The Evolution of Seasonal Adjustment at BLS: From the Ratio-to-Moving Average to Model Based Methods

Dick Tiller
Bureau of Labor Statistics

1st Seasonal Adjustment Practitioners Workshop

Washington, DC , November 4, 2016

Introduction

- My talk is about the extent of my career in SA at BLS with emphasis on modeling & research we do that is different from the conventional X-11 & SEATS approach
- My career at BLS spans 4 decades which covers all the SA methods used by BLS since the pre-computer era

Chronology of Seasonal Adjustment at BLS

• BLS Seasonal Factor Method	1960-80
• X-11	1972
• X-11 ARIMA	1980
• BLS Intervention Analysis	1988-95
• Seasonal Adjustment Accounting for Sampling Error	1994
• X-12 ARIMA	1996
• Variances for X-11 Accounting for Sampling Error*	1997
• X-13A-S	2011

*on going research

BLS Seasonal Factor Method (SFM), 1960-80

- BLS's entry into the era of computerized seasonal adjustment began in 1960
- Developed in reaction to widespread dissatisfaction with Census Method II's adjustment of the CPS labor force survey during 1958 recession (Raft & Stein, 1960, Shiskin, 1961)
- During its prime years used throughout BLS & outside by State agencies, foreign governments, & private firms, reviewed by the Gordon Committee (1962)

BLS Seasonal Factor Method

- Based on the ratio-to-moving average method as was Census Method but it did some things differently
  - One of the innovations, control charts with graduated weights to discount extreme observations was adopted by X-11 (Alan Young's preface to Ladiray & Quenneville, 2001)
- Could not keep up with the feature rich X-11, slowly phased out during the 1970's as X-11 became world-wide standard

X-11 ARIMA

- In 1979 the Levitan Commission recommended BLS
  - Adopt the use of ARIMA back-casts & forecasts
  - Adopt concurrent method of SA
- BLS adopted the forecast extensions of X-11 ARIMA in 1980
- Did not accept concurrent seasonal adjustment until 2004

BLS Intervention Analysis, 1989-95

- Frequent occurrence of large & persistence outliers primarily in Consumer Price & Producer Price Index series
- No way to deal with them in X-11 ARIMA
- Buszuwski & Scott (1988) used ARIMA models with intervention effects to estimate prior adjustments for X-11-ARIMA

BLS Intervention Analysis, 1989-96

- The development of X-12 ARIMA was of special importance with its ability to handle all kinds of aberrations in time series with RegArima & automated outlier detection option
- Later the incorporation in X-12 of automated true ARIMA model identification (TRAMO) made ARIMA modeling & outlier detection even easier

X-13A-S

- Of all of the other competitors to X-11 over the years, only SEATS was compelling enough for Census Bureau to include it as an alternative to X-11
- Provides a unified model-based pre-adjustment framework & a common set of diagnostic & evaluative tools
- BLS Seasonal Adjustment Methodology Team (2007) recommended adoption of X-13 as the BLS official software for SA & for the program offices to consider the SEATS option as an additional tool

Seasonal Adjustment Accounting for Sampling Error

- Although X-13 SA with its X-11 & SEATS options work well for a wide variety of economic time series, results less satisfactory for series generated from **periodic surveys where sample sizes are small**
- Many statistical agencies routinely seasonally adjust survey series with either option but by ignoring SE may lead to major distortions in the adjustments
- There is a 4th component, sampling error (SE), which needs to be added to the conventional decomposition

Small Area Estimation Problem

- Problem of seasonally adjusting survey data is closely related to the small area estimation problem studied in a time series context by Scott, Smith, & Jones (1977), Bell & Hillmer (1987, 1990), Binder & Dick (1989), & Pfeffermann (1991)
  - These studies apply signal extraction techniques to periodic survey data, where SE is treated as an unobserved component with known variances & autocorrelations
  - Objective: to improve on direct survey estimator or, equivalently, to purge the survey series of SE

Small Area Estimation Problem

- When seasonality is also present this approach has important implications for seasonal adjustment

- In 1994 BLS adopted this TS approach in the Local Area Unemployment Statistic program (LAUS) for reducing high variability in CPS State labor force series
  - Early LAUS models were not based on the classical TS decomposition
  - We used a two step approach to get SA estimates: first use our model to remove sampling error & then apply X-11

Seasonal Adjustment with Sampling Error

- In the classical decomposition, the time series, Y_t , is assumed to be observed without error & to be decomposable, given some type of model, into non-seasonal, N_t, and seasonal, S_t, components.
- We consider an additive decomposition (perhaps after logging)

$$Y_t = N_t + S_t, \quad N_t = T_t + I_t$$

Signal Extraction with Sampling Error

- When the observed series is generated by a periodic survey, a 4th unobserved component, survey error, e_t, independent of Y_t, is added to the classical decomposition

$$y_t = Y_t + e_t$$

where,

$$e_t \sim N(0, \sigma_{t,t}), \quad E(e_\tau, e_t) = \sigma_{\tau,t}$$

Signal Extraction with Survey Error

- When SE is present, there are two types of decompositions associated with different target values
  - **Ignoring Survey error (ISE)**,

$$y_t - S_t = N_t^* \neq N_t$$

$$N_t^* = N_t + e_t = T_t^* + I_t^*$$

  - **Accounting for SE (ASE)**

$$y_t - (S_t + e_t) = N_t$$

What happens when we ignore SE when its variance is large?

- Large sampling errors contribute variability to all components
- Some components may be more affected when survey error correlations are similar to the correlations in the true series
  - Strong positive SE autocorrelations at low lags generate spurious cycle-like variation in the trend
  - SE autocorrelations at 12 month lag increase variability of seasonal component

Labor Force Surveys

- Of special interest are the various labor force surveys conducted by government agencies around the world to collect statistics on employment & unemployment
- Most important real time indicators of current economic activity

Common Features of Labor Force Surveys:

- Repeated at regular intervals, monthly (U.S., Canada, Australia, Israel, Brazil) or quarterly (EU countries, New Zealand)
- Sampling units are households
- Units retained in sample over multiple periods according to a rotating panel scheme
- Small samples common for many sub-domains of interest
- Underlying true series have strong seasonal movements

Maryland CPS Employment Series

- To account for SE use a two step process

  Step 1: Estimate & remove SE from the survey series

  - Combine an ARIMA model of the true series with information on SE variances & covariances

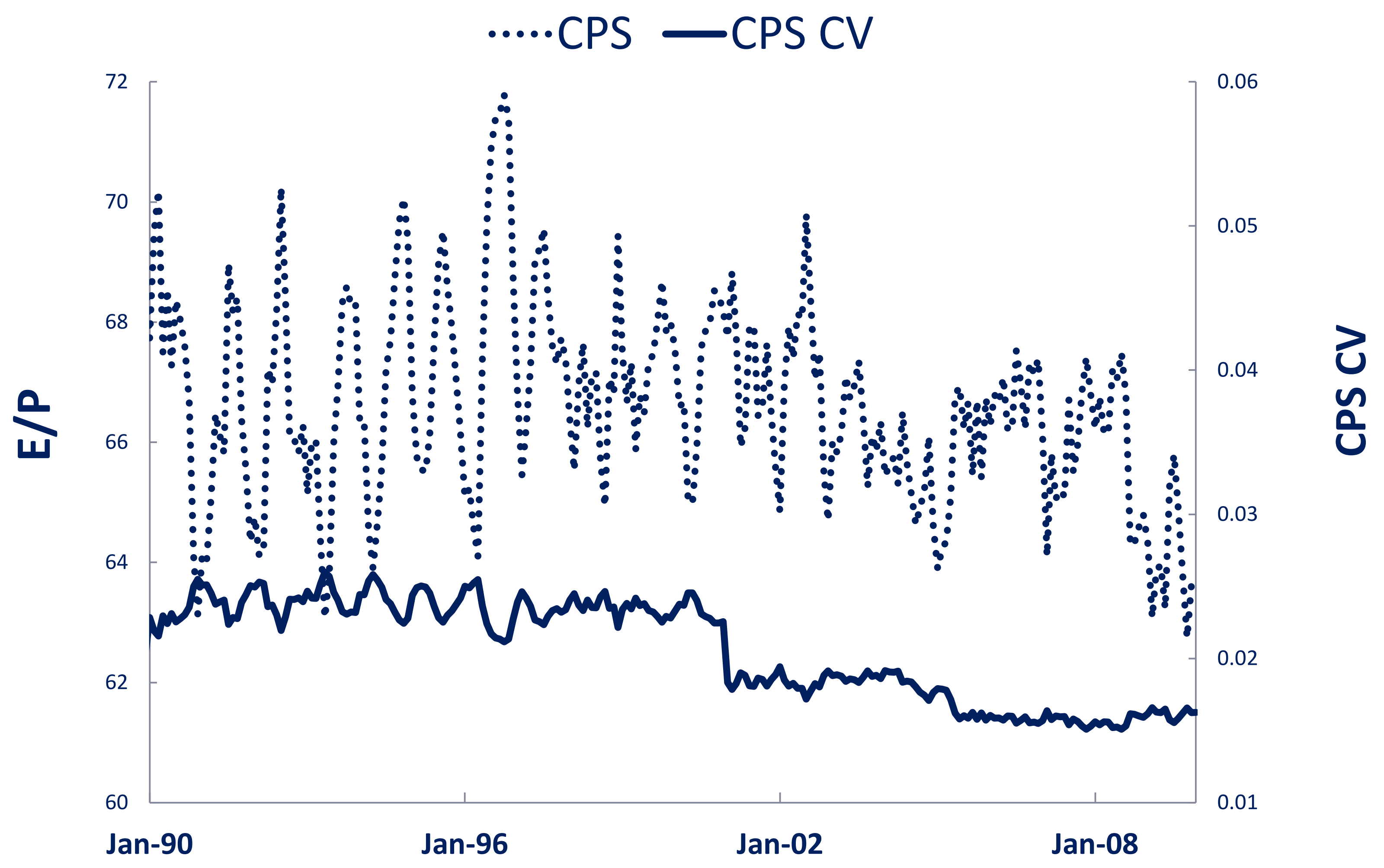
  Step 2: Decompose the estimated true series

  - Using either the X-11 or SEATS option where the ARIMA model is fixed to correspond to the estimated model of the true series

Maryland CPS Employment Series

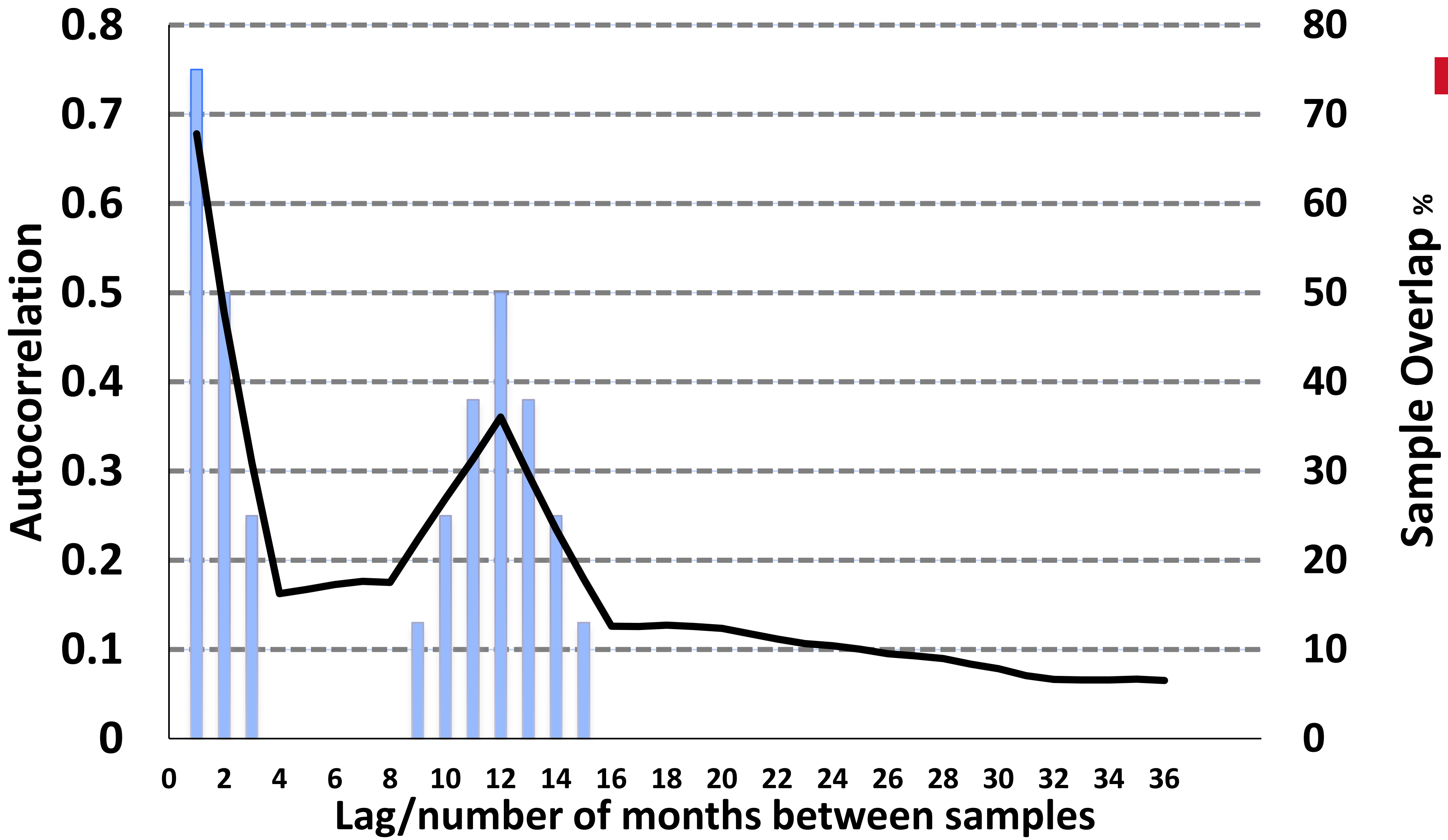
- Compare with decompositions that ignores SE
  - How important is ASE?
  - Does it make any difference which decomposition we use?

Example: CPS EP ratio for Maryland, 1990 to 2009



- Redesigns, sample size changes, & fluctuations in population proportions result in heteroskedastic SE

CPS Employment SE Autocorrelations



- 4-8-4 rotating panel design generates major overlaps between samples separated by 1 or more months

Sampling Error Model

- To account for both autocorrelation & heteroscedasticity the survey error is modeled as follows,

$$e_t = \sigma_{e,t} e'_t$$

$$e'_t = \sum_{i=1}^{15} \phi_i e'_{t-i} + \nu_t \quad \text{(standardized SE)}$$

ϕ_i's are computed from the SE autocorrelations

ASE Model of the CPS

- Data are logged which implies a multiplicative model where SE is still additive,

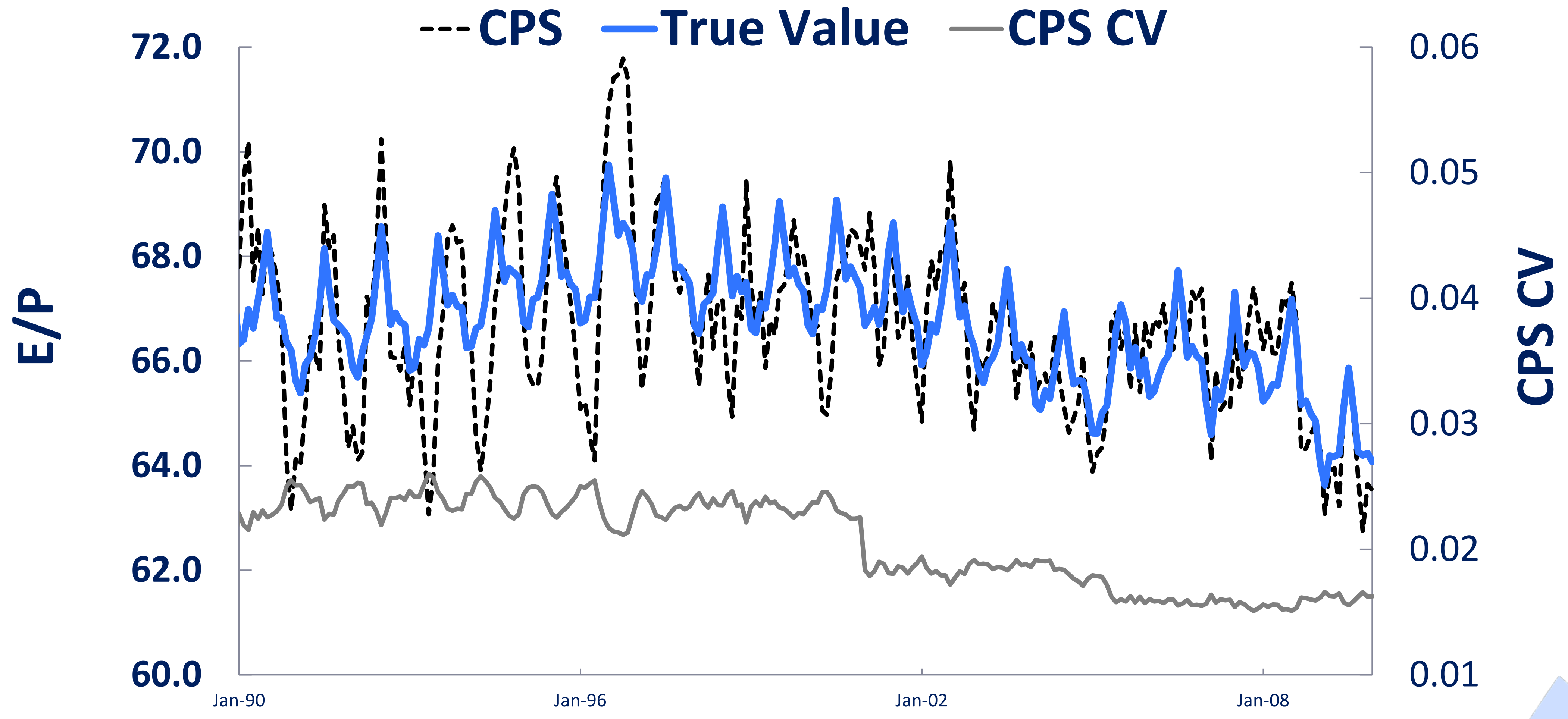
$$y_t = Y_t(1 + r_t), \qquad r_t = e_t / Y_t$$

- Model of true series with estimated parameter values

$$\nabla\nabla^{12} Log(Y_t) = (1 - 0.38B)(1 - 0.92B^{12})\nu_{Y_t}, \ \sigma^2_{\nu_Y} = 4.0 \times 10^{-5}$$

- Given these models, the Kalman filter & smoother is used to produce the MMSE estimates of the true series

Estimates of true series



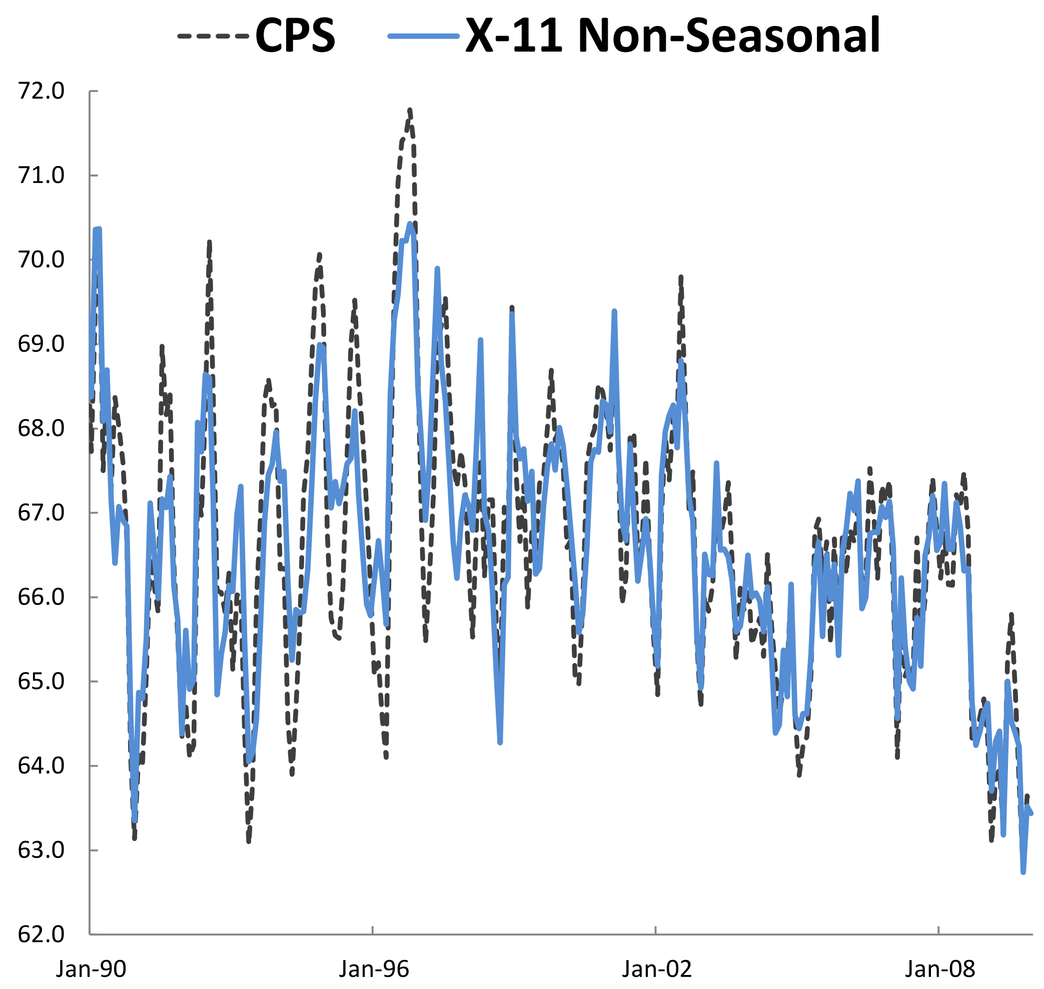
ISE Decomposition of the CPS

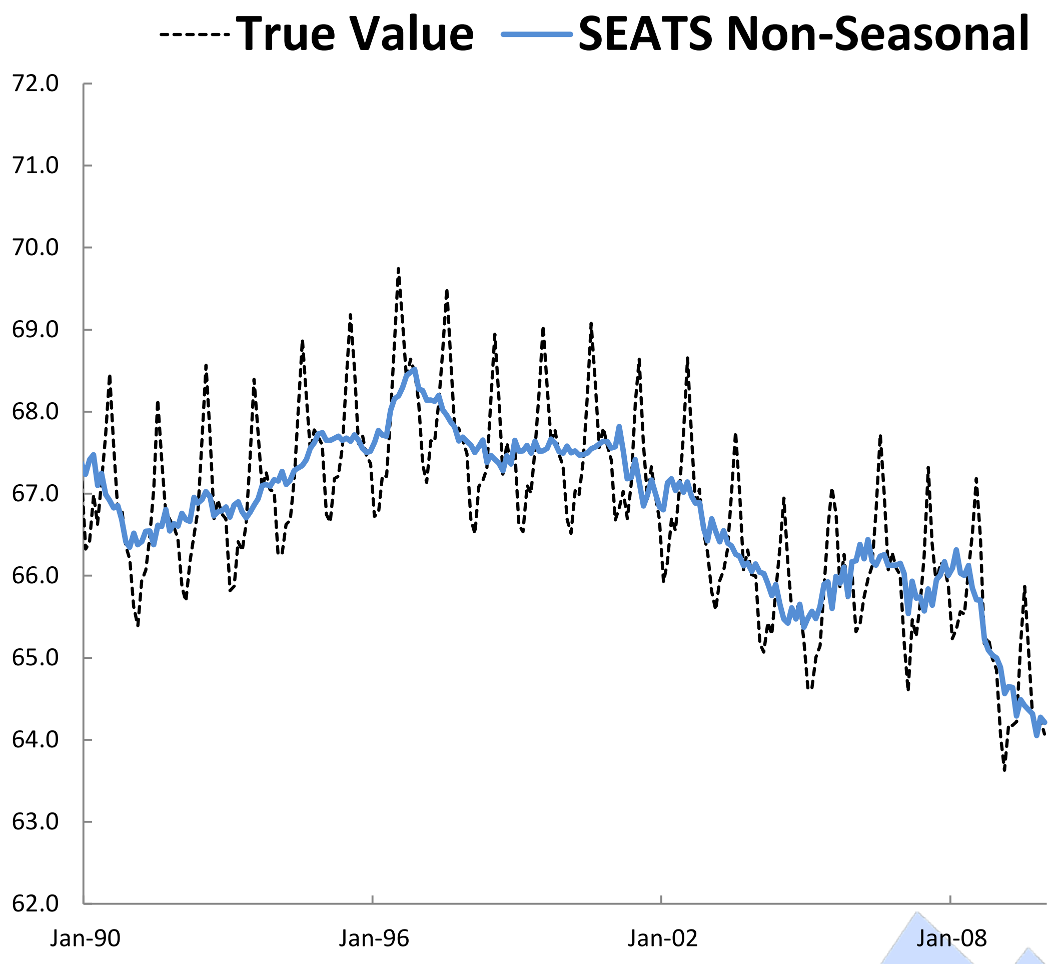
- Conventional approach: apply SEATS and X-11 options of X-13A-S directly to the CPS series
- Automodel option selects a (212)(011) ARIMA model
  - X-11 selects 3x5 seasonal & 13-term Henderson filters
  - SEATS substituted a (112)(011) model for the decomposition

ASE Decomposition of the CPS

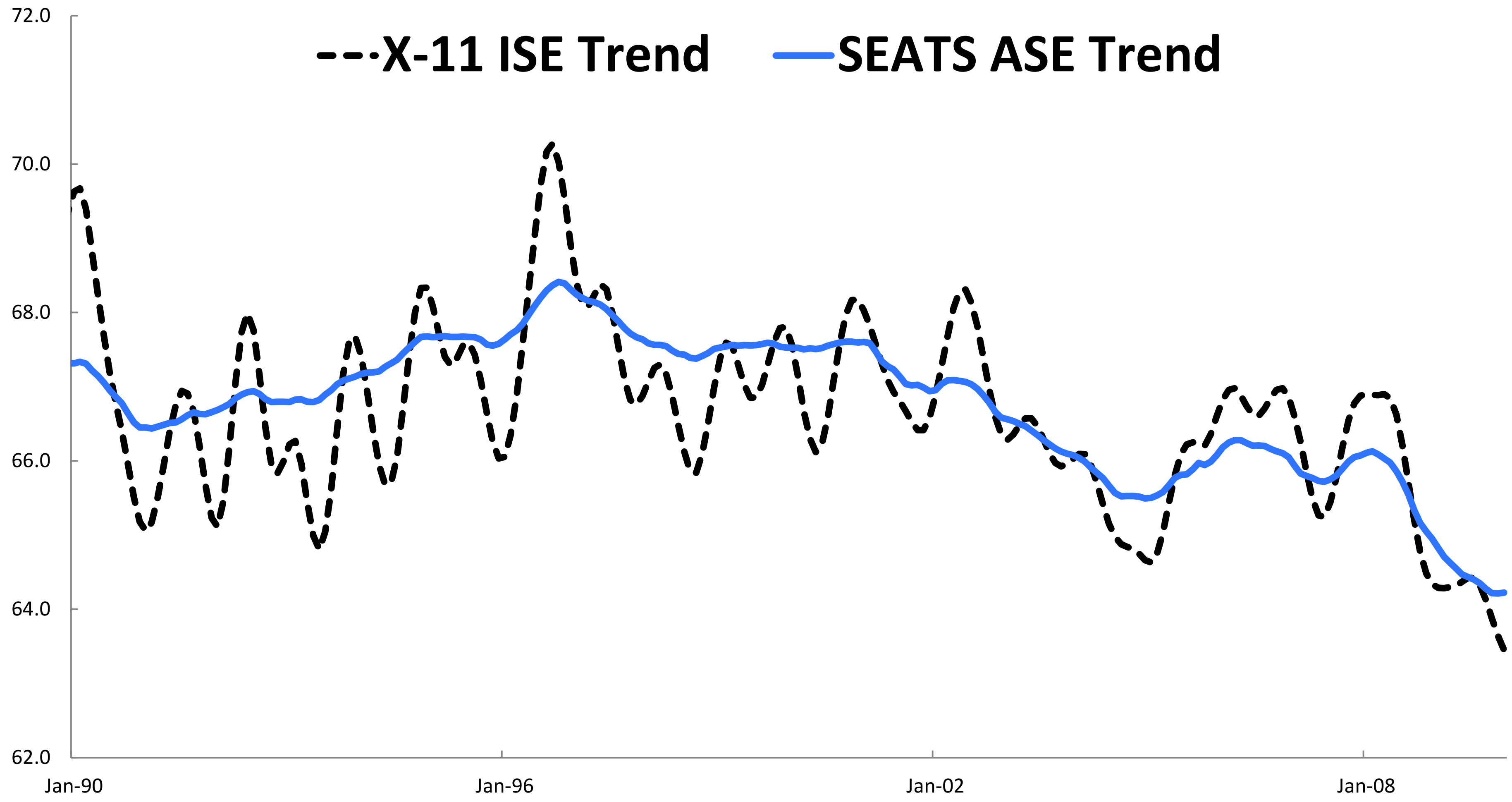
- Apply SEATS & X-11 options to the estimated true series
- For SEATS we fix the ARIMA model parameters to equal the ones estimated for the model of the true series—full model based approach
- We also use the same ARIMA model for the X-11 decomposition

X-11 ISE vs SEATS ASE, Non-Seasonal

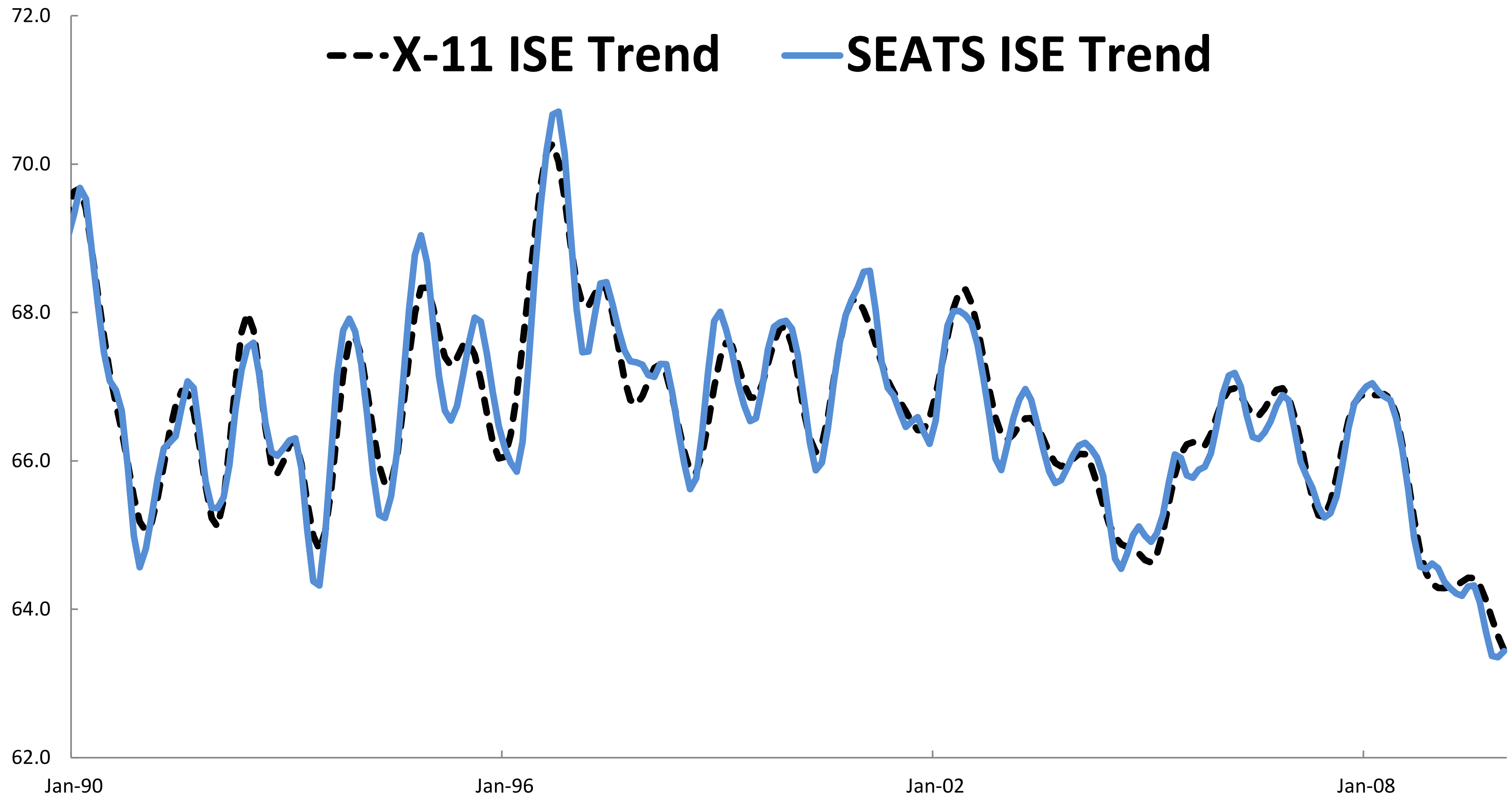




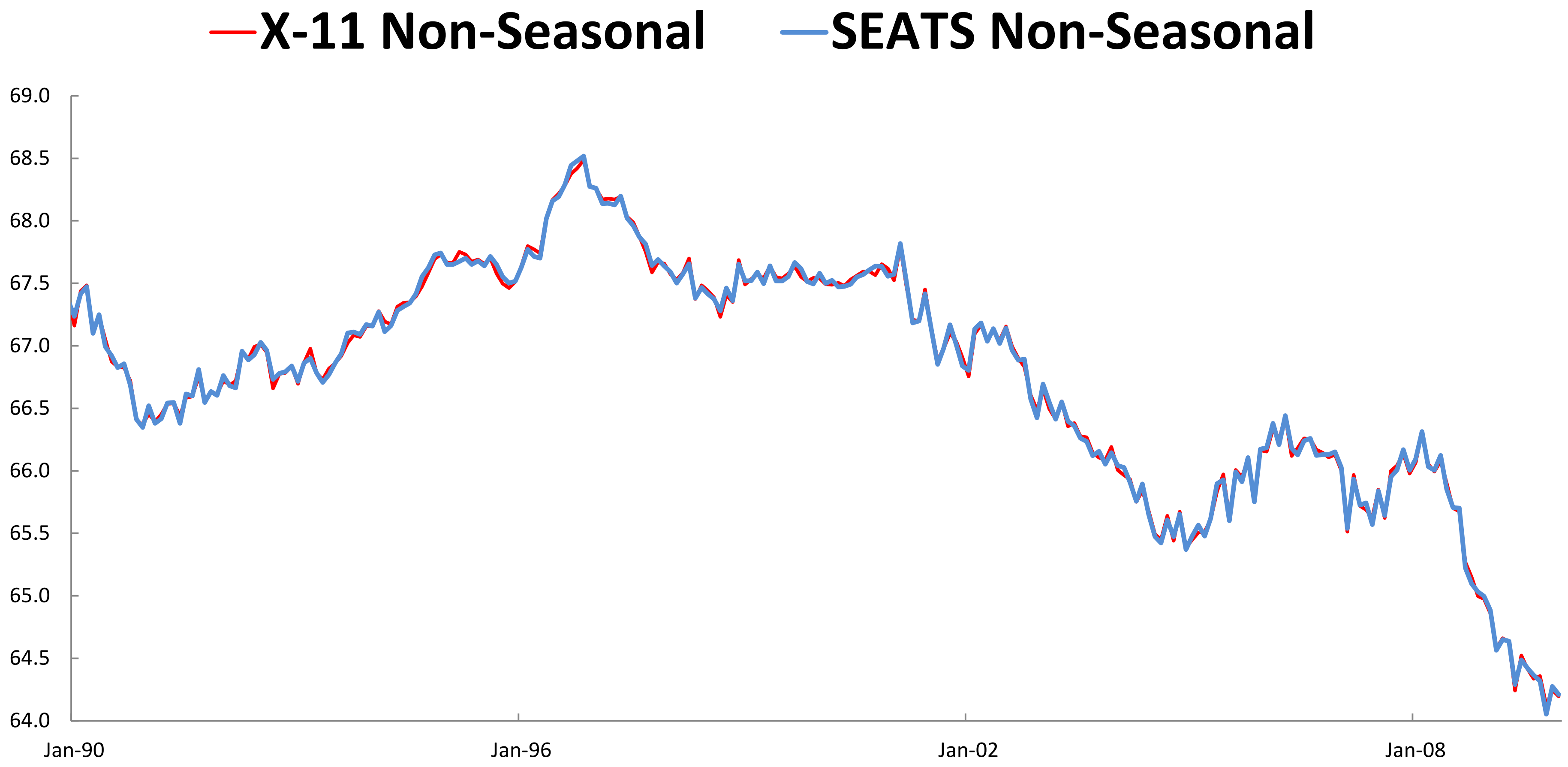
X-11Trend of Survey series, SEATS Trend of True series



X-11 & SEATS Trend for Survey Series



Non-Seasonal for True Series



Conclusions

- Accounting for survey error can be important in producing useful decompositions
- Which decomposition you use is less important

Variances for X-11 SA of Survey Series

- An often cited weakness of X-11 is that it does not provide variances
  - Model-based approach begins with models of the observed & unobserved components that define what is being estimated & then derive the estimators & their variances
  - X-11 skips the first step so that it is not obvious how to compute variances

Variances for X-11

- Current practice is to use standard errors reported for the direct survey estimators before seasonal adjustment
  - Assumes that SA does not affect the precision of the estimates
- Wolter & Monsour (1981) argued that this was wrong on several levels
  - Reporting a variance for an SA estimate based on classic finite sampling theory of true values being fixed is inconsistent with the conceptual basis of seasonal adjustment where true values are stochastic & strongly autocorrelated

Variances for X-11

- Even if we concede that a variance measure based only on sampling variability is useful, it is still not correct to treat the variance of the direct survey estimator as equivalent to the variance of its SA estimate

- Wolter & Monsour's (first) approach to variance estimation is based on the well-known property that X-11 can be closely approximated by a weighted moving average of the observed survey data

Variances for X-11

- Treating the true values as fixed, the error in the estimator is the same weighted moving average of the sampling errors from which its variance estimator can be easily computed given the survey error variances & lag covariances.

- No follow up to this insight until Pfeffermann (1994), which reignited interest (Bell & Kramer, 1999 & others)

Work at BLS on Pfeffermann's Approach

- Pfeffermann &Scott (1997),
- Pfeffermann, Scott, & Tiller (2000),
- Scott, Sverchkov, & Pfeffermann (2012),
- Pfeffermann & Sverchkov (2014)
  - MSE measures where the target values is the seasonally adjusted true series using the X-11 symmetric filters, almost unbiased at series center
  - Make a bias correction towards the ends of the series where asymmetric weights must be used in place of the symmetric weights

Pffermann's Approach

- Conceptually it produces a variance for seasonal adjustment of a survey estimator that conforms to the model-free design-based survey sampling approach familiar to statistical agencies
- Easy to implement given survey error variances & covariances

Replication Method

- Evans, McIllece, & Miller (2016) propose a similar approach to seasonal adjustment variances based on pseudo-replication method
- 160 replicate series for CPS employment and unemployment were created each month for 2003-2014
- Each replicate was SA, variances computed from the dispersion of the replicate SA's around the full sample SA

Replication Method

- Generalized variance functions used to smooth the monthly SA variances
- Not surprisingly, the variances for the seasonal adjusted CPS series are less than the variances for the unadjusted series & in some cases substantially less
- A comparison made to the Pfeffermann-Sverchkov method showed little difference between the two methods

Concluding Thoughts

- X-13 is a remarkable achievement building on 60 years of research & development since Julius Shiskin's first attempt to computerize SA in 1955
  - His goal was to make SA practical on a large scale for the entire world
  - Required strong commitment & dedication to overcome early technical problems & convince people to use it.
  - Luckily his successors were equally committed ---Estella Dagum, David Findley, Agustín Maravall

Concluding Thoughts

- Hopefully this history of steady improvements & major innovations will continue in the future
  - Series with sampling error
  - High frequency data (weekly, daily)

References

- Bell, W.R., & S.C. Hillmer (1987), "Time Series Methods for Survey Estimation," Research Report 87/20, Statistical Research Division, Bureau of the Census.
- Bell, W.R., and S.C. Hillmer (1990), "The Time Series Approach to Estimation for Repeated Surveys," *Survey Methodology*, pp. 195–215
- Buszuwski, J. A., & Scott, S. (1988), "On the Use of Intervention Analysis in Seasonal Adjustment," Proceedings of the Business and Economic Statistics Section, ASA
- Evans, T.D., McIllece, J.J., Miller, S.M. (2016), "Variance Estimation by Replication for National CPS Seasonally Adjusted Series," Proceedings of the Business and Economic Statistics Section, ASA
- Pfeffermann, D. & S. Scott. 1997. "Variance Measures for X-11 Seasonally Adjusted Estimators: Some Developments with Application to Labor Force Series." In Proceedings of the Section on Business & Economic Statistics: ASA, 211–216

References

- Pfeffermann, D. (1991), "Estimation and seasonal adjustment of population means using data from repeated surveys, " *Journal of Business and Economic Statistics*, 9, 163-177
- Pfeffermann, D., S. Scott, & R. Tiller (2000), "Comparison of Variance Measure for Seasonally Adjusted and Trend Series," *Proceedings of the 2nd International Conference on Establishment Surveys*, ASA, 755-764
- Scott, S., M. Sverchkov, & D. Pfeffermann. 2012. "Estimating Variance in X-11 Seasonal Adjustment." In *Economic Time Series: Modeling and Seasonality*, edited by William R. Bell, Scott H. Holan, & Tucker S. McElroy, 185–210. London: Chapman and Hall.
- Raff M. & R.L. Stein, (1960), New Seasonal Adjustment Factors for Labor Force Components, *Monthly Labor Review*, Vol. 83, No. 8, pp. 822-827.
- Scott, A.J., & Smith, T.M.F. (1974), Analysis of repeated surveys using time series methods to the analysis of repeated surveys, *International Statistical Review*, 45, 13-28

References

- Shiskin, J., (1961), “Tests and Revisions of Bureau of the Census Methods of Seasonal Adjustments, Bureau of the Census, Technical Paper No. 5, Washington, D. C., p.2
- 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

Contact Information

Dick Tiller
Bureau of Labor Statistics
www.bls.gov
202-691-6370
Tiller.Richard@bls.gov