Variance Measures for X-11 Seasonal Adjustment: A Summing Up of Empirical Work

August 2005

Stuart Scott & Michael Sverchkov, U.S. 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

Pfeffermann (1994) has proposed a solution to the long-standing problem of variance measures for time series seasonally adjusted by the X-11 method (Ladiray & Quenneville, 2001). Pfeffermann's approach builds from a method suggested by Wolter & Monsour (1981), using sampling error information and the linear approximation to X-11. The present paper pulls together results across the last decade, plus new results for variance estimation of seasonally adjusted change. The U.S. Bureau of Labor Statistics (BLS) is considering use of these measures for the analysis of employment and unemployment statistics as early as 2007.

After a brief preview of results in this section, methodology will be reviewed in Section 2. Section 3 presents basic results, both for pure X-11 seasonal adjustment and for the case of ARIMA extrapolation, with highlights from Pfeffermann & Scott (1997) and Pfeffermann, Scott, & Tiller (2000). Next, we compare our method with a method proposed by Bell & Kramer (1999), first treated in Scott & Pfeffermann (2003). Section 4 contains results for employment change, which expand and improve results presented in Scott, Sverchkov, & Pfeffermann (2004). The final section summarizes additional past work, outlines next steps to be taken, and offers conclusions.

We begin with the usual notion that an economic time series consists of a trend or trend-cycle, a seasonal component, and an irregular term,

$$Y_t = T_t + S_t + I_t .$$

Typically, our data come from a sample survey, leading us to describe an observed value in terms of the population value plus sampling error,

$$y_t = Y_t + \varepsilon_t .$$

We use the notations,

$$A_t = Y_t - S_t , \quad \hat{A}_t = y_t - \hat{S}_t$$

for the population seasonally adjusted value and its estimate at time *t*, respectively. The variance measures we consider are

$$SDA_t^2 = Var(\hat{A}_t - A_t)$$

$$SDH_t^2 = Var(\hat{A}_t - T_t) .$$

$$SDT_t^2 = Var(\hat{T}_t - T_t)$$

Our notation fits with the presentation of results in terms of standard deviations, rather than variances. For most applications, the most important of these measures is *SDA*, the standard deviation (SD) of the error in estimating the (population) seasonally adjusted value. *SDT* is useful, because some countries, including Australia and the U.K. (but not the U.S.), publish trend estimates. The hybrid measure SDH, the SD of the error in estimating the trend by the seasonally adjusted series is included mostly for methodological purposes and comparisons. We use corresponding notations for change, for example,

$$SDAC_t^2 = Var[(\hat{A}_t - \hat{A}_{t-1}) - (A_t - A_{t-1})] .$$

For a series seasonally adjusted with pure X-11, Fig. 1a plots *SDH*, along with *SDU*, the SD of the unadjusted series, that is, the sampling error SD. This picture fits with a characterization given by Wolter and Monsour (1981, p. 400): the SD for the seasonally adjusted series is likely to be below *SDU* in the center of the series, but tends to increase and exceed *SDU* at the two ends of the series. Fig. 1b overlays *SDH* and *SDU* when seasonal adjustment has been performed with ARIMA extrapolation. Here the seasonally adjusted measure lies below *SDU* throughout the series. Finally, for a different series, Fig. 1c presents an overlay of *SDA* and *SDU* , showing that an adjusted SD can exceed *SDU* throughout the time span. Thus, the relationship between *SDU* and our seasonally adjusted SD's depends considerably on characteristics of the series and the X-11 options used. The results in Figs. 1a and 1b are explored further in Section 3 and Figs. 4d and 4e; those for Fig. 1c in Section 4 and Fig. 5a. For now, we remark that the tendency of *SDA* or one of the other measures to lie below *SDU* over most or all of the time span is to be expected, since seasonal adjustment is a smoothing procedure. Furthermore, based on the assumed model, ARIMA extrapolation spreads the smoothing weights more broadly. This reduces the seasonally adjusted SD's near the ends of the series, compared to pure X-11.

2. Methodology

By combining the first equations presented above, the observed series is described as

$$y_t = Y_t + \varepsilon_t = T_t + S_t + I_t + \varepsilon_t . \qquad (1)$$

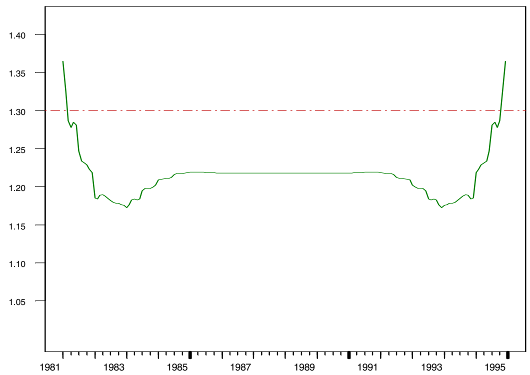
Important for us is the combined error

$$e_t = I_t + \varepsilon_t .$$

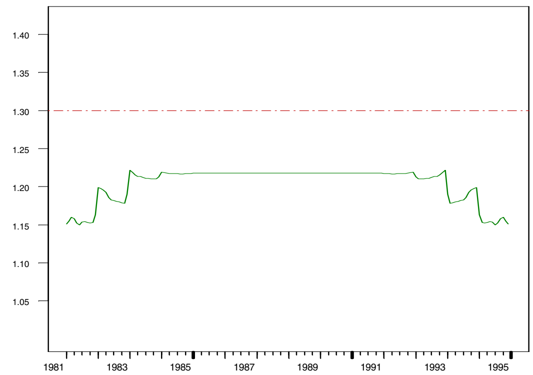
A basic assumption for seasonal time series is that T_t, S_t, and I_t are mutually independent, and we

Figure 1. Some Results for SD Measures for X-11 Seasonal Adjustment
Unadjusted SD – dash, Adjusted SD measure (SDA or SDH) – solid

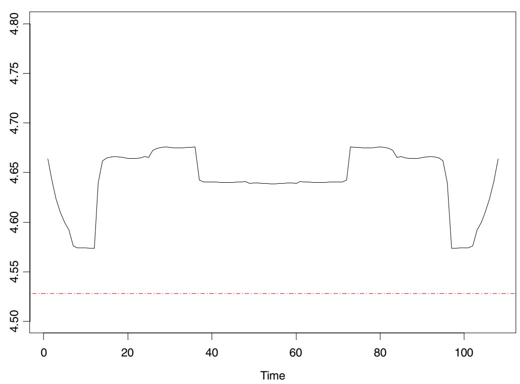
1a. AKEP SDH – Pure X-11



1b. AKEP SDH – Extrapolation



1c. EMPP SDA, Extrapolation



further assume mutual independence with ε_t. We assume that I_t and ε_t are stationary time series. Writing ν_k and λ_k for their respective autocovariances, the combined errors e_t have autocovariances

$$V_k = \nu_k + \lambda_k . \qquad (2)$$

Under the assumption that X-11 produces unbiased estimates for the trend and seasonal components, Pfeffermann (1994) develops the approximations

$$SDH_t^2 \approx Var(\sum_{k=1}^{N} w_{tk} e_k)$$
$$SDA_t^2 \approx SDH_t^2 + \nu_0(1-2w_{tt}) - 2\sum_{\substack{k=1 \\ k\neq t}}^{N} w_{tk}\nu_{t-k} \qquad (3)$$

where $\{w_{tk}\}$ represents the seasonal adjustment filter weights for time *t*. A formula similar to that for *SDH* applies for *SDT*. To estimate *SDH*, Formula (3) implies that we only need estimates $\hat{V}_k$ of the combined error autocovariances. For *SDA*, we also need estimates of the ν_k's, obtainable from (2) if we have estimates of the λ_k's. The latter estimates, which reflect the survey design, are available in our applications. We claim that a reasonable approximation for *SDA* is obtained by using only the sampling error variance and dropping the final summation, which is typically small due either to small magnitudes of the weights w_{tk} or to small or vanishing values of ν_k for $k \neq 0$.

For estimation of the V_k's in (2), consider the irregular component R_t produced by X-11. Pfeffermann (1994) derives the approximation

$$R_t \approx \sum\nolimits_{k=1}^{N} a_{tk} e_k , \qquad (4)$$

where $\{a_{tk}\}$ represents the irregular filter weights for time *t*. Notice that the X-11 irregular series is nonstationary because of the use of asymmetric, time-dependent weights when moving away from the center of the series. Taking autocovariances in (4), 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

$$\hat{U} = D\hat{V} = D(\hat{\lambda} + \hat{\nu}) . \qquad (5)$$

The matrix D is a known matrix built from the irregular weights. We assume $V_k = 0, k > C$, for some cut-off value C. When sampling error autocovariance estimates $\hat{\lambda}$ are available externally, we can solve the system (5) for $\hat{V}$. We model the irregular component as a low order moving average process, so (5) becomes a low order system. Often, the irregular is modeled simply as white noise. See Pfeffermann & Scott (1997) for more details.

Pfeffermann, Morry, & Wong (1995) and Pfeffermann, Scott, & Tiller (2000) treat ARIMA extrapolation. We may write

$$\hat{A}_t = \sum_{k=1}^{N} w_{tk} y_k + \sum_{k=1}^{N} w_{tk} y_k^{(f)},$$

where each value $y_k^{(f)}$ is a forecast or backcast derived from the ARIMA model fitted to the series. For given model coefficients, each of these values is a linear combination of the observed *y*'s, so that the seasonal adjustment has again a close linear approximation.

3. Results for monthly estimates

3.1 AKEP experiment

We have simulation results for seasonal adjustment with pure X-11 and with ARIMA extrapolation. The simulation is related to the Alaska Employment-to-Population ratio series (AKEP), which appears in Figure 2. The "EP ratio" and the unemployment rate are the key labor force statistics for U.S. states. They are derived from the Current Population Survey (CPS), a monthly household survey whose size has ranged from 47,000 to 60,000 in recent decades. The sample is composed of 8 panels, with a rotation scheme of 4 months in sample, 8 months out, and a final 4 months in sample. Figure 3 exhibits the per cent sample overlap, along with the estimated autocorrelations for AKEP, according to the lag in months. The per cent overlap is 75% at a one-month lag and 50% at a 12-month lag, with no overlap for lags 4 to 8 and beyond lag 15. The autocorrelations have the same general shape with lower peaks but with positive values even for lags 4 to 8. This is due to the replacement of households in the sample within small homogeneous geographical regions. The conclusion from Figure 3 is that one needs to account for high-order autocorrelations when computing the SD measures. A considerable amount of work in recent years (e.g., Pfeffermann, Tiller, & Zimmerman, 2000) attests to reliability of estimated autocorrelations, which allows us to use the low order system (5) to obtain the needed error autocovariances.

The simulation experiment uses data generated from a structural model (Harvey, 1989)

$$y_t = T_t + S_t + I_t + \varepsilon_t.$$

The trend T_t is a random walk with a constant drift. The seasonal effect S_t evolves according to the trigonometric relationship defined in Harvey (1989, Ch. 2) and the irregular component I_t follows an MA(2) model with coefficients .6 and -.3. The sampling error ε_t is generated from an AR(15) model with coefficients computed from the Yule-Walker equations, with standard deviation (SD) 1.14 and the sampling error autocorrelations displayed in Table 1. The remaining components have disturbance SD's of 1.96×10^{-5}, 0.02, and 0.61, respectively. This model differs from the model fit to the original series by including the irregular component and reducing the sampling error variance so that the combined error e_t has about the same variance as the sampling error in the official model used by BLS. We simulate 3000 series, apply to each the X-11 method with and without ARIMA extrapolation (as described below) using the X-12-ARIMA software (Findley *et al*, 1998), and compute the SD measures as described in Section 2.

Figure 2. AKEP, 1979-86

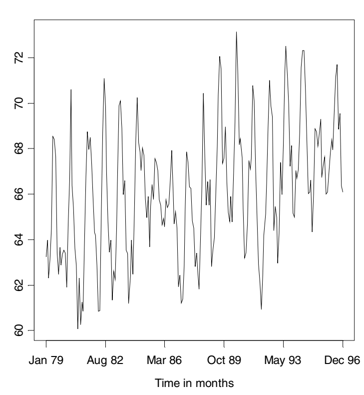


Figure 3. CPS Overlap and AKEP ACF

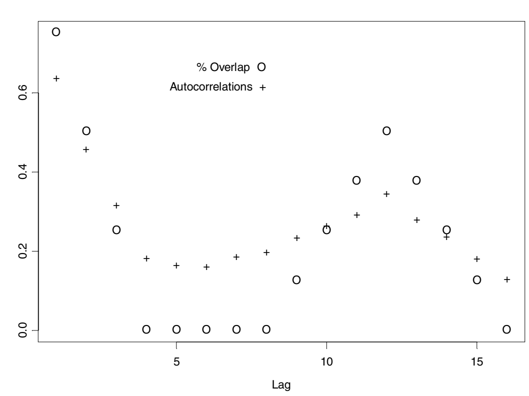


Table 1. AKEP Sampling Error Autocorrelations

Lag	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Value	0.64	.46	0.32	0.18	0.16	0.16	0.18	0.20	0.23	0.26	0.29	0.34	0.28	0.24	0.18

The simulations provide us with an approximation to the true standard deviations as follows. Fix a time point. For each replicate series, we have the true and estimated values, so, for example, we can compute the error $A_t - \hat{A}_t$. The empirical standard deviation of these 3000 errors is our approximation to the true value for the fixed time point, which we can compare to estimates from the method. We compare results for X-11 runs of 7 and 14 year spans, the short spans coming from the middle years of the longer series. (The results reported for this simulation in Pfeffermann & Scott, 1997, were based on only 300 replicates).

Table 2 shows mean estimates of the combined error autocovariances across the 3000 replicates. The top row shows the true values V_k, *k*= 0 to 3. (Recall that we solve (5) with order *q* for $\hat{\nu}$ and compute $\hat{V} = \hat{\lambda} + \hat{\nu}$, *q*=0 to 3. For example, with q=1, $\hat{\nu}_k$ is nonzero only for *k*=0 and 1.) For the long series, the table indicates underestimation of V_0 with *q*=0 and overestimation with *q*=1 or 3. The estimates are very close with the true value, *q*=2. For the short series, estimates are not quite as good: solving with *q*=2 gives the best estimates, with mild overestimation.

Figure 4 shows estimation results for SDA and SDH. (Recall that SDA estimates the SD of the error in estimating the true seasonally adjusted value; SDH estimates the SD of the error in estimating the trend by the seasonally adjusted value). Each part of Figure 4 shows mean SD estimates for MA(q) models, *q*=0 to 3, the solid symmetric curves. The red dot-dashed horizontal line is SDU, the SD of the unadjusted estimate. The rough black dashed line represents the empirical estimates of the truth from the simulation. Notice the variation in these values even at the center of the series where the true SD's are the same.

Table 2. Estimates of AKEP Combined Error Autocovariances, According to the Fitted MA(q) Model and Series Length

	V_0	V_1	V_2	V_3
True	1.85	.98	.48	.41
Estimates				
Long				
q=0	1.78	.83	.59	.41
MA 1	1.95	1.08	.59	.41
order 2	1.88	1.01	.51	.41
3	2.01	1.14	.61	.50
Short				
q=0	1.78	.83	.59	.41
MA 1	1.95	1.08	.59	.41
order 2	1.92	1.07	.56	.41
3	2.06	1.20	.69	.54

The estimates for SDH in Fig. 4d are 6.3% below SDU in the central part of the series. They increase toward the ends of the series, and the three points at each end of the series average 2.0% higher than SDU. The good performance of the estimates is seen in the extremely close agreement between the simulated truth and estimates when assuming the (correct) MA(2) model for the irregular. The bias is 1.7% near the ends and only 0.4% in the center. This provides strong evidence that both the shape and the magnitude of the estimates are correct. Table 3 presents the bias statistics; Table 4 reports the per cent reduction in moving from SDU to SDA or SDH.

Turning to Fig. 4e, the SDH estimates are the same in the center, but the estimates are actually lower at the ends of the series than in the center. The per cent reduction from SDU values is 6.4% in the center and 11.2% at the ends. The difference is that here seasonal adjustment has been applied with ARIMA extrapolation. Note that again the *q*=2 estimates agree very closely with the empirical estimates, even at the ends. The decrease in SDH toward the ends is a genuine property of X-11 with extrapolation and is due to the use of a model. The ARIMA extrapolation has been carried out with a (0,1,3) (0,1,1) model, fixing the model but estimating the model parameters independently for each replicate. Fig. 4f shows the results from the 7-year span. The results are similar to those for the long span, except that the *q*=2 estimates are larger than the simulated truth, 1.7% in the center, which fits with the mild overestimation of the $V(k)$'s shown in Table 2. For both long and short spans, the SDH estimates do vary with *q*, with the best estimates obtained under the MA(2) model for the irregulars.

Graphs for SDA in Fig. 4a-4c have the same shape as their counterparts for SDH, except that they are much lower. In the center of the series, the reduction from SDU is about 19%; at the ends the reduction is just above 20% with extrapolation and close to 10% with pure X-11. The choice of *q* makes little difference, since the estimates are so close.

Figure 4. AKEP Empirical and Mean Estimated Standard Deviations with MA(q) Irregulars, q=0 to 3
Empirical – black dash, q=0 – solid blue, q=1 – solid purple, q=2 – solid green, q=3 – solid gold, Unadjusted – red dot-dash

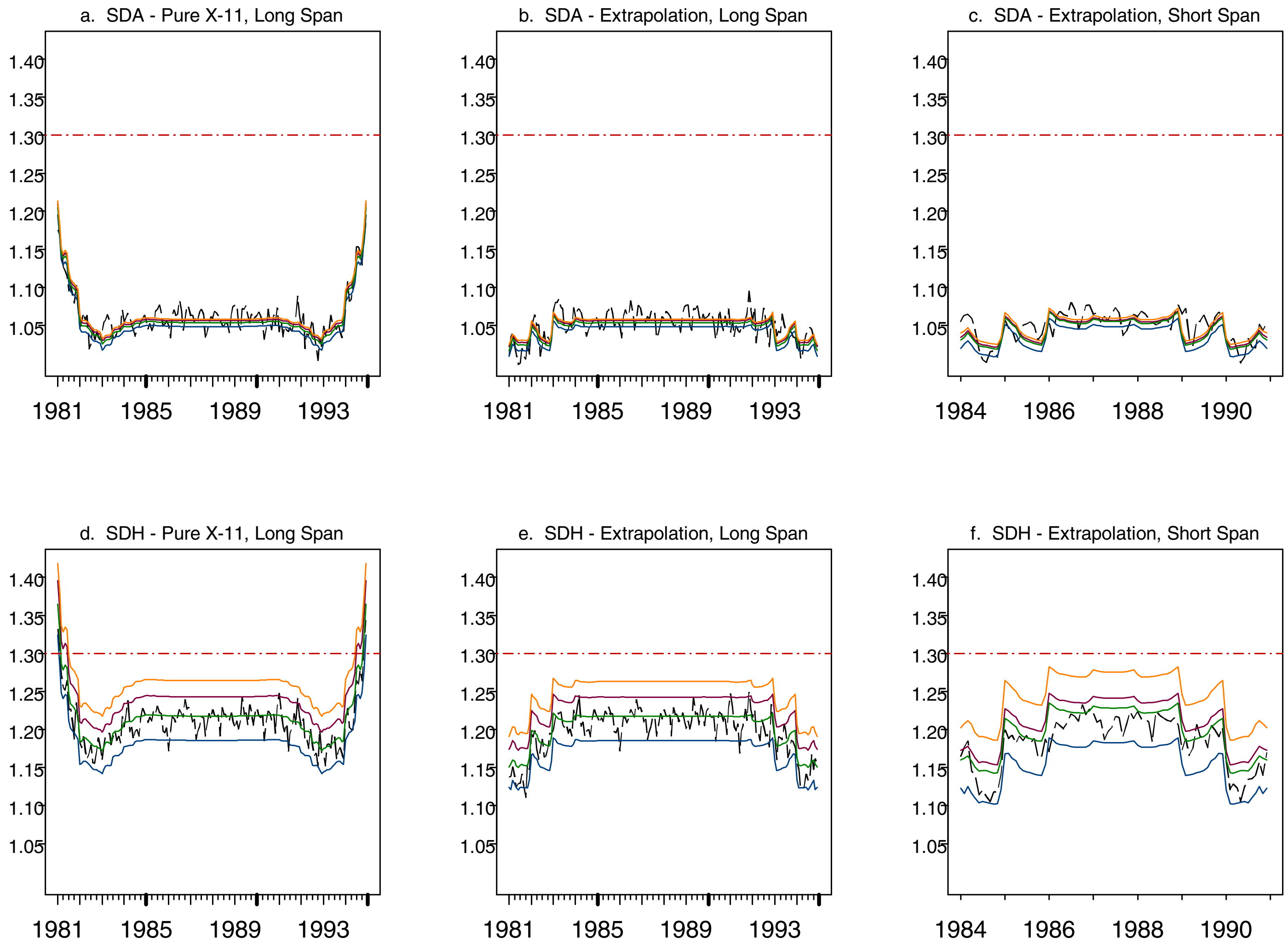


Table 3. Mean SDA and SDH Estimates and % Bias in Center and at Ends for Four Seasonal Adjustments of AKEP

		SDA			SDH		
		Est (q=2)	"True"	% Bias	Est (q=2)	"True"	% Bias
Ends							
Pure X-11	Long	1.17	1.16	1.07	1.33	1.30	1.70
	Short	1.18	1.18	-.65	1.34	1.33	1.00
Extrap	Long	1.02	1.03	-.46	1.16	1.15	.63
	Short	1.04	1.06	-1.90	1.16	1.17	-.25
Center							
Pure X-11	Long	1.05	1.06	-.34	1.22	1.21	.38
	Short	1.05	1.05	-.37	1.22	1.20	1.60
Extrap	Long	1.05	1.06	-.35	1.22	1.21	.36
	Short	1.06	1.06	-.26	1.23	1.21	1.73

Table 4. % Reduction from SDU with SDA and SDH with an MA(2) Irregular (negative values indicate values higher than SDU)

		SDA	SDH
Ends			
X-11	Long	9.7	-2.0
	Short	9.5	-3.2
Extrap	Long	21.2	11.2
	Short	20.4	10.6
Center			
X-11	Long	18.9	6.3
	Short	19.2	6.0
Extrap	Long	19.0	6.4
	Short	18.8	5.5

3.2 Comparison with the Bell-Kramer method

Bell & Kramer (1999) have developed an alternative variance measure, also building from the linear approximation to X-11. For a given time point, they define the target seasonally adjusted value as the value that would result from application of the *symmetric* linear seasonal adjustment filter to the series without the sampling error. The variance measure accounts therefore for two sources of error: the sampling error and the error arising from extending the observed time series with enough ARIMA forecasts and backcasts for application of the symmetric filter. The latter error is known as *revision error*. Suppose that the X-11 symmetric filter is of length $2m+1$. (For the default X-11 filter, m=84). Building from previous notation, denote by $y_{(b)} = Y_{(b)} + \varepsilon_{(b)}$ and $y_{(f)} = Y_{(f)} + \varepsilon_{(f)}$ the vectors of past and future values of length m. The target vector of seasonally adjusted values is $A^* = \Omega Y$ where $Y' = (Y_{(b)}', Y_{(obs)}' Y_{(f)}')$ with $Y_{(obs)} = (Y_1,...,Y_N)'$ defines the population values and Ω is the matrix of dimension $N \times (N+2m)$ of the symmetric X-11 filter weights used for the adjustment. The vector of the observed series augmented with the backcasts and forecasts can be written

$$\hat{y}' = (\hat{y}_{(b)}', y_{(obs)}', \hat{y}_{(f)}'),$$

with $\hat{y}_{(b)} = E(y_{(b)} \mid y_{(obs)})$ and $\hat{y}_{(f)} = E(y_{(f)} \mid y_{(obs)})$. Thus, the seasonal adjustment error is

$$A^* - \hat{A} = \Omega Y - \Omega \hat{y}$$

and the variance-covariance matrix of the error is

$$Var(A^* - \hat{A}) = \Omega Var(Y - \hat{y}) \Omega' .$$

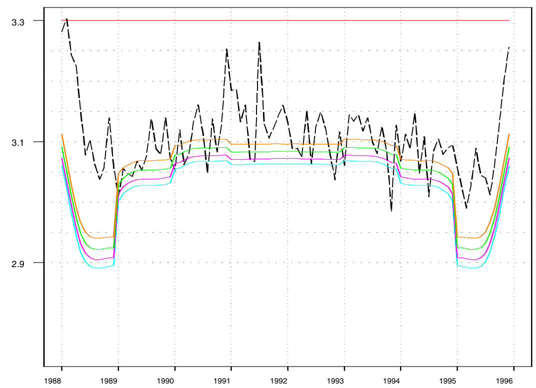
The matrix Ω is known, and Bell & Kramer (1999) provide the details for computing $Var(Y - \hat{y})$.

We briefly review a simulation experiment reported in Pfeffermann & Scott (2003), which compares the Bell-Kramer and Pfeffermann methods, denoted BK and DP, respectively. The simulation is based on models for Teenage Male Unemployment (TMU), which, like AKEP is derived from the CPS. TMU is one of eight series used to compute the seasonally adjusted civilian unemployment rate, the key household figure reported in the monthly BLS *Employment Situation* press release. Given the form of the models for the population series Y, and for the sampling error ε, William Bell's REGCMPNT program (Bell, 2003) estimates these components and the parameters for the component models. For TMU, this leads to an airline model for the signal, whose parameters θ_1 and θ_{12} are rounded to 0.6. Signal extraction of this model yields a (0,2,2) model for the trend, an (11,0,11) model for the seasonal component, and white noise for the irregular. An AR(15) model is adopted for the sampling error. These modeling steps are described in more detail in Pfeffermann & Scott (2003).

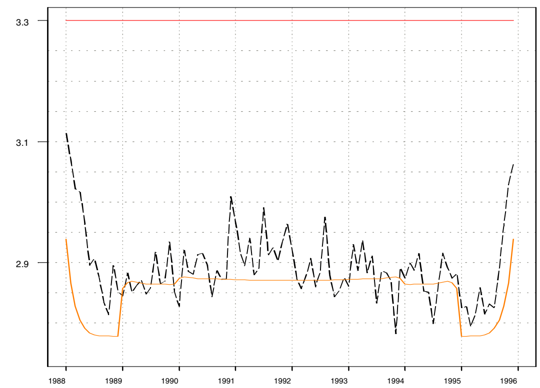
Three thousand replicates are generated from these models, each of length 22 years: 8 central years as the test period and 7 years at both ends in order to compute target values for the BK method. For the DP method, we carry out X-11 seasonal adjustment on the 8-year span with 5 years of forecasts and

Figure 5. Empirical and Mean Estimated Standard Deviations about Seasonal Adjustment Target

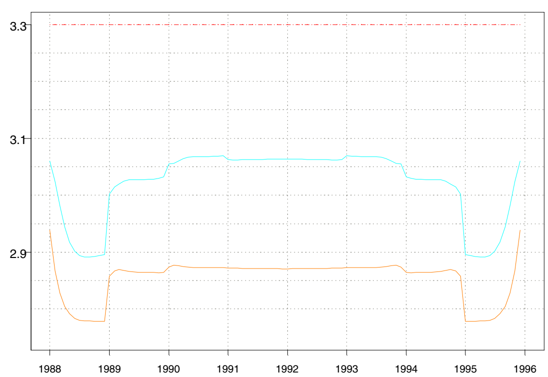
a. Pfeffermann Measure (q=0 to 3)



b. Bell-Kramer Measure



c. Pfeffermann (blue) and Bell-Kramer (gold) Measures



backcasts, the maximum available in the X-12 software, a close approximation to using the symmetric filter for the 8-year test period. In contrast to the AKEP experiment, the X-11 runs for TMU allow automatic selection of the model, as well as estimation of the model parameters. (Arguably, either approach reflects practice: sometimes, agencies allow automatic model selection; in other cases, an analyst selects and uses a particular model until it no longer fits well). SD results are restricted to replicates for which applicable models are found: 2584 replicates for the DP method, 2073 for Bell-Kramer. The figure is lower for Bell-Kramer, since signal model parameters estimated from REGCMPNT yielded nonstationary models in many cases; the DP method simply used the models estimated by X-12, which restricts models to be stationary. In both cases, we compute mean estimates across replicates and empirical estimates of the true SD's.

Fig. 5a contains mean DP estimates of SDA for MA(q) models for the irregulars, q=0 to 3, (solid lines), empirical estimates of the true values (black dash), and SDU (red dash). Estimates for q=0, the true value in blue, average 6.6% lower than the empirical values, based on the last three points at each end; they are 1.9% lower in the center, based on the central 24 time points. By comparison, the BK estimates, shown in Fig. 5b, are 5.4% lower at the ends and 1.2% lower in the center. Notice how the difference in target values affects the SD measures. Fig. 5c overlays mean estimates from the two methods. The DP estimates are 6.7% higher in the center and 5.0% higher at the ends. Unlike the BK method, the DP method is designed to capture the contribution of the irregular , which accounts for the sizable difference. In the presence of a time series irregular, the two methods can differ appreciably.

The BK method has less bias: on average it comes closer to its target. It also has less variability. Interestingly, it requires no output from X-11; it is calculated from estimated models for the population and the sampling error. The DP method is more flexible: it can be applied with pure X-11; it captures an extra source of variability, which can be appreciable, as will be seen in Section 4. While it benefits from sampling error information, it can provide SD estimates with little or no sampling error knowledge. This simulation provides evidence that both methods provide usable (but different) SD measures with good properties.

4. Results for month-to-month change

Each month the Commissioner of Labor Statistics must characterize change in employment and unemployment statistics to the Joint Economic Committee of the U.S. Congress. When seasonally adjusted change is +100,000, is employment "essentially unchanged" or has it "increased"? Assessing significance of change is of much greater practical importance for statistical agencies than having confidence limits for the level. While one-month change is far and away the most important case, we can consider more generally *h*-month change, $\hat{A}_t - \hat{A}_{t-h}$. Conceptually, handling change is quite easy. Since the difference of linear filters is again a linear filter, we can adapt the methods of Section 2 to calculate all three SD measures for *h*-month change. However, for the important employment application of this section, an alternative approach described in Scott, Sverchkov, and Pfeffermann (2004) seems more appropriate. A further advantage is that this alternative approach appears applicable to index number time series, such as price indexes.

BLS industry employment statistics come from its 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). 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. 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,

$$y_t = Y_0 \cdot r_1 \cdot r_2 \cdots r_t .$$

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 establishments reporting in both months.

Table 5. CES Series

Code		12/03 Level
EMPP	Total Private Employment	108.49
MFGD	Durable Manufacturing	8.87
MFGN	Nondurable Manufacturing	5.46
CONS	Construction	6.77
MING	Mining	0.50
WTRD	Wholesale Trade	5.60

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

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

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 $(1-B)\log(y_t)$. Summarizing, our variance for seasonally adjusted change comes from applying the basic method to the series $\log(r_t)$.

Using data for 1994-2003, we analyze employment change for six NAICS supersectors. Table 5 lists these industries, short-hand codes used subsequently, and their Dec 03 employment levels in millions. 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 a preview to the detailed results, Table 6a. contains the results for MFGD of applying the SDA measure to form nominal 95% confidence limits ($\pm 2 \cdot SDA$) for the seasonally adjusted log ratio. There are significant declines for the first nine months, but no significant change during the last three. Table 6b. has an extract from a single run on employment levels ending in Dec 03. It shows month-to-month per cent change in the seasonally adjusted series during 2003. These values range from over ½% to about $^1/_6$% during the first 9 months, deemed significant by our measure. While the numbers in Table 6a. and 6b. are not directly comparable, they are consistent with each other and indicate that the measure is very sensitive, since one-sixth per cent represents about 20,000 out of roughly 9 million.

As described above, the form of the link-relative estimator leads us to work with monthly log ratios. The payoff for this approach comes from relatively simple sampling error characteristics. First, monthly sampling error variance estimates for the log ratios show considerable variability but little relation to size of the log ratio (except possibly for CONS) or month of the year. Thus, we assume them to be constant

Table 6. Results for MFGD

a. Confidence Limits for Seasonally Adjusted Log Ratio (x10^4), 12 Runs

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
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

b. Month-month % Change in Seasonally Adjusted MFGD Level, 1/95-12/03 Run

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

and estimate λ_0 by the median value, reported in Table 8. Table 7 exhibits summary statistics for autocorrelations at lags 1 and 12. These are based on monthly estimates from 3/01 to 12/03, except that only a shorter span, 4/01-12/03, is available for EMPP. Median lag 1 autocorrelations are less than .10 in magnitude for these series, except for MING, and there is a great range in the estimates, seen from the minima and maxima. Still, the medians are consistently negative. Similarly, the lag 12 autocorrelations exhibit great variability, but all have the same sign (in this case, positive). Half have magnitude .10 or more. Median autocorrelations at other lags are all very low in magnitude. This leads us to carry out concurrent runs under two sampling error models for each series, white noise ignoring the autocorrelations and an MA(13) model based on fitting a (0,0,1) (0,0,1) model to the median lag 1 and lag 12 autocorrelations. While differences are not too large, we present SD results based on the MA(13) model for two reasons:

Table 7. Summary Statistics for Sampling Error Autocorrelations for Six Series

Lag	Industry	Median	Min	Max
1	EMPP	-.08	-.31	.27
	MFGD	-.05	-.46	.32
	MFGN	-.05	-.47	.29
	CONS	-.07	-.29	.34
	MING	-.15	-.57	.40
	WTRD	-.09	-.43	.38
12	EMPP	.10	-.08	.48
	MFGD	.17	-.26	.38
	MFGN	.18	-.26	.48
	CONS	.08	-.14	.24
	MING	.03	-.37	.41
	WTRD	.06	-.24	.55

(1) they appear to improve estimation of the $V(k)$'s,

(2) intuitively, low magnitude autocorrelations with the observed signs seem reasonable.

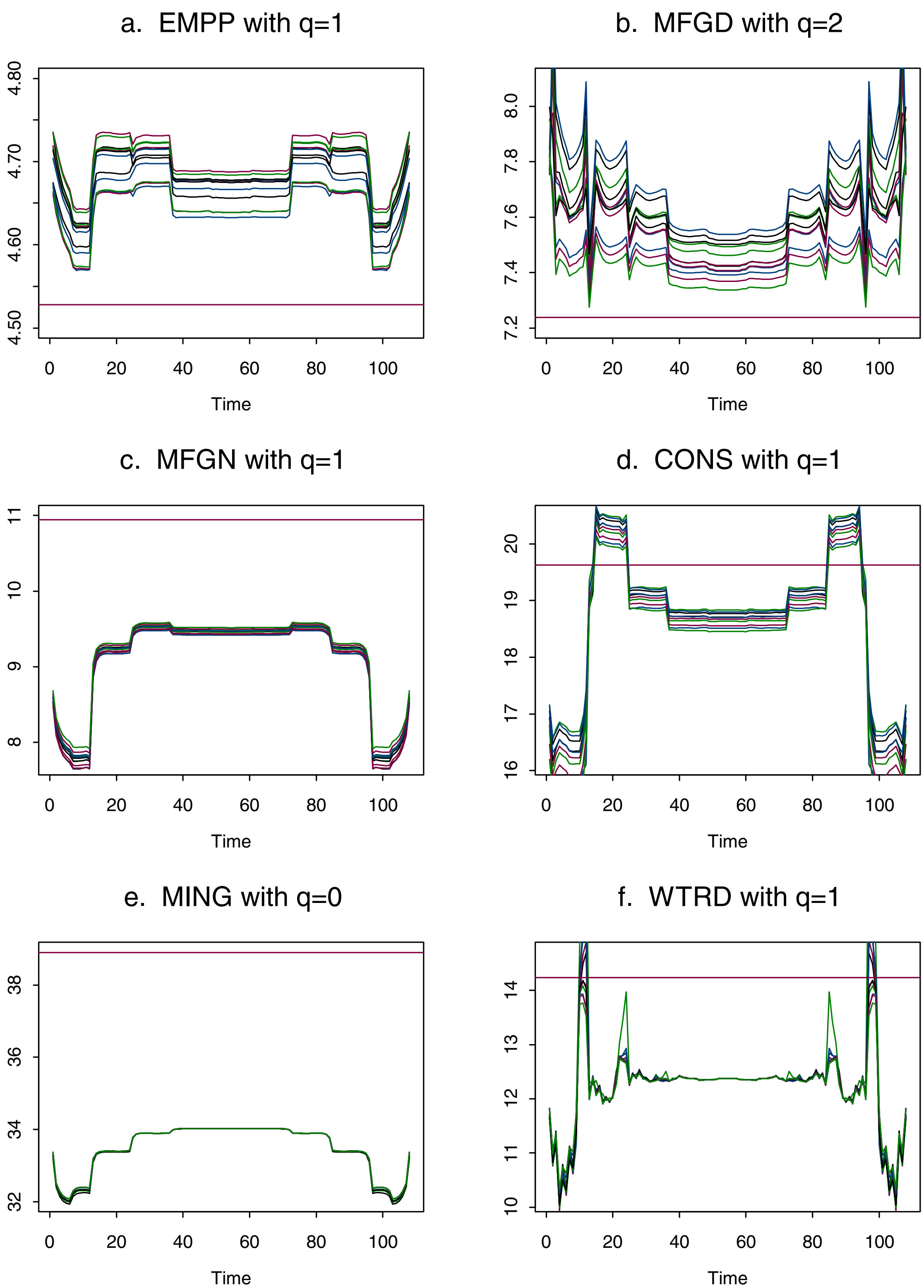
In addition to the sampling error variances λ_0, Table 8 contains variance estimates ν_0 for the irregular and V_0 for the combined error. It also shows the order selected for the MA(q) irregular. Note that for EMPP and MFGD, the irregular variance ν_0 is more than double the sampling error variance λ_0. These series have the largest samples and, consequently, the smallest sampling error. We obtain no valid estimates for ν_0 for MING, which has by far the largest λ_0, and so we take ν_0 to be 0. MFGN and WTRD also have small or vanishing ν_0. These five series exhibit a pattern of relatively large irregular variance ν_0 when sampling error variance λ_0 is small and small ν_0 for large λ_0. CONS is an exception, having sizable values of both.

Figure 6 contains SDA results from the 12 concurrent runs, along with a red horizontal line for SDU. To simplify the pictures somewhat, SDA and SDU are plotted according to time point in the X-11 run span rather than date. For example, point 108 is

Table 8. Variances for Sampling Error, Irregular, and Combined Error for Six Series (with selected orders q for MA(q) models for Irregular)

Series	λ_0	ν_0	V_0	q
EMPP	20.5	41.9	62.4	1
MFGD	52.4	129.7	182.1	2
MFGN	119.7	30.2	149.9	1
CONS	385.0	346.4	731.4	2
MING	1512.4	0	1512.4	0
WTRD	202.6	0	202.6	0

Figure 6. SDA for CES Series from 12 “Concurrent” Runs with
Unadjusted Standard Deviation (solid red horizontal line)



Jan 03 for the first run and Dec 03 for the 12th run. The variability across runs is due to changing estimates of ν_0 and model parameters. (As could be expected, this variability is least for MFGN, MING, and WTRD, where the variability comes almost entirely from the model parameters). For EMPP and MFGD, SDA exceeds SDU throughout the time span, with end values larger than the central values. As seen in Table 9, for EMPP, SDA is 3.4% higher at the ends and 3.1% higher in the center. MFGD's SDA is also higher than SDU, about the same percentage as EMPP in the center and an even greater percentage at the ends. Even so, we have already seen that the SDA measure is quite sensitive in detecting significant change for MFGD. The other four series all have positive reductions with SDA, in the 15-25% range at the ends and 5-15% in the center.

Table 9. % Reduction from SDU with SDA

Series	Ends	Center
EMPP	-3.4	-3.1
MFGD	-8.4	-2.9
MFGN	24.4	13.5
CONS	16.5	4.9
MING	16.0	12.5
WTRD	20.7	13.1

5. Conclusions and Additional Work

5.1. Conclusions for Pfeffermann's

$$SDA = (Var(\hat{A}_t - A_t))^{1/2}$$

Overall performance

Pfeffermann's method estimates SDA well, both for pure X-11 and for X-11 with ARIMA extrapolation in a variety of settings. The bias was mostly about 1% or less in magnitude for the Alaska simulations. For the teenage series, there was 1.9% underestimation in the center of the series and 6.6% underestimation at the ends.

Relation to SDU, the standard deviation of the unadjusted estimate

In most cases, SDA is less than SDU in the center of the series. When extrapolation is used with X-11, SDA also tends to be lower at the ends. However, in series with relatively low sampling error, a sizable irregular can be identified, and SDA can exceed SDU.

Sensitivity of measure

The Durable Manufacturing example shows that the measure is sensitive in practical settings. That is, the measure is able to identify quite small month-to-month change as significant.

Comparison with Bell-Kramer method

Pfeffermann's method has larger bias and is less stable in the simulation reported here. Recall again that the two methods estimate different errors. Pfeffermann's captures the effect of the irregular component, which makes a major contribution to series variability in cases such as our change measure for Durable Manufacturing.

Flexibility of the method

Although the method works best when sampling error autocorrelations are available, the method yields reasonable SDA approximations even when only sampling error variance estimates are available. The method can be applied in index number settings to obtain SDA for change.

5.2. Additional work

Multiplicative adjustment

A variant of the method is available for multiplicative seasonal adjustment. Simulation results have been obtained for a multiplicative seasonal adjustment with time-varying sampling error variances. The SDA estimates tended to track the simulated truth quite closely. Multiplicative adjustment with constant sampling error variances can also be handled. Since this mode occurs more frequently than additive, this very important case needs further study.

Change

Because of the importance of change, future work will focus on simulation experiments for change, both with the straightforward linear filter approach and with the method employed in Section 4.

Autocovariance estimation

Chen, Wong, Morry, & Fung (2003) have developed spectral approaches to estimation of the combined error autocovariances. These methods can handle additive and multiplicative seasonal adjustment modes and availability/nonavailability of sampling error autocorrelations. The method uses X-11 irregulars and assumes they follow a stationary process. In practice, this means cutting off at least a couple of years at each end of the series, a major disadvantage if the series is short. Some results show the method to have greater stability than the method of moments used in this paper. Further experimentation is planned.

X-11 filter options

For implementation, the method should be able to handle any combination of X-11's trend and seasonal filters. So far, our experiments have been restricted to the former default X-11 filter choices: the 13-point Henderson filter for trend, the 3×3 filter for preliminary estimation of the seasonal, and the 3×5 filter for final estimation of the seasonal. Given a series of length *N*, we currently run a special program to obtain the $N \times N$ matrix of linear filter weights for both the trend and seasonally adjusted series. Simple relations allow deriving from these the irregular and seasonal linear filter weights.

We plan to test the method for other X-11 filter options. The simplest approach will be to extend the series with enough forecasts and backcasts that the symmetric filter applies to the entire input span. In this case, only one vector of filter weights is needed for each set of X-11 filter options, rather than a matrix of weights depending on the series length.

Measure for levels of index-style estimators

We have derived a variant for handling level estimates in the employment estimation setting of Section 4. This procedure has yet to be tested.

Acknowledgments

We thank Julie Gershunskaya, Tamara Zimmerman, Steve Miller, and Ed Robison for providing sampling error information, and Shail Butani, Pat Getz, and George Werking for supporting this project. We also thank William Bell for helpful discussions.

References

Bell, William R. & Matthew Kramer (1999), "Toward Variances for X-11 Seasonal Adjustments," *Survey Methodology* 25, 13-29.

Chen, Zhao-Guo, Paul Wong, Marietta Morry, & Helen Fung (2003), "Variance Estimation for X-11 Seasonal Adjustment Procedure: Spectrum Approach and Comparison," Statistics Canada Report BSMD-2003-001E.

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 (2001), *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, Danny & Stuart 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.

Pfeffermann, Danny, Stuart Scott, & Richard Tiller (2000), "Comparison of Variance Measure for Seasonally adjusted and Trend Series," *Proceedings of the 2nd International Conference on Establishment Surveys*, American Statistical Association, 755-764.

Pfeffermann, Danny, Richard Tiller, & Tamara Zimmerman (2000), "Accounting for Sampling Error Autocorrelations: Towards Signal Extraction from Models with Sampling Error," *Proceedings of the ASA Section on Business & Economic Statistics*, 108-113.

Scott, Stuart, Michail Sverchkov, and Danny Pfeffermann (2004), "Variance Measures for Seasonally Adjusted Employment and Employment Change," *ASA Proceedings of the Joint Statistical Meetings*, 1328-1335.

Scott, Stuart and Danny Pfeffermann (2003), "Evaluation of Two Estimation Methods for Variance of X-11 Seasonally Adjusted Series," *ASA Proceedings of the Joint Statistical Meetings*, 3760-3767.

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.