September

2019

Employment Projections Technical Report

Employment Security Department
WASHINGTON STATE

2019 Employment Projections Technical Report

Published September 2019

Washington State Employment Security Department
Suzi LeVine, *Commissioner*

Daniel Zeitlin, *Director*
Policy, Data, Performance and Integrity

Steven Ross, *Director of Labor Market Information*
Labor Market and Economic Analysis

Prepared by

Alex Roubinchtein, *Projections and Statistical Analyses Manager*
Bruce Nimmo, *Economic Analyst*

For more information or to get this report in an alternative format, call the Employment Security Department Labor Market Information Center at 360-507-9621.

Contents

Summary

Annually, the Employment Security Department (ESD) creates 2-, 5- and 10-year employment projections.

Projection processes are guided by state and national requirements, available data and current software tools. The projections process consists of two major steps: the creation of industry projections and the conversion of industry to occupational projections. The conversion process is based on Occupational Employment Statistics (OES) survey data.

Employment projections start with time series of covered employment, processed within the unemployment insurance system at the county level. National forecasts from Global Insight are used as regressors for aggregated state employment forecasts.

Projection models for industry series are not predefined. This means assumptions are not made about which models are best for any given series. An optimization process selects the best combination of model outputs. The result is that model output selection may vary for each industry employment series.

Introduction

In this paper, we discuss the technical processes used to produce industry and occupational projections.

Data preparation and forecasting are done using R-software.

The projections process utilizes seven models (six models plus optimization). The seven models are: innovations state space exponential smoothing (a.k.a. ETS), auto regressive integrated moving average (ARIMA), complex exponential smoothing auto (CES), theta, hierarchy, constrained and an optimization process that combines outputs from the first four models. ARIMA and CES models use regressors. Consequently, regressors play a role in optimized forecasts.

Hierarchical time series forecasting functions are found in the R package hts. The hts package specializes in forecasting time series that can be disaggregated into hierarchical structures using attributes such as geography or levels of aggregation. Forecasts are generated for each series at each level of a hierarchy. These forecasts are then combined and balanced by an optimization function within this package. The combination approach optimally combines independent base forecasts and generates a set of revised forecasts that are as close as possible to the initial univariate forecasts, but also balanced within the hierarchical structure.

The constrained model was used for the first time last year. This model is based on complex log transformations of variables. Further details are available at: https://robjhyndman.com/hyndsight/forecasting-within-limits/

Important parameters used in model selection are **historical trend growth rates occurring after a major breaking point** for each series. To define these rates we used R's BFAST (Breaks for Additive Season and Trend) package.

Industry projections are produced at two levels: aggregated and detailed. The aggregated series are referred to as "series" and the detailed series are referred to as Industry Control Totals (ICTs). For each of the series (aggregated and detailed), we produce multiple forecasts.

Selected state series projections are used as regressors for state ICT projections.

In turn, state ICT projections are used as regressors for Workforce Development Area (WDA) ICT projections.

Base projections are benchmarked by the addition of noncovered employment (i.e., not covered by unemployment insurance). Noncovered employment comes from Current Employment Statistics (CES) data. A reconciliation process balances results of different levels of aggregation between regions and the state.

Staffing patterns are created and used to convert industry projections into occupational projections. Occupational openings include openings due to growth and to turnover. This year we used two types of turnover rates, state specific alternative rates and national separations rates. The alternative rates track openings created by turnover within occupations (i.e., workers stay within occupations but transfer to different companies) and when workers leave one occupation for another or leave the workforce. The separations method measures job openings created by workers who leave occupations and need to be replaced by new entrants. In this method, workers who exit the labor force or transfer to an occupation with a different Standard Occupational Classification (SOC) are identified as generating separations openings at the national level.

In addition to projections, we produce additional products:

- Skill estimations and forecasts based on job announcements from Help Wanted Online® (HWOL) skills/occupational data.
- The skill estimations are used to create matrices of related occupations based on skills. Such matrices are state specific.
- Occupations in Demand (OID) list. This list is used for determining eligibility for a retraining program (Training Benefits), as well as other education and training programs.

Industry projections

Data

- Covered employment time series
- Global Insight forecasts

Covered employment time series are based on Quarterly Census of Employment and Wages (QCEW) data. More information is available at: https://esd.wa.gov/labormarketinfo/quarterly-census.

Global Insight is an international economics organization well known for their data and forecasts.

Software used

The primary software used for forecasting is R-software (R). R is an open source object oriented language with advanced statistical and optimization features. It allows programmers to operate directly on vectors and matrices. This creates significant advantages over languages and software with sequential access to data, like SAS, when producing occupational projections.

State level aggregated industry forecast

Data preparation

Initial covered employment at the county level was aggregated into 42 industry groups (cells), presented in the file: allcodes file. Forty cells were aggregated for nonfarm employment, one for agriculture and one for private households. The cells for nonfarm employment are closely associated with employment related cells from Global Insight. However, to meet state employment projection requirements and Occupational Employment Statistic (OES) definition requirements, some cells were disaggregated for state projections. For example, we disaggregated transportation equipment to aerospace and other transportation equipment. The state and local government cell was disaggregated into three cells: government education, hospitals and other government. Two industries related to the information sector were also disaggregated.

We transformed some codes from Global Insight in order to match them with codes used in state projections. Due to these transformations, 40 state cells obtained matching relationships with Global Insight national forecasts. Two state cells, agriculture and private households, do not have related national forecasts.

A crosswalk between 4-digit North American Industry Classification System (NAICS) codes, Industry Control Totals (ICTs), aggregated series codes and common combined codes can be found at: allcodes file. As can be seen in the file, aggregated series do not in all cases represent an aggregation of ICT codes. The main reason is that aggregated series reflect commonly used definitions from the Current Employment Statistic (CES) classification system, while ICT codes reflect industry definitions used in the OES system. To match CES and OES systems, we created combined codes, which match aggregated series forecasts with detailed ICT forecasts.

Global Insight uses data with quarterly frequencies. In contrast, our historical and forecasted data use monthly frequencies. To make national forecasts usable as regressors for state forecasts, they must be interpolated from quarterly into monthly frequencies. To achieve this we used the **denton-cholette method** from the R-library **tempdisagg**. The **denton-cholette method** uses temporal disaggregation techniques to disaggregate low frequency time series to high frequency series. An in-depth discussion of disaggregation methods, is available at: https://journal.r-project.org/archive/2013-2/sax-steiner.pdf

Parallel processing

When processing time series, we use R's parallel processing capability. This capability reduces processing time by distributing processes over multiple cores within a computer. The preparation for using parallel processing includes: defining the number of cores in the computer and setting the number of used cores as the number of available cores minus 1. One core must be left to run general computer functions. After the number of cores are defined, core clusters need to be set up and registered with parallel processing functions. R-libraries need to be connected with registered clusters. Parallel processing has some limitations; interactive graphs are not available and failed iterations are not printed in error handling procedures. Parallel processing-time reduction can be computed (approximated) by multiplying the value 0.7 times the number of computer cores used in processing.

The main forecast procedure

The main industry projections process uses a batch process loop. Within this loop, models for each time series undergo training on full and hold-out samples. Model performance is used to create combined weighted forecasts. The four base models plus the combined forecasts constitute the output from this main procedure. The R-libraries used for data processing and forecasting are: **readxl**, **dplyr**, **forecast**, **foreach**, **doParallel** and **dfoptim**. Flexibility is built into code by using functions that measure the size of imported datasets. The only static element within the code is the preset maximum forecast horizon.

For each industry cell, which have regressors, we use the following four base models:

- Exponential smoothing: innovations state space autoregressive model with an optimized selection of smoothing parameters (criteria: minimum Mean Absolute Percent Error [MAPE]).
- ARIMA: The function **auto.arima** is used to optimize selection of parameters for ARIMA, seasonal ARIMA and periods of seasonality, etc., with regressors (criteria: AIC [Akaike's information criterion]).
- Complex Exponential Smoothing (CES) Auto. Function estimates CES in state space form with information potential equal to errors with different seasonality types and chooses the one with the lowest information criteria (IC) value.
- Theta model: decomposition approach to forecasting with optimized parameter of Box-Cox transformation. Model was the best performer in M3 forecasting competition data and used as one of the main benchmarks in M4 competition. Further details about the M-competitions are available at: https://www.mcompetitions.unic.ac.cy/

Autoregressive or external regressors

The exponential smoothing and theta models are autoregressive and only use historical employment time series to forecast employment. The auto ARIMA and CES models can include external regressors.

State space method

The state space method offers a unified approach to a wide range of models and techniques. In general, it includes equations for unobserved states and includes observation equations. Unobserved states (such as level, growth and seasonality) can be subject to change with time. Since the model can account for such changes, it is called **innovative**. The general model can be described as follows:

Let $x_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$, be a state vector, where l_t - stands for level; b_t - for growth; and s_t - for seasonality. State space equations can be written in the form:

$$y_t = w(x_{t-1}) + r(x_{t-1})\epsilon_t$$

$$x_t = f(x_{t-1}) + g(x_{t-1})\epsilon_t$$

where ϵ_t - error terms with mean zero and variance δ^2. The equation $\mu_t = w(x_{t-1})$ is a one-step-ahead-forecast for the states y_t - observed numbers (employment in our case). Other parameters are defined by the type of mode. For instance, models with multiplicative errors use $r(x_{t-1}) = 1$ resulting in $y_t = \mu_t(1 + \epsilon_t)$. Thus, relative errors for multiplicative models are represented by $\epsilon_t = (y_t - \mu_t)/\mu_t$. As can be seen in the state space model, the term “dynamic” refers to states, rather than to observed numbers as in traditional descriptions. More details about the state space model are available at: http://listinet.com/bibliografia-comuna/Cdu339-C8E1.pdf

In R’s *forecast* package, similar state space models for 30 exponential smoothing variations are subject to internal optimization. In our model specifications, we chose to allow a damping parameter as a variable. This choice improved the quality of model estimations compared to the use of a default value of one.

Technical details about the models, which are used in the *forecast* package, can be found at: http://robjhyndman.com/papers/automatic-forecasting/

Complex exponential smoothing

CES is based on conventional exponential smoothing and the notion of the “information potential,” an unobservable part of time series that the classical models do not consider. In contrast to exponential smoothing, CES can capture both stationary and non-stationary processes, giving it greater modelling flexibility. Any measured time series contains some information, which may be less than the time series in all its totality and potentially unobservable. It was shown that there is a convenient way to write the measured series and the information potential using a complex variable: $y_t + ip_t$ where y_t is the actual value of the series p_t is the information potential on the observation and is the imaginary unit, the number that satisfies the equation: $i^2 = -1$. For more information see: https://www.researchgate.net/publication/322468244_Complex_Exponential_Smoothing_for_Seasonal_Time_Series

Theta model

The theta method of Assimakopoulos and Nikolopoulos (2000) is equivalent to simple exponential smoothing with drift. This is demonstrated in Hyndman and Billah (2003). A series is tested for seasonality. If deemed seasonal, the series is seasonally adjusted using a classical multiplicative decomposition before applying the theta method. The resulting forecasts are then re-seasonalized. Prediction intervals are computed using the underlying state space model.

Full sample and hold-out sample

For each time series and each model, two forecasts are produced:

- one based on a full sample; and
- one based on a 24-month hold-out sample.

For the full sample forecast, we used all available historical data from January 1990 to June 2018 for parameter estimations. We then forecast for the period from July 2018 to December 2027. Estimations for the hold-out forecast are based on historical data from January 1990 to June 2016 and then forecast from July 2016 to June 2018. As a result of this method, for each time series we have four fittings on a full sample and four hold-out sample forecasts for the four models: innovations state space exponential smoothing, ARIMA, CES and Theta.

Optimization

We use an optimization procedure to define weights for combining the four full forecasts. The weights are based on the performance (fitting results) of the models on both full sample and hold-out sample forecasts. We used mean absolute scale errors (MASE) as a measure of performance, which allows comparisons of different series. MASE is a measure of forecast accuracy proposed by Koehler & Hyndman (2006).

$$MASE = \frac{MAE}{MAE_{in-sample.naive}}$$

where

$$MAE = \frac{\sum_{i=1}^{n} | x_i - \hat{x}_i |}{n}$$

expresses the average absolute difference between each point of time n series x and $\hat{x}$ forecast of x. $MAE_{in-sample,naive}$ is the mean absolute error produced by a naive one-step-ahead forecast, calculated on the in-sample data. For seasonal data, a step consists of 12 months.

We calculate two mean absolute scaled errors for each of the four models: for full sample fitting $MASE_{full}$ and hold-out sample forecast $MASE_{hold}$.

We define the optimum four weights $z = (z_1, z_2, z_3, z_4)$ for combining forecasts for four model $(i = 1, \dots, 4)$ classes by solving the problem, find unknown $z = (z_1, \dots, z_4)$, for which:

$$\sum_{i=1}^{4} (MASE_{full}(i) + MASE_{hold}(i)) * z_i \rightarrow min$$

The combined optimum forecast is defined as $\sum_{i=1}^{4} f\,or(x_t^i) * z_i$.

For two series without regressors we use the same optimization procedure, but exclude regressors from the AMRIMA and CES models.

Variations of the main procedure can be applied to different levels of aggregations.

Forecast outcomes

The main procedure produces five forecasts for each time series: four base models plus a combined optimum forecast. We repeat this procedure for log transformed series and thus potentially have 10 forecasts for each series.[1]

After creation of the main forecast output, we make use of two additional forecasting models: hierarchy and restricted.

Hierarchy forecast

Hierarchical time series forecasting functions are found in R's hts package. The hts package specializes in forecasting time series that can be disaggregated into hierarchical structures using attributes such as geography. Forecasts are generated for each series at each level of the hierarchy. These forecasts are then combined and balanced by an optimization function within this package. This approach combines independent base forecasts and generates a set of revised forecasts that are as close as possible to the initial univariate forecasts, but also balanced within the hierarchical structure. Hierarchy forecasting was applied to both aggregated series and detailed Industry Control Totals (ICTs):

- State series, or ICT, to state total.
- WDA series, or ICT, to state series, or ICT, to state totals.

We used two model options available in the *hts* package, *arima* and *ets*.

Technical details about hierarchy forecasting are available at:
https://www.otexts.org/fpp/9/4

[1] Estimations for some models can fail for a variety of reasons. The chance for failure increases for unstable series with small numbers involving some zeros. To avoid interruptions in loop processing, for failed series, we use **tryCatch** loops, rather than the default **do** loop. Using a **foreach** loop allows us to have all successful forecasts in output lists as well as identification of all failed series.

Restricted forecast

To impose a positivity constraint (limit), we used the log scale, by specifying the Box-Cox parameter. More details for this procedure are available at: https://robjhyndman.com/hyndsight/forecasting-within-limits/

Supplemental parameters used for forecast selections

Directly calculated parameters

The following parameters were calculated and used for forecast selections:

1. Historical growth rate for the last two years (2016Q2-2018Q2).
2. Historical growth rate for the last 10 years (2007-2017)
3. Historical annual growth for all years (1990-2017)
4. Absolute difference in average growth rate between June and July for the last three years of history (2015-2017) and the first projected year (2018).

Statistical parameters

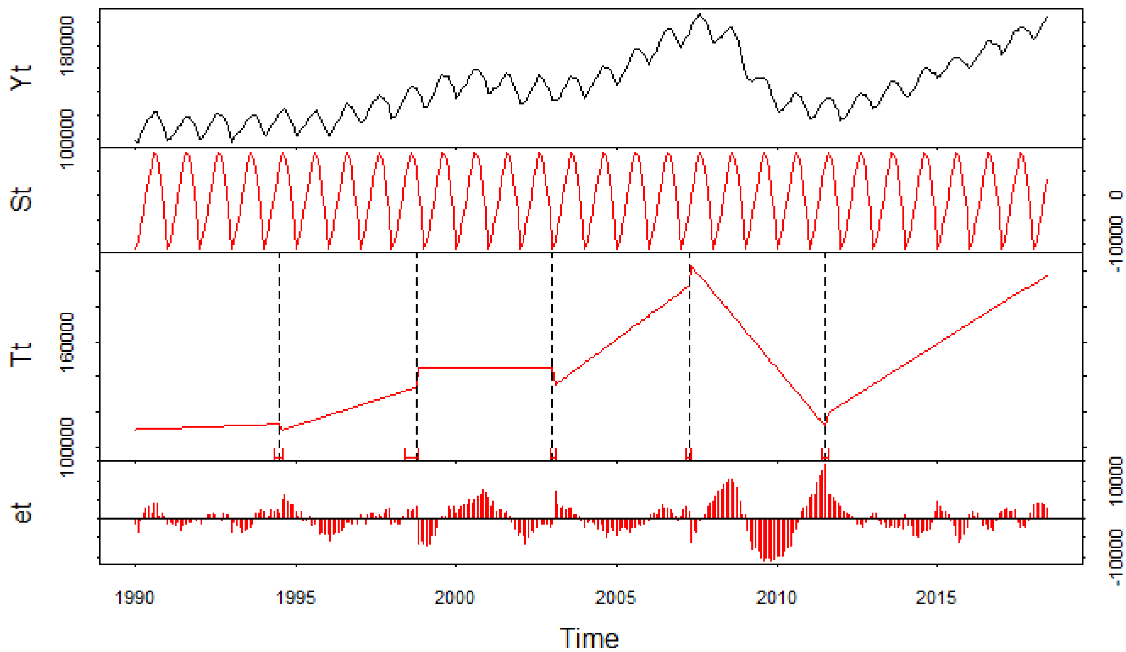
Statistical parameters created in this round of projections are historical trend growth rates occurring after a "major breaking point," with confidence levels for estimations, for each series. These parameters are historical trend growth rates. To define these rates we used R's BFAST (Breaks for Additive Season and Trend) package.

The main goal of the package is to integrate the decomposition of time series with methods for detecting and characterizing change within time series. *BFAST* estimates the time and number of abrupt changes within time series. The base decomposition of time series Y_t for time t, from the beginning to the end of a period of interest, is:

$$Y_t = f(S_t, T_t, e_t), \quad where: S_t - seasonal, \quad T_t - trend \quad and \quad e_t - remainder$$

For instance, a graph of breaking points for construction employment from January 1990 to June 2018 is presented in *Figure 1*.

Figure 1: Breaking points for construction employment



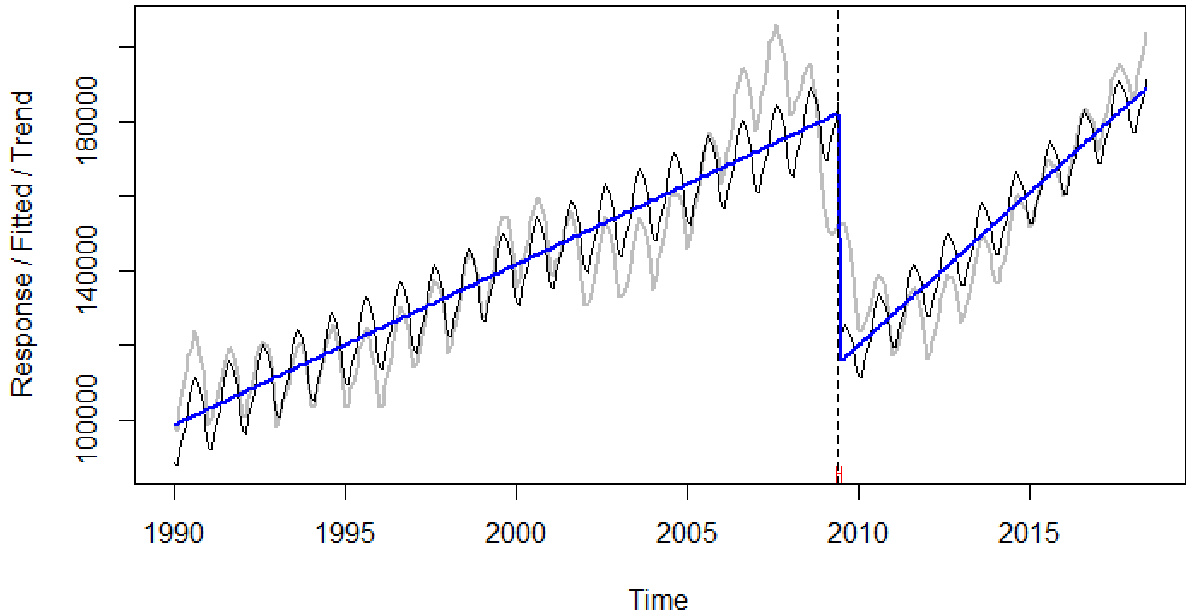
In *Figure 1*, there are five breaking points for the period under consideration: February 1995, March 1999, April 2003, May 2007 and June 2011. The confidence intervals (red marks on the T_t axis) are one month before and one month after, for the majority of breaking points except for February 1995. The February breaking point consists of two months before and one after.

The most significant atypical behavior of this time series is in the interval between May 2007 and June 2011. The remainders on the e_t axis are most significant. They are the largest at the last breaking point.

This construction example gives an idea of how the *BFAST* package can be used for time series evaluation. The package also has the useful function *bfastmonitor* which can be used to monitor the consistency in new data, based on observed evaluated data. Evaluated data can include all available historical data, custom specific intervals and model definitions of largest historical stable intervals. The intervals for evaluation cannot be less than 25 percent of all observed data points.

In this round of projections, we used the function *bfast1* to identify one major breaking point for each series. One of the custom control features in this function is the ability to set the minimum share of time points for each of two intervals. We set our share at the level of 0.25. The graph for the same construction employment as in *Figure 1*, but with only one major breaking point is in *Figure 2*.

Figure 2: The major breaking point for construction employment



In *Figure 2*, the major breaking point for construction employment occurred in May 2009, with confidence intervals between May and July 2009. By supplementing the output from *bfast1* with the function *bfast01classify*, we can produce annualized growth rates for both intervals (before and after the major breaking point). In addition, *bfast01classify* can create significance levels for fitted models.

In our example, the growth rate on the first interval was 3.22 percent and on the second 5.64 percent. Both estimations have extremely high levels of significance. In our evaluations, we mainly used growth rates for the second intervals as long as they had high significance levels.

To evaluate the "smoothness of transition" between historical and forecasted numbers, we calculated the average value for the last three years of changes between June and July and compared the results with the changes between the last month of historical data of June 2018 and the first month of forecast data, July 2018. Any big discrepancies between averaged values and the transition from last historical to a first forecasted value identifies forecasts that are not good candidates for selection.

Selection of aggregated state forecasts

At this stage of the projections process, we select just one of 13 state aggregated series forecasts. Selected series are used as regressors in later steps. It is possible that a selected series represents a linear combination of a few forecasts. This selection process is an *informal process* and is based on the *supplemental parameters used for forecast selections* described above.

The following considerations were used in the informal selection of forecasts:

- historical growth rates for the entire history period and for the last interval after a major breaking point (if significance for the second interval is high);

- the latest aggregated long-term employment forecast from the Office of Financial Management (OFM) and short-term forecast from the Economic and Revenue Forecast Council (ERFC);
- previously published forecasts: our forecasts, OFM and ERFC forecasts;
- smoothness of transitions between the last month of historical data and the first month of forecasted data;
- general knowledge of underlying trends in specific industries and;
- avoidance of extreme growth and decline rates.

We attempt to select forecasts with growth rates close to those used by OFM and ERFC. We do this unless we have convincing evidence that the OFM and ERFC forecasts are inconsistent between themselves or have significant differences with previously produced results. ERFC forecasts are used for budgetary planning purposes and require the use of adaptive controls. Consequently, forecasts should be updated often to reflect the most current data. In such cases, up-to-date data takes priority over long stable forecasting time periods. Our forecasts are mainly used for career development. Prospective students need forecasts that are stable for medium-range time periods. Frequent updates of forecasts in such cases would be disruptive.

In addition, we think it is important to apply general economic analysis to forecasts of industry employment. An example of such analysis can be seen in our pre-recessionary analysis of productivity trends. Specifically, in the construction industry, our analysis demonstrated that a combination of high employment growth coexisted with a high rate of declining productivity. The declining productivity was compensated for by large price index growth. The combination of high employment growth, low productivity and high prices could not last indefinitely and pointed to a high probability of a downward correction. Therefore, our employment projections for construction growth were more pessimistic. In fact, this type of correction happened during the great recession and created a large drop in construction employment. This drop was the largest among all major industry sectors.

A similar situation occurred in the forecasting of aerospace employment. The delay of Boeing's Dreamliner aircraft, combined with high demand, created an artificial boom in aerospace employment trends. Our projections were more in line with normal aerospace long-term declining cyclical trends. In both cases, our declining trends were subject to strong criticisms. Subsequent events affirmed the practice of applying knowledge of underlying trends. The artificial aerospace conditions eventually ceased and declining cyclical trends continued.

In total, combining base and log transformation models, among 42 series, optimum models were selected 25 times. Models from class ets were selected 10 times. Models from class arima were selected in four cases. Regression models were selected in six cases and hts in two cases.

Selected series forecasts are used in the following three independent, but related steps.

Step 1. Draft of state level aggregated benchmarked forecast

Actual historical covered employment numbers for the last 18 months are combined with noncovered employment from the CES program. These numbers are aggregated to two base points used for forecasts resulting in a change of frequencies. The two points are: annual average for the year 2017 and quarterly average for second quarter 2018 (2018Q2). This procedure is called *benchmarking*. Unlike benchmarking by the CES program, we do not use wedging or other adjustments to incorporate code changes. Thus, our benchmark numbers can be slightly different from CES numbers. The growth rates from selected forecasts are applied to benchmarked base numbers. The result is that we produce three required points for industry forecasts: 2020Q2, 2022 and 2027.

The results of benchmarked forecasts are rolled up to create multi-level tables that are somewhat comparable with CES tables. The table is submitted to regional economists, state agencies specified in state law and the Economic Revenue Forecast Council for their feedback.

Step 2. Detailed state level forecasts

For the most part, we repeat all procedures from **Step 1**'s aggregated series for the detailed state level forecasts. We used selected aggregated forecasts as regressors for state detailed forecasts. We use the same formal adjustments and supplemental parameters for selection. However, in the selection process, we do not use aggregated external forecasts (ERFC and OFM). Instead, in some cases, we use common combined codes rolled up from aggregated series forecasts.

Eight combined codes are the same as aggregated series and ICT codes and selected aggregated series forecasts were used in such cases. One ICT code for education, 6100, is a combination of two aggregated series: education services and government education services. We combined them to come up with an ICT forecast. The combined ICT directly matches with one combined code.

In nine cases where ICT codes equaled aggregated Ser codes, the aggregated Ser forecasts were used. Among the remaining 274 selected ICT forecasts, the largest number, 93, belongs to the ets model. The optimum combination of models was second with 91. Arima was third with 52 selections. Theta came in fourth with 20. CES was fifth with 12. The simple naive model was selected six times.

Step 3. Local workforce development area (WDA) forecasts

The procedures for producing and formally adjusting local level aggregated and detailed forecasts, in a mathematical sense, are the same as for the state.

We use state aggregated and detailed forecasts from previous steps as regressors for WDA aggregated and detailed forecasts.

Three possible outcomes are possible for each series:

3.1 - All options did not fail and thus we have 13 outputs for each of the series or ICT.

3.2 - All options failed and thus we do not have any forecasts.

3.3 - Some options failed and we have fewer than 13 outputs.

If all options fail,[2] we assign statewide growth rates for those series.

For outcomes within steps 3.1 to 3.3, we use formal adjustment procedures similar to those described in **Step 1**.

Creating weighted WDA forecasts

The purpose of this step is to select for each available forecast, among all available options in each area, a weighted forecast. As in previous steps, we want a forecast that produces growth rates for periods of interest closest to the ones at the state, i.e., regressor, level. Adjustments in this step are completely formal, conducted in R and exclude interventions.

Let's define g_s, g_m and g_l as short-term (2018Q2-2020Q2), medium-term (2017-2022) and long-term (2022-2027) average annual growth rate employment projections for each state forecast (aggregated or detailed).

Let's also define t $(t = 1, \dots, 456)$ as a time index for all series with monthly frequencies and j as an index for forecast options for each available series i. For outcome 3.1, it will be 13 series. The numbers will be less if some forecast failed (outcome 3.3). Let's define n_i as a subset of non-failed forecasts for each series i. Then, the optimization problem for each of the available series i can be written as follows.

Find the weights w_j of aggregation for forecast options from the following conditions:

$$0 \le w_j \le 1, \quad j = 1, \dots, n_i$$

$$\sum_{j=1}^{n_i} w_j \left(\left(\left(\sum_{t=364}^{366} y_j^t \Big/ \sum_{t=340}^{342} y_j^t\right)^{0.5} - 1 - g_s\right)^2 + \right.$$

$$\left. + \left(\left(\sum_{t=385}^{396} y_j^t \Big/ \sum_{t=325}^{336} y_j^t\right)^{0.2} - 1 - g_m\right)^2 + \left(\left(\sum_{t=445}^{456} y_j^t \Big/ \sum_{t=385}^{396} y_j^t\right)^{0.2} - 1 - g_l\right)^2 \right) \to min$$

where $y_j = (y_j^1, \dots, y_j^{456})$ vectors of employment numbers for option j.

After weights are determined, the weighted forecasts for each series i are simply calculated as:

$$y_t^f = \sum_{j=1}^{n_i} w_j * y_j^t$$

[2] Some local (aggregated and detailed) series might not exist (i.e., have zero covered employment). These too can be interpreted as failed series.

Step 4. Draft of industry forecasts

Formally adjusted detailed industry forecasts are benchmarked in the same manner as described in **Step 1**. After benchmarking, numeric discrepancies between the state and WDAs are resolved through an informal process. Since discrepancies are normally very small, due to formal adjustments, for the most part state numbers are slightly modified to meet WDA totals. In some cases, the inverse is required and WDA numbers are subject to adjustments to meet state totals.

Minimal informal interventions are possible at this stage of the projections process. Interventions are based on known discrepancies between state and local area trends. For industries with atypical behavior (e.g.; large employers that dominate a given industry), interventions are applied (e.g.; smoothing of extreme growth rates).

Industry tables for each WDA are created in the same manner as in **Step 1** and submitted for internal review by regional economists.

Step 5. Final adjustments and output of industry employment projections

Generally, this step is informal and involves processing feedback from state agencies and regional economists. There are generally two types of responses:

1. Responses based on expected, event-based information.
2. Responses on the level of suggestions related to major trends.

Event-based information is related to expected closures and layoffs or expected new hirings due to business expansions or relocations. In general, forecasts predict ongoing trends. Scenario analysis can be used to forecast possible shocks.

Event-based adjustments are applied, distributed and balanced between aggregated and detailed forecasts.

Suggestions related to major trends are evaluated based on available data and underlying economic trends and each receive a response. We either provide reasons for rejecting suggestions or inform sources that suggestions will be incorporated into projections. Accepted suggestions are incorporated in the most conservative manner.

The main way to incorporate trend-based suggestions at the state level is by returning back to the **Selection of aggregated state forecasts** and repeating all subsequent steps for affected industries. In addition, it is possible to modify models for affected industries. For suggestions related to local areas, we return back to **Step 4**.

After the process is complete and all aggregated and detailed projections are benchmarked, informal adjustments are required to meet the following balancing requirement. These balancing requirements must be met for each of the three projected time periods (2020Q2, 2022 and 2027):

- For each industry, totals for local areas for aggregated and detailed industries should be equal to state numbers.
- For each area, a balance between detailed and aggregated forecasts should be achieved at the aggregated combined series level.

Satisfying the above two conditions leads to a balance between state and local area forecasts at the combined series aggregation level.

While **Step 5** processes may seem complicated, for the most part, they are not difficult or time consuming after all automated adjustments have been made. Discrepancies are normally not large and are handled by either a bottom-up approach where state totals are made equal to areas totals, or by using a top-down approach with proportional adjustments to local area numbers so that they meet state totals.

For a few series with multiple cross-match-adjustments at the combined series level, the process is more complicated. In this round of projections, we mainly used a bottom-up approach for all adjustments: from detailed areas to detailed state and aggregated areas and then to aggregated state.

The industry projections process produces two major outputs:

1. Aggregated industry projections for the state and all areas, which are rounded to the closest 100 and rolled up to create a multi-level table that is somewhat comparable with CES tables (as in **Step 1**).
2. Industry Control Total (ICT) files for the state and all areas, which are not rounded.

The aggregated projections output is published and is used for analyses in projections reports, but is not used for producing occupational projections. The non-rounded ICT output is used in subsequent steps for producing occupational forecasts. ICT industry projection numbers, rounded to integers, are also published.

Occupational projections

Data used

Occupational employment projections result from the conversion of industry employment into occupations. These conversions are based on occupation/industry ratios (i.e., staffing patterns) from the Occupational Employment Statistics (OES) survey. The survey is conducted by the Labor Market and Economic Analysis (LMEA) division of the Employment Security Department (ESD) in cooperation with the U.S. Bureau of Labor Statistics (BLS).

The full OES survey is conducted over a three-year cycle. One-third of the survey is completed each year. Occupational estimations and projections are subject to the limitations of the OES survey. The survey includes nonfarm employment and agriculture services, but excludes noncovered employment, self-employment and unpaid family members, major agriculture employment (except services), and private households.

The sample for the OES survey is designed for metropolitan statistical areas (MSAs). From the perspective of statistical accuracy for occupational projections, this level of aggregation is the most appropriate. However, for different applications like the Training Benefits Program, we use WDA aggregation levels for regional details. The direct use of OES staffing patterns for WDAs can create significant bias for a variety of reasons.

The data source for the creation of staffing patterns was almost entirely raw survey data. The majority of data comes from the BLS final (i.e., not preliminary) files. These files include employment, employment distributions by wage intervals, final weights and indicators showing whether original survey responses or imputed responses are used. Response imputations come from other similar in-state areas or from other states. Imputations can have a significant influence on the OES-based staffing patterns.

The process of selecting staffing patterns for each industry and area includes calculating industry totals from raw files. Totals are calculated for weighted employment with imputation and without imputation. Totals are then compared with Industry Control Totals (ICTs) for the base year periods 2017 and 2018Q2. Our preference is to use data without imputations, but in some cases, they do not represent significant shares of employment in ICT files. In such cases, either samples with imputations or substituted staffing patterns are used. These substitutions are introduced using statewide staffing patterns. In some cases, substitutions may come from other similar in-state areas. Staffing patterns can create significant bias for industries with high shares of noncovered employment, which are not part of the survey (e.g., religious organizations).

For a few industries, combined staffing patterns were used between areas. For 2019, this only occurred for the King County and Snohomish County WDAs. This was a necessary step because King and Snohomish counties were combined in the OES survey sample. National staffing patterns are used as a last resort and for this year's projection cycle, it was only necessary for three industries: iron and steel mills and ferroalloy manufacturing, cutlery and hand tool manufacturing and religious organizations.

Some problems are unavoidable and significantly influence final occupational estimations and projections. For example, doctors can be employed by clinics or hospitals, but often are employees of independent associations or are self-employed. For this reason, staffing patterns for medical institutions are bound to be biased. Also noteworthy this year was the limited use of some results from the 2012 OES green supplemental survey for agricultural industries. The green supplemental survey allowed us to create staffing patterns for agriculture, based on weighted sample responses. This year we also used older survey responses and current covered employment to account for a major employer, which has been missing from the latest surveys.

To manage the staffing pattern process, we used two additional columns in our ICT files. The columns indicate the type and area of origination for staffing patterns. For instance, if an original staffing pattern is used, the area of origination will be the same as the area for industry employment. If an original is not used, the area of origination might be the numeric indicator zero (0) for statewide substitution, or the numeric indicator 5.5 for the combination of King and Snohomish counties, etc.

Occupational projections use some national inputs. The inputs are self-employment and unpaid family worker ratios, separation rates for each occupation and change factors (which we modify).

Step I. Making staffing patterns with selected change factors

The source files are:

- Raw survey data with and without imputations.
- Extracts from the 2012 OES green supplemental survey for agricultural industries.
- National staffing patterns (used only for private households).
- Hiring data from HWOL.

Each selected vector $a_{i,j}^{v}$, where $i = 1,..,m$, index for occupations, $j = 1,\ldots,n$, index for industries and $(v \in V)$ index for areas (can be the number or combination of numbers) is normalized:

$$\sum_{i=1}^{n} a_{i,j}^{v} = 1,\ \mathit{for\ all\ industries\ j\ and\ areas\ v}.$$

The combination of vectors $a_{i,j}^{v}$ is often called a *matrix of staffing pattern*. It represents the normalized structures for the distribution of industry employment between occupations for the base period(s).

We define the index for base and projected periods as $t = 1,..,5$ and for this round of projections it represents the years 2017, 2018Q2, 2020Q2, 2022 and 2027. The base staffing patterns are used for the years 2017 and 2018Q2 (t=1,2). For other periods, patterns are modified with the incorporation of limited change factors.

Change factors $c_{i,j}$ come from national data. They predict expected changes in occupational shares for each industry over 10 years. The reliability of change factors tends to be low because unlike industry employment, there are no historical time series for occupational employment.

Due to the lack of historical trends upon which to base future changes, BLS uses researchers' expectations about structural occupational changes within industries to create change factors. Within this BLS process, there is a high degree of subjective judgment.

This is especially true since change factors must be developed for each occupation within an industry. Occupational outputs are very sensitive to these change factors. It is very important to evaluate the adequacy of change factors before use. Incorrect change factors can drastically increase errors in projections.

We used national change factors in combination with historical state data to create change factors for a limited number of cells. The factors were created only where state historical series were available and were consistent with the suggested change factors from national files. In such cases, we used the most conservative estimations. Change factors reflect expected changes over 10 years, and staffing patterns for projected periods must be modified accordingly:

$$c_{i,j}^{2} = (c_{i,j})^{0.2}, \quad c_{i,j}^{5} = (c_{i,j})^{0.5}, \quad c_{i,j}^{10} = c_{i,j}.$$

For the two base periods, change factors are not used. The value for all missing change factors can be assumed to be one, and modified staffing patterns are calculated as:

$$a_{i,j}^{v,t} = c_{i,j}^{t} * a_{i,j}^{v} \quad where\ period\ index\ t = 1,..,5,$$

where $a_{i,j}^{v}$ are the staffing patterns for the base period. Staffing patterns for each period, industry and area need to be normalized to totals of 1.

Step II. Calculation of occupational projections

The results from the previous calculations, for each component $x_{j,v}^{t}$ of the original ICT vectors, in each time period, give as output normalized vectors for occupational distributions $a_{i,j}^{v,t}$. The base occupational employment for each period is simply calculated as:

$$e_{i,t}^{v} = \sum_{j=1}^{n} a_{i,j}^{v,t} * x_{j,v}^{t}, \quad i = 1,..,n, \quad t = 1,..,5 \quad v = 0,\dots,12$$

due to:

$$\sum_{i=1}^{m} a_{i,j}^{v,t} = 1 \quad for\ each\ j,\ v\ and\ \ t\ we\ have \sum_{i=1}^{m} e_{i,t}^{v} = \sum_{j=1}^{n} x_{j,v}^{t}.$$

The totals of occupational employment for each area in each point of time equals the totals of industry employment.

The numbers for the base period 2018Q2 represent distributions of industry employment between occupations according to normalized staffing patterns. Often, these too are called staffing patterns. These staffing patterns are convenient for publications, but need to be normalized for any calculations outside base periods or with modified ICT files.

Step III. Calculations of self-employment and unpaid family members

Raw self-employment ratios s_i for each occupation come from national data. Based on these ratios, we calculate unadjusted estimated self-employment totals for each area for the base year period 2017 as:

$$se_l = \sum_{i=1}^{m} s_i * e_{i,1}^{v}$$

We use estimated numbers of self-employed from the American Community Survey to adjust national self-employment ratios for each area. Let's define the survey numbers for each area as $self_l$. The ratio of adjustment is defined as $ratio_l = self_l/se_l$. The ratio is assumed to be the same for all periods and occupations and in this way, adjusted numbers of self-employed for each area v and occupation i are defined as:

$$ase_{i,t}^{v} = se_{i,t}^{v} * ratio_l$$

Then the total of occupational employment is defined as:

$$et_{i,t}^{v} = e_{i,t}^{v} + ase_{i,t}^{v}$$

Step IV. Adding openings due to separations and state specific alternative rates

The Bureau of Labor Statistics (BLS) separations method measures job openings created by workers who leave occupations and need to be replaced by new entrants. In this method, workers who exit the labor force or transfer to an occupation with a different Standard Occupational Classification (SOC) are identified as generating separations openings at the national level. This method does not track turnover within occupations. Turnovers within occupations occur when workers stay in occupations, but change employers. This also means that jobs filled by interstate movement, when workers stay within occupations, are not identified as new jobs.

Beginning with the 2017 projections cycle, ESD created a new Washington state specific *alternative* occupational method to the BLS *separations* method. The objective was to track job openings that occur when workers transfer within occupations. For simplicity, we refer to this method as the *alternative* method and to the rates as the *alternative* rates.

The *alternative* method is based on Washington state wage records, making the resulting rates specific to Washington state.

The *alternative* rates track openings created by turnover within occupations (i.e., workers stay within occupations but transfer to different companies) and when workers leave one occupation for another or leave the workforce.

The method consists of three major steps:

1. Estimating the total number of annual industry transfers that include:
    a. Transfers between industries
    b. Transfers inside industries
    c. New individuals in Washington state wage records (wage file)
    d. Exits or individuals who are no longer in the wage file

2. Converting industry transfers to occupational transfers using occupation-to-industry staffing patterns (shares of occupations for each industry).
3. Calculating alternative rates as total transfers, minus growth or decline, divided by estimated occupational employment for a base period.

Information about the separations methodology is available at: https://www.bls.gov/opub/mlr/2018/article/occupational-separations-a-new-method-for-projecting-workforce-needs.htm and information about the *alternative* methodology is available at: https://esd.wa.gov/labormarketinfo/projections.

From a mathematical point of view, calculations are the same for all three rates. Let's define the rates as r_i. Then the openings due to replacement or separations for each occupation for each period are defined as follows:

$$rep_{i,v} = \frac{et_{i,b}^{v} + et_{i,f}^{v}}{2} * r_i,$$

where $et_{i,b}^{v}$ and $et_{i,f}^{v}$ are employment totals for the beginning and end of the period. We calculate replacements for periods between 2018Q2 and 2020Q2, 2017 and 2022, 2022 and 2027.

Step V. Making final outputs

Final outputs include the following results. Calculations are rounded to integers and aggregations to totals for two and three digit SOC levels:

- Total occupational employment estimations $et_{i,t}^{v}$ for all five periods.
- Average annual growth rates for three periods: 2018Q2-2020Q2, 2017-2022 and 2022-2027.
- Average annual number of openings due to growth $gr_{i,v}$ for each period, which are calculated by subtracting starting points from end points and then dividing the results by the number of years in the period (two or five).
- Average annual openings due to replacements $arep_{i,v}$ calculated by dividing $rep_{i,v}$ by the number of years in the period.
- Total openings due to growth and replacements are calculated as follows:

$$tot_{i,v} = max((arep_{i,v} + gr_{i,v}),0)$$

We initially round employment estimations and then aggregate them to totals and to two and three digit SOC codes. In this way, results are additive for each column. However, the above formula for calculating total openings introduces non-additivity into the calculations. As a result, the aggregated level of total openings might not equal the total of growth plus replacement.

Some detailed occupational employment estimations are suppressed due to confidentiality. Data are aggregated to three levels, including totals, and then suppressed. Suppressed data normally impact only the most detailed data. Higher levels of aggregations are normally not impacted by suppression.

General use, employment projection byproducts and tools

Employment projections provide a general outlook for industries and occupations in Washington state. Appendices to the main 2019 projections report describe how occupational projections are used as the basis for the Occupations in Demand (OID) list, covering Washington's 12 workforce development areas and the state as a whole. It also describes how we converted occupational projections to skill projections using specific skills extracted from Washington state job announcements. During the creation of skills projections, we produced skills-to-occupation matrices. These matrices allowed us to create a state-specific tool useful for matching any given target occupation to related occupations (see *Appendix 3* in the main *2019 Employment Projections Report*).

Some of the models and functions (i.e., tools) developed during the production of employment projections are very applicable to other fields related to time series analyses. For example, R's innovations state space models and CES models are more effective for seasonal adjustment than the so-called "one level" model, which currently is the most used model for seasonal adjustments. The innovations state space models allow for dynamic changes in seasonal parameters, and in this way, makes the implementation of additive adjustment factors very effective. In the one level model, such factors lack the ability to reflect trend changes. The CES model incorporates regular and non-regular seasonal variations. The use of additive parameters, in conjunction with hierarchy forecasting tools, would allow users to create seasonally adjusted series, balanced between different levels of aggregation, as a co-integrated process. Commonly used seasonal adjustment models do not allow for the direct balancing of different levels of aggregations. Such balancing is normally achieved by a top-down disaggregation or a bottom-up aggregation of seasonally adjusted series.

Useful capabilities for time series analyses are contained in R's *bfast* package. This package makes possible a better understanding of historical trends and the impacts of specific events, like recessions, on such trends. The *bfast* package is very effective for identifying pro-cyclical and counter-cyclical industries. One of the most useful features of *bfast* is its ability to monitor the consistency of new data, based on observed evaluated data. Evaluated data could include all available historical data, custom specific intervals or the largest historical stable intervals defined by models. This package could be used, for instance, for evaluating the typical or atypical behaviors of Current Employment Statistic (CES) samples or employers' reports that are processed within the unemployment insurance system. The use of *bfast*'s automated tools could significantly increase the speed, quality and consistency of analyses within any organization's processes.

References

Assimakopoulos, V. and Nikolopoulos, K. (2000). The theta model: a decomposition approach to forecasting. International Journal of Forecasting 16, 521-530.

Commandeur, JFJ and Koopman, SJ (2007). *An Introduction to State Space Time Series Analysis* Oxford University Press, Oxford. Accessed on 07/24/2019.

Hyndman, RJ and Khandakar, Y (2008). *Automatic time series forecasting: the forecast package in R*. Journal of Statistical Software 27(3). Accessed on 07/24/2019.

Hyndman, R.J. and Athanasopoulos, G. (2013) *Forecasting: principles and practice*. OTexts: Melbourne, Australia, Section 9.4. Accessed on 07/24/2019.

Hyndman, R.J., and Billah, B. (2003) Unmasking the Theta method. International J. Forecasting, 19, 287-290. Accessed on 07/24/2019.

Quarterly Census of Employment and Wages, https://esd.wa.gov/labormarketinfo/quarterly-census Accessed on 07/24/2019.

Sax, Christoph and Steiner, Peter (2013) *Temporal Disaggregation of Time Series*, Contributed Research Article. Accessed on 07/24/2019.

Svetunkov, Ivan and Kourentzes, Nikolaos (2016). Complex Exponential Smoothing for Seasonal Time Series, https://www.researchgate.net/publication/322468244_Complex_Exponential_Smoothing_for_Seasonal_Time_Series. Accessed on 07/26/2019