

VARIANCE MEASURES FOR SEASONALLY ADJUSTED EMPLOYMENT AND EMPLOYMENT CHANGE

Stuart Scott and Michail Sverchkov, Bureau of Labor Statistics;
Danny Pfeffermann, Hebrew University & University of Southampton
2 Massachusetts Ave, NE, rm 1950, Washington, DC 20212, scott.stuart@bls.gov

1. Introduction

Some of the most closely-watched U.S. economic statistics are monthly industry employment from the Bureau of Labor Statistics (BLS) survey of business establishments. This article uses the method developed by Pfeffermann (1994) to obtain a variance measure for monthly seasonally adjusted change, typically the focus of interest for employment and other key economic indicators. The method applies to seasonal adjustment with X-11 and its extensions (Ladiray & Quenneville, 2000), used at BLS for most of its series. This application is geared to an index-style estimator, and appears to be applicable in common index number settings. Consideration of sampling error is an important part of the proposed method.

Due to the large sample size of the survey and the high correlation between establishment employment in adjacent months, our results show that the variance measure can identify very small changes as significant. Indeed, application of the measure finds significant employment declines for Manufacturing Durable Goods (MFGD) in 9 of 12 months during 2003.

The next section presents our methodology. Section 3 presents results for MFGD and four other major industries, and the final section provides a summary.

2. Description of proposed method

2.1 Pfeffermann's method

Let y_t , $t=1, \dots, N$, denote the survey estimates of a time series, assumed to have an additive decomposition of the form

$$y_t = Y_t + \varepsilon_t = (L_t + S_t + I_t) + \varepsilon_t = (L_t + S_t) + e_t, \quad (1)$$

where $Y_t = (L_t + S_t + I_t)$ is the population value, L_t the trend level, S_t the seasonal effect, I_t the irregular component, and $\varepsilon_t = (y_t - Y_t)$ the sampling error. The seasonally adjusted estimator is $\hat{A}_t = (y_t - \hat{S}_t)$ where \hat{S}_t is the X-11 estimate of the seasonal component. The series $\{I_t\}$ and $\{\varepsilon_t\}$, and hence the combined error $\{e_t\} = \{I_t + \varepsilon_t\}$, are assumed to be stationary with zero mean. For their autocovariances, we write

$$v_k = Cov(I_t, I_{t+k})$$

$$\lambda_k = Cov(\varepsilon_t, \varepsilon_{t+k})$$

$$V_k = Cov(e_t, e_{t+k}).$$

The form of the sampling error autocovariances is typically a function of the survey design.

The basic measure of interest is the variance of the error in estimating a seasonally adjusted value, for which Pfeffermann (1994) develops the following approximation:

$$Var_c(\hat{A}_t - A_t) = Var_c[\hat{S}_t - S_t] - \varepsilon_t \approx$$

$$Var_c\left(\sum_{k=1}^N \tilde{w}_{tk} e_k\right) + \lambda_0(1 - 2\tilde{w}_{tt}) - 2\sum_{k \neq t} \tilde{w}_{tk} \lambda_k, \quad (2)$$

where the subscript c signifies the joint distribution of the combined error terms $\{e_t\}$ and the weights \tilde{w}_{tk} define the linear approximation to \hat{S}_t . When X-11 is applied with ARIMA extrapolation, the weights are also a function of the ARIMA model parameters.

The use of (2) requires estimation of the vectors λ and V of autocovariances of the sampling error and the combined error. Let R_t be the X-11 residual, that is, X-11's estimate of the irregular component, and write its linear approximation as

$$R_t = \sum_{k=1}^N a_{tk} y_k = \sum_{k=1}^N a_{tk} M_{t+k} + \sum_{k=1}^N a_{tk} e_k, \quad (3)$$

where $M_t = L_t + S_t$. The first term on the right hand side is ordinarily close to zero for all t , assuming that the X-11 estimators of M_t are unbiased (cf. Pfeffermann (1994) and Pfeffermann & Scott, 1997). With this assumption, the approximation $R_t \approx \sum_{k=1}^N a_{tk} e_k$ holds throughout the span of the series. Notice, however, that the X-11 residual series is nonstationary because of the use of asymmetric, time-dependent weights near the ends of the series. Taking autocovariances in (3), we obtain an expression for $U_{tm} = Cov(R_t, R_{t+m})$ in terms of the V_k 's. Estimating U_{tm} by $R_t R_{t+m}$ and averaging over t leads to a linear system for estimating V of the form

$$\hat{U} = D\hat{V} = D(\hat{\lambda} + \hat{v}).$$

When sampling error autocovariance estimates $\hat{\lambda}$ are available externally, we solve the system

$$\hat{U} - D\hat{\lambda} = D\hat{v} \quad (4)$$

for \hat{v} . See Pfeffermann & Scott (1997) for more details. The true irregular component usually follows

a low-order MA(q) model (and is possibly white noise), implying that (4) is a low-order system.

2.2 Variance measures for employment

Industry employment statistics come from BLS's Current Employment Statistics (CES) program, a monthly survey of over 300,000 establishments. As described in Morisi (2003), in recent years this large survey has become a probability survey with industry coding switched to the North American Industrial Classification System (NAICS). With these changes in place, variance and covariance estimates for the unadjusted series are computed monthly using the balanced repeated replication (BRR) method. The survey has the further advantage of having an annual population figure from an external source, the Unemployment Insurance program. Quarterly business tax forms collected in this program include monthly employment data which are assembled first at the state and then the national level. With a 10-month lag, these benchmark population values become available and are incorporated into estimation. An employment estimate y_t comes from a "link-relative" estimator,

$$y_t = Y_0 \cdot r_1 \cdot r_2 \cdots r_t. \quad (5)$$

Y_0 is the latest available benchmark, subsequent subscripts denote number of months away from the benchmark, and

$$r_j = \frac{\sum_{i \in M_j} w_{ij} y_{ij}}{\sum_{i \in M_j} w_{i,j-1} y_{i,j-1}}$$

is the ratio of weighted employment in months j and $j-1$, with y_{ij} representing the employment of establishment i in month j and M_j the set of units reporting in both months.

Remarks.

1. As already suggested, this estimator capitalizes on the high correlation between an establishment's employment in adjacent months.

2. The estimator is in fact a separate ratio estimator. For an industry aggregate of subindustries h

$$y_t = \sum_h y_{ht} = \sum_h Y_{h0} r_{h1} \cdots r_{ht}.$$

An analysis of variance for these separate ratios over a three-year period shows strong effects of month and a limited effect of subindustry. In this case, separate and combined ratio estimators are fairly close, a justification for treating the estimator more simply as a combined ratio estimator.

3. Current month estimates are 10 to 21 months away from the most recent benchmark. For example, for Dec 2003, the last available benchmark is Mar 2002, 21 months away; a month later, Jan 2003 data are derived using the Mar 2003 benchmark, 10 months away.

4. Each month the estimator in (5) is multiplied by another ratio, so we can expect the variance of employment level to increase, an instance of nonstationary sampling error.

Traditionally, all CES national employment series have been seasonally adjusted multiplicatively. This leads us to consider monthly change on the log scale.

$$\log(y_t) - \log(y_{t-1}) = \log \frac{y_t}{y_{t-1}} = \log(r_t).$$

This simple form looks promising for deriving a sampling error model. We may write

$$\log(y_t) - \log(y_{t-1}) = (\log(Y_t) - \log(Y_{t-1})) + \log \left(\frac{\varepsilon_t}{\varepsilon_{t-1}} \right)$$

to express monthly change in terms of a signal part and a sampling error part. If we can find an ARIMA model for the logarithm of the series with at least one regular difference, then we will have an ARIMA model for use in applying X-11 with extrapolation to $(1-B)\log(y_t)$. In the next section, we examine properties of the sampling error, and adopt a simple but reasonable model for use in computing our variance measure. Summarizing, our variance for seasonally adjusted change comes from applying the basic method to the series $\log(r_t)$.

Can we derive a variance measure for employment levels, given its nonstationary sampling error? From the basic form of the estimator,

$$\log(y_t) = \log(Y_0) + \sum_{j=1}^t \log(r_j)$$

we can write decompositions for the two components

$$\log(Y_0) = l_0^{(r)} + s_0^{(r)} + i_0^{(r)}, \quad i_0^{(r)} \sim N(\mathbf{0}, \eta^{(r)})$$

$$\sum_{j=1}^t \log(r_j) = l_t + s_t + i_t + \varepsilon_t$$

with

$$i_t \sim N(\mathbf{0}, \eta^{(r)}), \quad \varepsilon_t \sim N(\mathbf{0}, t\lambda^{(r)}).$$

We are assuming white noise (WN) processes for each irregular component and for the sampling error component of each $\log(r_j)$. Then, the combined error e_j for $\log(y_j)$ has variances and autocovariances

$$\tilde{V}_j(\mathbf{0}) = \eta^{(r)} + \eta^{(r)} + j\lambda^{(r)}$$

$$\tilde{V}_j(k) = j\lambda^{(r)}, \quad k > 0.$$

Inserting these formulas into a covariance equation and averaging across time, in place of (4) we obtain the linear system

$$U - D_1\lambda = D_2v .$$

Using estimates \hat{U} and $\hat{\lambda}$, we can solve for

$$\hat{v} = \hat{\eta}^{(v)} + \hat{\eta}^{(r)} .$$

Following Pfeffermann (1994), we use properties of the lognormal distribution to obtain a variance measure on the original scale. We may write this as

$$\hat{Var}\left(\frac{\hat{A}_j}{A_j} - 1\right),$$

the variance of percentage error in estimating the seasonally adjusted employment level.

Remarks.

1. The above formulation assumes that the nonstationary contribution to error comes from the sampling error only, but an alternative model that permits $Var(i_t)$ to grow linearly in time like the sampling error can also be accommodated.

2. At the end of the series, it will be appropriate to compute variance measures for the last 10 to 21 points, depending on the time of the last benchmark. March benchmark employment figures have no sampling error, so the method doesn't apply.

3. Results

Using data for 1994-2003, we analyze employment change for five NAICS supersectors: Construction, Manufacturing Durable Goods, Manufacturing Nondurable Goods, Wholesale Trade, and Mining. The timing for beginning data collection under NAICS varies according to industry, with the earliest switch coming in 2000 for Wholesale Trade. Early portions of the series are reconstructions. All results are for the change measure; we haven't yet tested the measure related to employment level.

3.1 Manufacturing, Durable Goods

Figure 1 contains the observed series for Manufacturing, Durable Goods (MFGD). A leveling off in the late 90's is followed by a steep decline during the 2001 recession. After accounting for three additive outliers (AO's) and a calendar effect of varying intervals between reference weeks, we find that a (111) (011) ARIMA model fits $\log(\text{MFGD})$. To apply the method developed in Sec. 2, we modify the series for these regression effects, compute log ratios, and apply X-11 seasonal adjustment with ARIMA extrapolation. Figure 2 shows the observed and seasonally adjusted log ratio series. Employment

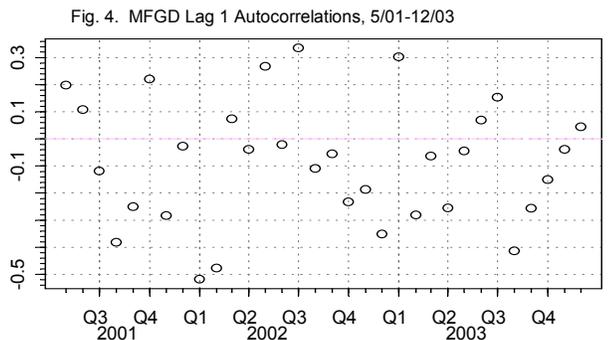
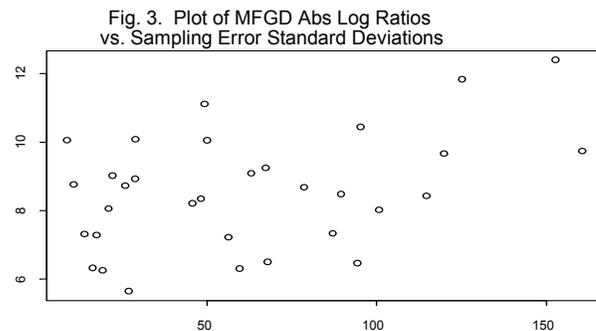
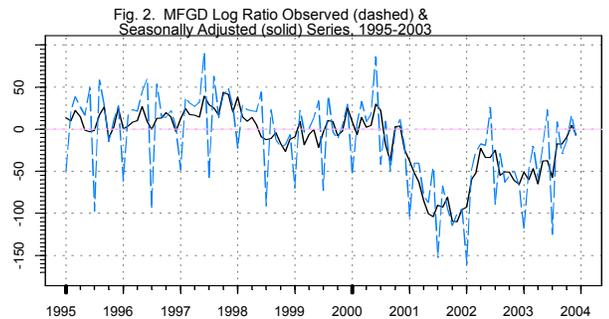


Fig. 5. MFGD Standard Deviations for Error in Seasonal Adjustment, 12 runs

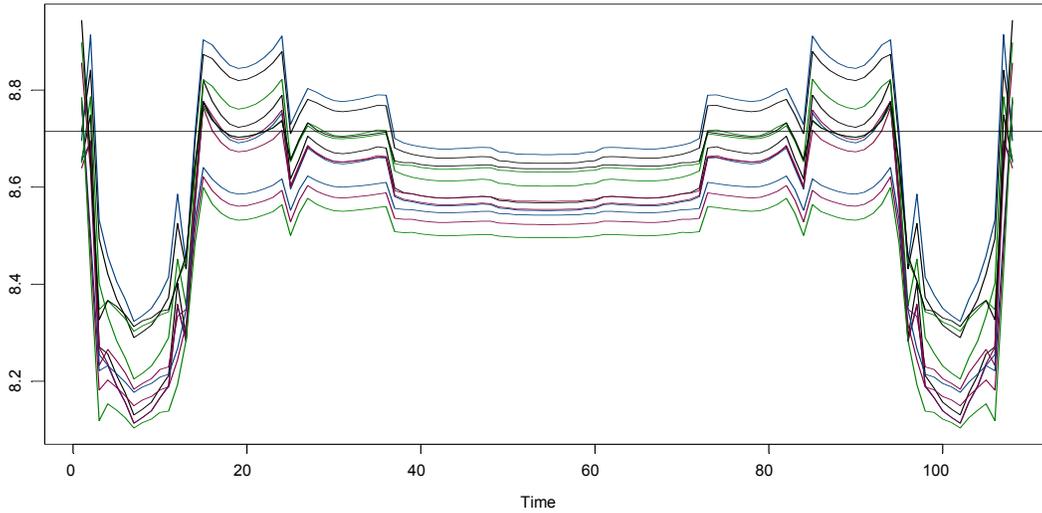


Table 1. Mean Estimates from $U - D\lambda = Dv$ for MFGD

	U	λ	v			
			q=0	q=1	q=2	q=3
Lag						
0	80.5	76.0	82.0	117.7	100.5	97.1
1	-12.4	0	0	52.8	37.2	34.1
2	-35.5	0	0	0	-16.0	-18.4
3	-7.1	0	0	0	0	-2.1

declines correspond to negative log ratios, which are lowest during the official recession period, 3/01-11/01. We see that the log ratios are highly seasonal, as we could expect from the ARIMA model.

As mentioned in Sec. 2.2, sampling error standard deviations and autocorrelations are computed each month using the BRR method. Figure 3 is a scatterplot of absolute log ratios and the sampling error standard deviations from Apr 2001 to Dec 2003. The largest two standard deviations occur for large log ratios, but, otherwise, there is very little pattern. This leads us to assume a constant variance for the sampling error. Figure 4 shows estimated lag 1 sampling error autocorrelations for May 2001 – Dec 2003. The preponderance are negative, but there is considerable variability, with values ranging from -0.52 to +.34 (cf. Table 5). Lag 12 autocorrelations are mostly positive, but again highly variable, ranging from -.32 to +.41. Means are -.09 and +.10, respectively. Autocorrelations for other lags are also variable, with means close to 0 (-.03 for lag 2). We adopt a white noise model, given the modest magnitudes and large variability in autocorrelations.

In further testing, we plan to try an MA model with nonzero coefficients for lag 1 and/or lag 12.

To test our variance measures, we make “concurrent” runs for 2003 to mimic a production setting. Based on seasonal adjustment specifications derived from the 1994-2002 span, we carry out 12 runs on 9-year spans ending in successive months of 2003, and apply the method to the results of each run.

As described in Section 2, based on estimates \hat{U}_k computed from the X-11 irregulars and $\hat{\lambda}^{(r)}$ of the sampling error variance (assumed constant), autocovariances v_k are estimated for different MA(q) models for the irregular component, q=0 to 3. Table 1 shows mean estimates from the 12 runs for the error autocovariances. We select q=2. There is good agreement between the estimates for q=2 and q=3. Subsequent autocovariances U_k are predicted well with q=2. (See Pfeffermann, 1994, for use of this diagnostic for determining the order of the MA model for the irregular component.)

Table 2. Confidence Limits for MFGD Log Ratio, 12 Runs

Month	1	2	3	4	5	6	7	8	9	10	11	12
upper limit	-29	-42	-31	-52	-25	-25	-49	-3	-7	6	23	13
lower limit	-61	-74	-64	-85	-58	-58	-82	-37	-41	-27	-10	-20
change?	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N

Table 3. Month-month % Change in Seasonally Adjusted MFGD

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-0.51	-0.60	-0.47	-0.65	-0.38	-0.37	-0.58	-0.17	-0.18	-0.10	0.04	-0.07

For all 12 runs, Figure 5 shows the SDA measure, the standard deviation for the error in estimating the seasonally adjusted log ratio. To facilitate analysis, each graph is plotted with respect to its position in the time span, 1 to 108, rather than date. (Recall each run is for a 9-year span). So, for the first run, time point 108 corresponds to Jan 03; for the last run, it corresponds to Dec 03. Overall, the shapes are similar to shapes in previous studies: symmetric for additive adjustments with constant sampling error variances, relatively flat in the middle, a major dip during the end years, with an increase toward the very end. In the middle years, there is a 2% range in values across the runs, due to variability in estimating the U_k 's. In the end years, estimates vary by 2½-3%, with extra variability coming from changing estimates of the model parameters used for extrapolation.

The horizontal line is the constant sampling error standard deviation, which represents the standard deviation of the unadjusted estimate. Our SDA measure is on average 1½% lower in the middle of the series and about 5½% lower during the end years. At the very end of the series, it actually exceeds the standard deviation of the unadjusted series by ½%. This contrasts with earlier studies where the variance for error in estimating the seasonally adjusted series with ARIMA extrapolation usually has been lower throughout. This is due to the presence of a large time series irregular. Going back to Table 1, we find that the variance of the MA(2) model for the irregular is 100.5, about one-third higher than the sampling error variance.

Next we apply our SDA estimates to test for significant change, which is of interest to analysts. For each span we form a 95% confidence interval for the true change at the last time point, namely

$$(\hat{a}_N - \hat{a}_{N-1}) \pm 2 \cdot SDA$$

Table 2 shows the confidence limits from the 12 runs. (Here, \hat{a}_N represents the seasonally adjusted log ratio

at the last time point N , and the log ratios have been multiplied by 10^4). There are significant declines for the first 9 months, but change is not significant during the last 3. Table 3 gives an extract from a single run on employment levels ending in Dec 03. It shows month-to-month per cent change during 2003. These values range from over ½% to about 1/6% during the first 9 months, deemed significant by our measure. Thus, the measure is very sensitive, since 1/6% represents about 20,000 out of 9 million.

3.2 Results across series

X-11 seasonal adjustment has been carried out for log employment for all five industries. Table 4 contains the selected models, along with regression effects. All the models are relatively simple except for Wholesale Trade; the Ljung-Box goodness-of-fit statistic at lag 24 indicates adequate fits. Summary X-11 Q statistics are all below 0.3, providing evidence that seasonality is present and that the seasonal adjustments are of good quality. All the models contain a regular difference, so we do have ARIMA models for the log ratios as well.

Figure 6 contains scatterplots of absolute log ratios vs. sampling error standard deviations for the other four series, MFGN, CONS, MING, and WTRD. Except for MFGN, each has 1 or 2 large log ratios for which the standard deviations are relatively large. Otherwise, three of the four series have little, if any, discernible pattern. Construction shows a limited positive relationship. In this article, we assume constant sampling error standard deviations for all five series.

Table 5 contains summary statistics for $\log(r_t)$ sampling error autocorrelations for all five series. These estimates are computed monthly using BRR from data Mar 01–Dec 03. Average lag 1 autocorrelations are negative in all cases, with the MING series value of -.15 the only value exceeding .10 in magnitude. In all cases, the variability is considerable, ranging from -.61 to +.37 for MING.

Table 4. Seasonal Adjustment Results

	Series	ARIMA Model	Calendar effect	AO	LS or ramp	Ljung-Box Q (p)	X-11 Q
Manufacturing, Durable Goods	MFGD	111,011	Y	3	0	18.2(.64)	.22
Manufacturing, Nondurable Goods	MFGN	111,011	N	2	0	23.1(.34)	.27
Construction	CONS	011,010	Y	6	2	22.8(.47)	.20
Mining	MING	111,011	N	1	3	18.6(.61)	.17
Wholesale Trade	WTRD	014,011	Y	1	1	25.7(.14)	.18

Note: AO=Additive Outlier, LS=Level Shift, X-11 Q=alternative Q₂ summary measure from X-12-ARIMA

Figure 6. Absolute Log Ratios vs. Sampling Error Standard Deviations

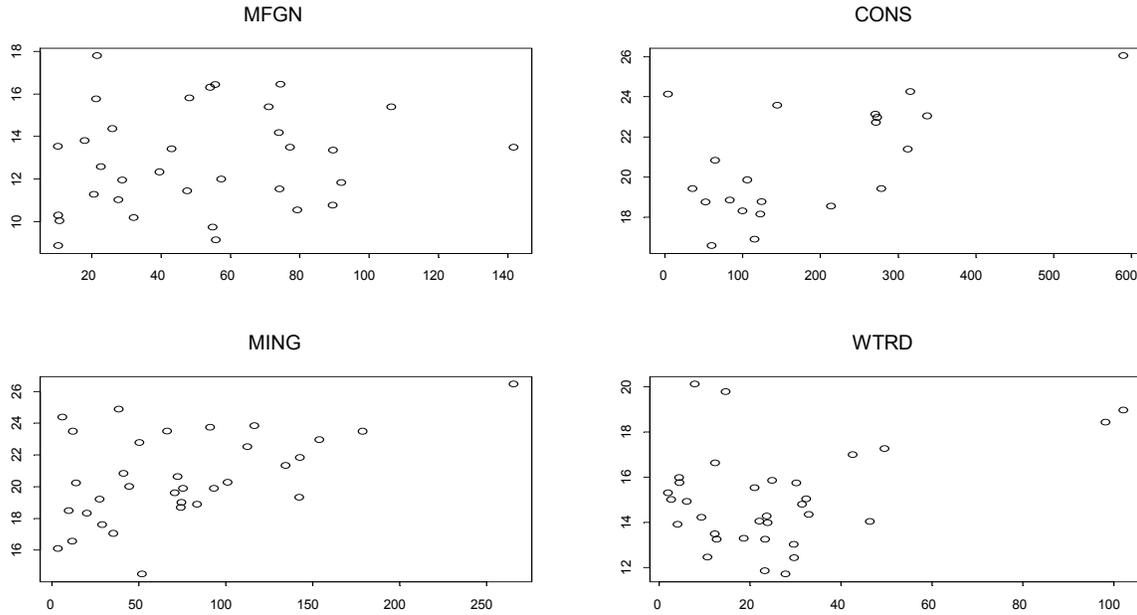


Table 5. Summary Statistics for Sampling Error Autocorrelations from 3/01 – 12/03 Data

Industry	Lag	Mean	Min	Max
MFGD	1	-.09	-.52	.34
	2	-.03	-.39	.47
	12	.10	-.32	.41
MFGN	1	-.08	-.51	.36
	2	-.04	-.44	.31
	12	.16	-.28	.53
CONS	1	-.07	-.28	.32
	2	-.07	-.41	.33
	12	.04	-.14	.22
MING	1	-.15	-.61	.37
	2	-.02	-.33	.42
	12	-.00	.43	.41
WTRD	1	-.06	-.38	.39
	2	-.04	-.46	.32
	12	.06	-.24	.55

Mean lag 2 autocorrelations are also negative, but very close to 0. The largest mean magnitude is .07 for Construction, where values range from -.41 to .33. Lag 12 values are mostly positive, but again with much variability. MFGN's mean is .16, the only mean exceeding .10; its values range from -.28 to .53. Given that the autocorrelations are not very strong, we will not include any autocorrelations in our sampling error models. Rather, we adopt simple white noise models. This will facilitate application of the methodology for a variance measure for level. In further work, we plan to test these series with MA sampling error models having nonzero coefficients at lag 1 or 12.

Turning to our main results, we compare the SDA measure to the standard deviation of the unadjusted series. Table 6 shows average % reduction with SDA across the 12 concurrent runs in the center of the series and at the endpoints. Also shown is the

maximum % reduction, which always occurs during the dip in the end years. As for MFGD, a large time series irregular is estimated for CONS. The % reduction of the SDA is only 1.0% in the center, similar to the 1.5% value for MFGD. However, unlike MFGD, CONS has a reduction at the endpoints, with magnitude 8.5%. In fact, all the series except MFGD have greater reductions at the ends than in the center. In an earlier simulation experiment of seasonal adjustment with ARIMA extrapolation, SDA exhibited this property.

MFGN has the largest reductions overall, 11.4% in the center and 24.1% at the endpoints. While the ARIMA model is the same as for MFGD, either no irregular component or a very small one is identified in the 12 runs. This occurs when (4) has no valid solution for $q=0$ to 3. Again, no irregular is found for MING or WTRD. Figure 7 shows the SDA measure and the unadjusted standard deviation for MING. Note the similar % reduction values in the center for the three series with little or no irregular. The reduction is purely the smoothing effect of the X-11 filter.

Table 6. Summary Statistics across 12 Runs for % Reduction from the Standard Deviation for the Unadjusted Series Achieved with the SDA Measure

Industry	Center	Ends	Max
MFGD	1.5%	-0.4%	5.9%
MFGN	11.4	24.1	32.8
CONS	1.0	8.5	11.1
MING	11.4	13.6	16.8
WTRD	11.3	16.6	27.6

3.3 Estimating the X-11 irregular autocovariances

One of the two major steps in the method is solving the linear system $\hat{U} - D\hat{\lambda} = D\hat{v}$. An issue arises in the computation of \hat{U} , the autocovariances of the X-11 irregular. In principle, one would wish to avoid using extreme irregulars, which could unduly affect the estimates. On the other hand, X-11's default identification of extreme irregulars eliminates too many values, which tends to cause severe underestimation of \hat{U} . For these series, X-11's identification has been suppressed, but a separate test for outliers identifies two additive outliers for MFGD, which have been removed in computing the results reported above. Additional work is needed on the identification of extremes or other techniques for stabilizing estimation of \hat{U} . One alternative is to use a frequency domain approach proposed by Chen (2004). In addition, the

estimation of a large time series irregular for MFGD and CONS and little or no irregular for the other series suggests further study of the sampling error variances.

4. Summary

Given an index number type of estimator for industry employment, a straightforward application of the Pfeffermann method provides a variance measure for seasonally adjusted month-to-month change. This variance estimate is quite sensitive in assessing significance of change for large industries such as Manufacturing, Durable Goods. A variance measure for employment levels has also been proposed. These measures may be applicable in many index series settings.

Future work includes

- (1) testing the proposed method for employment levels,
- (2) testing both change and level measures by simulation experiments,
- and
- (3) testing other sampling error models and different methods for estimating X-11 irregular autocovariances for the five industries studied here.

BLS is considering implementing these variance measures for both industry employment and labor force statistics. We are presently streamlining computer programs and planning to supply BLS analysts with these measures on a trial basis.

ACKNOWLEDGMENTS

We thank Julie Gershunskaya for providing sampling error variances and autocovariances, and George Werking, Pat Getz, and Shail Butani for supporting this project.

REFERENCES

Chen, Zhao-Guo (2004), "Variance Estimation for X11 via Model-based Spectral Approaches under the Additive and the Multiplicative Decompositions," *Proceedings of the ASA Section on Business & Economic Statistics*

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, & Bor-Chung Chen (1998), "New capabilities and methods of the X-12-ARIMA seasonal-adjustment program," *Journal of Business and Economic Statistics*, 16, 127-152

Ladiray, Dominique and Benoit Quenneville, *Seasonal Adjustment with the X-11 Method*, Springer, New York

Morisi, Teresa L. (2003), "Recent Changes in the National Current Employment Statistics Survey," *Monthly Labor Review* 126, 3-13

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., M. Morry, and P. Wong (1995), "Estimation of the Variances of X-11 ARIMA Seasonally Adjusted Estimators for a Multiplicative Decomposition and Heteroscedastic Variances," *International Journal of Forecasting* 11, 271-283

Pfeffermann, D. & S. Scott (1997), "Variance Measures for X-11 Seasonally Adjusted Estimators; Some New Developments with Application to Labor Force Series", *Proceedings of the ASA Section on Business & Economic Statistics*, 211-216

Wolter, K. M. & 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, & J. N. K. Rao, Academic Press, New York, 367-403

The views expressed are those of the authors and do not represent official positions of BLS.

Fig. 7. MING Standard Deviations for Error in Seasonal Adjustment, 12 runs

