Testing Models for Weight Smoothing in the Current Employment Statistics Survey November 2017

Christopher Grieves, Julie Gershunskaya

Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC, 20212

Abstract

Survey estimates may be susceptible to the influence of sample units having large design weights associated with unusual observed values. Especially in smaller samples, these sample units can influence estimates disproportionately causing them to be very unstable. In this paper, we consider several model-based approaches for weight smoothing where the design weights are modeled as a function of observed survey quantities. Using these modeled weights, one hopes to reduce volatility in the weights, thus producing better estimates. In this paper we extend prior work on the Current Employment Statistics Survey (CES). Several prospective models are used for the weights, including LOESS curves and Bayesian methods. The new "smoothed" weights are then used to create new survey estimates and we compare these estimates to the true value. Analysis of the fitted weights is performed in the end to find cases where "smoothed" weights may give worse estimates.

Key Words: sampling weights, extreme observation, small area, Bayesian

1. Introduction

Weight smoothing has been proposed as a method of adjusting survey weights to reduce variance of survey estimates. In the classical design-based framework, sample weights are viewed as fixed non-random quantities reflecting unequal sample inclusion probabilities as well as possible adjustments usually related to survey nonresponse or frame deficiency. It is well known that survey estimates may be inefficient if the design weight is not related to the variable of interest or if this relationship is not strong. Alternatively, under the model-based approach to inferences from survey sampling, reviewed in Pfeffermann and Sverchkov (2009), or the generalized design-based inference approach of Beaumont (2008), the sample weights are viewed as realizations of a random vector. The advantage of viewing the problem in this manner is that it gives the opportunity to model the survey weights conditioning on observed sample quantities.

With regard to the CES survey, we are interested in estimating the relative over-the-month change of employment, R_t, defined as the ratio of the total employment in the current month, Y_t, to total employment in the previous month, Y_{t-1}. The variable of interest, reflecting the influence of individual measurements on the target estimate, is the "residual", $r_{j,t} = y_{j,t} - R_t y_{j,t-1}$, where $y_{j,t}$ and $y_{j,t-1}$ are the observed unweighted employment for the current and previous month of sample unit j. (For the moment, we suppose that R_t in the above residual is known.) Using these quantities, we are interested in models of the form $w_j = f(r_{j,t}) + e_{j,t}$, with w_j being the survey weight for unit *j*, f() is some function of the residuals and $e_{j,t}$ is some random error . After modeling, we use these new "smoothed weights" to create estimates of R_t, which we hope are more efficient than estimates based on the original weights. In previous work on weight smoothing for CES, Gershunskaya and Sverchkov (2014) took this approach with some success.

In this paper, we expand on the number of models and how their tuning parameters considered in the previous CES work. The models include LOESS, Penalized B-Spline models with restrictions on their tuning parameters and a Bayesian Model. We compare the results to the currently used CES estimator that employs a two tail Winsorization method.

2. The CES Survey

2.1 CES Frame and Sample Selection

The CES survey derives its frame from Quarterly Census of Employment and Wages (QCEW) program. The QCEW is an administrative program that collects employment and wage information from all establishments covered under the unemployment insurance (UI) on a quarterly basis.

From the derived frame, CES chooses a stratified simple random sample of UI accounts, that is, when a UI account is chosen all establishments under that UI account are included in the sample. Stratification is performed by state, industry supersector (a grouping of North American Industrial Classification System codes), and total employment size. Optimal allocation at a fixed constant cost per a unit is used to minimize the variance of over the month change is used to allocate a fixed state sample size to the strata.

2.2 CES Estimator

The primary estimate of interest for the CES survey is the over the month change $\boldsymbol{R_t}$. The estimator used is defined as follows:

$$\widehat{\boldsymbol{R}}_{\boldsymbol{t}} = \frac{\sum_{j \in S_t} w_j y_{j,t}}{\sum_{j \in S_t} w_j y_{j,t-1}},$$

where *j* denotes the establishments, *t* is the current month, $y_{k,t}$ and $y_{k,t-1}$ denote the employment of sample units in the current and previous months, and S_t is the "matched sample" or the set of sample units reporting positive employment in the current and previous months.

To produce monthly estimates of levels, we use the annual census value produced from the QCEW, Y_0, and apply the ratio with $\hat{Y}_{t=1} = Y_0 \hat{R}_{t=1}$ and subsequent months estimated as $\widehat{Y_t} = \hat{Y}_{t-1} \hat{R}_t$. For more details see the BLS Handbook of Methods.

2.3 Challenges of Estimation

As we described above, the optimal allocation used at the CES survey sample design stage is aimed to minimize the variance of the over the month change estimate. Ideally, such allocation strategy should produce "optimal" weights for the efficient survey weighted estimator. However, in the realized sample, a sample unit with a large weight may grow at a rate much faster than expected during the design stage; as a result, it may not necessarily represent other units in the population to the degree its sampling weight might suggest. This large change in employment in conjunction with the large weight overly influences the ratio estimate. We can see this in the first order Taylor expansion on the

$$\hat{R}_t \approx R_t + \frac{1}{Y_{t-1}} \sum_{j \in S_t} w_j \left(y_{j,t} - R_t y_{j,t-1} \right),$$

where Y_{t-1} is total employment in month *t-1* and R_t is the change of employment from month *t-1* to *t* and w_j is the design weight. Supposing that there is a large change in employment, in the Taylor expansion a single unit can influence the ratio estimate by shifting it $\frac{w_j \left(y_{j,t} - R_t y_{j,t-1} \right)}{Y_{t-1}}$ units, which can be disproportionate in some cases causing large end of the year revisions or highly variable over the month changes.

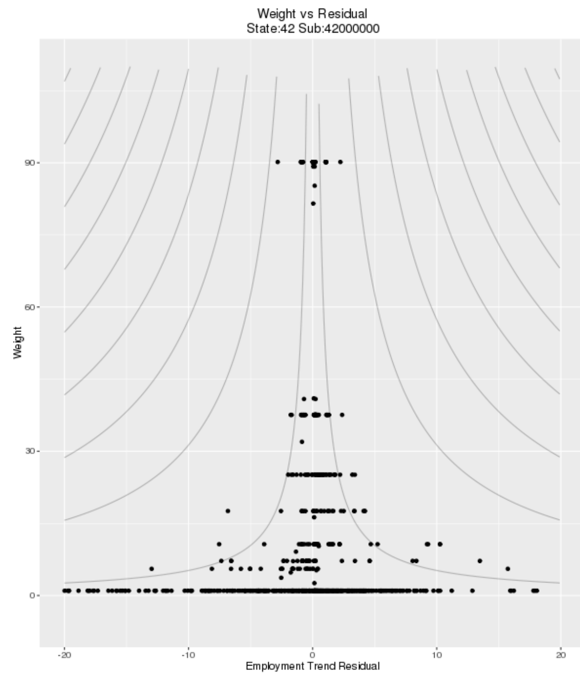
The current solution on the CES survey is a form of Winsorization where the cut offs are determined by a method devised by Kokic and Bell (1994) and adapted to CES in Gershunskaya and Huff (2004), Gershunskaya (2011). Weights are then either censored to the cutoff values or in more extreme cases removed from the ratio altogether.

3. Weight Smoothing

In this section, we begin to consider weight smoothing as a solution to the challenges we presented in the previous section. We start by considering the survey weights as random rather than fixed quantities. We may then model the weights as a function of some response variable, $w_j = f\left(r_{j,t}\right) + e_{j,t}$, where *f* is some function we fit, $r_{j,t}$ is the residual from the previous section, and $e_{j,t}$ is some error term. Our new smoothed weights would then be $v_j = f\left(r_{j,t}\right)$. The hope is that we can produce this new set of "smoothed" weights with reduced variation in the weights and better aligned with the survey response to give us increased efficiency in our estimates. Theoretical justifications and empirical evidence for this approach have been presented in Beaumont (2008).

3.1 Application to CES

In our application of weight smoothing to the CES survey, we consider modeling the design weights conditioned on our "residuals", $r_{j,t} = y_{j,t} - R_t y_{j,t-1}$. We use an estimate of R_t since we do not observe the true value. A typical scatter plot of these two variables is presented below. Some observations to make is that the variance of residuals tends to decrease as weight increases, some amount of skewness in the scatter plot, and some observations have a relatively large change in employment given its weight.



For CES, the weight smoothing approach was first considered by Gershunskaya and Sverchkov (2014). Their work considered using non-parametric LOESS models fit to the data with encouraging results; the new smoothed weights produced estimates with typically lower revisions and tracked the "true" changes in employment closer. We use the values from the QCEW as a proxy for truth.

3.2 Models

Below is a survey of the models we used in our application of weight smoothing and a brief description.

3.2.1 LOESS

LOESS, or locally weighted regressions (Cleveland 1979), is a non-parametric model that creates regressions at each point using q nearest neighbors. The regressions are weighted as a function of the distances from that point to its q nearest neighbors. To fit the model, a few tuning parameters must be chosen. In general a "smoothing" parameter $s \in (0,1]$ must be chosen. s is the percentage of data to be used in each regression.

As stated in the previous section, these models were first considered in the original CES work performed by Gershunskaya and Sverchkov (2014). The models were fit in SAS using their automatic parameter selection technique.

We will consider some restrictions on the smoothing parameter. The SAS Proc LOESS procedure gives the user the option to set an upper and lower bound on the potential smoothing parameter. SAS will perform its model selection based on the restricted domain of smoothing parameters, choosing the one that minimizes some criteria. Please see the SAS website for more details

3.2.2 Penalized B-Splines

Penalized B-splines (Eilers and Marx 1996) are another non-parametric model we tested in our weight smoothing application. B-splines are piecewise polynomials connected at x values called knots. We demand that adjacent polynomials be continuous at knots. Furthermore with Penalized B-Splines we add a penalty on the estimates of coefficients to not over fit the data. This penalty is controlled by the tuning parameter λ, with larger values of λ producing smoother curves.

We fit the splines to the data using the SAS Transreg procedure letting SAS choose the smoothing parameter with its automatic selection procedure. We consider different upper bounds on the smoothing parameter λ in our research.

3.2.3 Bayesian Model

We finally consider one Bayesian model to condition our weights on residuals using the typical Bayesian formulation

$$P(\omega_j|r_j) \propto P(r_j|\omega_j)P(\omega_j)$$

Where $P(\omega_j)$ is our prior distribution holding our beliefs about the design weight of unit j. $P(r_j|\omega_j)$ is our likelihood function for the residuals deciding how likely a residual is given our weight. We placed a truncated normal prior on our weights with mean of the design weight.

$$\omega_j \sim N(w_j, \tau) \ \ \omega_j \epsilon [1,100]$$

Where w_j is the original design weight and τ is its precision. We use precision, defined as the inverse of variance $\frac{1}{\sigma^2}$, in lieu of variance as is tradition with much Bayesian statistics. We place a prior on τ as $\tau \sim Uniform(.1,1)$. The idea is that the design weight w_j is a good initial guess and this information should be included into the prior.

We define our likelihood on the residuals as a Student's t-distribution, once again using precision for the scale parameter.

$$r_{j,t} \sim t(\mu, \frac{{\omega_j}^{\alpha}}{\sigma}, \nu)$$

We finally attempted to fit flat priors on the other parameters μ, ν with some trial and error done to decide the parameters.

$$\mu \sim Normal(0, .001)$$
$$\nu \sim Uniform(20,100)$$
$$\sigma \sim InvGamma(.01, .01)$$
$$\alpha \sim Uniform(.01, 2.5)$$

We finally used the JAGS (Just Another Gibbs Sampler) software package to perform our MCMC (Markov chain Monte Carlo) sampling from the posterior distribution, $P(\omega_j | r_j)$, for each of the weights. From the samples we took the mean for each weight as our new smoothed weight.

4. Results

4.1 Evaluation Criteria

Estimates were made for publication cells at the State, MSA, and Industry Classification for the 2011,2012 and 2013 benchmark years. There were 2351 publication cells for 2011 and 2012 and there were 2199 publication cells for the 2013 benchmark year.

We fit models monthly by state and industry classification. We did not fit at MSA level due to small sample sizes. Using the smoothed weights from the models, we produced CES estimates as described in section 2.2. To evaluate the performance of our estimators we used the following criteria. We use the QCEW as our proxy for truth in evaluation. We denote $\hat{Y}_{\mathrm{i},12}$ as our estimator at month 12 for publication cell *i* and $Y_{\mathrm{i},12}$ as the QCEW value at month 12 for publication cell *i*. First we consider the end of the year benchmark revision.

$$\boldsymbol{rev_i} = \left(\hat{Y}_{\mathrm{i},12} - Y_{\mathrm{i},12}\right)$$

We also consider the relative benchmark revisions for publication

$$\boldsymbol{relRev_i} = \frac{\left(\hat{Y}_{\mathrm{i},12} - Y_{\mathrm{i},12}\right)}{Y_{\mathrm{i},12}}$$

Denoting $\hat{Y}_{i,t}$ as our estimator of month *t* for publication cell *i* and $Y_{i,t}$ as the QCEW value at month *t* for publication cell *i*, we consider the average absolute difference in over the month change to see how well the estimator matched changes in the QCEW.

$$\boldsymbol{c_i} = \frac{\mathbf{1}}{\mathbf{12}} \sum_{t=1}^{12} \left| (\hat{Y}_{i,t} - \hat{Y}_{i,t-1}) - (Y_{i,t} - Y_{i,t-1}) \right|$$

We calculate these criteria for all publication cells and present summary statistics of them in the next section.

4.2 Model Results

We tried different models and restrictions on their tuning parameters, however for the sake of brevity, we only include selected models that performed well. There also did not appear to be a significant improvement by using one set of tuning parameters versus another, Most differences were between types of models. We did note worse results for 2011 and 2013 for LOESS when restricting the lower bound of the smoothing parameter.

In the tables below we label the LOESS model as **LOESS (*LB, UB*)** where *LB* is the lower bound on the tuning parameter and *UB* is the upper bound on the tuning parameter. We label the Penalized B-Spline as **Spline Lambda < *UB*** where *UB* is the upper bound of the smoothing parameter. No other variations on the Bayes model were attempted due to time constraints, we simply label it **Bayes**.

Table 1 below contains summary statistics of benchmark revisions. One can see that in most cases the absolute mean of the LOESS model is the smallest.

Table 1: Benchmark Revision Summary Statistics for Chosen Models

2011 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-27140	-863	-101	-161	601	16970
LOESS (0, 0.8)	-17770	-649	-14	2	558	35210
Spline Lambda < 1000	-19220	-650	-32	-96	462	23570
Bayes	-21890	-897	-177	-368	357	11020
2012 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-48470	-1008	-109	-340	592	25500
LOESS (0, 0.8)	-50200	-950	-128	-480	491	33540
Spline Lambda < 1000	-50160	-952	-165	-589	404	27110
Bayes	-49540	-932	-179	-462	381	24610
2013 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-15490	-816	-81	-111	650	14270
LOESS (0, 0.8)	-20820	-599	10	11	622	64350
Spline Lambda < 1000	-18000	-640	-32	-138	499	21030
Bayes	-14080	-776	-121	-283	386	10090

On the next page we include the summary statistics for Relative Benchmark Revision and Average Absolute Difference of Over the Month Change Summary Statistics. …

Table 2: Relative Benchmark Revisions Summary Statistics for Chosen Models

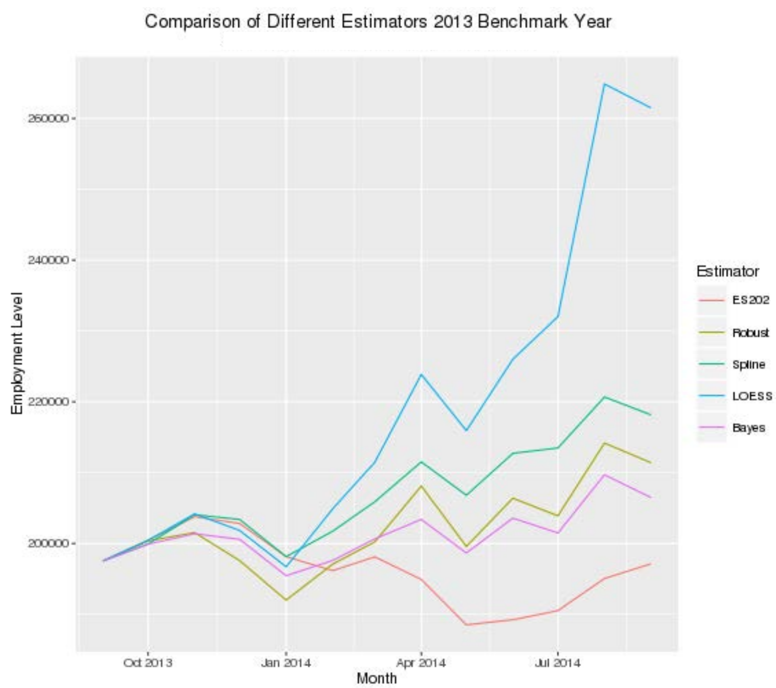
2011 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-38.97%	-3.47%	-0.45%	-0.36%	2.50%	96.97%
LOESS (0, 0.8)	-43.73%	-2.58%	-0.10%	0.08%	2.33%	208.50%
Spline Lambda < 1000	-43.73%	-2.53%	-0.16%	-0.09%	2.01%	192.50%
Bayes	-35.75%	-3.25%	-0.82%	-0.69%	1.61%	73.25%
2012 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-65.14%	-3.77%	-0.58%	-0.49%	2.40%	121.90%
LOESS (0, 0.8)	-64.75%	-3.58%	-0.60%	-0.61%	2.31%	114.60%
Spline Lambda < 1000	-64.23%	-3.58%	-0.80%	-0.88%	1.90%	108.00%
Bayes	-63.97%	-3.40%	-0.82%	-0.67%	1.58%	133.90%
2013 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	-48.58%	-2.92%	-0.44%	-0.27%	2.39%	46.97%
LOESS (0, 0.8)	-48.24%	-2.07%	0.05%	0.19%	2.37%	48.37%
Spline Lambda < 1000	-47.84%	-2.24%	-0.17%	-0.08%	1.96%	49.92%
Bayes	-48.14%	-2.67%	-0.55%	-0.49%	1.65%	37.56%

Table 3: Average Absolute Difference of Over the Month Change Summary Statistics for Chosen Models

2011 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	7	183	341	454	574	8930
LOESS (0, 0.8)	7	149	264	396	472	18040
Spline Lambda < 1000	6	151	282	479	542	7457
Bayes	8	161	272	377	462	4474
2012 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	6	206	389	653	725	13920
LOESS (0, 0.8)	5	162	309	601	641	13710
Spline Lambda < 1000	4	150	294	581	629	13540
Bayes	6	173	322	594	636	13750
2013 Benchmark Year						
	Min	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Max**
Robust Estimator	5	200	378	566	666	8353
LOESS (0, 0.8)	5	160	300	504	570	7507
Spline Lambda < 1000	6	151	282	479	542	7457
Bayes	5	174	315	497	564	7945

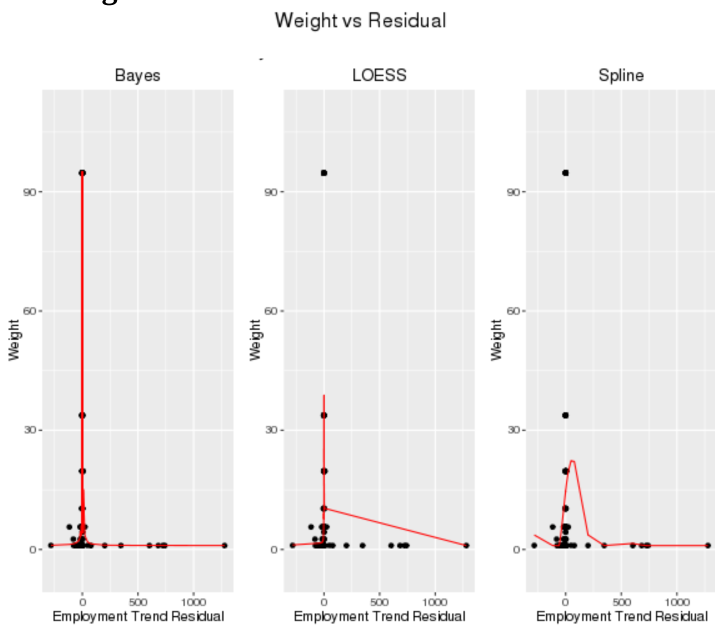
We will briefly look at some of the extreme cases to see where improvements could be made. Below is a monthly time series of all our estimates including the QCEW labeled as ES202 in graph below. We can see that the LOESS, the blue line in Figure 1, model tended to overestimate over the month change, with one extreme happening in August.

Figure 1: Bad LOESS Estimate



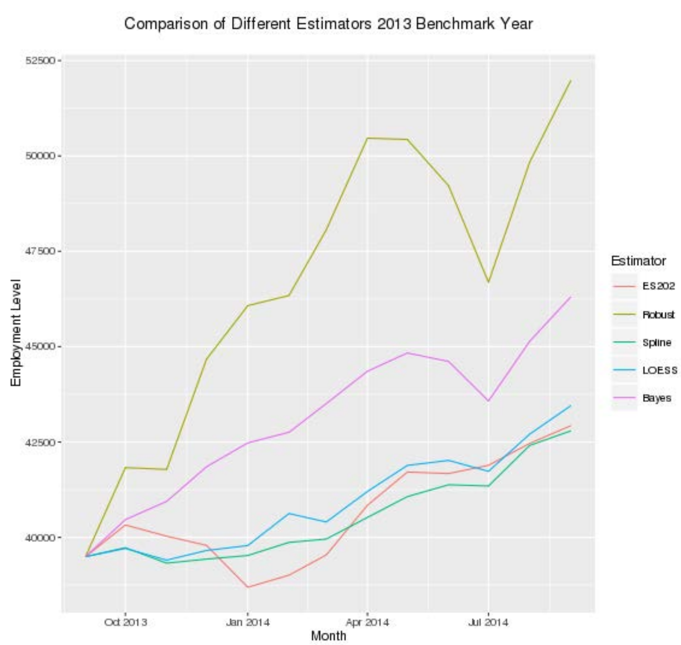
To understand why this occurred, we plotted the model fit of weight vs residual for each of our three models below. The **black** dots are the design weights and the red line is our model fit. One should note that the LOESS model up-weighted large positive residuals in this model fit causing over estimation of the ratio. This is a model problem that would need to be remedied in some manner either through weight trimming or analyst review.

Figure 2: Model Fits for Bad LOESS Model



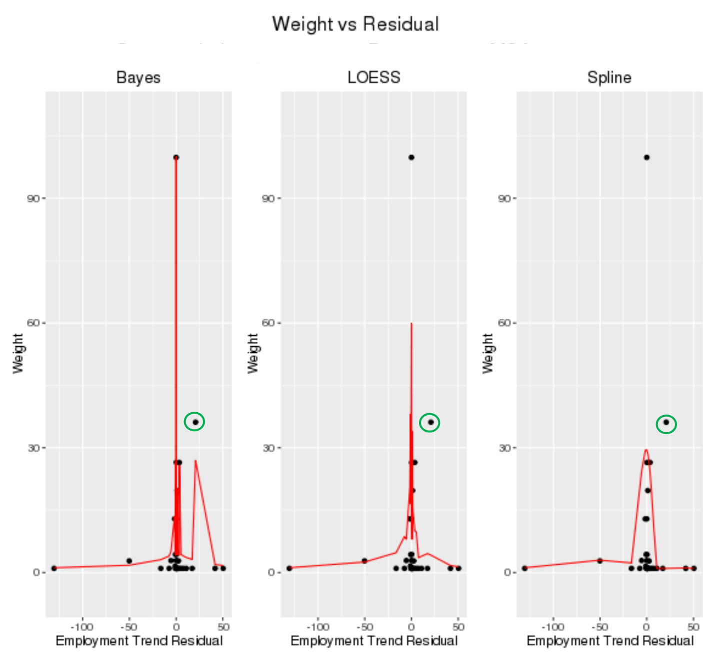
The next example is one where the Bayes estimator fails to perform well. We will look at the change between October to November due to it estimating a gain in employment rather than a drop. You can see this in the time series below.

Figure 3: Time Series of Estimators, Bad Bayes Estimate



We can once again look at the graph of the model fit. We can see the one outlier circled in green that failed to be down weighted by the Bayes model, unlike the other models. This caused the positive jump in employment rather than decline reported by the other models.

Figure 4: Model Fits for Bad Bayes Model



5. Conclusions

We make the following observations and conclusions about our results:

- Fitting a model to our weights generally reduced benchmark revisions, reduced the variance of revisions across estimation cells, and better estimated over-the-month change than the robust estimator.
- We consider the LOESS model the best due to it outperforming the robust estimator in most evaluation statistics.
- Penalized B-Splines and the Bayes Model generally under estimated the true employment value though still reducing the variance of the revisions across estimation cells.
- The Bayesian Model mitigated large outliers the best and reduced variance of estimates the best. However a large consistent negative bias and large computation time make the model impractical.

- Idiosyncrasies of the model fits can cause large outliers and would need to be mitigated through some process as seen in the time series of estimates.

5. Future Work

Although current results look promising, some work still remains to be done before one considers using a process such as this in production. This work includes:

- Current work is being done on calculating the variance of estimates from weight smoothed weights.
- Is there any benefit to an MSA level “effect”? If an MSA is different than the others the modeled weights may give poor results for that MSA.
- Can we check to see how well two models agree on their estimates of weights to detect poor fits? As in the case of the bad LOESS model we presented, perhaps this problem can be caught by checking against the Spline model in some fashion.
- Is there any benefit to further trimming weights or removing them from the ratio either before or after model fitting?

References

Beaumont, J.-F. (2008). A new approach to weighting and inference in sample surveys. *Biometrika*, 95, 3, *pp*. 539–553

Beaumont, J.-F., and Rivest, L.-P (2009). Dealing with outliers in survey data. Chapter 11 in D. Pfeffermann and C. R. Rao (eds.), *Handbook of Statistics. No. 29A, Sample Surveys: Inference and Analysis*, pp. 247-280

Bureau of Labor Statistics (2011). Employment, hours, and earnings from the establishment survey. Chapter 2 of BLS Handbook of Methods, U.S. Department of Labor, http://www.bls.gov/opub/hom/pdf/homch2.pdf

Eilers, Paul H. C.; Marx, Brian D. Flexible smoothing with B -splines and penalties. Statist. Sci. 11 (1996), no. 2, 89--121. doi:10.1214/ss/1038425655. http://projecteuclid.org/euclid.ss/1038425655.

Gershunskaya, J. (2011) Treatment of influential observations in the Current Employment Statistics survey. Unpublished doctoral dissertation, University of Maryland, College Park.

Gershunskaya, J. and Sverchkov, M. (2014) On Weight Smoothing in the Current Employment Statistics Survey. JSM 2014 – Survey Research Methods Section.

Kokic, P. N., and Bell, P. A. (1994), “Optimal Winsorizing Cutoffs for a Stratified Finite Population Estimator,” *Journal of Official Statistics*, 10, 419-435.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data, *Sankhya* B, 61, Pt.1, 166 – 186

Pfeffermann, D. and Sverchkov, M. (2003). Fitting generalized linear models under informative sampling. Chapter 12 in R. L. Chambers and C. Skinner (eds.), *Analysis of Survey Data,* Chichester: Wiley , pp. 175 – 195

Pfeffermann, D. and Sverchkov, M. (2009). Inference under informative sampling. Chapter 39 in D. Pfeffermann and C. R. Rao (eds.), *Handbook of Statistics. No. 29B, Sample Surveys: Inference and Analysis*, pp. 455-487