Seasonal Adjustment Methodology for Weekly Unemployment Insurance Claims Data

by Thomas D. Evans and Connor J. Doherty

Thomas D. Evans and Connor J. Doherty are mathematical statisticians on the Statistical Methods Staff, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics.

Seasonal adjustment is a statistical technique that attempts to measure and remove the influences of predictable seasonal patterns to reveal how weekly UI initial and continuing claims change from week to week.

Over the course of a year, the amount of UI claims undergoes fluctuations due to seasonal events including changes in weather, holidays, and school schedules. Because these seasonal events follow a more or less regular pattern each year, their influence on statistical trends can be eliminated by seasonally adjusting the data from week to week. These seasonal adjustments make it easier to observe the cyclical, underlying trend, and other nonseasonal movements in the series.

The Bureau of Labor Statistics (BLS) seasonally adjusts two weekly unemployment insurance (UI) claims series for the Employment Training Administration (ETA) at the Department of Labor (DOL). These series are for initial and continued claims. Seasonal factors for these series can be found on the BLS website at Seasonal Adjustment for Weekly Unemployment Insurance Claims : U.S. Bureau of Labor Statistics (bls.gov) and on the DOL website at Unemployment Insurance Weekly Claims Data, Employment & Training Administration (ETA) - U.S. Department of Labor (doleta.gov).

Seasonally adjusted data for the current year are produced with a technique known as *projected seasonal adjustment*. Under this practice, seasonal factors are forecasted out from the end of the series and applied to the series as new observations become available. Beginning with 2002, we use claims data through the month of January in the current year to make the forecasts. By using data beginning and ending at the end of January, we avoid having holiday effects directly at the end or beginning of our series. Before 2002, BLS incorporated data only through December of the prior year in the development of new seasonal factors.

Utilizing data through the end of January each year, BLS reestimates the seasonal factors for both time series by including another full year of data in the adjustment process. Based on this annual reestimation, BLS revises historical seasonally adjusted data for the previous 5 years. As a result, each year's data are generally subject to five revisions before the values are considered final. The fifth and final revisions to data for the earliest of the 5 years are usually quite small, while the first-time revisions to data for the most recent years are generally much larger. For the major aggregate labor force series, however, the first-time revisions rarely alter the essential trends observed in the initial estimates.

Adjustment Methods and Procedures

Prior to 2002, the method of seasonal adjustment for weekly data-was developed by staff of the Federal Reserve Board (FRB)[1]. This method only allowed for fixed seasonal factors, but when seasonality is changing over time this method failed to remove all of the seasonality in the series, which users found confusing. BLS evaluated an alternate method[2] developed by the FRB and found it to improve the weekly seasonal adjustment over such periods. BLS and ETA introduced this alternate method to develop seasonal factors on April 11, 2002, effective with the release of claims data for the week ending April 6, 2002.

The first FRB method of seasonal adjustment assumed that the claims series had fixed seasonality, suggesting the claims data reflected a holiday or regular seasonal event the same way each year and the seasonal factors change only from the effects of the calendar. The second method assumes that the claims series exhibit variation in response to a seasonal event (moving seasonality). The second method allows the coefficients that determine the factors to change over time, in addition to reflecting the change based on calendar effects. (As part of testing the second method, it was confirmed that the two claims series do in fact exhibit moving seasonality.)

Starting with the release of new projected factors in 2024, BLS moved to a structural time series (STS) model. The model is set up in state-space form using the Kalman filter/smoother for estimation. There are many advantages to the new model. We use the Proc SSM[3] package in SAS/ETS® software to simplify the programming. SSM is capable of modeling monthly, quarterly, and high-frequency data (weekly, daily, hourly, et al.), and automatically outliers as described in De Jong and Penzer.[4] Similar models have been used by BLS in the Local Area Unemployment Statistics program since 1988, as they are also capable of accounting for sampling error in the Current Population Survey employment and unemployment series for states.[5] Structural time series models can also easily handle calendar effects as well as various types of outliers.

An obvious question might be as to why we are not using the standard X-13ARIMA-SEATS program for weekly seasonal adjustment. X-13 expects the data to have constant periodicity, but weekly series will have 52 or 53 weeks in a year. Also, the position of the dates of the series changes from year to year and the seasonal patterns rarely recur due to leap years. For more detailed information about STS models and weekly data, see Harvey, et al.[6]

For the two national UI series that are seasonally adjusted by BLS, the main steps of the seasonal adjustment process proceed in the following order:

1. Time series modeling
2. Model decomposition
3. Evaluation
4. Variances

1. Time series modeling

Time series models play an important role in seasonal adjustment. They are used to identify and adjust the series for atypical observations and other external effects.

Structural time series models

A classical additive decomposition model for an STS model in our case is:

$$Y_t = T_t + S_t + I_t + O_t + H_t$$

where Y_t is the observed series (at time t), T is a trend component, S is a seasonal component, I is an irregular component, O is an outlier component, and H is a holiday component. The time series components are assumed to be stochastic, and the trend, seasonal, and irregular components each have disturbance terms which are uncorrelated from each other. Variances for the disturbance terms are also known as hyperparameters. A positive variance for a component means that it is stochastic, while a zero variance shows deterministic behavior.

Trend-Cycle

The trend-cycle component consists of a local linear trend with a level and slope to form an integrated random walk. Trends represent the long-term evolution of a series and are modeled here as:

$$T_t = T_{t-1} + \beta_{t-1} + \omega_{T_t}, \quad \omega_{T_t} = NID(0, \sigma^2_{T_t})$$

$$\beta_t = \beta_{t-1} + \omega_{\beta_t}, \quad \omega_{\beta_t} = NID(0, \sigma^2_{\beta_t})$$

where ω_{T_t} and ω_{β_t} are disturbance terms.

This simple trend-cycle model can accommodate patterns ranging from an irregular cyclical series to a linear trend with a fixed rate of growth. Shifts up or down in the level give the trend-cycle a jagged appearance while changes in slope are inherently more gradual, causing acceleration, deceleration or change in direction. Overall smoothness, therefore, depends on the magnitude of the level variance relative to the slope variance. A small level variance relative to the slope variance implies a smooth trend (i.e. few turning points). In contrast, if the slope variance is small relative to the level variance, the trend will frequently change direction. In general, the trend-cycle is a combination of a long-run trend and more variable cyclical fluctuations.

Seasonal

The stochastic seasonal component is specified in terms of s seasons in a year:

$$S_t = \sum_{j=1}^{\lfloor s/2 \rfloor} S_{j,t}$$

where s=365 or 366, j=1,…,

$\lfloor s-2 \rfloor$ and t=1,…,T. When s is even $\lfloor s/2 \rfloor = s/2$, but when s is odd, $\lfloor s/2 \rfloor = (s-1)/2$.

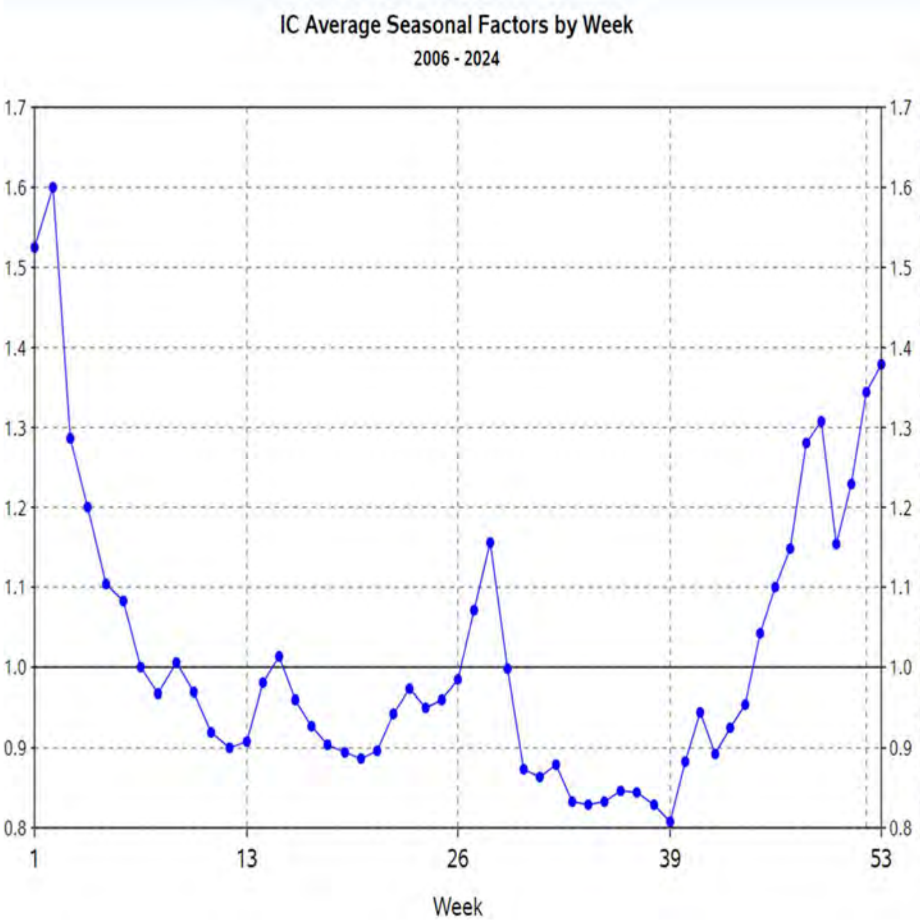
Each $S_{j,t}$ is found by:

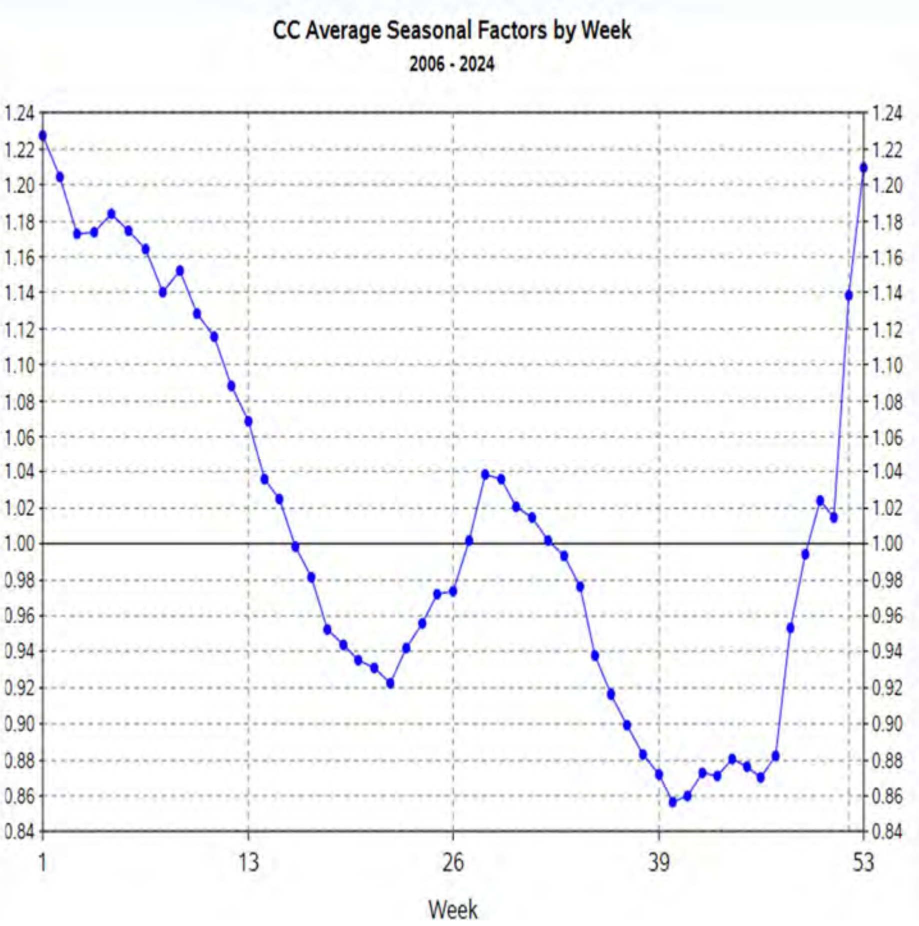
$$\begin{bmatrix} S_{j,t} \\ S^*_{j,t} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} S_{j,t-1} \\ S^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} \eta_{j,t} \\ \eta^*_{j,t} \end{bmatrix}$$

$$\eta_{j,t} \sim NID(0,\sigma_S^2),\ \ \eta^*_{j,t} \sim NID(0,\sigma_S^2)$$

where $\lambda_j = 2\pi j/s$ is the frequency in radians, and $\eta_{k,t}$, $\eta^*_{k,t}$ are disturbance terms that allow the seasonal effects to evolve stochastically over time. If seasonality is stochastic, only the expected value of the sum of the effects will equal zero over a year. Also, $S^*_{j,t}$ will be redundant $j=s/2$. More details can be found in Harvey[7]. All the possible sine and cosine pairs are not required to reasonably model the seasonal component for weekly data, so we use a combination of statistical tests and graphs to help determine the optimal number.

The below plots show the average multiplicative seasonal factor by week for Initial Claims (IC) and Continued Claims (CC).





Irregular

The irregular component is assumed to be white noise and includes erratic fluctuations not captured by the other components. It consists of a single white noise disturbance:

$$I_t = \upsilon_t, \quad \upsilon_t \sim NID(0, \sigma_t^2)$$

where υ_t is a disturbance term that has a zero mean and constant variance.

Holiday

The holiday component is made up of effects caused by special events across the year. Holidays usually fall yearly and are defined as follows for each year unless a special weighting is required. The holiday component is a linear combination of all the holiday effects in the model:

$$H_t = \sum_k \tau_{H,k} H_{k,t}$$

A specific holiday can be:

$$H_{k,t} = \begin{cases} 1 \; if \; t = k \\ 0 \; if \; t \neq k \end{cases}$$

where $t=1,2,\ldots,T$, k represents a week with a holiday from a list determined by BLS, and $\tau_{H,k}$ is a coefficient for the change in level of the series at time k. Note that k can vary for a specific holiday from year to year. For most holidays, the variable will be all zeroes during a year except for one week with a 1. There are sometimes special holidays or events that do not occur yearly. In these cases, the formula is the same as above but only occurs in certain years. These special holidays tend to be either 5, 6, or 11 years apart.

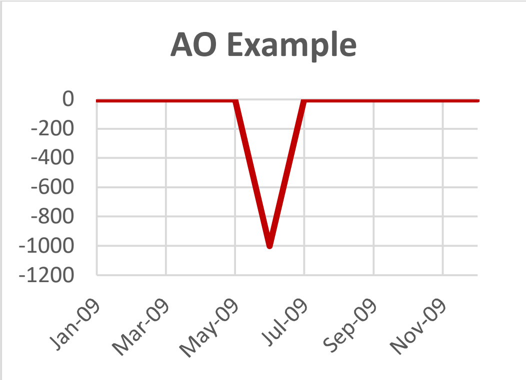
These effects are temporary breaks in a series that result from events such as moving holidays or the differing composition of weekdays in a month between years. These effects have different influences on the same week across years, thereby distorting the normal seasonal patterns for the given week.

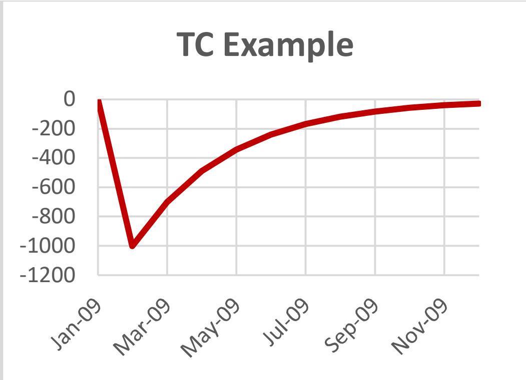
Some special holidays can be required at different times across the UI series. Examples can be a late Thanskgiving, Christmas on Monday in week 51 or 52, July 4th on Wednesday, or Christmas on Friday. Note that most holidays in monthly or quarterly data do not move from a certain month or quarter, but holidays vary in the week where they fall. For example, Labor Day falls in either week 36 or 37 and Easter moves between weeks 12 and 18. Note that BLS now removes these moving holiday effects along with the seasonal effects for the final seasonally adjusted series. See Cleveland, et al.[8] for more information on holiday modeling with UI series.

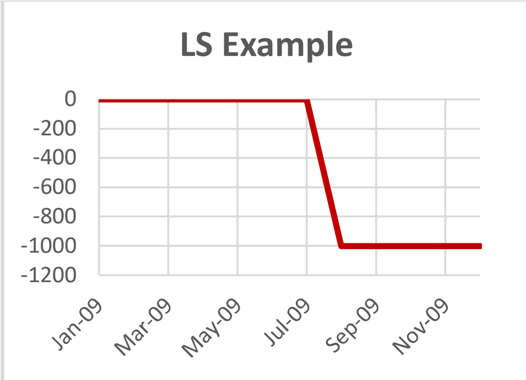
Outlier

A common form of outlier that presents a special problem for seasonal adjustment is an abrupt shift in the level of the data that may be either transitory or permanent. Three types of outliers are usually distinguished:

- An additive change that affects only a single observation (AO)
- A temporary change (TC) that has an effect that diminishes to zero over many weeks depending on the decay factor
- A level shift (LS), or a break in the trend of the data, which represents a permanent increase or decrease in the underlying level of the series







These three main types of outliers, as well as other types of external effects, may be handled by the time series modeling component of the SSM program. This is done by adding to the STS model appropriately defined regression variables based on intervention analysis originally proposed by George E.P. Box and George C. Tiao.[9] Outliers are detected through knowledge of the series, examination of plots, automatic outlier detection in SSM, or by checking the standardized prediction errors.

The outlier component is a linear combination of all the outlier effects in the model as below:

$$O_t = \sum_k \tau_{O,k} \zeta_{k,t}$$

where $\zeta_{k,t}$ is an indicator variable identifying when the outlier effect first occurred and its duration, and τ_k is a coefficient for the change in the level of the series at time k.

The outlier types are modeled as follows:

$$\text{AO: } \zeta_{k,t} = \begin{cases} 1 \ if \ t = k \\ 0 \ if \ t \neq k \end{cases}$$

$$TC\text{: } \zeta_{k,t} = \begin{cases} 0 \ if \ t < k \\ \alpha^{t-k} \ if \ t \geq k \end{cases}$$

$$LS\text{: } \zeta_{k,t} = \begin{cases} 1 \ if \ t \geq k \\ 0 \ if \ t < k \end{cases}$$

where k is the time point where the outlier effect first occurred, and α is the rate of decay back to the previous level $(0 < \alpha < 1)$.

While STS models can represent a wide class of evolving time series patterns, they do not account for the presence of occasional outliers and other special external effects. An outlier represents a sudden break in the normal evolutionary behavior of a time series. Ignoring the existence of outliers may lead to serious distortions in the seasonally adjusted series.

Outliers can occur in UI series for various reasons. Before UI claims could be filed on the internet, a government holiday could affect the claims series as it represents the loss of a day for filing claims. While claimants can now file through the internet for all states, fewer claims might still be filed in a particular holiday week as family events and vacations may delay filings. Significant weather events such as hurricanes and winter events may lead to an unusual increase in claims. The effects of the 9/11 attacks also affected the series for several weeks in 2001. Strong effects from the pandemic over 2020-2021 seriously impacted the UI series. For example, initial claims rose quickly from a typical level of 200,000-300,000 to over 6 million in April 2020.

__Model adequacy and length of series__

The preference is to use relatively long series in fitting time series models, but with some qualifications. Sometimes, the relevance of data in the distant past to seasonal adjustment is questionable, and a shorter series might be preferred.

Even though the filters have limited memory, there are reasons for using longer series. First, for homogenous time series, the more data used to identify and estimate a model, the more likely it is that the model will represent the structure of the data well and the more accurate the parameter estimates will be. The exact amount of data needed for time series modeling depends on the properties of the series involved. Arbitrarily truncating the series, however, may lead to more frequent changes in model identification and to large changes in estimated parameters, which in turn may lead to larger-than-necessary revisions in forecasts.

Second, although level shifts and other types of outliers tend to occur more often in longer series, SSM has the capability of automatically controlling these effects. Finally, attempting to fit longer series often provides useful insights into the properties of the series, including their overall quality and the effects of major changes in survey design.

Intervention analysis is used extensively to estimate the magnitude of known breaks in UI series and of automatic outlier detection to identify and correct for the presence of additional atypical observations. Once a model is estimated, it is evaluated in terms of its adequacy for seasonal adjustment purposes. The criteria essentially require a model to fit the series well (there should be no systematic patterns in the prediction errors). When there is a tradeoff between the length of the series and the adequacy of the model, a shorter series is selected. In this case, the identification of the model is not changed with the addition of new data unless the model fails diagnostic testing.

Kalman filtering/smoothing

SSM makes a forward and backward pass through the data. The forward pass uses a recursive algorithm and is known as the Kalman filter (KF). This updates our knowledge of the system every time a new observation is added.

The actual model error is the difference between the true values of the signal (the true population values) and the model's estimate of that value. Since the predicted values reflect all the available information up to time *t-1*, the corrected estimates for time *t* reflect all the available information from both the historical and current values of the series. As we don't know the true values, we cannot compute the actual model error. The one-step-ahead prediction errors are the difference between the UI observed values and their predictions at each step from the model.

Because the prediction errors represent movements not explained by the model, they should not contain any systematic information about the behavior of the signal or the noise component of the UI data. The prediction errors, when standardized, should approximate a randomly distributed normal variate with zero mean and constant variance. The model diagnostics test the standardized prediction errors for departure from these properties.

While the KF is a natural way to forecast the seasonal component to get projected factors or to make concurrent seasonally adjusted estimates, it is not well suited for producing historical estimates for a fixed set of data observations since it is designed to produce a current period estimate only and not to revise any earlier estimates. To achieve this, a method known as "smoothing" is used. This process revises each of the KF estimates for a period running from *t=1* to the last available observation at *t=N*. These "retrospective" estimates are obtained for the "Kalman smoother," which makes a backward pass through the data from *t=N* to *t=1*. Smoothing is batch processing in the sense that it operates on all the data at once in contrast to the KF, which processes one observation at a time. SSM produces useful diagnostic measures in the smoothing phase for identifying outliers and structural breaks such as AO effects in the measurement equation and in the structural disturbances based on cross-validation errors.

Not surprisingly, the estimates from the smoother typically look "smoother" than those from the filter. This is because the variances of the smoothed estimates are never larger than the variances for the filtered estimates and are usually much smaller toward the center of the series. But it is important to note that since these smoothed estimates use data from the entire series, they do not correspond to estimates that are available to data users in real time. Note that smoothing is only performed once a year during the annual review. See Harvey[10] for more details about Kalman filtering/smoothing.

2. Model decomposition

The STS method of seasonal adjustment assumes that the original series is composed of three main components: trend-cycle, seasonal, and irregular. Depending on the relationship between the original series and each of the components, the mode of seasonal adjustment may be additive or multiplicative. Formal tests are conducted to determine the appropriate mode of adjustment.

The multiplicative mode assumes that the absolute magnitudes of the components of the series are dependent on each other, which implies that the size of the seasonal component increases and decreases with the level of the series. With this mode, the monthly seasonal factors are ratios, with all positive values centered around unity (1.0). The seasonally adjusted series values are computed by dividing each month's original value by the corresponding seasonal factor.

In contrast, the additive mode assumes that the absolute magnitudes of the components of the series are independent of each other, which implies that the size of the seasonal component is independent of the level of the series. In this case, the seasonal factors represent positive or negative deviations from the original series and are centered around zero. The seasonally adjusted series values are computed by subtracting the corresponding seasonal factor from each month's original value.

A multiplicative seasonal effect is assumed to be proportional to the level of the series. A sudden, large change in the level of the series will be accompanied by a proportionally large seasonal effect. In contrast, an additive seasonal effect is assumed to be unaffected by the level of the series. In times of relative economic stability, the multiplicative option is generally preferred over the additive option. However, in the presence of a large level shift in a time series, multiplicative seasonal adjustment factors can result in systematic over- or under-adjustment of the series; in such cases, additive seasonal adjustment factors are preferred since they tend to track seasonal fluctuations in the series more accurately and have smaller revisions.

Prior to 2020, the UI series were modeled multiplicatively, and diagnostic tests verified this choice. However, beginning in March 2020, it was clear that the multiplicative factors could cause distortion to the size of the seasonal components. Unfortunately, due to the practice of using projected seasonal factors, changes could not be made immediately. However, in the summer of 2020, BLS made changes to create additive projected factors for official release. Later, we switched to a hybrid adjustment in which additive factors were only used during the worst part of the pandemic period. Starting in 2023, BLS switched back to multiplicative projected factors except for the period from March 2020 through July 2021. No revisions will be made to those seasonal factors even when we revise factors back 5 years in our annual review. In accordance with our usual practice, seasonal adjustment models and factors are always reviewed at the beginning of the calendar year, and necessary changes are made to the seasonal adjustment settings as required. Although there are possibly very large level shift effects early in the pandemic, they are not permanent shifts and had to be replaced by a combination of AOs and TCs. Otherwise, pandemic LSs can cause distortions in the seasonally adjusted data in future years.

3. Evaluation

A series should be seasonally adjusted if three conditions are satisfied: the series is seasonal, the seasonal effects can be estimated reliably, and no residual seasonality is left in the adjusted series. A variety of diagnostic tools is available in SSM, and others were added to our program to test for these conditions, including frequency-spectrum estimates, heteroscedasticity tests, goodness-of-fit tests, numerous diagnostic plots, etc. If diagnostic testing shows that any of the three conditions listed fails to hold for a given series, a series is deemed not suitable for seasonal adjustment. The possibility of changing outliers or holidays in the model is also tested as needed.

4. Variances

Week-to-week changes for UI seasonally adjusted series can be both relatively large and variable, so the changes can be difficult to interpret. Using a parametric bootstrap to estimate variances for initial and continued claims week-to-week changes, Evans and Sverchkov[11] showed that most changes are not statistically significant. Recently, Evans, et al.,[12] used a different method to estimate variances that utilizes linear filter weights based on the seasonally adjustment model for CPS labor force series. This method can be applied to the STS models introduced for UI weekly series and is a current research project at BLS.

[1] Cleveland, W.P. (1986), "Calendar Adjustment and Time Series," Special Studies Paper No. 198, Division of Research and Statistics, Federal Reserve Board, Washington, DC.

[2] Cleveland, W.P., Scott, S., and Evans, T. (2018), "Weekly Seasonal Adjustment: A Locally-weighted Regression Approach," in the *Handbook on Seasonal Adjustment 2018 edition*, Eurostat, Luxembourg: Publications Office of the European Union.

[3] SAS Institute Inc. (2020), SAS/ETS® 15.2 User's Guide, Cary, NC: SAS Institute Inc. Available online at https://go.documentation.sas.com/api/docsets/etsug/15.2/content/ssm.pdf?locale=en#nameddest=etsug_ssm_toc.

[4] De Jong, P., and Penzer, J. (1998), "Diagnosing Shocks in Time Series," *Journal of the American Statistical Association*, v. 93, pp. 796-806.

[5] Tiller, R. (2006), "Model-Based Labor Force Estimates for Sub-National Areas with Large Survey Errors." Available online at https://www.bls.gov/osmr/research-papers/2006/st060010.htm.

[6] Harvey, A., Koopman, S.J., and Riani, M. (1997), "The Modeling and Seasonal Adjustment of Weekly Observations," *Journal of Business and Economic Statistics*, v. 15, No. 3, pp. 354-368.

[7] *Ibid.*

[8] Cleveland, W.P., Scott, S., and Evans, T. (2018), *op.cit*.

[9] George E.P. Box and George C. Tiao (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, v. 70, n. 353, pp. 71-79.

[10] Harvey, A. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, U.K.: Cambridge University Press.

[11] Evans, T.D., and Sverchkov, M. (2016), "Variance Estimation for Weekly Seasonally Adjusted National UI Claims Series," in JSM Proceedings, Business and Economics Statistics Section. Available online at https://www.bls.gov/osmr/research-papers/2016/st160330.htm.

[12] Evans, T.D., Sverchkov, M., Rottach, R.A., and Doherty, C.J. (2023), "Evaluation of Variance Methods for Seasonally Adjusted Series," in JSM Proceedings, Business and Economics Statistics Section. Available online at https://www.bls.gov/osmr/research-papers/2023/st230060.htm.