J

Variance Estimation by Replication for National CPS Seasonally Adjusted Series October 2015

Thomas D. Evans, Justin J. McIllece, Stephen M. Miller
Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212

Abstract
There is much interest in month-to-month changes for Current Population Survey (CPS) seasonally adjusted labor force series at the national level. Much of this interest focuses on producing confidence intervals around the monthly change in the seasonally adjusted national unemployment rate. Unfortunately, variances for those series are not currently available due to the complexity of the X-11 seasonal adjustment process. Many studies are available on how to estimate those variances. A common approach is to utilize sampling error information and a linear approximation to X-11. Less work has been applied to an alternative approach of deriving variances from seasonally adjusted replicate series that also accounts for sampling errors. An adequate time series of consistent CPS replicate weights is now available which allows for an examination of new variance estimates for seasonally adjusted series by the replication approach. A description of our methodology and results are presented for the national employment, unemployment, and unemployment rate series.

Key Words: SEATS; seasonal adjustment; replication; GVF; sampling errors

1. Introduction

Variances are not officially available for the Current Population Survey national seasonally adjusted series. Variances for the unadjusted series are used instead based on an assumption that they are close enough. However, we know that variances for seasonally adjusted series will generally be smaller than those for unadjusted series—the problem is finding a method that works well and is relatively easy to implement.

Researchers have undertaken various approaches to calculate variances for seasonally adjusted series. A brief description of a few of those studies follows. In the late 1950s and early 1960s, the Census Bureau investigated variances by replication for seasonally adjusted estimates using 20 replicate series (Wolter and Monsour, 1981) but never published any results. Armstrong and Gray (1986) of Statistics Canada found lower variances from replication for some Labour Force Survey seasonally adjusted series but found them too unstable to use without further work. Wolter and Monsour (1981) utilized the linear approximation to X-11 and sampling error information for variances of seasonally adjusted series. The estimates were smaller for central observations but larger at the ends of the series. The work by Wolter and Monsour has been extended in various ways by Pfeffermann (1994), Bell and Kramer (1999), Scott, Sverchkov, Pfeffermann

[1] Disclaimer: Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

(2012), and Pfeffermann and Sverchkov (2014). The Pfeffermann-Sverchkov method allows for the irregular component to be in the signal or included with the sampling error in the error term. We compare Pfeffermann-Sverchkov results to ours in Section 3.

Another common approach for variances of seasonally adjusted series is model-based. While there are several worthwhile studies on model-based approaches, one of the most significant is found in Tiller (2012). Tiller applies the Kalman filter and a state-space model for state CPS series that accounts for sampling error. Thus, the resulting seasonal adjustment variances account for time series and sampling error variation and are often much lower than those for their respective unadjusted state series. These variances are officially published by the U.S. Bureau of Labor Statistics (BLS).

Why has BLS not yet produced variances for national seasonally adjusted series even though theory should tell us that seasonally adjusted variances will almost always be lower? (See Appendix.) One reason is that work on variances for X-11 has progressed slowly and early methods were not so simple to implement in production. Another issue with using one of the X-11 approaches is that variances for the unadjusted and seasonally adjusted series may not be directly comparable. Variances from time series models are another possibility, but that would be another large project. As we now have a reasonably long time series of CPS replicate weights, researching variances by replication seems a reasonable next step.

The research in this paper explores the use of replicates to create variances for CPS national seasonally adjusted series. Section 2 discusses the methodology used in this paper to create variances; Section 3 covers our results and a comparison to the Pfeffermann-Sverchkov method; and Section 4 offers a summary. Figures and tables are in the Appendix followed by references.

2. Methodology

Every month, the Census Bureau provides BLS with 160 replicate weights for CPS national series. The method for creating those weights has been consistent since 2003 and we can easily reproduce the full sample estimates from the replicate weights back to that point. Since 2003, differences between the full sample estimate and the average of the 160 replicate series have been almost zero for each month. It is desirable to have at least 8-10 years of monthly data for seasonally adjusted procedures so we feel this approach is now doable.

As the replicate weights are readily available, creating variances by replication for seasonally adjusted series is fairly straightforward. Seasonal adjustment for most directly adjusted CPS national series uses the X-11 procedure, but the SEATS procedure was used for newly adjusted series starting in January 2015 (Tiller and Evans 2015). Since it is likely that the use of SEATS for official national CPS seasonal adjustment will increase over time, we chose to use SEATS as implemented in the X-13ARIMA-SEATS program (U.S. Census Bureau 2015) for this study.

One hundred and sixty replicate series were created with the replicate weights for total employment, total unemployment, and the total unemployment rate series over the single period of 2003-2014. (Note that data near the end of the series will obviously change as new months are added due to the use of moving averages in seasonal adjustment). All model settings, model parameters, and any detected outliers in X-13 were held fixed for each series based on the seasonal adjustment for the respective full sample series. We also computed replicate variances for employment and unemployment as indirect seasonal adjustments following the BLS official method (see Tiller and Evans for details). Differences were small from estimating the variances using an indirect or direct approach.

Variances for both the unadjusted and seasonally adjusted replicate series were calculated using the balanced repeated replication method for both the unadjusted and seasonally adjusted replicate series:

$$Var(\hat{Y}_0) = \frac{1}{160(1-K)^2} \sum_{r=1}^{160} (\hat{Y}_r - \hat{Y}_0)^2$$

where K=0.5, $\hat{Y}_r$ is the estimate for the r$^{\text{th}}$ set of replicate weights, and $\hat{Y}_0$ is the full sample estimate using the estimator weights. BLS uses the same equation for replicate variances of unadjusted national CPS series (BLS and U.S. Census Bureau, 2006).

Due to changes in the CPS sample design from the 2000 and 2010 Decennial Censuses, some replicate variances were unreasonably large in certain months during the phase-in for these changes. To eliminate these artificial effects, ARIMA models were fit to the variance and standard error series and any detected outliers were treated as missing values. For example, replicate variances for the employment level originally have large spikes in April 2004 and April 2014 where changes based on the 2000 and 2010 censuses were introduced.

Since replicate variances tend to be noisy, the CPS fits generalized variance functions (GVFs) to them using a two-step process (BLS and U.S. Census Bureau, 2006). Initially, similar labor force series are grouped together using a clustering algorithm. Then an iterative, weighted least squares regression model is fit to each group. The model parameters are constrained in such a way that they reflect the properties of a binomial distribution, which is appropriate for CPS series that count the number of people in certain subgroups, such as the employment and unemployment series described above.

Similarly, GVFs were fit to the replicate variances for both the unadjusted and seasonally adjusted series in this research. However, the GVF model used for this project was not the same as the historical model used by the CPS. Since certain properties have been fairly stable since 2003, such as the national sample size and the population growth rate, each series was fit individually with a simple regression model constructed to retain the desired binomial properties.

3. Results

The replicate variances and the month-to-month change standard errors for total employment, unemployment, and the unemployment rate are seen in Figures 1-6. As expected, the replicate variances are noisy, which clearly justifies modeling them with a GVF approach.

Figure 1 shows the replicate variances for the unadjusted and seasonally adjusted total employment levels over 2003-2014. The GVF variances roughly cut through the respective variances, but there are times where sampling and/or replication error might drift high or low for an extended period. Note that our GVF variances do not follow the extreme changes in replicate variances and are smooth even though we do not explicitly apply any smoothing filters. The standard errors for month-to-month changes in Figure 2 also shows a similar story.

The plots in Figures 3 and 4 are the same as for Figures 1 and 2, except they are for total unemployment. As expected, the seasonality is more pronounced for the GVF unemployment variances than for employment. Also, the variances in Figure 3 begin to rise in 2008 with the onset of the Great Recession and have gradually fallen as the recovery slowly continues. The total unemployment rate plots in Figures 5 and 6 show similar findings as for unemployment.

The average percent reduction in month-to-month change GVF standard errors from unadjusted to seasonally adjusted for employment, unemployment, and the unemployment rate are in Table 1. While the average reductions for the series are around 8 percent, there is more variation over the months for unemployment and the unemployment rate. For unemployment and unemployment rate, months that have stronger seasonality tend to have larger reductions in standard errors. Several of the monthly reductions are over 10 percent, and the overall range in reductions by month is 3.2% to 14.3%. This is important as it means that series with stronger seasonality are more likely to show gains from calculating variances for seasonally adjusted series instead of simply using the variances from the unadjusted series.

Overall, using unrounded data, there are several months when, applying 90% confidence intervals (CIs) from the unadjusted series for the seasonally adjusted series, there are differences for employment (twice), unemployment (four times), and the unemployment rate (twice) over the 2003-2014 period. BLS currently posts the differences needed for a statistically significant change for 31 national seasonally adjusted series at http://www.bls.gov/web/empsit/cpssigsuma.pdf. Beginning with August 2015 data, the new method for GVF variances was applied. (GVF factors used in the published table for seasonally adjusted series are those for the unadjusted series.) Significance levels for total unemployment rate changes in the table are shown to the hundredths place.

Finally, a comparison was made to the Pfeffermann-Sverchkov method in Section 1. CIs at the 90% level for unemployment rate month-to-month change seasonally adjusted of the two methods are plotted in Figure 7. The differences are small and those for our approach are slightly narrower. The seasonally adjustment results for Pfeffermann-Sverchkov in the plot was X-11. However, utilizing SEATS for these data makes little difference in this case. A small dip in the last observation for the Pfeffermann-Sverchkov CIs disappears with SEATS. Though the program for the Pfeffermann-Sverchkov method can account for an irregular term, we did not include it in the error term since the effects on the variances were very small.

4. Summary

Our study analyzed variances for seasonally adjusted employment, unemployment, and unemployment rate from replication and newly developed GVF models. In general, we find that our method works well for topside national CPS series and produces smoother and lower variances for seasonally adjusted series. Major findings are:

- The new generalized variance function procedure is consistent over time and smooths the replicate variances but clearly shows that the variances for the unadjusted series are seasonal.
- The reductions in variances and their differences across months for seasonally adjusted series show that this effort is worthwhile.
- A comparison of confidence intervals for month-to-month change for seasonally adjusted series shows little difference when compared to the Pfeffermann-Sverchkov method.
- More variance series need to be produced for highly disaggregated CPS series that are noisier to complete our evaluation.

Acknowledgements

The authors thank Michael Sverchkov for his assistance creating variances using the Pfeffermann-Sverchkov method.

Appendix

We should expect variances to generally be lower for seasonally adjusted series than unadjusted series and the following additive adjustment equation for a particular month helps demonstrate this:

$$\hat{A}_t = \hat{U}_t - \hat{S}_t$$

where $\hat{A}_t$ = *Seasonally adjusted series*, $\hat{U}_t$ =*Unadjusted series*, and $\hat{S}_t$ = *Seasonal component*.

By taking the variance of the adjusted series,

$$V\left\{\hat{A}_t\right\} = \sigma_{AA} = \sigma_{UU} - 2\sigma_{US} + \sigma_{SS} = \sigma_{UU} + \sigma_{SS}\left(1 - 2\frac{\sigma_{US}}{\sigma_{SS}}\right)$$

$$\sigma_{AA} - \sigma_{UU} \;<=>\; 0 \quad where$$

$$\left(1 - 2\frac{\sigma_{US}}{\sigma_{SS}}\right) <=> 0$$

$$\Rightarrow \quad 1 \;<=>\; 2\frac{\sigma_{US}}{\sigma_{SS}}$$

$$\Rightarrow \quad \frac{1}{2} \;<=>\; \frac{\sigma_{US}}{\sigma_{SS}}$$

where

$$V\left(\hat{A}_t\right) = \sigma_{AA},\; V\left(\hat{U}_t\right) = \sigma_{UU},\; C\left(\hat{U}_t, \hat{S}_t\right) = \sigma_{US},\; V\left(\hat{S}_t\right) = \sigma_{SS},$$

and the operator <=> represents the respective comparison statement less than, equal to, or greater than. All of these variance and covariance parameters are time varying, but we dropped the explicit reference to time for notational convenience.

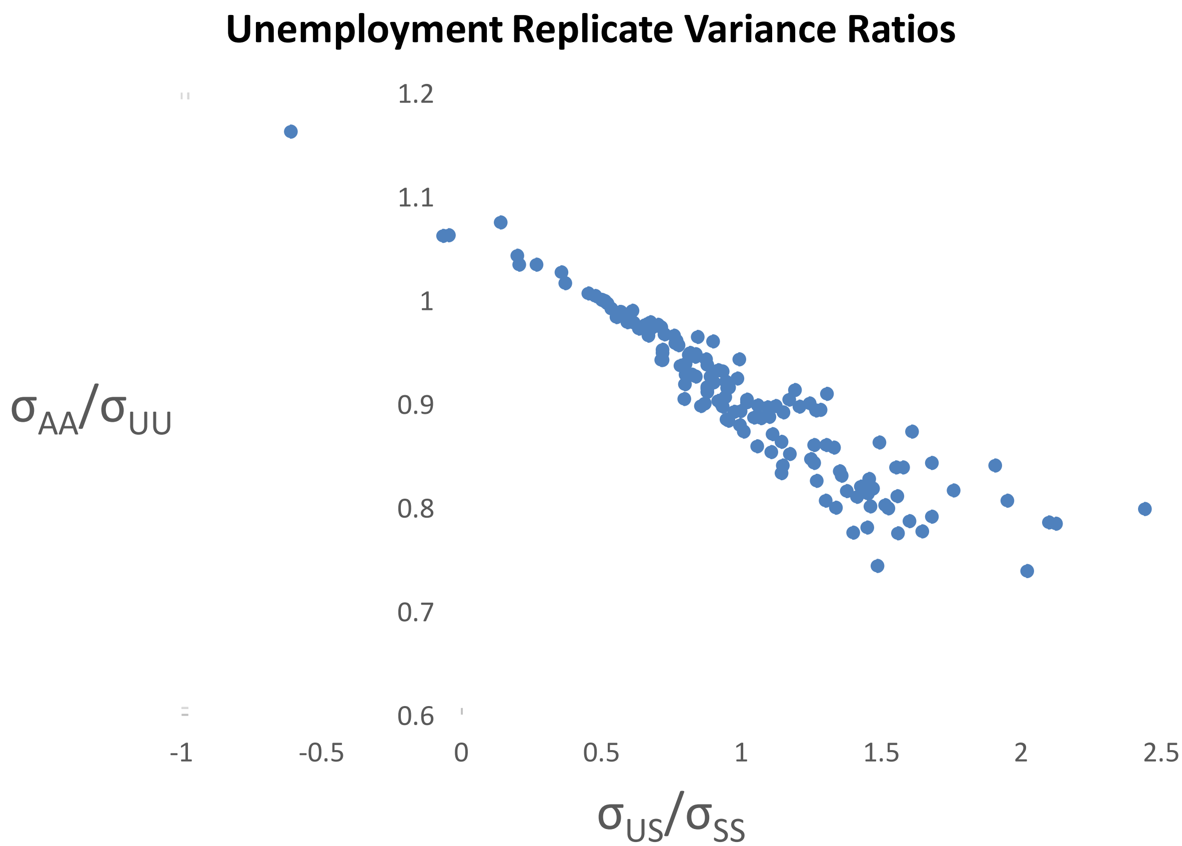
From the above result,

$$\sigma_{AA} > \sigma_{UU} \quad if \quad \frac{1}{2} > \frac{\sigma_{US}}{\sigma_{SS}}$$

$$\sigma_{AA} < \sigma_{UU} \quad if \quad \frac{1}{2} < \frac{\sigma_{US}}{\sigma_{SS}}.$$

The term $\frac{\sigma_{US}}{\sigma_{SS}}$ can be thought as similar to a regression coefficient β if one were to model $U_t = \beta S_t + e_t$. Generally, if the variation is mostly seasonal, then the term $\frac{\sigma_{US}}{\sigma_{SS}}$ will be close to 1. Conversely, if the series is not seasonal or if the model fit is poor, then the term will be close to zero.

The below plot of replicate variance ratios shows this to be the case with unemployment.



Unemployment Replicate Variance Ratios
1.2
1.1
1
0.9
0.8
0.7
0.6
σAA/σUU
-1
-0.5
0
0.5
1
1.5
2
2.5
σUS/σSS

Figure 1: Employment Variances

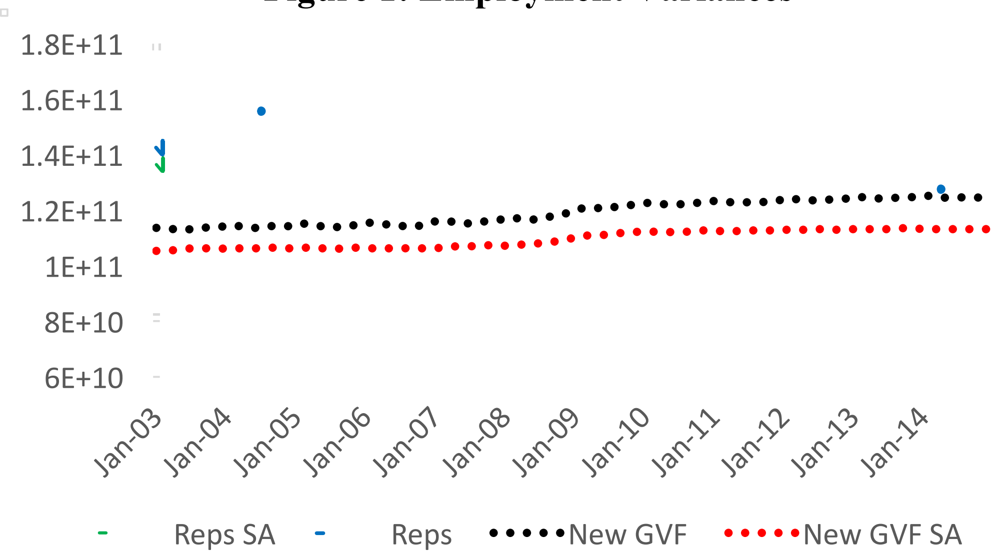


Figure 2: Employment Month-to-Month Change Standard Errors

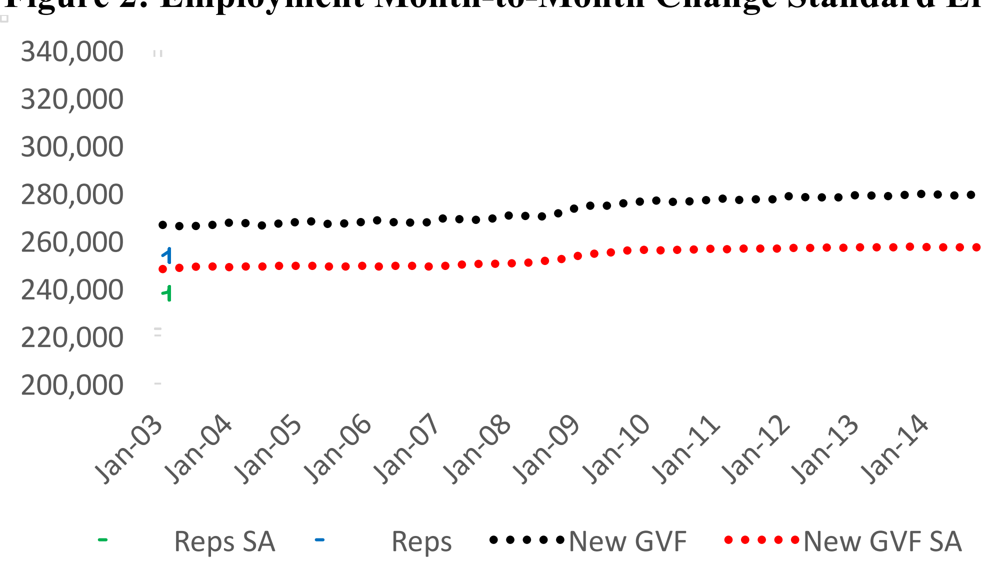


Figure 3: Unemployment Variances

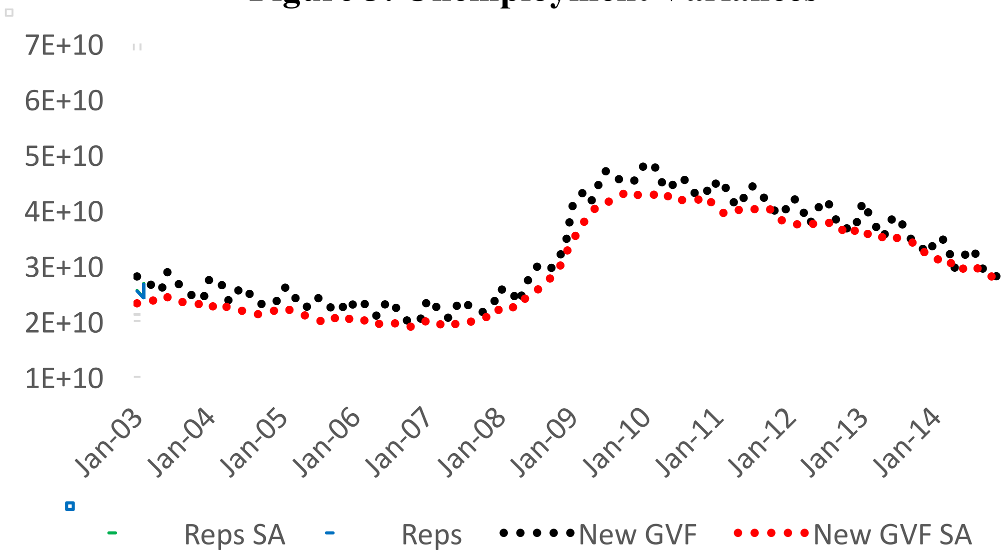


Figure 4: Unemployment Month-to-Month Change Standard Errors

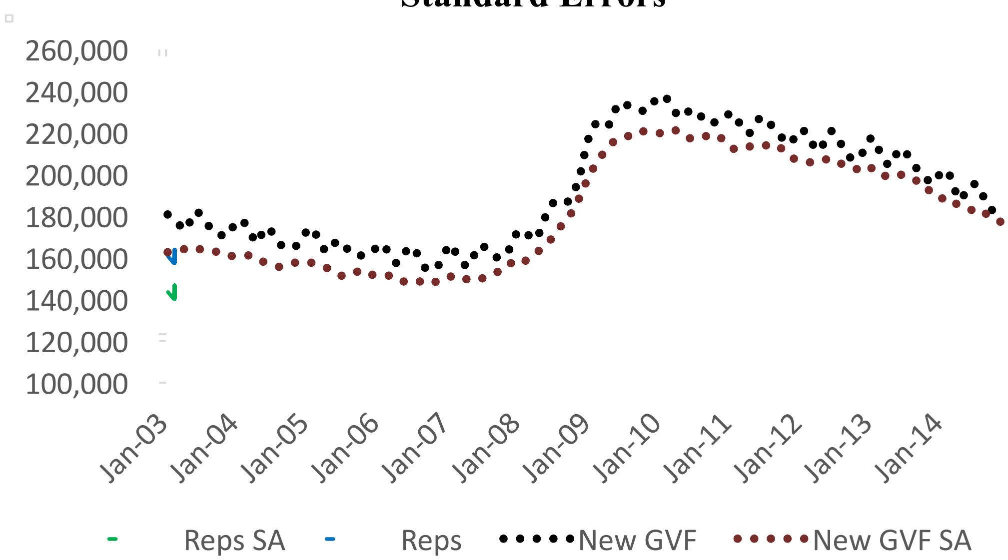


Figure 5: Unemployment Rate Variances

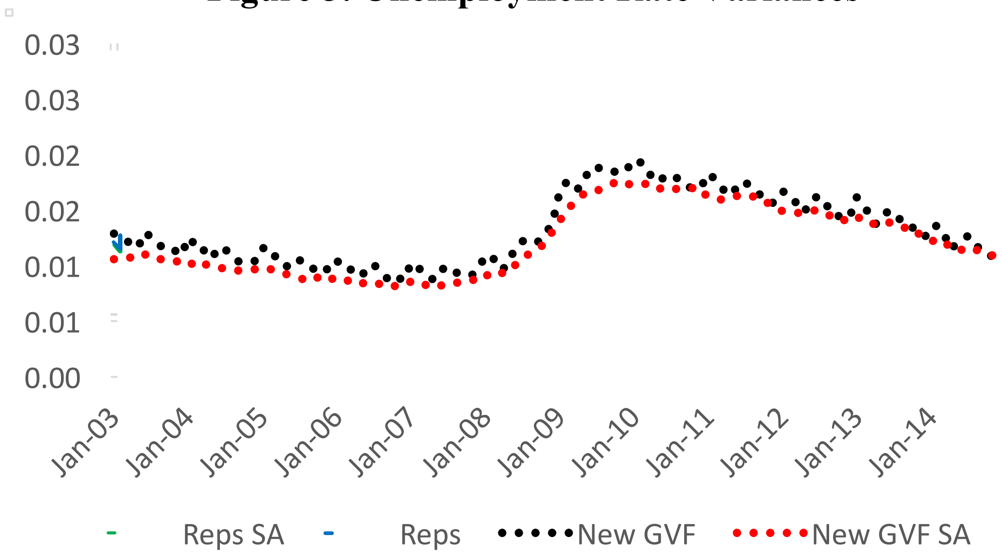


Figure 6: Unemployment Rate Month-to-Month Change Standard Errors

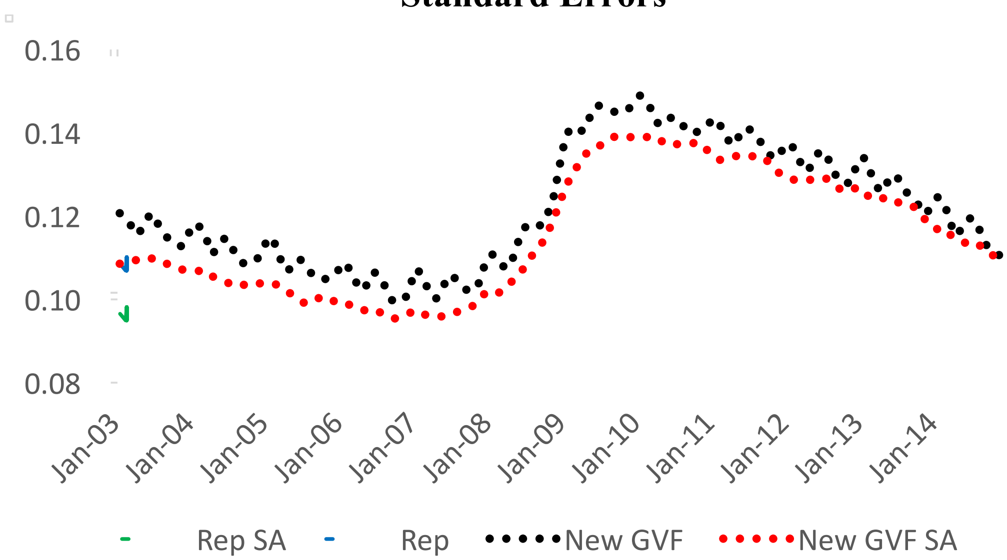
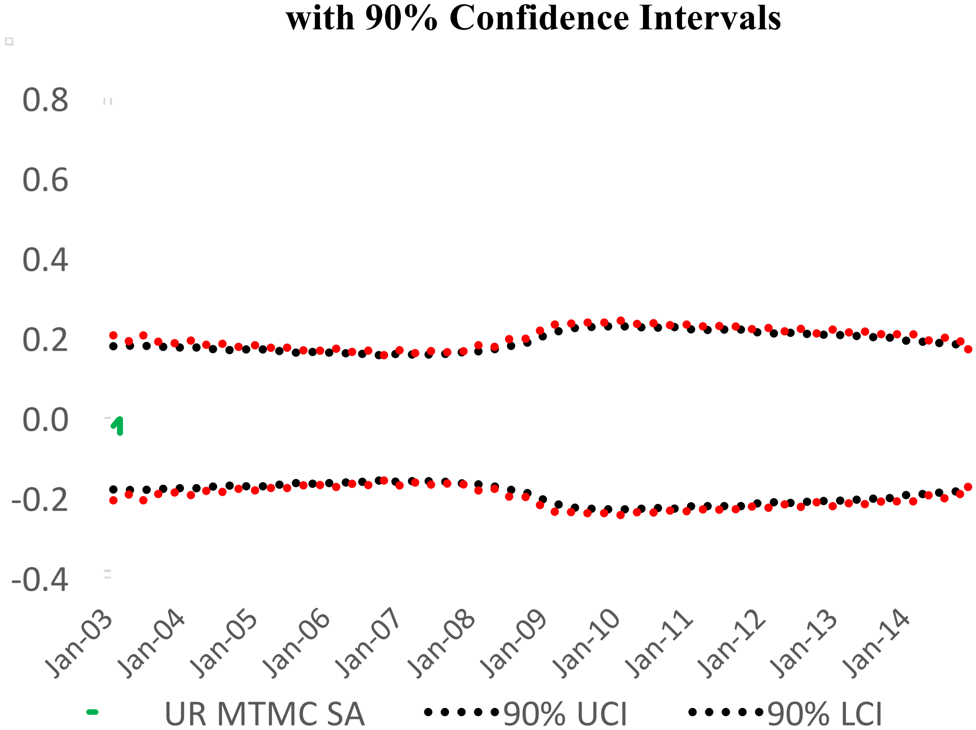


Table 1: Average Percent Reduction in Month-to-Month Change for GVF Standard Errors from Unadjusted to Seasonally Adjusted			
Month	**EM**	**UN**	**UR**
Jan	8.4	14.2	14.3
Feb	8.3	13.2	13.2
Mar	8.2	10.5	10.5
Apr	8.0	3.2	3.5
May	7.9	4.9	4.8
Jun	7.6	12.2	11.1
Jul	7.6	13.2	12.0
Aug	7.8	9.3	8.7
Sep	7.9	5.2	5.0
Oct	7.8	3.6	3.3
Nov	7.9	4.2	3.9
Dec	8.1	4.5	4.5
Average	**8.0**	**8.2**	**7.9**

Figure 7: Comparison with Pfeffermann-Sverchkov Seasonally Adjusted Month-to-Month Change with 90% Confidence Intervals



References

Bell, W.R., and M. Kramer (1999), "Toward Variances for X-11 Seasonal Adjustments," *Survey Methodology*, 25:1, 13-29.

Pfeffermann, D. (1994), "A General Method for Estimating the Variances of X-11 Seasonally Adjusted Estimators," *Journal of Time Series Analysis*, 15, 85-116.

Pfeffermann, D., and M. Sverchkov (2014), "Estimation of Mean Squared Error of X-11-ARIMA and Other Estimators of Time Series Components," *Journal of Official Statistics*, v. 30:4, 809-836.

Scott, S.M., Sverchkov, and D. Pfeffermann (2012), "Estimating Variance in X-11 Seasonal Adjustment," in *Economic Time Series: Modeling and Seasonality*, edited by W.R. Bell, S.H. Holan, and T.S. McElroy, CRC Press, Boca Raton, FL.

Tiller, R.B. (2012), "Frequency Domain Analysis of Seasonal Adjustment Filters Applied to Periodic Labor Force Survey Series," in *Economic Time Series: Modeling and Seasonality*, edited by W.R. Bell, S.H. Holan, and T.S. McElroy, CRC Press, Boca Raton, FL.

Tiller, R.B., and T.D. Evans (2015), "Methodology for Seasonally Adjusting National Household Survey Labor Force Series with Revisions for 2015," available online at http://www.bls.gov/cps/cpsrs2015.pdf.

U.S. Bureau of Labor Statistics and U.S. Census Bureau, *Design and Methodology Current Population Survey Technical Paper 66* (2006). Washington, DC: Authors. Available online at http://www.census.gov/prod/2006pubs/tp-66.pdf.

U.S. Census Bureau (2015*), X-13ARIMA-SEATS Reference Manual* (Version 1.1), Washington, DC: Author. Available online at http://www.census.gov/ts/x13as/docX13AS.pdf.

Wolter, K.M., and N.J. Monsour (1981), "On the Problem of Variance Estimation for a Deseasonalized Series," in *Current Topics in Survey Sampling*, ed. D. Krewski, R. Platek, and J.N.K. Rao, Academic Press, New York, 367-403.