Outlier Review during Concurrent Seasonal Adjustment of CES State and Area Series November 2017

Jonathan Creem[1]

U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract

The Current Employment Statistics (CES) State and Area program publishes seasonally adjusted data for over 2000 series each month, covering over 400 subnational geographic areas. The seasonal factors used to make these adjustments have traditionally been developed using historical employment data forecasted for one year. A particular challenge in concurrent adjustment of these series is the outsized impact of various events on localized areas. These events range from natural disasters and strikes to unusual weather, and are often difficult to properly identify and model at the end of a time series. We describe the monitoring, screening, and review process used to model these events during concurrent adjustment.

The empirical research tests how a stricter critical value policy would affect simulated time series data after an exogenous event is added to the concurrent adjustment process. This exercise provides insight into how the procedure would fare under realistic circumstances and how to handle potential outliers in the data. The results provide evidence that a stricter critical value policy may be beneficial.

Keywords: Concurrent Seasonal Adjustment, outlier identification, RegARIMA modeling

[1] Any opinions express in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistic

1. Introduction

The Current Employment Statistics State and Area (CESSA) program publishes seasonally adjusted data for over 2000 series each month, covering over 400 subnational geographic areas. The seasonal factors used to make these adjustments have traditionally been developed using historical employment data forecasted for one year. As of March 2018, the way in which CESSA develops its seasonal adjustment factors will transition from this method based solely on historical employment to a concurrent seasonal adjustment.

Concurrent seasonal adjustment is the process of recalculating the seasonal factors each month using the most recent data that are available. RegARIMA models are typically set during an annual review. Parameters are re-estimated during concurrent processing, but rarely are the models changed (i.e. by adding an outlier). A particular challenge in concurrent adjustment of these series is the outsized impact of various events on localized areas. These events range from natural disasters and strikes to unusual weather, and are often difficult to properly identify and model at the end of a time series.

This research investigates how to improve the detection of these outsized impacts by adjusting the critical value used to evaluate potential outliers during the automatic outlier detection process in the X-13ARIMA-SEATS.[2] Simulated data reflecting four different ARIMA models are randomly generated and an exogenous event is replicated and applied. The simulated data is then passed through concurrent seasonal adjustment and is compared to a baseline series. A root mean squared error statistic of the seasonal adjustment is then calculated to quantify the amount of error involved in the concurrent process with the event at the end of the time series. The outlier detection critical value is then adjusted and the simulation is repeated.

Efficient outlier detection is also essential to the CESSA program. Reviewing seasonal factors is labor intensive with analysts needing to review over 2000 series. Better insight on setting a critical value policy will minimize the error of the adjusted series and will assist in the identification of events that should be set as an outlier in the model.

Section 2 gives background on seasonal adjustment in CESSA and details the exogenous events of interest. Section 3 covers the methodology used in the research. Section 4 details the results gathered from the experiment and explains how the insights can be leveraged going forward. Finally, Section 5 offers some concluding remarks.

2. Background

CES State and Area currently relies on a forecast method to derive the seasonal factors for each month of the estimation year. Ten years of employment data are used to calculate these factors and five years of factors are replaced back in history. CESSA is also unique because of the "Two Step" seasonal adjustment method that must be employed.[3] The Two Step method incorporates two sets of historical data, one based on data from the Quarterly Census of Employment and Wages (QCEW) and the other on data from the historical CES estimates. The pair of historic data is necessary due to the potential differences in seasonality as demonstrated by Berger and Phillips (1993). The forecasted seasonal factors

[2] See Findely et al. (1998) for details regarding the X13ARIMA-SEATS methodology
[3] See Scott et al. (1994) for an analysis of seasonal adjustment of a hybrid series.

that are applied to the estimate data for the upcoming year cannot be determined by the seasonality of the benchmark data.

Once the calculation of the seasonal factors transitions from an annual process to a monthly process the CESSA program will adjust its review procedure to accommodate. Model setting is initially done during the annual processing timeframe. Annual review determines the order of the ARIMA model along with the moving average filters. These specifications remain unchanged when the series is processed through a concurrent seasonal adjustment.

Programs that use concurrent seasonal adjustment at the Bureau of Labor Statistics (BLS) have generally been hesitant to add outliers to models mid-year except under extreme circumstances. It is difficult to determine whether an extreme value at the end point of a time series is a true outlier or whether it reflects, for example, changing seasonality without prior knowledge of the source. If a true outlier occurs, it is also difficult to determine what type of intervention should be added to the model since the future values of the time series are at that point unknown.

An additional aspect that must be considered is how to properly identify a data point that is an outlier during concurrent seasonal adjustment. A sudden labor strike in a large company or impactful weather related events may affect the level of an employment series for a short period. Not accounting for these and other outliers may have an adverse influence the seasonal adjustment factors. More specifically, the outliers will have an effect on the moving averages that are used to isolate the irregular component from the de-trended SI ratios. Allowing the extreme value of a labor strike to persist will distort the series and potentially lead to a moving average that is unrepresentative of the seasonal effect. In the case of the level shift, when the moving averages are applied to the SI ratios, the estimates of the seasonal component before the shift will be underestimated and those after the shift will be overestimated. The inaccurate information from these two scenarios may then lead to the production of unreliable factors to be applied to future estimates.

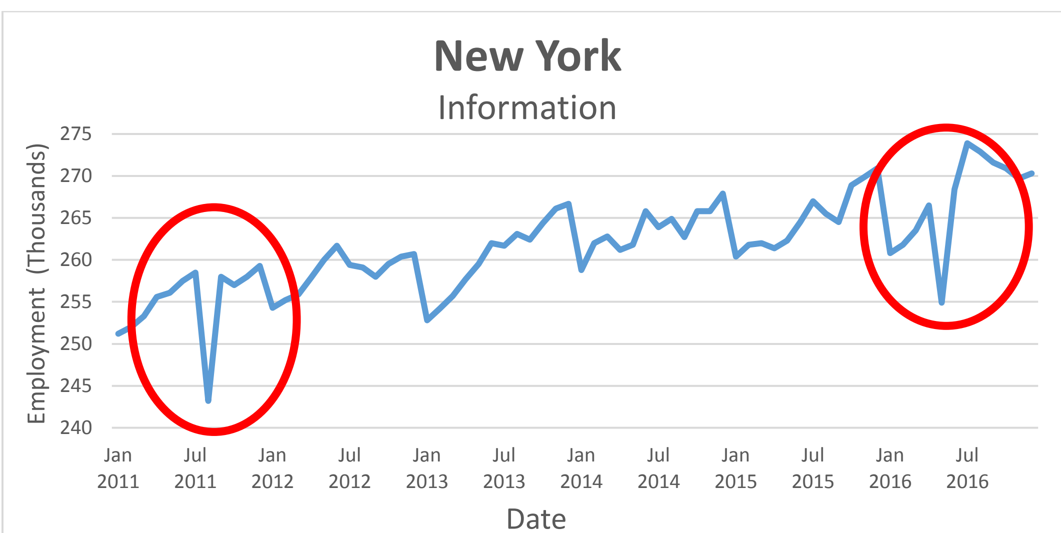


Figure 1:

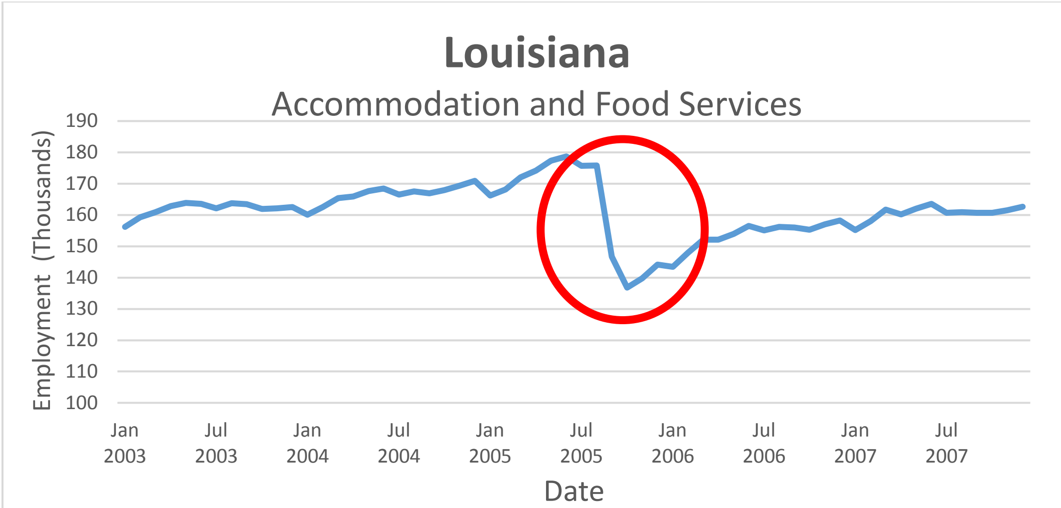


Figure 2:

Figure 1 and Figure 2 above depict two actual events that have affected CESSA series in the past. In Figure 1, a labor strike disrupted the routine fluctuations of the Information series in New York. On two occasions, in 2011 and 2016, workers for a large telecommunication company in the state went on strike for one month. Figure 2 depicts the Accommodation and Food Service series in Louisiana that was heavily affected by Hurricane Katrina in 2005. Like what occurred in Louisiana, a major hurricane can cause the level of an employment series to shift for an extended period of time.

2.1 Concurrent Seasonal Adjustment Review

Review of concurrent seasonal adjustment will be incorporated in the monthly estimation production cycle done by CESSA. Each month CESSA produces two estimates of the industries employment level. The first estimate is a preliminary estimate that is based off of the sample microdata collected at that time. A review is done by both national office analysts, as well as by analysts at the regional at state level. Preliminary estimates are then locked and released to the public. The final estimate cycle begins a few weeks later after additional microdata is incorporated into the sample.

Seasonally adjusted data have always been reviewed during both estimation cycles with the projected factors developed during annual processing being applied to the newly estimated employment level. Concurrent seasonal adjustment, however, will only run during the preliminary estimation cycle. At this time, the current month’s factors will be developed along with the previous month’s final estimation factors.

Although the quantity of concurrent factors an analyst must review is fewer than what is necessary during the annual production review of the seasonal adjustment models and forecasted factors, various monthly production deadlines require efficient review of the data. It is often the case that adjustments to the not seasonally adjusted data come late in the production cycle. In the concurrent adjustment environment where the seasonal factors are based off of the most recent data, late adjustments such as these can have an effect on the process.

Given the short timeframe an analyst has to review the concurrent factors, establishing a set of tools that can be used to effectively consider whether an outlier should be declared is desirable. The literature on outlier detection has generally considered knowing the source of an outlier as one of the best strategies in selection.[4] Analysts in CESSA are trained to accumulate local knowledge of the area's employment dynamics to use in monthly estimation. Knowledge of a local strike or a substantial weather related event, as described above, could be attained very early in the process. The analyst, in this case, would have an ample amount of time to decide whether the event should be treated as an outlier.

In a case where the analyst does not have adequate knowledge of the source of an outlier or is not sufficiently accustomed to the trend and seasonal patterns of the series, a strategy such as automatic outlier detection may be helpful. McDonald-Johnson and Hood (2001) looked closely at various methods of outlier selection in the seasonal adjustment process. They tested a combination of automatic outlier detection with a default critical value versus various visual identification methods and found that the automatic detection was the superior method of selection. It is not clear at this point, however, how sensitive the CESSA program should set the critical value on the automatic outlier detection to adequately model outliers that would negatively affect the concurrent factors without causing too many outliers to be identified. A real risk to consider would be whether the outlier process is mistakenly identifying true seasonal changes in the data as atypical events.

Testing in concurrent seasonal adjustment in the CESSA program has not utilized the automatic outlier detection capabilities thus far. Instead, the determination of an outlier when the concurrent factors are reviewed are based primarily on the analyst's knowledge of the potential event in question.

3. Empirical Simulations

A series of simulations were conducted to gauge the extent in which an exogenous event would be identified in the concurrent adjustment process and if enhancing the detection capabilities would improve the seasonally adjusted data. To begin, 1000 time series were randomly generated using the Forecast package in R for four different ARIMA models. The models used, along with their autoregressive and moving average coefficient parameters are listed in Table 1.

Table 1.

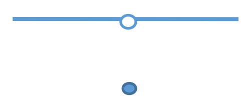
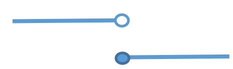
ARIMA Models and Parameters						
ARIMA Model	**AR**	**MA**	**AR12**	**MA12**	**d**	**d12**
(0,1,1)(0,1,1)	**-**	**0.23**	**-**	**0.68**	**1**	**1**
(0,2,1)(0,1,1)	**-**	**0.72**	**-**	**0.71**	**2**	**1**
(1,1,0)(0,1,1)	**0.43**	**-**	**-**	**0.71**	**1**	**1**
(0,1,0)(1,1,0)	**-**	**-**	**-0.54**	**-**	**1**	**1**

[4] See McDonald-Johnson and Hood (2001) for an overview on selecting outliers in the seasonal adjustment process

The columns titled “AR” and “MA” reflect the non-seasonal autoregressive and moving average coefficients. Columns “AR12” and “MA12” reflect the seasonal autoregressive and moving average coefficients, and columns “d” and “d12” reflect the non-seasonal and seasonal difference indicators in the model. The coefficients used reflect common ARIMA parameters derived during the CESSA model setting procedure.

Next, a set of baseline seasonally adjusted data was generated to compare to the concurrent adjustment procedure. The simulated time series data was seasonally adjusted from January 1997 to December 2016. This 20 year timeframe was meant to maximize the number of observations before and after the date (June 2007) in which the impact was applied. An exogenous event meant to simulate the type of variability that an employment series can encounter was then added to each newly seasonally adjusted series. These events include an Additive Outlier (AO), and a Level Shift (LS). Table 2 describes the type of impact applied.

Table 2.

Exogenous Events				
Event	**Description**	**Scenario**	**Model**	**Graphic**
Additive Outlier (AO)	Series is impacted at a single point in time (t)	Labor Strikes	1 for $t = t_0$ 0 for $t \neq t_0$	
Level Shift (LS)	Series is impacted and continues at new level	Hurricane Katrina	-1 for $t < t_0$ 0 for $t \geq t_0$	

The magnitude of the event was controlled for by the equivalent t-value from the outlier detection process at the concurrent point. X-13 uses a t-test in the detection process. The t-value of a given effect size is dependent on the ARIMA model, residual variance, and series length. The effect sizes were numerically derived effect to scale against equivalent t-values used in the detection process.

Once the baseline seasonally adjusted series were developed, the next step was to simulate the concurrent process with the calculated impacts applied to the simulated time series. The concurrent process only ran for the 12-month span from January 2007 to December 2007. For each month in the timeframe, the concurrent process applied factors for the current month, as well as, update the concurrent factors for the previous month before moving on to the next period where the process repeated. This method of updating the seasonal factors for two months is consistent with established CESSA concurrent seasonal adjustment methodology. The exogenous event is realized at the end of the time series once June 2007 is initiated and it remains in the simulated time series for the remainder of the 12-month span.

Each simulation is then tested with the outlier detection critical value set at the t-value level used in calculating the impact level. To further understand this methodology, Figure 3 outlines the combinations of impact levels with outlier detection critical values.

Event / Model		**Impact (In terms of equivalent t-value)**					
		0.00	3.85	5.00	8.00	10.00	99.00
Critical Value	3.85		X				
	5.00			X			
	8.00				X		
	10.00					X	
	No Detection						X

Figure 3:

The columns in Figure 3 depict the t-values used to calculate the impacts applied to the simulated series for each combination of the event and ARIMA model. The rows depict the critical values that were set in order to test each impact level. The impacts range from a t-value of zero, which represents a time series not affected by an event, to 99.00, which represents a time series affected by an extremely large event. The initial set of simulations, represented by the bottom row of the figure, were tested without a critical value set. The diagonal line denoted by "X" indicates the simulated time series tested at a critical value equivalent to the t-value used to impact the series. The impact level of 99.00 is implemented as a control simulation. Large improvements are expected to occur when simulating this impact event even at the relatively large critical value levels of 10.00 and 8.00. Outlier detection is limited to the same 12-month span that was tested the concurrent process.

3.1 Measure of Gain

The baseline seasonally adjusted data described above is established as the most accurate seasonal adjustment of the simulated series. In other research on concurrent seasonal adjustment a benchmark or "final" value is leveraged to contrast the data stemming from the concurrent process. For example, Pierce and McKenzie (1987) find that gain from concurrent seasonal adjustment is realized when initially published estimates of the Federal Reserve Board's Industrial Production Index is revised once a year. Kropf et al. (2002) studied the replacement of projected factors to concurrent factors in the CES National employment data. They found that the concurrent factors had less error when a revision to final seasonal values was calculated.

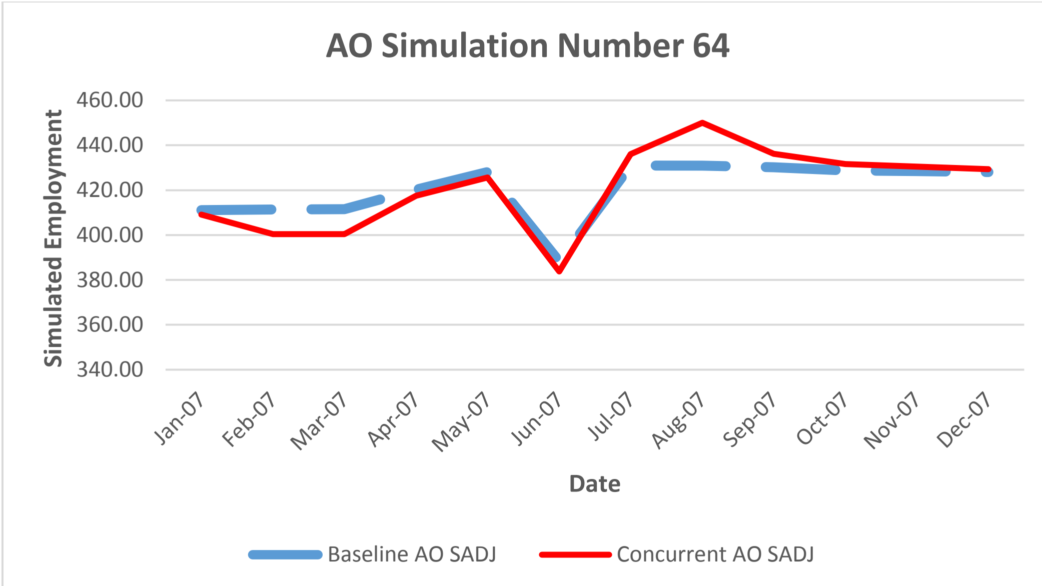


Figure 4:

Figure 4 above provides an example of a simulated AO series. The dotted line depicts the baseline series with the event applied after data is seasonally adjusted. The solid line depicts the concurrent adjustment of the series. The space between the two sets is the error adjusting the critical value policy aims to minimize.

A root mean square error statistic that measures the difference between the two sets of data for each combination of events tested is derived. The measure to calculate the gains from having a stricter critical value policy is a ratio of these root mean square errors taking the form:

$$RMSE\ ratio_{e,m}\ (r) = \frac{RMSE(r_{e,m}^{sv})}{RMSE(r_{e,m}^{nv})}$$

The numerator in this formula represents a statistic *(r)* calculated from the concurrent adjustment having a stricter critical value *sv* of event *e* in model *m*. The denominator reflects the same event and model combination when no critical value policy *nv* is applied. Intuitively speaking, the ratio depicts how much could be improved on the error between the baseline data and the concurrent data if a progressively lower critical value is instituted. If no gain is to be made the ratio would be close to a value of 1.

4. Results

The measure of gain is presented for each combination of event and model in Figure 5. For all combinations of model and event most ratios indicate that there are gains to be made by instituting a stricter critical value policy. In particular, the data shows that as the impact gets larger, the gains from the stricter value policy get more and more significant. For example, consider model number 1 in which ARIMA model (0,1,1)(0,1,1) was impacted with an AO event. The bottom row, reflecting simulations with no outlier detection, have a ratio of 1. When the impact levels are small, gains only begin to be realized when the critical value is equally as small. For an impact of 3.85, gains made by having a critical value of 5.00 improves the root mean square error by 3.5%. By having the critical value set at 3.85 the root mean square error is improved by 5.7%.

AO	(0,1,1)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 1	3.85	1.001	0.943	0.902	0.868	0.857	0.297
	5.00	1.000	0.965	0.922	0.868	0.856	0.297
	8.00	1.000	1.000	0.994	0.915	0.866	0.297
	10.00	1.000	1.000	1.000	0.979	0.926	0.297
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

LS	(0,1,1)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 2	3.85	1.001	0.981	0.920	0.826	0.781	0.196
	5.00	1.000	0.997	0.965	0.827	0.781	0.196
	8.00	1.000	1.000	1.000	0.928	0.805	0.196
	10.00	1.000	1.000	1.000	0.990	0.910	0.196
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

AO	(0,2,1)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 3	3.85	1.000	0.984	0.974	0.946	0.928	0.464
	5.00	1.000	0.986	0.977	0.948	0.928	0.464
	8.00	1.000	0.992	0.982	0.966	0.943	0.464
	10.00	1.000	0.999	0.992	0.974	0.959	0.464
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

LS	(0,2,1)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 4	3.85	1.000	0.997	0.990	0.963	0.946	0.424
	5.00	1.000	0.999	0.993	0.965	0.946	0.424
	8.00	1.000	1.000	0.998	0.986	0.961	0.424
	10.00	1.000	1.000	1.000	0.997	0.985	0.424
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

AO	(1,1,0)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 5	3.85	1.000	0.951	0.911	0.851	0.835	0.328
	5.00	1.000	0.971	0.930	0.853	0.835	0.328
	8.00	1.000	1.000	0.992	0.911	0.862	0.328
	10.00	1.000	1.000	0.999	0.979	0.938	0.328
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

LS	(1,1,0)(0,1,1)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 6	3.85	1.000	0.995	0.955	0.852	0.806	0.212
	5.00	1.000	1.002	0.986	0.856	0.806	0.212
	8.00	1.000	1.000	1.000	0.964	0.864	0.212
	10.00	1.000	1.000	1.000	0.996	0.953	0.212
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

AO	(0,1,0)(1,1,0)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 7	3.85	1.001	0.934	0.861	0.762	0.720	0.490
	5.00	1.000	0.973	0.893	0.763	0.720	0.489
	8.00	1.000	1.001	0.994	0.812	0.732	0.489
	10.00	1.000	1.000	1.000	0.881	0.774	0.489
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

LS	(0,1,0)(1,1,0)	Impact (In terms of equivalent t-value)					
		0.00	3.85	5.00	8.00	10.00	99.00
Model Number 8	3.85	1.001	0.974	0.879	0.705	0.634	0.096
	5.00	1.000	1.000	0.956	0.710	0.633	0.096
	8.00	1.000	1.000	1.000	0.914	0.690	0.096
	10.00	1.000	1.000	1.000	0.993	0.891	0.096
	No Detection	1.000	1.000	1.000	1.000	1.000	1.000

Figure 5: The green shaded areas indicate the critical values level that result in the most gain.

Once the impact levels begin to become larger in magnitude, gains begin to appear at less strict critical values. For an impact of 8.00, an immediate improvement of 2.1% in the root mean square error is realized once we incorporate the outlier detection at a critical value level of 10.00. At the three stricter critical value levels, the measure of gain reflects an 8.5%, 13.2% and 13.2% improvement, respectfully. Finally, the most gains are made when the impact level is 10.00 and the critical value becomes very strict. An immediate gain of 7.4% is realized once the least strict critical value policy considered (10.00) is implemented. The root mean square error is progressively improved upon at the critical value level of 8.00 and 5.00 but then improves the data slightly less at the 3.85 critical value level. At the critical value of 5.00, the measure of gain reflects an improvement of 14.4% but an only an improvement of 14.3% at the 3.85 critical value level.

Finally, there is an obvious and immediate gain when a critical value in implemented to test an outlier with an impact level of 99.00. These impacts are meant to simulate an extremely large event so it is expected that these measures of gain to be equally as large even at the critical value level of 10.00. If the counterfactual was observed and the gain at this level was relatively insignificant further investigation would take place.

Model 8, in which ARIMA model (0,1,0)(1,1,0) was impacted with an LS event, shows the most gain from the simulations. When the level shift impact is relatively large at the 10.00 level, an immediate gain of 10.9% is realized once a critical value policy is implemented. Another large gain is measured at the 8.00 critical value level where an improvement of 31% is realized. The measure of gain is improved again slightly at the 5.00 critical value level but is then depressed once then critical value level of 3.85 is tested. The reduction in gain as the critical value gets stricter may indicate an outlier detection policy that is too aggressive and is over fitting the seasonally adjusted data.

Model 3 and Model 4 which depict an ARIMA model of (0,2,1)(0,1,1) saw the least amount of measured gains for their respective event. When the AO event was tested, a gain of 7.2% was realized when the impact level was at 10.00 and the critical value was set at 3.85. This was the scenario where the most gain was seen for this model. Similarly, when the LS event was tested, a gain of 5.4% was realized when the impact level was at 10.00 and the critical value was set at 3.85.

In some cases, the results show losses to the root mean squared error when a stricter critical value policy is applied compared to the simulation where no critical value is applied. For example, when testing an impact level of 3.85 in Model 6, there seems to be no improvement on the root mean squared error until the simulation is tested at the 3.85 critical value level. When the concurrent process was tested using a critical value of 5.00, however, our measure of gain shows a loss of 0.2%. It remains unclear why this would be the case. It is possible that rounding abnormalities are causing the ratio to become greater than 1.

In total, the results suggest that a stricter critical value policy may be beneficial, especially when the impacts are larger in magnitude. A less strict critical value policy may allow quantitatively insignificant events in the time series to influence the concurrent factors that ultimately allot for more error in the resulting seasonally adjusted data.

5. Conclusion

The transition to concurrent seasonal adjustment in the CESSA is set to be completed with the release of the 2018 benchmark. Mance (2015) evaluated how effective the transition would be for the program and discovered that the application of concurrent factors would reduce the over-the-month change of the seasonally adjusted series, as well as reduce the error when revised to seasonally adjusted benchmark data. Once concurrent adjustment goes live, analysts will be responsible for review of the concurrent factors each estimation cycle.

Large labor strikes and natural weather events have had a noticeable impact on regional employment levels in the past. Accounting for these outsize impacts is critical in order to produce accurate concurrent factors. Automatic outlier detection is not utilized for the time being. The identification and selection of outliers is currently based on having knowledge of the source of the event affecting the time series.

The results of this research suggest that setting a stricter critical value policy for outlier detection may be beneficial. There are slight distinctions in the amount of gain comparing across the ARIMA models tested. Having a strict critical value can improve the root mean square error by more than 30 percent in the best scenario. The simulations in general, however, find that improvement may be had across all model and event combinations.

Further research is needed if setting a precise critical value is desired. An application using actual data should be tested but the infrequency of outliers may potentially hinder a robust study of its effects under different situations. Testing of CESSA concurrent process began in 2015. As of September 2017, only one potential outlier has been identified.

References

Berger, F. D., & Phillips, K. R. (1993). Reassessing Texas employment growth. *Southwest Economy*, (Jul), 1-3.

Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B. C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, *16*(2), 127-152.

Kropf, J., Manning, C., Mueller, K., & Scott, S. (2002, August). Concurrent Seasonal Adjustment for Industry Employment Statistics. In *Joint Statistical Meetings of the American Statistical Association, the International Biometrie Society, the Institute of Mathematical Statistics, and the Statistical Society of Canada, New York*.

McDonald-Johnson, K., & Hood, C. C. (2001). Outlier Selection for RegARIMA Models. *Proceedings of the Business and Economic Statistics Section, American Statistical Association, Alexandria, VA.*

Mance, S. Concurrent Seasonal Adjustment of State and Metro Payroll Employment Series October 2015.

Pierce, D. A., & McKenzie, S. K. (1987). On concurrent seasonal adjustment. *Journal of the American Statistical Association*, *82*(399), 720-732.

Scott, S. (1994). Seasonal Adjustment of Hybrid Economic Time Series Stuart Scott, George Stamas, Thomas J. Sullivan, and Paul Chester, US Bureau of Labor Statistics Stuart Scott, 2 Massachusetts Ave., NE, Rm. 4915, Washington, DC 20212 Key Words: Employment statistics, revisions, X-11-ARIMA, smoothness. In *Proceedings of the Business and Economic Statistics Section* (p. 163). The Association.