DISCUSSION

Stuart Scott, Bureau of Labor Statistics
441 G St., N.W., Room 2126, Washington, D. C. 20212

Overall Comments:

Estela Dagum quoted the Gordon Commission and its recommendations on U.S. employment and unemployment statistics:

> ...that estimates of standard errors of seasonally adjusted data be prepared and published as soon as the technical problems are solved.

That was 1962. The current set of papers is very encouraging for offering new and creative suggestions for tackling this problem, and I am pleased to have the opportunity to read and study these papers.

Model-based approaches provide appealing solutions from a mathematical viewpoint. Hausman and Watson (1985) broke new ground with an unobserved component models (UCM) approach which incorporated information from the survey design. Bell and Hillmer have worked to integrate time series modeling with the survey design perspective, including efforts to develop greater capability for estimating correlations stemming from the rotating panel design for the Census Bureau's Current Population Survey. However, seasonal adjustment based on time series models has now been carried out for about 15 years without supplanting X-11, and it appears unlikely to do so in the near future.

ARIMA models represent the most worked-out method for modeling seasonal time series, but in practice the path to a fully acceptable model is not trivial: two statisticians can arrive at different models; modeling is still time-consuming. State-space models form a more general class, but only limited models are tried so far and model identification is not satisfactory yet. David Findley's empirical work, discussed further below, prefers ARIMA models to the unobserved components. Government agencies, chief producers of seasonally data, typically wish to limit the use of models for reasons of objectivity and reproducibility. Finally, experience with Producer Price Index energy series has shown me that economic time series models do not always conform well to simple univariate time series models (cf. Buszuwski and Scott, 1988). In particular, differencing alone can be inadequate to achieve stationarity.

This suggests to me that the adaptive nature of X-11 is attractive. I think Wayne Fuller has been quoted as saying, "Seasonal adjustment is what X-11 gives me." Thus, it is a positive development to me that the papers by Pfeffermann and Dagum and Quenneville in this session, plus a paper presented by Findley and Monsell (1990), present methods for variances of X-11 seasonally adjusted series. The above comments on the problems with models are meant to suggest humility in the face of reality, rather than full satisfaction with X-11. The papers by Findley and Hausman and Watson offer practical suggestions for working with models.

All the papers contain good applied statistics. The knowledge and experience of the authors are reflected in the detailed work in their examples and the attention to realism.

I'll now turn to more specific comments, first on the two X-11 related papers, then the model-based. Since this has been a learning experience for me, the emphasis will be on summarizing rather than critiquing.

Pfeffermann:

The basic decompositions for Y_t the true series and y_t the observed series are

$$Y_t = T_t + S_t + \varepsilon_t = N_t^* + S_t$$

and

$$y_t = Y_t + e_t = T_t + S_t + I_t = N_t + S_t,$$

where $I_t = \varepsilon_t + e_t$ includes both a random fluctuation and a sampling error term, assumed to be uncorrelated. Starting from a linear approximation to X-11,

$$\hat{N}_t = \Sigma\ w_{jt} \cdot y_{t+j},$$

three different variances are considered,

(1) $\mathrm{Var}\,(\hat{N}_t | T,S)$

(2) $\mathrm{Var}\,(\hat{N}_t - N_t | T,S)$

(3) $\mathrm{Var}\,(\hat{N}_t - N_t^* | T,S)$

(1), a conditional variance of the seasonally adjusted series, includes ε_t and e_t. (3) represents the variance with respect to the realized population value $T_t + \varepsilon_t$. Supporting (3), we may say that what counts is the realization that we get, say the unemployment for July, 1990, rather than some mean across the infinite realizations that are possible for each month. On the other hand, we may argue that two series are not equally reliable, when one series has a larger e_t component over time, suggesting the usefulness of (1). Even conditioning on T_t and S_t, N_t is a random variable, so (2) is not the same as (1). (2) is more limited,

$$\mathrm{Var}(\hat{N}_t - N_t | T,S) = \mathrm{Var}(\hat{S}_t - S_t | T,S).$$

This form often appears in model-based papers, except that the conditioning is with respect to the data $\{y_t\}$. Healthy aspects of this paper are airing these alternatives and the issue of conditioning and reviewing earlier work starting with Wolter and Monsour (1981).

Using the linear approximation to X-11, in the central section of the series, Pfeffermann obtains

$$R_t \approx \Sigma\, a_j \cdot I_{t+j},$$

where I_t is the irregular component as defined above and R_t is the estimated irregular from X-11. This equation exhibits dependencies among $\{R_t\}$, implying that X-11 irregulars cannot be expected to behave like white noise. From this basic equation, assuming higher order autocorrelations are 0, he obtains a simple linear system

$$U = A\ V,$$

U and V vectors of nonzero autocovariances for R_t and I_t, respectively. Using the usual estimate for the k^{th} order sample autocovariance for R_t, V is estimated from the linear system.

The paper provides an iterative procedure for estimating the cutoff C for nonzero covariances, plus some overall advice in the context of rotating panel designs for unemployment surveys , where the examples are drawn.

The proposed method gives reasonable U-shaped variance curves in the simulations and examples. For Total Unemployed Men in Canada, (3) gives standard deviation values smaller than standard deviations for the unadjusted series in the center of the series and larger at the ends, and about 10% higher than values from one of the Wolter-Monsour estimates, consistent with (3)'s accounting for the sampling error term e_t. The method is appealing in its simplicity and in not requiring modeling. In the basic form presented here, it would suffer when a series experiences larger fluctuations in the most recent year or two.

Dagum & Quenneville :

The authors set up a common, simple unobserved components model (UCM) to approximate X-11. They use Kalman filter and fixed interval smoothing techniques to estimate model parameters. If a comparison of estimates from X-11 and the model indicate a reasonable goodness-of-fit, then an estimated covariance matrix from the UCM formulation is used for the series being adjusted with X-11. (For simplicity in this discussion, "X-11" is intended to include X11ARIMA, Statistics Canada's widely used extended version of the basic U. S. Census X-11). Nice features of the paper include (1) a detailed discussion of the numerical optimization routine carried out in obtaining maximum likelihood estimates for signal-to-noise ratios, a spelling out of the differences in handling logarithmic or multiplicative decompositions, and a detailed analysis of the 1981-82 recession period. In particular, the authors find that the UCM seasonally adjusted series behaves similarly to the X-11 series in this time interval, probably the most sensitive part of the series under examination.

Turning to some points for further consideration, it is seen that the method matches X-11 with a UCM, while an ARIMA model may be used for extrapolation in the seasonal adjustment program. It seems a little unsatisfying to use two differing model formulations in the method.

A "compromise" UCM is used, in part because earlier work on models approximating X-11 refer to the standard filter options. Consideration of this method provides additional incentive for deriving approximating models for each of the limited number of combinations of X-11 trend and seasonal filter options. Also, the seasonal part of the UCM model seems "simpler" in some sense than X-11, suggesting that a check for residual seasonality in the UCM seasonal adjustment be made.

A chi-square test on standardized differences between the two seasonally adjusted series is applied to check for suitability of the UCM formulation. As the authors state, this is considered an indicator, not a formal statistical test. While such a goodness-of-fit test is natural to consider in this case, some additional thought might be given on the appropriate test or suitability check.

As stated above, there appears to be a reasonable overall fit between the two seasonally adjusted series in the two unemployment examples. However, from the graphs, it appears that there are some seasonal patterns to the differences prior to the recession period. The estimated variance curves for level and change are U-shaped, except that the curve for U. S. unemployment level is increasing.

Findley:

This paper on time series modeling does not discuss variance estimation, but has relevance to the topic. One of the drawbacks stated for model-based seasonal adjustment, which permits variance estimation, is identification and selection of models. Findley has worked for some time on model selection criteria and this paper represents his *current* advice. Previously, he has recommended AIC. In this paper, he compares AIC to two new criteria. (Also, as mentioned above, Findley and Monsell, 1990, present a resampling approach to variance estimation in the X-11 setting).

Still based on examining likelihoods, the paper presents much technical work on likelihood functions in rather general settings and derives statistics for comparing nonnested models. Basically, given parameter spaces $\Theta^{(1)}$ and $\Theta^{(2)}$ associated with two families of models, the test examines differences in log likelihoods evaluated at maximum likelihood values for the parameters. The model from family (1) or (2) is preferred according as the difference is positive or negative, with magnitude exceeding a critical value defining an inconclusive band around zero. The associated limit theorem is general enough to include cases where neither family includes the true model. The results suggest a graphical technique, in both a full-blown and shortcut form,

(F) plot $L_M(\hat{\theta}_M^{(1)}) - L_M(\hat{\theta}_M^{(2)})$, $N/2<M\leq N$,

(S) plot $L_M(\hat{\theta}_N^{(1)}) - L_M(\hat{\theta}_N^{(2)})$, $N/2<M\leq N$.

(F) is computer-intensive, in requiring maximum likelihood estimates for parameters for a large number of values of M. (S) uses likelihoods based on partial samples of size M, but with maximum likelihood parameters based on all N observations. With iid observations the values L_M are available as partial sums; for time series models computed with a Kalman filter algorithm, the L_M's can also be generated in a single pass of the data. A limit theorem provides justification for (S).

The most specific limit theorem states:

$$2 \cdot \frac{L_N(\hat{\theta}_1) - L_N(\hat{\theta}_2)}{\sqrt{N}} + \sqrt{N} \cdot \log\frac{\sigma^2(k^1)}{\sigma^2(k^2)} \xrightarrow{\mathcal{L}} N(0,\upsilon^2)$$

Under the null hypothesis H_0: $\sigma^2(k^1)=\sigma^2(k^2)$, the second term drops out. Thus,

$$Z = 2 \cdot \frac{L_N(\hat{\theta}_1) - L_N(\hat{\theta}_2)}{\sqrt{N}\cdot\hat{\upsilon}}$$

can be used as a test statistic for comparing two competing models. The above null hypothesis, called "weak equivalence" in the paper, can be described as the competing models having equal mean square forecast error. The statistics Z_{YW} and Z_{GM} applied in the paper correspond to using estimates $\hat{\upsilon}$ based on Yule-Walker type estimates and more robust estimates (based on an S-PLUS computing routine), respectively.

The graphical procedure and tests using Z_{YW} and Z_{GM} are applied to a set of 43 economic time series modeled by Bell and Pugh(1989), and results compared to their use of AIC for model comparisons. The table below summarizes results. Z_{YW} is not sensitive to differences in models, giving justification for Z_{GM} with its robust standard error estimate. The graphical analysis and Z_{GM} are quite consistent, as might be expected, and are also close to the AIC results. Z_{GM} and the graph have the advantage of indicating whether a difference in models is significant or not more clearly than AIC. The four series I (rather arbitrarily) designated as inconclusive from AIC have differences less than one in magnitude. Thus, in spite of the results appearing in the table, the graph and Z_{GM} may not really be less sensitive.

The above limit theorem and test are time series analogues of results of Vuong(1989) for iid observations. Issues remain in the choice of $\hat{\upsilon}$, which is a problem in estimating a spectral density. The theoretical results of the paper are daunting to one tending to focus on the applied side. The new statistics still appear very similar to AIC, and lack its appealing feature of a penalty function for the number of parameters in a model. Both the theoretical results and the graphs for some of the series motivate and may add to the reader's intuition for any of these likelihood statistics. Also appropriate is the emphasis on formulating limit theorems not requiring either class to contain the true model. Interesting to me is the preference in most cases for ARIMA models over UCM's. While UCM's are really more general, the forms in use right now don't seem to be rich enough, particularly the part for the seasonal ($\Sigma S_{t-j}=\varepsilon_t$).

Hausman & Watson:

As in their 1985 JASA paper, the authors tackle a specific problem and propose solutions which go beyond standard treatment. In this paper, they seek time series models for the error in preliminary survey estimates. This enables them to carry out model-based seasonal adjustment and obtain variances of seasonally adjusted estimates in a more careful way. The models offer potential for improving the (unadjusted) preliminary estimates. As they point out, this is an important topic, since it is the preliminary estimates for key series such as the building permit series they study which are used by policymakers and forecasters. Some work of a similar nature has been initiated at BLS for the monthly establishment survey of employment. An additional feature of the paper is permitting heteroscedastic innovations in the seasonal component.

Census' building permits series lend themselves to this analysis, since an annual census is available for benchmarking monthly survey values. A substantial proportion of the population is covered by the survey, which has a cut-off design covering all large permit-issuing offices. Even after accounting for nonresponse, preliminary and revised figures include roughly 60% and 70% of the population, respectively.

Before focusing on the modeling, let me comment on the main assumptions. I don't propose changing them, but would like to have seen a little more discussion of them. The final census annual value is assumed to contain no error. Often, data from a census are less accurate than data from a survey. In this case, I have no reason to believe that response or measurement error is very large. ("Measurement error" as treated in the paper basically refers to sampling error). The second assumption is that the monthly figures derived from the annual census benchmark contain no error. The final monthly series is constructed by constraining the estimates to sum to the census annual total subject to minimizing revisions in the month-to-month changes. There may be some biases stemming from response patterns for individual months.

Here are the models for y_t, y_t^p, and y_t^r, the final, preliminary, and revised estimates, respectively:

$$y_t = N_t + S_t,$$

$$N_t = \tau_t + \varepsilon_t,\ (1\text{-}B)\tau_t = e^{\tau}_t,\ \Sigma S_{t-j} = e^{S}_t,$$

$$y^p_t = y_t + u^p_t,$$

$$u^p_t = -.01 + .24\, u^p_{t-1} + .77\, a^r_t - .19 a^r_{t-1} + e^p_t,$$

$$y^r_t = y_t + u^r_t, \quad u^r_t\ \text{AR(1)}.$$

The trend model is simple, since the series is described as "volatile and trendless." Some care is taken before settling on the forms for u^r_t and u^p_t, plus treating the errors as uncorrrelated with $\{y_t\}$.

Changing the seasonal model to permit different variances by month leads to a fair amount of improvement in the likelihood. A heteroscedastic model based on three groups of months (one group of six having zero variance) achieves nearly all the improvement in the likelihood with one additional parameter rather than 11 with the unrestricted heteroscedastic model.

I am pleased to see some treatment of this issue. Most of the time seasonality is concentrated in a few months, with several months having negligible contributions to seasonality. X-11 addresses this issue to some extent by permitting different seasonal filters by month, but this feature is not commonly used. Findley and Monsell (1990), in their resampling approach to variance estimation, found it useful to do some grouping of months in carrying out their resampling.

Hausman and Watson report substantially lower variances for their model-based seasonal adjustment, compared to X-11. They rather casually refer to their X-11 variance estimates as "using standard calculations." It appears that they have used something similar to Pfeffermann's (2), to put it on a similar footing to the model-based calculations. I accept the finding of a lower variance, but would like to see comparisons of other aspects of the adjustment, e. g., the factors themselves. Good modeling skills are employed to achieve the improvement.

References (not cited by authors):

Buszuwski, James A. and Scott, Stuart (1988), "On the Use of Intervention Analysis in Seasonal Adjustment," Proceedings of the ASA Business & Economic Statistics Section, pp. 337-342

Findley, David F. and Monsell, Brian C. (1990), "Standard Errors for Seasonal Adjustment: A Resampling Approach," Proceedings of the ASA Business & Economic Statistics Section

Table (Findley application). Model Preferences for 43 Series from Alternate Likelihood Statistics

	AIC	Graph	Z_{GM}	Z_{YW}
ARIMA model preferred	36	27	27	12
UCM preferred	3	3	4	0
Inconclusive	4	13	12	31