Statistical Agency Use of Macro Editing in Industry-Area Employment Estimation

Lee Baker, Taylor Le, and Nicholas Rose[1]
U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract

The Current Employment Statistics State and Area (CESSA) program produces monthly industry employment estimates for subnational areas based on a survey of about 634,000 nonfarm worksites. Before estimates are published, they go through several screening procedures at the micro (individual report) and macro (estimation cell) level. CESSA adapted a process based on the Fay-Herriot model for extreme outlier detection at the macro level. The standardized difference between the sample-based estimate (Y1) and the synthetic part (Y2) of the model are used to identify significant deviations as candidates for macro editing. In those cases where the standardized difference exceeds a given threshold and analysts cannot find economic reasons to support the extreme movement, the modeled estimate is used to replace the direct sample-based estimate. This paper examines the process that CESSA uses to identify extreme macro outliers and its application to employment estimates at the state and area level. The effect of this procedure on error variance and bias when adjusting extreme estimates at different standardized cut-off levels is explored in an empirical study.

Key Words: Small area estimation, monthly employment data, outlier detection**.**

1. Introduction

Before publication, every statistical agency verifies whether its estimated figures seem plausible (De Waal 2009). The Current Employment Statistics State and Area (CESSA) program uses the most prevalent method, known as the aggregation method, whereby estimates are compared to prior years of the same or related time series. This method is effective for validating employment trends, which can be highly seasonal and area-specific. But this method can also lead to bias in the direction of historical levels and subsequently miss economic turning points. Gershunskaya (2012) proposed formalizing aggregation review in terms of measured deviations from historical trends. This paper builds off her work by empirically studying the effects of replacing a sample-based estimate with a model-based estimate in months where the sample-based estimate deviates greatly from the synthetic estimate.

Editing is a significant part of survey estimates for national statistical agencies. While the primary focus of this paper is on aggregation review of estimates, understanding the editing process of CESSA survey data that is done in advance of estimation will provide further context. Each month, the CESSA surveys about 147,000 businesses and government agencies nationwide, representing approximately 634,000 individual worksites in order to provide detailed industry data on nonfarm payroll employment, hours, and earnings for all states, the District of Columbia (D.C.), Puerto Rico, the Virgin Islands, and about 450

[1] Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the Bureau of Labor Statistics.

metropolitan areas[2] and divisions. The majority of CES data collection instruments (CATI, TDE, EDI, and Internet collection) use real-time screening to control the quality of the collected data. Following collection, the data are run through further screening parameters and reviewed at the item level by subject-matter experts. Finally, observations with a significant effect on estimates are reexamined by analysts as part of aggregation review. Only after the survey data are thoroughly reviewed and all less invasive interventions are considered would the analyst resort to replacing the sample-based estimate with a modeled estimate.

CESSA monthly news releases are among the timeliest (typically eight weeks after the reference month at the state level, and ten weeks after at the metro area level) economic indicators for subnational areas. Government officials, the Federal Reserve, and other market observers rely on CESSA estimates to monitor possible economic turning points. True economic turning points will tend to profile as outliers, and for the CESSA program, it is essential to our mission to capture and report these turning points. On the other hand, CESSA is a sample-based survey and, by extension, many of the directly estimated extreme values are simply false positives representing statistical error. In formalizing the approach to handling extreme values, the goal is to minimize the number of false positives in the publication while protecting against rejecting data reflecting true economic turning points.

The CESSA program is aided by the unemployment insurance (UI) system, which mandates a quarterly count of nearly all employment within the scope of the program. These data are tabulated by the Quarterly Census of Employment and Wages (QCEW) program quarterly and annually and made available between five and six months after the end of the reference quarter. The administrative UI data are used as a sampling frame for CESSA and, once a year in March, CESSA replaces its estimates with the adjusted[3] QCEW data up to the third quarter of the previous year through a process called benchmarking. The CESSA and QCEW universes are not exactly the same, but there is enough overlap that their trends are about the same. CESSA annual benchmark revisions are the program's standard for measuring survey error. This paper will examine the variance and bias effects of a proposed macro editing method using four years of benchmarked estimates.

A modified Fay-Herriot model is used to test and replace extreme movement in the sample-based estimates. While some subnational industries are estimated using the modified Fay-Herriot model, the majority of statewide and larger MSA levels are estimated using a sample-based estimator. To macro edit the sample-based estimates, the standardized difference (z-score) between the sample-based estimates ($\widehat{\boldsymbol{Y}}_{\mathbf{1}}$) and the synthetic part of the model ($\widehat{\boldsymbol{Y}}_{\mathbf{2}}$) is used to mark significant deviations. In practice, these extreme deviations are reviewed by analysts and only replaced if found to be economically unreasonable; however, the following research data replaces any and all sample-based estimates that exceed stated thresholds.

[2] Metropolitan areas include Metropolitan Statistical Areas (MSAs), New England City and Town Areas (NECTAs), divisions thereof, and a small number of nonstandard areas. Throughout this paper, all sub-state areas within the scope of the CESSA program are referred to as "MSAs."
[3] Participation in state unemployment insurance (UI) programs is mandatory for the vast majority of U.S. workers (approximately 97%); however, about 3% of workers are exempt from UI. Common examples are student workers, elected officials, legislators, railroad employees, and religious organizations. For benchmarking purposes, CESSA supplements the QCEW with imputed employment derived from other sources.

2. Survey methodology

The CES sample is stratified by state, industry, and employment size. In order to minimize the overall variance or sampling error in the statewide total private employment level, a fixed number of sample units (at the UI account level) to be drawn from each stratum is determined using optimum allocation. Sample weights are assigned at the time the sample is selected and are inversely proportional to the UI's probability of being selected from the population. In general, the less employment a UI has, the higher the weight assigned to the UI.

For monthly estimation, CESSA uses a matched sample concept. A matched sample comprises of sample units that reported positive employment for both the reference month ($\boldsymbol{t}$) and the prior month ($\boldsymbol{t-1}$). Units that report out-of-business are excluded from the matched sample and a net birth-death factor ($\widehat{\boldsymbol{N}}_{\boldsymbol{i,t}}$), which is based on the historical relationship between business births, deaths, and continuing units, is added at the cell level[4].

In cells with adequate sample, CESSA uses its sample-based "Robust" estimator to estimate employment. The reliability of the Robust estimator is limited in cells with small sample so they are instead estimated using either a composite estimator or a modified Fay-Herriot model (estimators are reevaluated annually). For the purpose of this paper, only the Robust and modified Fay-Herriot (MFH) estimators will be discussed in detail. The MFH uses the cell's Robust estimate and its five-year average monthly trend as inputs as well as the relationship between those inputs for all fifty states and D.C. Which estimator is used for a particular cell depends on the characteristics of both the cell and the available observations.

2.1 Micro editing

CESSA dedicates considerable resources to the editing and screening of sample data. All micro-data (individual survey records) are subject to three rounds of edits. First data are edited at collection. Then the data go through an automated edit and screening process before finally being reviewed by analysts.

During survey collection, the submitted employment, hours, and earnings data must pass two checks before being accepted. First, the entered micro-data must fall within a predetermined range of acceptable values. If data fall outside the range, the respondent must enter a reason code explaining the unusual movement. The second, a check of impossible values, will automatically reject the data. An example would be to check that the total number of all employees is less than the total number of production workers or that the average number of hours worked is greater than the number of hours in a week.

The next round happens after collection and before analyst review. This round flags data in daily batches and flagged data are sent to a data collection center for a re-contact of the respondent to validate and explain the change. This round of screening has three parts: another logic check, reasonableness parameters, and comparison to the establishment's history.

[4] Further information on the CES net birth/death model is provided by Mueller (2006).

After passing the first two rounds of review, the data are manually screened by analysts in two distinct stages before and after estimates are generated. During both stages, data thought to be erroneous are removed from the monthly sample and sent to the data collection center for reconciliation.

2.2 Robust estimator

The Robust estimator is used to generate approximately 55% of the statewide employment estimates and 28% of the MSA employment estimates. In order for a cell to use the Robust estimator, the cell must pass the p-percent test[5] and the average responding sample must contain either 30 UI accounts or unweighted sample coverage employment must represent at least 50% of the population and have population employment of at least 3,000. A cell must meet these adequacy requirements for two consecutive years to qualify for a sample-based estimator. The Robust estimator derives employment levels ($\hat{\boldsymbol{Y}}_{\boldsymbol{i,t}}$) using a weighted link relative ($\boldsymbol{L_i}$) formula, which is the ratio of the weighted employment sums for the matched sample ($\boldsymbol{S}$) at the cell level ($\boldsymbol{i}$).

$$\hat{\boldsymbol{Y}}_{\boldsymbol{i,t}} = \left(\hat{\boldsymbol{Y}}_{\boldsymbol{i,t-1}} - \sum_{\boldsymbol{j=1}}^{\boldsymbol{S}} \boldsymbol{y}_{\boldsymbol{j,t-1}}^{*} \right) \hat{\boldsymbol{L}}_{\boldsymbol{i}} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{S}} \boldsymbol{y}_{\boldsymbol{j,t}}^{*} + \hat{\boldsymbol{N}}_{\boldsymbol{i,t}}$$

2.2.1 Weighted link relative

The weighted sample link relative is the sum of weighted current month employment divided by the sum of weighted previous month employment for all observations ($\boldsymbol{j}$) in the matched sample ($\boldsymbol{S_{i,t}}$). This ratio accounts for the rate of change in the current month employment from the previous month employment. Ratios less than one cause the cell's employment level to decline from previous to current month, while ratios greater than one cause the cell's employment level to increase (where $\hat{\boldsymbol{Y}}_{\boldsymbol{i,t}}$ and $\hat{\boldsymbol{Y}}_{\boldsymbol{i,t-1}}$ are current and previous month employment levels, respectively).

$$\hat{\boldsymbol{L}}_{\boldsymbol{i}} = \frac{\sum_{j \in S_{i,t}} \boldsymbol{w_j d_j y_{j,t}}}{\sum_{j \in S_{i,t}} \boldsymbol{w_j d_j y_{j,t-1}}}$$

2.2.2 Weight adjustment

A cell's monthly sample can contain a small number of observations that may have a large and adverse effect on the employment estimate. The influence of such observations may be due to large sample weights, a significant change in the reported employment levels, or a combination of these factors. If left untreated, influential observations may cause level shifts in the monthly estimates, especially at the most detailed publication levels.

The Robust estimator addresses this effect using a variation of the "winsorization" weight reduction method (Kokic and Bell, 1994). Weight adjustments, just as the macro adjustments at the center of discussion in this paper, reduce the variance of the estimate, but may also introduce bias. For example, in the context of the matched sample, an establishment's employment change may appear unique compared to the rest of the sample, but when the population becomes available, it is known to be truly representative of the

[5] The p-percent test, implemented in 2006, is the BLS standard for identifying cells at risk for a respondent's identity to be reasonably inferred, by either direct or indirect means, in compliance with the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).

change in employment for similar establishments in the non-sampled portion of the population. In this way, moderating the effect of the sample's influential units can lead to a reduction in the representativeness of the sample.

The Robust estimator is designed to reduce the volatility of estimates due to extreme outlying reports in a controlled procedure that protects against incurred bias. This procedure identifies influential observations at the sampling unit (i.e. UI account) level for each estimating cell's matched sample ($\boldsymbol{S}$). An influential UI tends to have a relatively high weight ($\boldsymbol{w_j}$) and/or a large over-the-month change in reported employment. For the purpose of this procedure, the influence of a UI on the weighted link relative is measured as a weighted residual where the residual is$(\mathbf{r_j})$:

$$\boldsymbol{r_{j,t} = w_{j,t}y_{j,t} - L_j w_{j,t} y_{j,t-1}}$$

Influential UIs are those with large, positive or negative, residuals relative to other units of the same size class in the matched sample. A weight adjustment factor $(\boldsymbol{d_j})$ is used to reduce the influence of UIs with extreme residuals.[6] Applying the weight adjustment factor to the Robust estimator effectively removes or reduces the UI's influence from the weighted link relative ($\boldsymbol{\hat{L}_i}$). The procedure will remove some UIs from the weighted link relative by applying a weight adjustment factor of zero and treating the movement as "atypical" (more below). Others will have their influence reduced by applying a weight adjustment factor greater than zero but less than one. The remainder (and, in practice, majority) receive a factor of one allowing full influence on the weighted link relative.

2.2.3 Atypical observations
Atypical UIs only represent themselves in the Robust estimator by design. Influential UIs identified by the Robust winsorization ($\boldsymbol{y^*}$) are not included in the link relative; instead, the unweighted sum of the previous month's atypical observations is removed from the previous month's employment before multiplying by the weighted link relative. Then, the sum of the current month's atypical observations is added back into the estimate.

$$\boldsymbol{\hat{Y}_{i,t} = \left(\hat{Y}_{i,t-1} - \sum_{j=1}^{k} y^*_{j,t-1}\right) L_i + \sum_{j=1}^{k} y^*_{j,t} + \hat{N}_{i,t}}$$

2.3 Modified Fay-Herriot estimator
The MFH estimator is used to calculate employment for MSA-level cells that do not meet Robust estimator sample criteria or are known to have a high variance$(\boldsymbol{V_1})$. The MFH estimator is used to generate approximately 22% of the MSA employment estimates. The MFH estimator "borrows strength" from the historical data and from across states. The MFH estimator calculates a cell's employment as a weighted average of two separately derived estimates. The MFH estimator uses two employment estimate inputs: the current month Robust estimate ($\boldsymbol{\hat{Y}_1}$), covered above, and a synthetic input ($\boldsymbol{\hat{Y}_2}$) adjusted by a beta coefficient. The synthetic input is based upon a population historical average of the sector, for the same month, across all states and D.C.. The beta coefficient is intended to transform

[6]For a complete explanation of weight adjustment calculation see the BLS Handbook of Methods, Chapter 2.

the synthetic input to the historical average for the subnational cell. The MFH formula used to estimate a cell's employment is:

$$\widehat{\boldsymbol{Y}}_{i,t} = \boldsymbol{g}_1\widehat{\boldsymbol{Y}}_1 + \boldsymbol{g}_2\widehat{\boldsymbol{Y}}_2$$

2.3.1 Synthetic input

Monthly changes in employment have industrial and geographic specific trends (e.g. seasonality). The MFH estimator builds on this assumption by projecting such trends and using them as a model input. First, the monthly change is projected using the cell's five year average ratio of current $\boldsymbol{Y}_{t,\bar{n}}$ and previous month ($\boldsymbol{Y}_{t-1,\bar{n}}$) benchmarked employment. The projected monthly change$\left(\boldsymbol{Y}_{t,\bar{n}}/\boldsymbol{Y}_{t-1,\bar{n}}\right)$ is regressed on the reference period ratio of current and previous month estimated employment $\left(\widehat{\boldsymbol{Y}}_t/\widehat{\boldsymbol{Y}}_{t-1}\right)$ for all states and D.C. with the associated industrial supersector. This yields a beta coefficient ($\boldsymbol{\beta}$) that "corrects" area forecasts based on the current tendency across all areas.

$$MFH\ Regression\ Equation\colon \left(\frac{\widehat{\boldsymbol{Y}}_t}{\widehat{\boldsymbol{Y}}_{t-1}}\right) = \boldsymbol{\beta}\left(\frac{\boldsymbol{Y}_{t,\bar{n}}}{\boldsymbol{Y}_{t-1,\bar{n}}}\right) + \boldsymbol{\varepsilon}$$

The coefficient $\boldsymbol{\beta}$ transforms the cell's projected monthly change into the synthetic input ($\widehat{\boldsymbol{Y}}_2$), as shown in the formula below. This transformation produces an input that improves on $\boldsymbol{Y}_{t,\bar{n}}$ by incorporating recent trends and so reducing bias while also decreasing variance of $\widehat{\boldsymbol{Y}}_{i,t}$.

$$\widehat{\boldsymbol{Y}}_{i,t} = \boldsymbol{g}_1\left[\left(\widehat{\boldsymbol{Y}}_{i,t-1} - \sum_{j=1}^{k} \boldsymbol{y}^*_{j,t-1}\right)\widehat{\boldsymbol{L}}_i + \sum_{j=1}^{k} \boldsymbol{y}^*_{j,t} + \widehat{\boldsymbol{N}}_{i,t}\right] + \ \boldsymbol{g}_2\,\boldsymbol{\beta}\left(\frac{\boldsymbol{Y}_{t,\bar{n}}}{\boldsymbol{Y}_{t-1,\bar{n}}}\right)\widehat{\boldsymbol{Y}}_{i,t-1}$$

2.3.2 MFH model weights

The weight that is placed on the sample-based portion of the MFH model is calculated using variance $\boldsymbol{V}_1$ (derived based on a generalized variance function) of $\widehat{\boldsymbol{Y}}_1$, variability $\boldsymbol{V}_2$ of the historical monthly changes, and the average lack of fit ($\boldsymbol{A}$) from the regression. The regression's average lack of fit ($\boldsymbol{A}$) is the degree of the dispersion of data points around the regression's best-fit line. The calculation of the $\widehat{\boldsymbol{Y}}_1$ weight ($\boldsymbol{g}_1$) is shown below. Since $\boldsymbol{g}_2$ is calculated as $(\boldsymbol{1} - \boldsymbol{g}_1)$, the lower the variance of $\widehat{\boldsymbol{Y}}_1$, the more the resulting MFH weighted average reflects the sample-based estimate and the higher the $\widehat{\boldsymbol{Y}}_1$variance the more the result reflects the synthetic estimate.

$$\boldsymbol{g}_1 = \frac{\boldsymbol{A}}{\frac{\boldsymbol{V}_1}{\boldsymbol{V}_2} + \boldsymbol{A}}$$

2.3.3 MFH z-score

The macro editing procedure uses a z-score ($\boldsymbol{Z}$) to test for deviations between the observed sample-based values ($\widehat{\boldsymbol{Y}}_1$) and the historical values adjusted by $\boldsymbol{\beta}$ ($\widehat{\boldsymbol{Y}}_2$). In the empirical study that follows, Robust estimates with z-scores greater than or equal to prescribed thresholds are replaced by MFH modelled estimates. In practice, this replacement would only be done if deemed necessary by an analyst.

$$\mathbf{Z}_{i,t} = \frac{\widehat{Y}_{1_{i,t}} - \widehat{Y}_{2_{i,t}}}{\sqrt{\frac{V_1}{(1-g_1)}}}$$

3. Empirical study

The proposed method to identify and adjust extreme macro outliers builds on the long-standing assumption that current sample-based estimates should not substantially differ from corresponding historical sample values. Under this method, deviations are identified using the MFH z-score described above. The following empirical analysis examines the effect of this procedure on a proxy for total survey error (benchmark revision) and seeks to determine whether the procedure is reasonable to use for production of published estimates.

Using the full set of Robust estimates for all states, D.C., and all MSAs, four annual[7] sets of experimental monthly estimates were created (benchmark years[8] 2012 to 2015), where Robust estimates were replaced with MFH models at different z-score thresholds from 2.0 to 4.0 in increments of 0.1. For example, if a Robust estimate has z-score of 2.35, then the z-score groups with thresholds of 2.3 and below will replace that Robust estimate with the MFH model estimate, whereas that same estimate uses the Robust estimator in groups 2.4, 2.5, 2.6, and so on. Included for reference is a control group, referred to as “raw”, which has no adjustments for extreme movements.

Trade-off is expected between monthly prediction error and aggregate yearend revisions; a tighter threshold tends to reduce monthly revisions but tends to come at the cost of increasing aggregate yearend revisions, an indication of bias. Conversely, higher thresholds tend to reduce bias over time since they allow extreme monthly variance, which in many cases is detrimental to the reliability of month-to-month changes. Finally, the number of adjustments a threshold allows is an important consideration. While the analyst will only use macro adjustments as a last resort interventions, adjusting a large number of series increases the risk of missing economic turning points.

3.1 Monthly variance

In the CES survey, the over-the-month change is closely monitored. As such, monthly prediction error must be considered in evaluating the effect of using the model-based estimator in place of the Robust estimator. Since CESSA annually benchmarks estimates to adjusted monthly QCEW counts, the benchmark revision is a strong proxy of total survey error[9]. The effect on accuracy of replacing robust estimates with MFH models is measured using root mean square error (RMSE) of the difference between estimated and benchmarked (i.e. population) monthly changes. Figures 1, shows the aggregate of all Robust estimates by each benchmark year and figures 2-5 show the aggregates by supersector[10].

[7] Each year starting with October and ending with September.
[8] The start of a benchmark year is the period following benchmarking to adjusted QCEW data. For the purpose of this paper, the period starts in October and ends in September of the following year. For example, benchmark year 2012 begins with October 2012 and ends in September 2013. Data prior to October 2012 have been replaced with the adjusted QCEW data.
[9] QCEW is not subject to sampling error, but it does contain other types of administrative error.
[10] Supersectors are groupings of NAICS industry sectors and are defined at: https://www.bls.gov/sae/saesuper.htm

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{l}\left(\left(Y_{i,t}-Y_{i,t-1}\right)-\left(\hat{Y}_{i,t}-\hat{Y}_{i,t-1}\right)\right)^2}{l}}$$

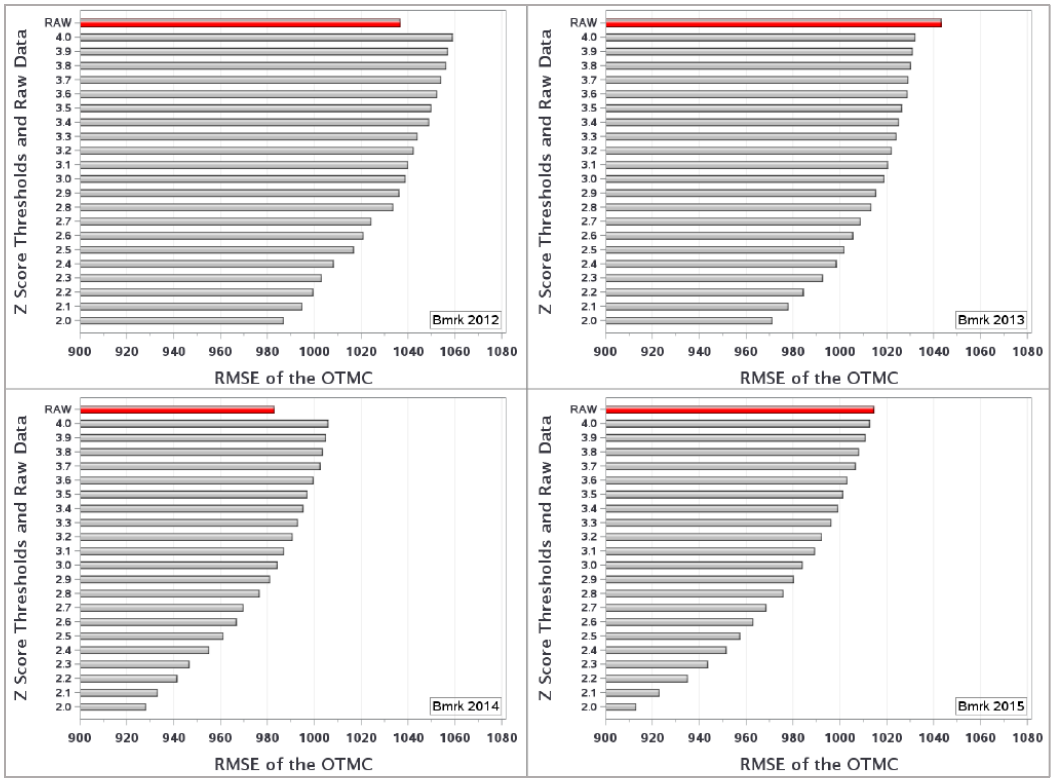

Figure 1: RMSE of the OTMC for all Robust series

In the results for Figure 1, RAW represents estimates that have not been altered. The results show that adjusting for macro outliers at lower z-score thresholds tends to lower monthly error. The pattern is consistent across all four benchmark years where replacing Robust estimates with MFH estimates at lower z-score thresholds yields OTMCs that are closer to the population values. The results also show that in two cases, the raw data out-perform adjusting only at the highest z-scores evaluated. This could indicate unusual events in the population, which should not be edited, and are a fair representation of the risk associated with controlling monthly variance.

RMSE is sensitive to outliers; therefore a further breakdown at the industrial sector level is conducted to see if any persistent trend occurs in any of the subcategory that moves against the rest of the sample.

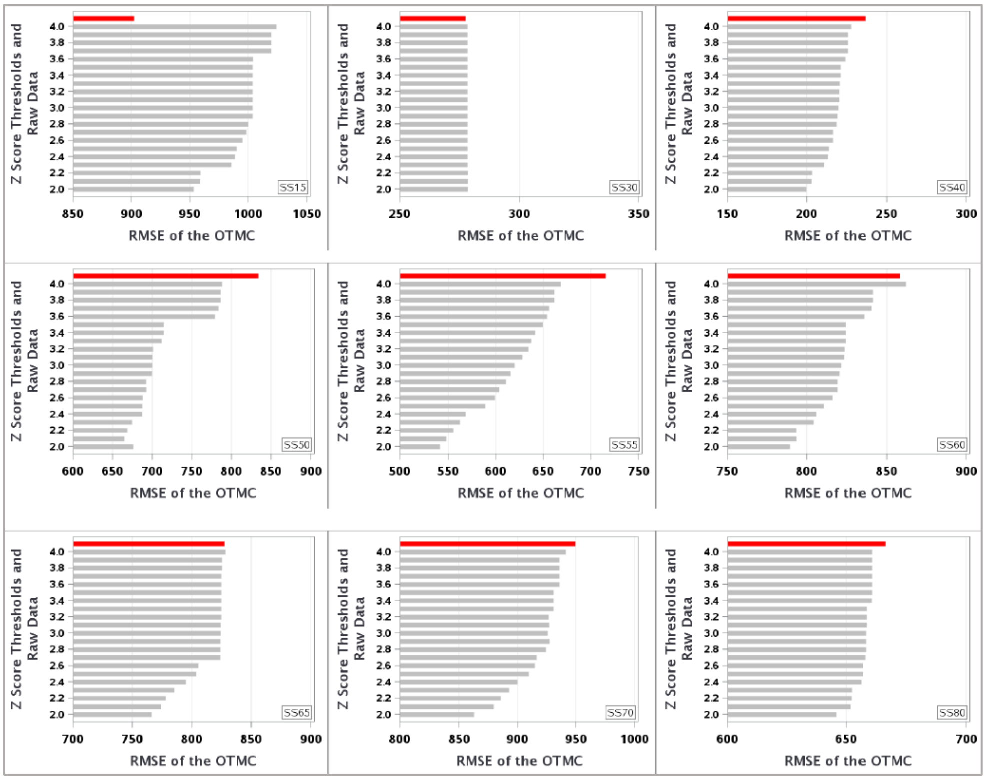

Figure 2: RMSE of the OTMC for all Robust series by Super Sector (benchmark 2012)

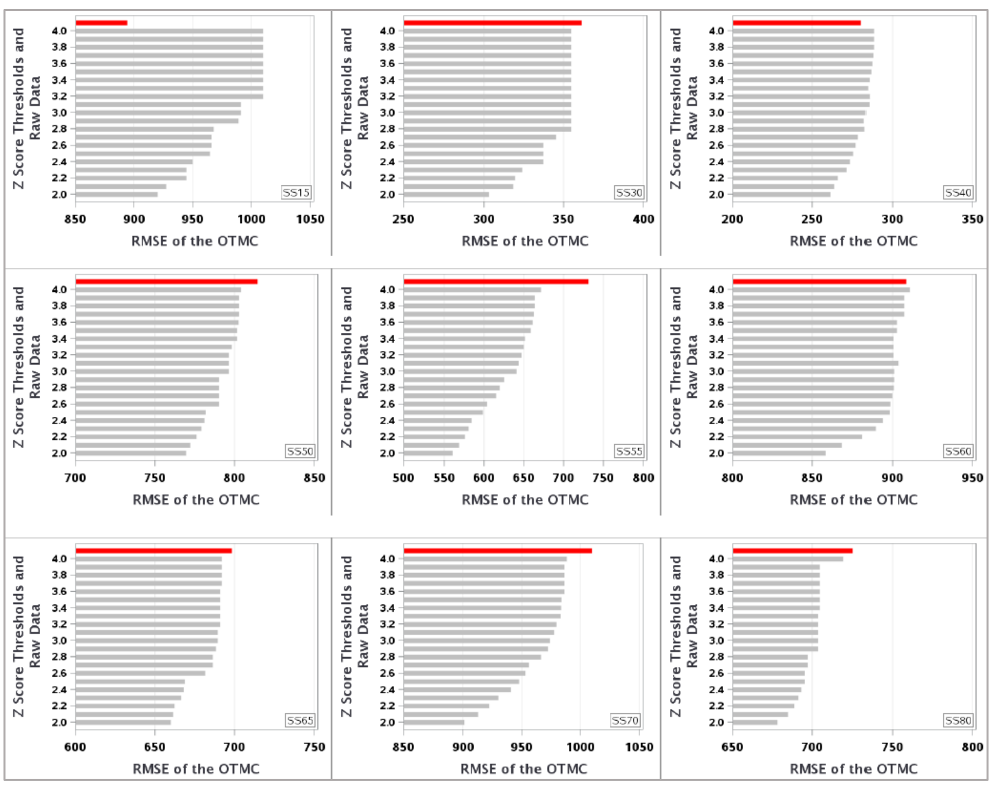

Figure 3: RMSE of the OTMC for all Robust series by Super Sector (benchmark 2013)

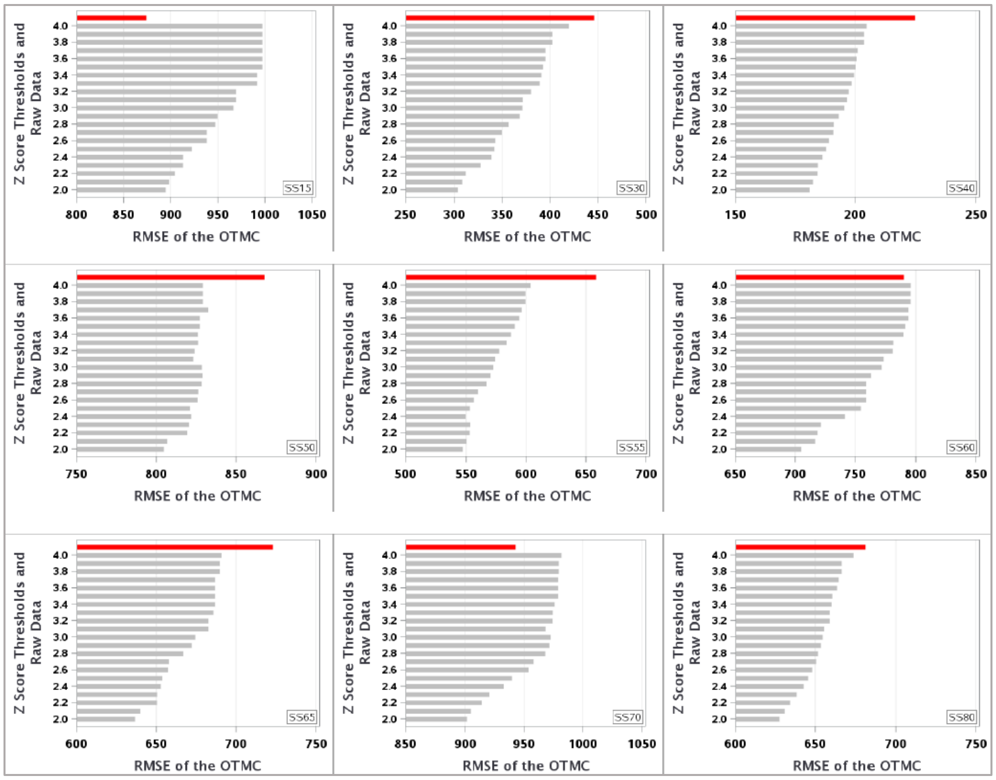

Figure 4: RMSE of the OTMC for all Robust series by Super Sector (benchmark 2014)

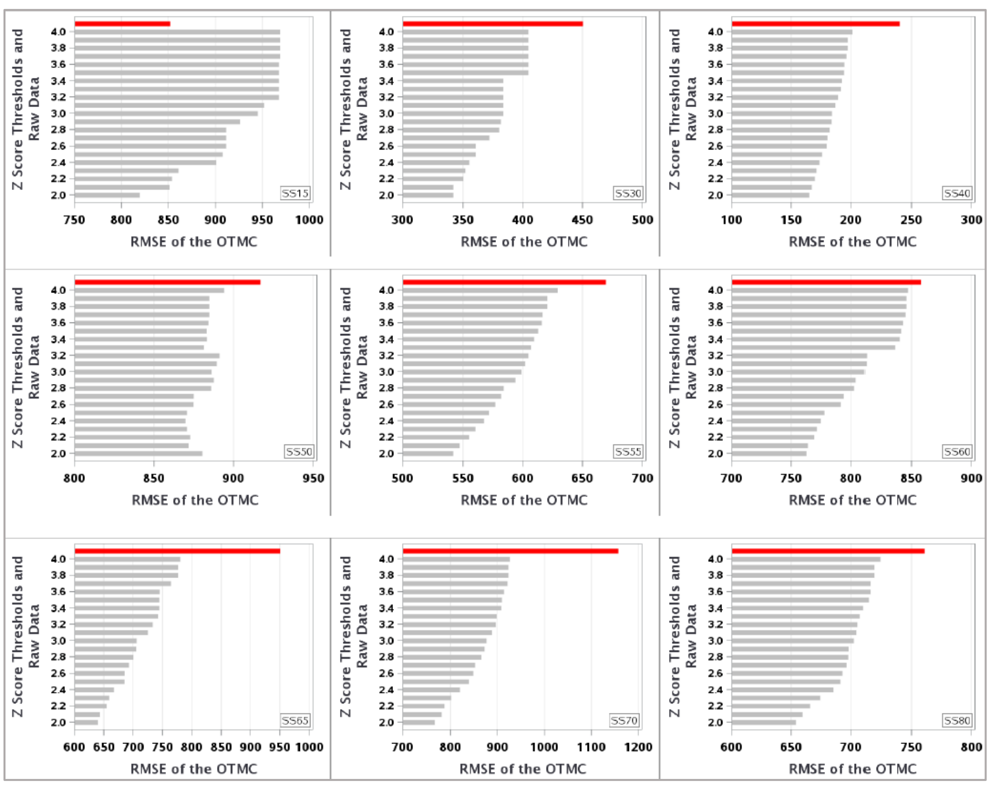

Figure 5: RMSE of the OTMC for all Robust series by Super Sector (benchmark 2015)

When the data are examined by supersector the pattern of low z-score, low RMSE remains the same for the most part. In most cases, some degree of macro editing seems to be better than none (i.e. raw). Supersector 15 (Mining, Logging, and Construction) appears to present an exception, however there are only 15 observations with absolute z-scores greater than or equal to 4 across all four benchmark years so it's not surprising to find a high degree of "noise" in this relatively small data set. By comparison, supersector 80 only had 3 Robust estimates with z-scores greater than or equal to 4, supersector 50 had 53, and the rest averaged 267.

3.2 Yearend bias

Ideally, monthly CESSA estimates would accurately capture both monthly changes in employment and employment levels, however, to do both simultaneously is challenging at times due to the bias-variance tradeoff dynamic. While the Robust estimator offers low bias, it also allows for a considerable amount of variance; whereas the MFH risks biasing individual estimates toward their historical trends. For bias considerations, this study uses the aggregated yearend revision ($\hat{A}$) or the absolute difference in the benchmark over-the-year change (OTYC) and the OTYC of estimates adjusted at cascaded z-score thresholds. These yearend revisions are aggregated across all series[11] for statewide and sum of all MSAs in order to assess the bias effects on statewide and area estimates separately.

$$\hat{A} = \left|\left(\sum_{i=1}^{l} Y_{i,t} - Y_{i,t-12}\right) - \left(\sum_{i=1}^{l} \hat{Y}_{i,t} - \hat{Y}_{i,t-12}\right)\right|$$

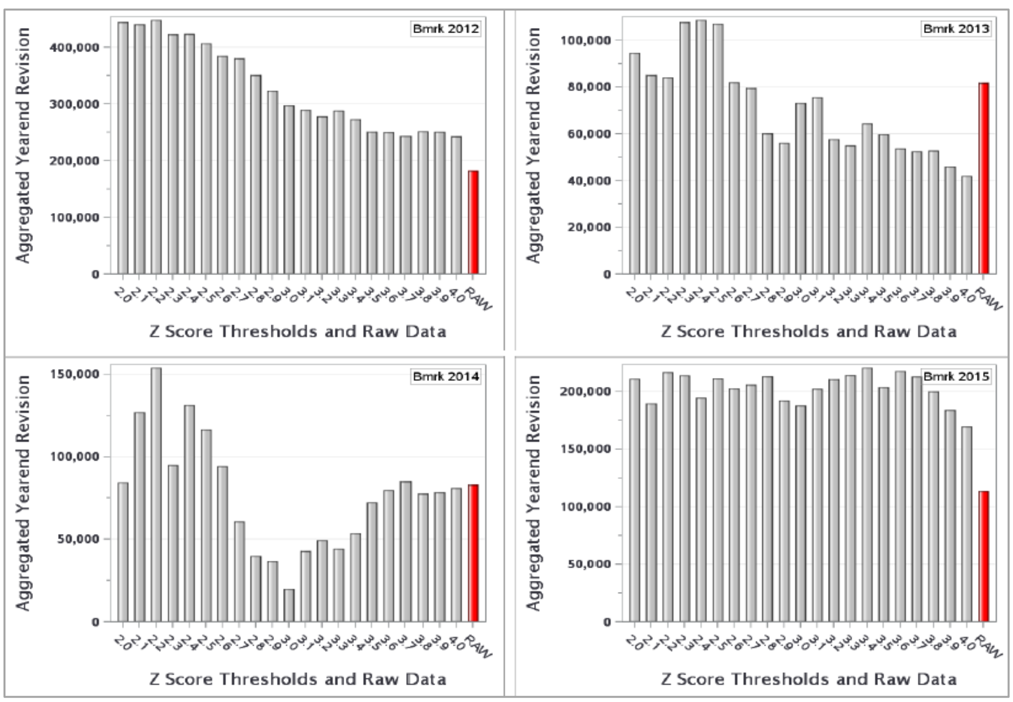

Figure 6. Statewide OTY Revisions

---

[11] Aggregates contain only Robust estimates. Since many CESSA estimates are directly modeled, the aggregates will not match aggregation done independently with a full set of publicly available data.

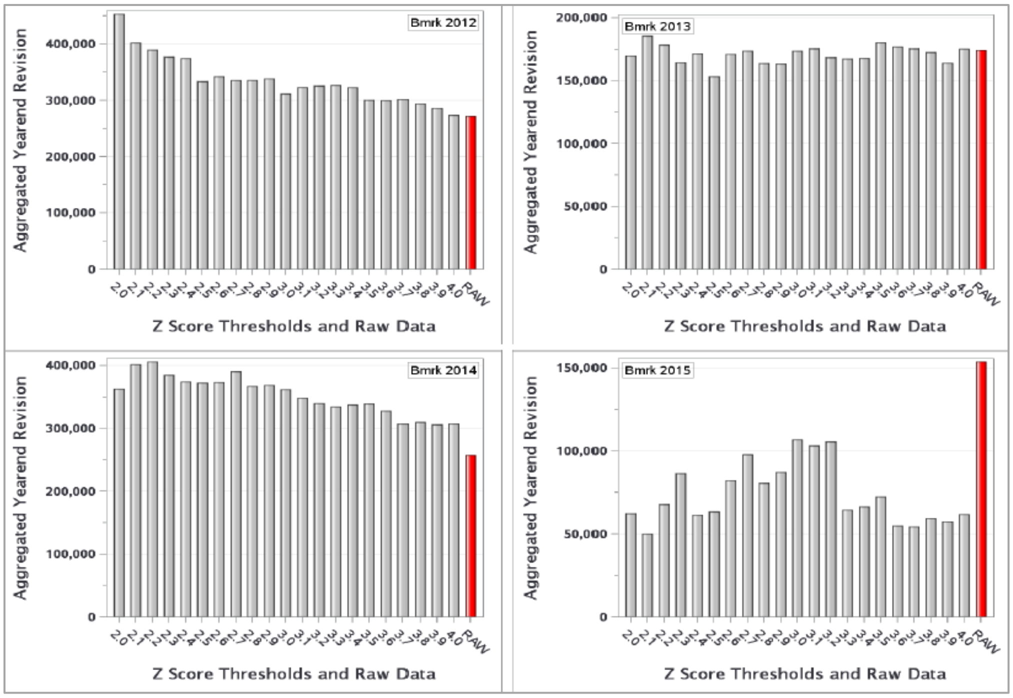

Figure 7. MSA OTY Revisions

In figures 6 and 7, note the difference in scale across benchmark years for both statewide and MSA aggregates. Yearend revisions ranged from 19,974 (2014) to 447,799 (2012) for the statewide aggregates and from 50,159 (2015) to 453,467 (2012) for the MSA aggregates. Also, notice the evidence of the theoretical variance-bias tradeoff. For the most part, modeling extreme monthly changes tends to increase bias for both the statewide and MSA aggregates. There is not a distinct pattern in the yearend revisions as in the monthly RMSE, but this is, perhaps, partially attributed to sets with exceptionally low bias. For example, in 2013, the yearend statewide revision for the raw Robust estimates was 174,356, which was 35.9% and 32.4% lower than 2012 and 2014, respectively. The relatively larger MSA aggregate yearend revisions in 2012 and 2014 do exhibit a pattern consistent with the expected tradeoff where the controlled variance comes at the expense of increased bias. This pattern can also be seen in statewide for 2012 where yearend revisions were at their largest.

While bias is expected to increase when Robust estimates are replaced with modeled estimates, at some extreme level it does provide considerable improvement. Consider statewide 2013 and MSA 2015, notice the extreme yearend revisions for the raw data. These extreme cases are usually the result of a small number of especially erratic estimates, which despite best efforts to manage through micro editing and less invasive adjustments, are highly errant. Cases such as this are why CESSA requires a procedure for macro adjustment.

3.3 Distribution diagnostics

In Figures 8, z-scores for the “raw” Robust statewide estimates are plotted against theoretical normal distributions ($\mu = 0$; $\sigma = 1$) using quantile-quantile plots. The plots show that all but a few estimates fall on or near the theoretical distributions. While this overall distribution suggests that the Robust estimator has a reasonable probability of accurately predicting monthly change, the distributions also has a “fat-tailed” property that show there are a number of observations with extreme z-scores. Building from the analysis of the RMSE and yearend revisions, we can reasonably conclude that these extreme observations have a considerably lower likelihood of accurately predicting monthly change and more often than not will adversely affect yearend revisions. Again, it is important to note that analysts research events affecting local employment and seek to validate survey data before modeling the estimate.

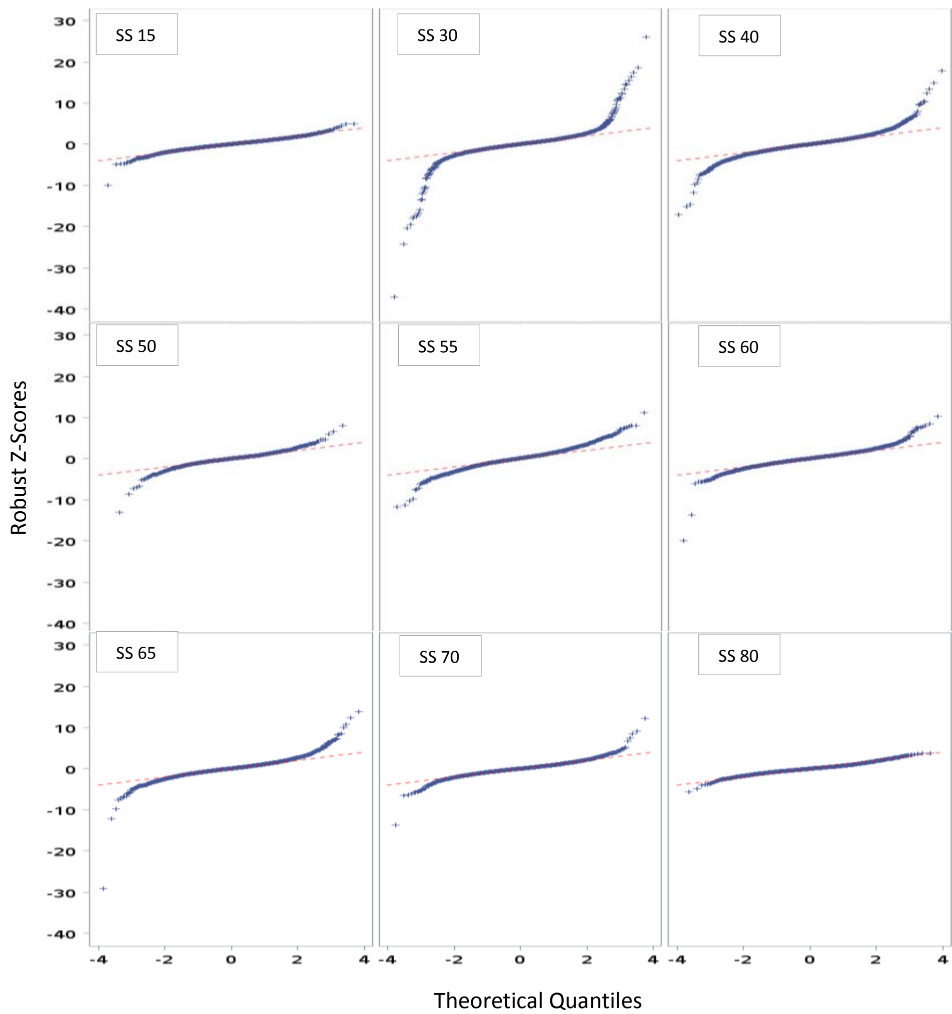

Figure 8. Q-Q Plot of Robust Z-Scores by Supersector

3.4 Number of Adjustments Flagged

Macro adjustment should only be done as a last resort remedy for unexplainable extreme movements; however, editing could certainly have an adverse effect on estimates. Therefore, flagging 10% of series may cause 10% of the estimates to be adjusted in the extreme scenario. A separate analysis is done on the percentage of series that analysts macro adjust in comparison to the number of adjustments that are flagged in the later section. Figure 9 shows the relationship that the lower the threshold, the more series will be flagged as outliers. It also shows that the number of series that are flagged as positive outliers is about the same as the number that are flagged as negative outliers.

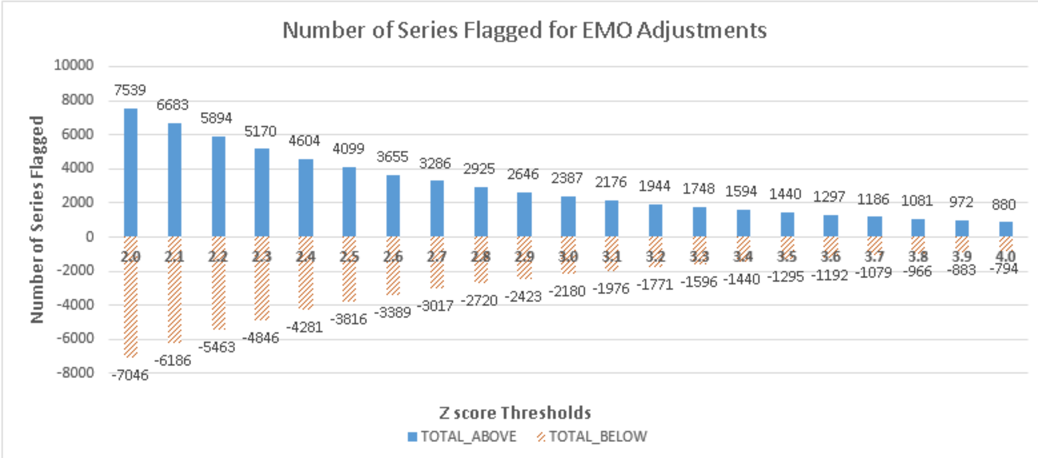

Figure 9: Number of Macro Adjustments Flagged

4.0 Summary

In this paper we offer an empirical review of the macro editing method proposed by Gershunskaya (2012). While macro editing is a long-standing practice in the CESSA survey, prior methods were based more on the application of analyst knowledge using the long-standing assumption that current sample-based estimates should not substantially differ from corresponding historical values. This proposal builds on that assumption, but formalizes the approach, allowing CESSA to hedge the risk of “bending” estimates to match history. Additionally, formalizing the macro editing procedure offers consistent treatment of estimates and furthers review efficiency in the tight constraints of the monthly production environment.

It is reasonable to assume that replacing some small number of monthly Robust estimates with model-based estimates, based on their deviation from historical trends, will improve both monthly error variance and yearend bias. While there is no z-score threshold that will always lower both monthly RMSE and yearend revisions, the fat-tail distribution of the Robust estimates (as shown in the Q-Q plots) suggests that there are outliers in the data. CESSA chose to establish a macro editing absolute z-score threshold of 3.0, with those above 2.5 eligible under certain circumstances with additional scrutiny. Not all Robust estimates with an absolute z-score greater than or equal to 3.0 will be replaced. Only those that an analyst deems economically unreasonable will actually be replaced. This procedure is not infallible, but it is relatively rigorous and, put simply, it is a preferred alternative to publishing monthly estimates that indicate major shifts in employment when the CESSA analyst believes the change is most likely statistical error.

CESSA implemented the proposed macro-editing procedure beginning with the October 2016 re-estimates. Dedicated analysts continue to review pre-published estimates as they always have, but now to supplement their traditional techniques of aggregation review they have access to a macro-edit report that identifies Robust estimates with an absolute z-score greater than or equal to 2.5. The analyst will first consider whether any observations have “slipped through” prior data editing procedures and evaluate all other micro level adjustments within approved methodology. After these considerations, if the analyst believes the estimate is economically unreasonable, they are authorized to replace the sample-based estimate with the modeled estimate for observations with an absolute z-score greater than or equal to 3. As of the July final production cycle, of the full set of sample-based estimates (54,325 observations), 1,254 (2.3%) qualified for adjustment with an absolute z-score greater than or equal to 3 and 1,029 (1.9%) with an absolute z-score between 2.5 and 3. Of those, 367 observations (29.3%) and 172 (16.7%), respectively, were replaced by model-based estimates.

This empirical study outlines CESSA considerations prior to adopting formalized macro-editing. The Fay-Herriot model is widely used in the statistical community for small-area estimation and is used by CESSA to estimate subnational industry estimates with small sample sizes. Given that the Robust estimates are known to contain outliers, using the Fay-Herriot model for monthly estimates that profile as outliers is a logical approach. When relying on history as an indicator of current trends there is always the risk of “bending” estimates to history, but the results of this study suggest that the risk in modeling relatively few accurate estimates with absolute z-scores of 3 or greater is out-weighed by the overall reduction in error.

References

Bureau of Labor Statistics. (2004). *Employment, hours, and earnings from the Establishment survey* (BLS Handbook of Methods Chapter 2). Washington, DC: U.S. Department of Labor. http://www.bls.gov/opub/hom/pdf/homch2.pdf

De Waal, T. (2009). Statistical Data Editing, Handbook of Statistics, Sample Surveys: Design, Methods and Applications, Eds. D. Pfeffermann and C.R. Rao, Amsterdam:Elsevier BV. *Vol. 29A*, Ch. 9.

Gershunskaya, J. (2012). Model Based Macro-Editing Approach to State and Area Estimates. *Proceedings of the Fourth International Conference on Establishment Surveys, American Statistical Association*. https://ww2.amstat.org/meetings/ices/2012/papers/301918.pdf

Kokic, P. N., & Bell, P.A. (1994). Optimal Winsorizing Cutoffs for a Stratified Finite Population Estimator. *Journal of Official Statistics, Vol. 10* (No.4), 419-435. http://www.jos.nu/Articles/abstract.asp?article=104419

Mueller, K. (2006). Impact of business births and deaths in the payroll survey. *Monthly Labor Review*, *May*, 28-34. http://www.bls.gov/opub/mlr/2006/05/art4full.pdf