the overall US Economy

Although ADP processes the payrolls for 1 in every 6 U.S. workers, the composition of those workers in different size classes and industries does not exactly match the size and industrial composition of the firms outlined in the CES. Therefore, the Quarterly Census of Employment and Wages, which is a complete count of payroll employment derived from unemployment insurance tax records, is used to adjust the ADP data. The QCEW data by the nine size classes are available only for March of every year. This is the month to which BLS benchmarks its CES estimates against the 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 QCEW March values. For the time period after the last benchmark, the weights are extrapolated forward to the latest month using the ADP matched-pair growth rates within each size class.

Regression and results

We use a restricted structural vector autoregression model to estimate the monthly change in private payrolls by regressing the most recent payroll growth reported by the BLS for each super sector on a constant term and (1) ADP matched-pair growth rates by industry, (2) lagged values of BLS estimates of growth of employment by industry with industry specific restrictions, (3) unemployment insurance claims (UNI_US), (4) oil prices, (5) the Michigan Consumer Sentiment Index (CSENT_US), (6) and the Composite Index of Leading Indicators (LEAD_US) . Unemployment insurance claims are used as a common driver across all industries. The claims enter the equations with a negative sign, which signals lower expected employment growth rates when unemployment insurance claim levels increase. The remaining three external drivers are restricted to affect selected industries. Oil prices drive natural resources and mining industry. The Composite Index of Leading Indicators drives the manufacturing and administrative and support services and has a positive impact on employment projections. The Michigan Consumer Sentiment Index enters the construction equation with a positive sign. The equations for all industries are estimated simultaneously. Thus, the model is made up of 13 endogenous variables—the 13 BLS super sector industries—and 17 exogenous variables—the 13 ADP super sector industries, unemployment insurance claims, oil prices, the Michigan Consumer Sentiment Index and the composite index of leading indicators.

The equation for each industry is constructed as follows:

For example,
% change in construction = -1.17 +0.47 * % change in size-weighted ADP construction +0.203 one-month lagged growth in natural resource and mining -0.206 * two-months lagged growth in natural resources and mining +0.08 * three-months lagged growth in natural resources and mining +1.09 * one-month lagged growth in manufacturing -0.35 * two-months lagged growth in manufacturing -0.40 * three-months lagged growth in manufacturing +0.25 *

one-month lagged growth in construction -0.24 * two-months lagged growth in construction +0.13 * three-months lagged growth in construction -0.29 * one-month lagged growth in transportation and utilities +1.16 * two-months lagged growth in transportation and utilities -0.27 * three-months lagged growth in transportation and utilities -0.19 * one-month lagged growth in information -0.10 * two-months lagged growth in information +0.087 * three-months lagged growth in information -0.30 * one-month lagged growth in finance and insurance -0.99 * two-months lagged growth in finance and insurance +1.03 * three-months lagged growth in finance and insurance +0.12 * one-month lagged growth in education -0.21 * two-months lagged growth in education +0.001 * three-months lagged growth in education -1.02 * one-month lagged growth in healthcare +1.85 * two-months lagged growth in healthcare -0.54 * three-months lagged growth in healthcare +0.62 * one-month lagged growth in leisure and hospitality +0.24 * two-months lagged growth in leisure and hospitality +0.76 * three-months lagged growth in leisure and hospitality -0.37 * one-month lagged growth in professional, scientific and technical services +0.48 * two-months lagged growth in professional, scientific and technical services -0.63 * three-months lagged growth in professional, scientific and technical services +0.34 * one-month lagged growth in management of companies and enterprises +0.53 * two-months lagged growth in management of companies and enterprises -0.52 * three-months lagged growth in management of companies and enterprises -0.41 * one-month lagged growth in administrative and support services -0.15 * two-months lagged growth in administrative and support services -0.41 * three-months lagged growth in administrative and support services +0.15 * one-month lagged growth in other services -1.31 * two-months lagged growth in other services -0.22 * three-months lagged growth in other services -0.0136 * Michigan Consumer Sentiment Index -0.0216 * percentage change in unemployment insurance claims.

This method allows different trends of employment by size of payroll within industries, while assuming that the other industry-wide relationships implied by the regressions hold for all size classes within an industry. Inclusion of lagged employment in the regressions controls for shifting differences between the BLS sample and the ADP payroll database, and inclusion of other employment-related variables controls for differences in the definitions of employment used by BLS and ADP and the difference in sampled firms. The industry-level data are then aggregated to total private employment using industry weights from the BLS' QCEW. Then the percentage change estimates are calculated back to employment level change in thousands.

The final step of the process is to use the estimation results to predict the current month's employment changes by super sector and size class, which can then be aggregated into total private sector employment.

Company-level estimates

ADP data are used to construct employment by five company size classes: small companies (1-19, 20-49), medium companies (50-499), and large companies (500-999, 1,000+).

Each control unit in the cleaned ADP datasets with outliers removed is matched to a company, identified by the employer identification number. The size of the control unit may coincide with the size of the company, or the company can comprise many control units. The employment of all the control units with a common EIN is summed up to generate the company-level employment.

The companies in each super sector are then grouped into nine cells according to their sizes, and the matched pair employment growth for each of the 90 industry-company cells is calculated. The matched employment growth for each industry-company cell is then seasonally adjusted and the long-week effects are removed.

The QCEW establishment-based industry cell sizes and shares of each company cell within every establishment cell are used to construct the company-based industry cell size in the each year's March benchmark month. The benchmarked company cells are then extrapolated to grow at the matched-cell employment growth rates. Monthly distributions of company cells in every super sector are calculated from the extrapolated company cell employment levels.

The monthly industry distribution of company cells and the forecasted industry employment growth are then combined to produce the forecasted employment growth for company cells in the NER.