Shrinkage Estimators for Damping X12-ARIMA Seasonals

Don M. Miller

Virginia Commonwealth University

School of Business, Box 844000

1015 Floyd Avenue, Room 3112

Richmond, Virginia 23284-4000

(804) 828-8410

Fax: (804) 828-8884

dmmiller@vcu.edu

and

Dan Williams

Baruch College

17 Lexington Ave. C-301

New York, NY 10010

(212) 802-5983

Fax: (212) 802-5968

daniel_williams@baruch.cuny.edu

May 23, 2003

This research was supported by a grant from VCU School of Business Faculty Excellence Fund.

Shrinkage Estimators for Damping X12-ARIMA Seasonals

We examine the effect of damping X-12-ARIMA's estimated seasonal variation on the accuracy of its seasonal adjustments of time series. Two methods for damping seasonals are proposed. In a simulation experiment, we generated time series data for each of 90 distinct experimental conditions that, in aggregate, characterize the variety of monthly series in the M3-competition. X-12-ARIMA consistently overestimated the actual seasonal variation by an amount consistent with statistical theory. Damping seasonals reduced X-12-ARIMA's estimation error by as much as 73%, and under no conditions was estimation error increased beyond a trivial amount. Improvement depended primarily on the degree to which random variation in a series dominated seasonal variation. One of the proposed methods was somewhat more accurate, and is somewhat more complex, than the other. Other factors examined include the presence or absence of trend, asymmetry in the seasonal pattern, and constant vs. increasing seasonal variation over time. In an analysis of real data -- the 1428 monthly series of the M3-competition -- damping X-12-ARIMA seasonals prior to forecasting (1) reduced the average forecasting MAPE by 5.4% to 2.1% and (2) improved forecasting accuracy for 59% to 64% of the series, depending on the forecasting horizon. This research suggests that damping X-12-ARIMA seasonals leads to more accurate seasonal adjustments of time series, thus providing a more reliable basis for policy-making, forecasting, and the evaluation of forecasting methods by researchers.

Keywords: Damped seasonals, decomposition, empirical Bayes, seasonal variation, seasonality, shrinkage estimators, time series, X-12-ARIMA

1. Introduction

The X-12-ARIMA program is the primary method used for seasonal adjustment of government and economic time series in the United States, Canada, and the European Union. For example, it is used by the U.S. Census bureau and other U.S. government agencies to deseasonalize economic series before their official government publication. Originally developed by the U.S. Census Bureau, it is based on ratio-to-moving-average (classical) decomposition (Macauley, F.R., 1930; also described in Makridakis, et. al., 1998) and includes a great number of improvements that have been developed through empirical testing over the years, with the X-12-ARIMA variant having being released in 1996. Pierce (1980) cited the "widespread success" of the X-11 variant and stated that it was "the most common seasonal adjustment procedure in current use." His observations would seem to apply equally today.

X-12-ARIMA accomplishes seasonal adjustment through the development of a set of factors that account for the seasonal variation in a series. In this paper, we investigate the possibility that X-12-ARIMA tends to overestimate the amount of seasonal variation in a time series. This would in turn suggest the possibility of improving the accuracy of seasonal adjustment by "damping" the estimated seasonal variation, that is, using shrinkage methods to adjust the estimated seasonal factors toward a central value.

Armstrong (1978) concluded from prior research that damped seasonals would improve accuracy. In a previous paper (Miller and Williams, forthcoming), we show that damping classical decomposition seasonal factors improves accuracy. Since X-12-ARIMA has its roots in classical decomposition, it is reasonable to speculate that damping X-12-ARIMA seasonals might provide similar improvements. On the other hand, one might justifiably be skeptical, since X-12-ARIMA is widely considered to be substantially more accurate than classical decomposition.

In examining the impact of damping X-12-ARIMA seasonal factors on the accuracy of seasonal adjustments, we employ two shrinkage methods for damping seasonals and address the following questions:

1. Does X-12-ARIMA exaggerate seasonal variation? If so, how much and under what conditions?
2. Does damping of X-12-ARIMA seasonal factors improve estimation accuracy? How much, and under what conditions?
3. Does damping of X-12-ARIMA seasonal factors for seasonal adjustment prior to forecasting improve forecasting accuracy?
4. Which shrinkage method performs better? Under what conditions does one outperform the other?

In Section 2, we discuss X-12-ARIMA briefly, then the two shrinkage methods. In Section 3, we use two real time series to illustrate the effect of damping estimated seasonal variation. In Section 4, we present the results of a simulation in which, under a variety of controlled conditions, the use of damped seasonals is compared to using X-12-ARIMA alone in regard to the accuracy of estimated seasonal factors. In Section 5, we investigate the effect on forecasting accuracy of using damped seasonals rather than X-12-ARIMA alone for seasonal adjustment prior to forecasting. In Section 6, we summarize our findings and offer concluding remarks. Some important but more technical information is provided in the Appendix rather than the main text to promote accessibility to readers who are less interested in the mathematics and/or detail. These include the justification for shrinkage and details of X-12-ARIMA, our implementation of the shrinkage methods, the simulation, and the forecasting analysis.

2. Methods for Estimating Seasonal Factors

2.1. X-12-ARIMA

A brief overview of the X-12-ARIMA algorithm is provided in the appendix. A more thorough overview is provided by Makridakis, Wheelwright, and Hyndman (1998). More comprehensive descriptions of X-12-ARIMA and its earlier versions are provided by Ladiray and Quenneville (2001), U.S. Census Bureau (2000), Findlay, et. al. (1998), Dagum (1988, 1980), and Shiskin, et. al. (1967). X-12-ARIMA offers a choice among several decomposition models. The multiplicative model, in which the original value in a given period is decomposed into a trend-cycle value, a seasonal factor, and randomness (which X-12-ARIMA calls *irregular*) factor, is probably the most widely used. It can also include factors representing the effects of trading days and the Easter holiday, respectively.

X-12-ARIMA proceeds through three iterations whose primary purpose is to adjust for disruptive effects such as extreme values and calendar effects such as trading-days and holidays. In the final iteration, a series of final seasonal factors is produced, normalized so that their average over each 12-month period is approximately 1.0 (Ladiray and Quenneville, 2001). As described in the Appendix, the calculation of seasonal factors in the last iteration is similar statistically to the estimation of multiple population means, as in a randomized complete block experiment (one-way analysis of variance). Stein (1955) showed that, provided the sampling distributions of the sample means are normal distributions with equal variances, estimation error associated with the sample means is reduced by adjusting (i.e., *shrinking*) them toward their global mean. Since the X-12-ARIMA development of final seasonal factors is similar, though not identical, to the Stein/analysis of variance conditions, we suspected that damping X-12-ARIMA seasonals (shrinking them toward 1.0) would improve accuracy.

2.2 Two proposed shrinkage estimators of seasonal factors

Shrinkage estimators, sometimes called empirical Bayes estimators, were introduced by Robbins (1955) and Stein (1955) and further developed by many authors, notably Rutherford and Krutchkoff (1969), Efron and Morris (1973, 1975), and Morris (1983). Applications of shrinkage methods to forecasting problems have been developed by Greis and Gilstein (1991) and Bunn and Vassilopoulos (1999).

2.2.1 Global damping: the James-Stein shrinkage estimator

Following up on Stein (1955), James and Stein (1961) presented a simple shrinkage estimator for which improvement in estimation accuracy is substantial in certain situations. One of the shrinkage methods employed in this paper is an adaptation of the James-Stein estimator. The James-Stein estimator of the seasonal factor for period k of year j is

$$S_{jk}^{JS} = W_j^{JS}(1) + (1 - W_j^{JS}) S_{jk}, \quad j = 1, 2, \ldots, J; \; k = 1, 2, \ldots, K \qquad (1)$$

where $0 \le W_j^{JS} \le 1$, S_{jk} is the X-12-ARIMA seasonal factor for period k of the year j, and W_j^{JS} is the damping factor for year j, K = number of periods in one year (i.e., K = 12 for monthly data), and J = number of years. The higher the damping factor, the greater the damping. That is, no seasonal damping takes place if $W_j^{JS} = 0.0$ and complete damping (i.e., no estimated seasonal variation) is accomplished by using $W_j^{JS} = 1.0$. After damping, the seasonal factors remain normalized.

Armstrong and Collopy (2000) have proposed an ad hoc estimator in the form of expression (1), in which the damping parameter is determined by the number of observations, subjective information about seasonal variation, historical errors in estimating seasonal factors, the number of periods ahead in the forecasting horizon, and the assumed rate of deterioration of seasonal variation over the forecasting horizon. In contrast, the damping parameter W_j^{JS} is based entirely on the data and is defined as follows:

$$W_j^{JS} = \left(\frac{K-3}{K-1}\right) * \frac{\hat{V}}{\hat{V} + \hat{A}_j} \tag{2}$$

where V = the variance of the sampling distribution of the individual seasonal factor S_{jk} and A_j = the variation among the true seasonal factors for year j. (Explicit definitions of V and A_j -- and how they are estimated from the data -- are provided in the Appendix.) The coefficient (K– 3)/(K– 1) is a modifier suggested by Morris (1983). For simplicity, we have assumed that all seasonal factors (for all years) have the same variance V. Under this assumption, V is determined by the amount of random variation in the data (see Appendix). Because A_j can vary from year to year (as assumed in the X-12-ARIMA model), W_j^{JS} can also vary from year to year.

An examination of (2) suggests that the amount of damping needed is greater when V is relatively large compared to A_j, that is, when random variation dominates seasonal variation. We refer to this James-Stein-based method as *global damping* to convey the idea that the same amount of shrinkage (i.e., the same damping parameter) is applied to all of the seasonal factors in a given year. The average value of W_j^{JS} plays a central role in the interpretation of results in this research and is used as an important characteristic of each series examined.

2.2.2 Local damping: the Lemon-Krutchkoff shrinkage estimator

The second shrinkage estimator that we propose is adapted from a method developed by Lemon and Krutchkoff (1969), which builds on Robbins (1955). The LK seasonal factor for a given period (say, period *k**) in a given year *j* is a weighted average of the twelve X-12-ARIMA seasonal factors for that year, as follows:

$$S_{j,k^*}^{LK} = \sum_{k=1}^{12} W_{j,k,k^*} S_{j,k}\,, \qquad j = 1, 2, \ldots, J;\;\; k = 1, 2, \ldots, K \qquad (3)$$

These factors do not necessarily average 1.0 for a given year and therefore were normalized.

For example, the LK seasonal factor for July 2002 would be a weighted average of the twelve X-12-ARIMA seasonals for 2002. The weight given to each of the twelve X-12 seasonals is proportional to its proximity (in magnitude) to the July 2002 X-12 seasonal factor. Thus, the greatest weight is given to the July 2002 X-12-ARIMA seasonal itself. All other X-12-ARIMA seasonals are given smaller weights. Months whose seasonal factors differ substantially from the X-12-ARIMA July seasonal receive negligible weight. The determination of weights is explained in greater detail in the Appendix. Note that the seasonal factor for each month is based on a unique set of weights. To reflect this characteristic, we refer to this adaptation of the Lemon-Krutchkoff method as *local damping*.

To better understand local damping conceptually, suppose you wished to estimate the mean annual income for adults in your county, but you had no data. However, data are available for 11 other counties, some of which are thought to be quite similar to your county. Then it would be natural to use an average of the means for the similar counties as the estimate for your county. Now, suppose you also had a small, inadequate sample of data for your county. Rather than throw away the data from the other counties, it makes sense to form an estimate by averaging their means and the mean for your county, perhaps assigning a weights to each county based on its perceived similarity to your county. Now, suppose you have no domain knowledge for judging similarity. The Lemon-Krutchkoff estimator uses the sample means themselves for this purpose. The weight assigned to a county is determined by the difference between its sample mean and your county's sample mean. With this approach, the greatest weight would necessarily be assigned to your county. If we substitute the twelve X-12-ARIMA seasonals for the twelve counties' sample means, this is essentially how the Lemon-Krutchkoff method dampens seasonal variation.

Thus, damping helps when (1) the X-12-ARIMA seasonal estimate is subject to considerable sampling error (i.e., large random variation) and (2) the X-12-ARIMA seasonals for some other months are of similar magnitude (i.e., small seasonal variation among some subset of months). For example, suppose monthly sales outcomes exhibit great random fluctuation. On average, they are about the same for the first eleven months but increase sharply each December for the holiday season. In this case, the X-12-ARIMA seasonal factors for the first eleven months would be similar but would differ sharply from the December seasonal. With local damping, their estimated seasonals would be adjusted toward their mean, while the December seasonal would remain virtually unchanged.

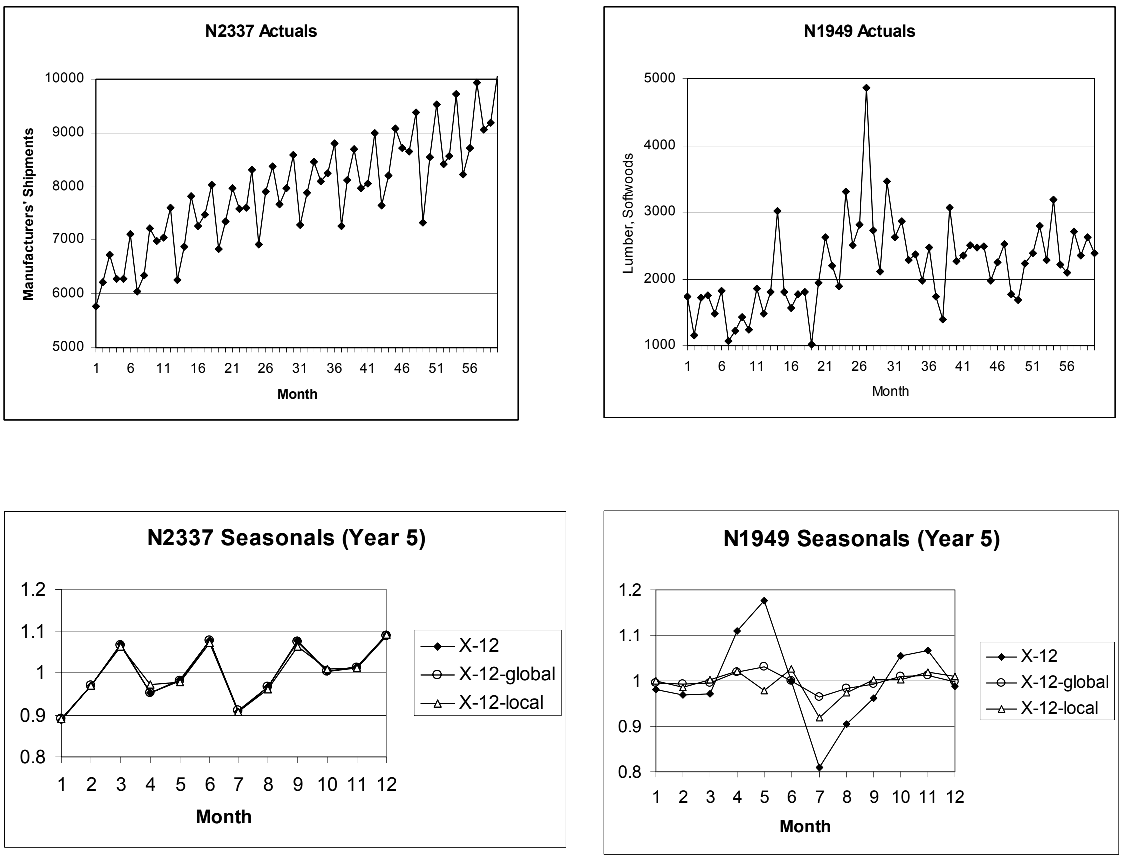
3. Examples with Real Data

The two monthly series depicted in Figure 1 illustrate the need to damp seasonal variation and the effect of using shrinkage estimators for this purpose. Both are taken from the M3-competition (Makridakis and Hibon, 2000). Neither series has been adjusted in any way (e.g., for outliers). Figure 1(a) depicts the first five years' actuals for 'Manufacturer's shipments, instruments and related products' (series N2337). This series is dominated by seasonal variation and exhibits little random variation. Figure 1(b) depicts the first five years' actuals for 'Lumber, softwoods, southern pine, exports total sawmill' (series N1949). It appears to be dominated by random variation with little apparent seasonal variation. Corresponding to each series are plots of the seasonal factors (fifth year only) using X-12-ARIMA alone and damped (global and local). Surprisingly, the amount of seasonal variation estimated by X-12-ARIMA (without damping) is slightly greater for 'Total Sawmill' than for 'Manufacturers' Shipments'! Damping the X-12-ARIMA seasonals for 'Manufacturers' Shipments' leaves them almost unchanged. The variation among seasonal factors for 'Manufacturers' Shipments', measured by their standard deviation, was .064 with X-12-ARIMA, .063 after global damping, and .061 after local damping. In contrast, damping reduces the estimated seasonal variation for 'Total Sawmill' considerably: The standard deviation for seasonal variation was reduced from .080 to .015 with global damping and .030 with local damping.

Figure 1 Two examples of the effect of damping seasonal variation

(a) 'Manufacturer's shipments, instruments and related products' (Series N2337, Years 1 – 5)

(b) 'Lumber, softwoods, southern pine, exports total sawmill' (Series N1949, Years 1 – 5)



4. The effect of the proposed methods: A simulation study

4.1 Simulation Design

In Miller and Williams (forthcoming), we found that the improvement in accuracy that results from damping classical decomposition seasonals (with either shrinkage method) depends primarily on the damping factor W_j^{JS}, which in turn is determined by two factors: (1) the variance of the seasonal estimator and (2) the amount of seasonal variation. The variance of the seasonal estimator was determined primarily by the amount of random variation and the length of the series. The presence or absence of a constant trend had little or no effect. Asymmetry of the seasonality pattern (e.g., three seasonals well above 1.0 balanced by nine seasonals slightly below 1.0) affected the relative performances of the two proposed shrinkage methods. In this simulation study we consider these same simulation factors, plus the effect of seasonal factors whose magnitudes increase over time (an X-12-ARIMA capability not usually incorporated within classical decomposition).

It is not possible to determine the error associated with a seasonal adjustment method for a real-data time series because its underlying seasonal factors are unknown. For this reason, we generated simulated time series with known seasonal factors. The simulation design reflects the characteristics of the 1428 monthly series of the M3-competition in regard to the factors mentioned above. For each M3 series, we performed classical decomposition and recorded its length in months, the amount of seasonal variation [standard deviation of the estimated seasonal factors, after using global damping (per Miller and Williams, forthcoming)], amount of random variation (standard deviation of the "irregular" component), trend (average of six different estimates of overall trend, expressed as a percentage of the series mean), and the asymmetry of the seasonal pattern (absolute value of the coefficient of skewness for the estimated seasonal factors). Additionally, we determined the optimal values of the smoothing coefficients when both simple exponential smoothing and Holt's linear exponential smoothing models were applied to the seasonally adjusted series. These results are summarized in Table 1.

Table 1 Summary of features for 1428 M3-Competition monthly series

	length (months)	Seasonal variation: StDev of CD seasonals	Random variation: StDev of CD irregulars	Abs. trend (% per month)	Skewness (abs value) for CD seasonals	Optimal values for Holt α	Holt β
5th Pct.	51	0.001	0.005	0.025	0.036	0.05	0
Q1	78	0.011	0.019	0.135	0.200	0.125	0
Median	115	0.060	0.060	0.271	0.437	0.5	0.010
Mean	99.34	0.089	0.100	0.431	0.569	0.477	0.056
Q3	116	0.127	0.148	0.570	0.795	0.5	0.1
95th Pct.	126	0.282	0.304	1.398	1.471	0.9	0.2

Based on the results shown in Table 1, we generated simulated data which reflect the range of outcomes observed for each factor in the monthly M3 data. The simulation design is summarized in Table 2. We generated the series in two steps: We first created a nonseasonal series, then multiplied by a set of known seasonal factors. We generated no-trend series in such a way that simple exponential smoothing (SES) provides optimal forecasts, and we generated trend series in such a way that Holt's linear exponential smoothing is optimal. (See the Appendix for more detail.) For the random component of these models, we used 5 levels of random variation that represented the 5th, 25th, 50th, 75th, and 95th percentiles of random variation observed for the M3 series. Thus we generated ten types of nonseasonal series, consisting of two levels of trend (trend or no-trend), combined with five levels of randomness.

The no-trend models used an initial level of 100. The trend models were initialized with level = 52.5 and trend = 1.0 per month. Thus, at the series midpoint, the expected level is 100, and the initial trend is 1% of that level. Ninety percent of the monthly M3 series had estimated trend values of less than 1%.

We observed that, as might be expected, M3 series with greater random variation tended to have smaller optimal smoothing coefficients when SES and/or Holt models were fit to the deseasonalized data. For example, for the group of M3 series whose random component had a standard deviation of approximately .06, the median optimal smoothing coefficients were approximately .5 (level) and .015 (trend), respectively. Thus, for each level of random variation, we generated nonseasonal data using the median values of the optimal smoothing coefficients for the group of M3 series that exhibited approximately that amount of random variation.

As stated previously, the primary factors governing the effect of damping are the variance of the seasonal estimator and the amount of seasonal variation. In order to simplify the simulation design, we controlled the variance of the seasonal estimator by using the five levels of random variation cited above and a fixed series length. Thus, each series consisted of 96 simulated monthly observations (i.e., eight years), which is approximately the mean length of M3 series.

Finally, we multiplied by a specific set of seasonal factors to produce the final series. We used nine different seasonal patterns as provided in Table 6 of the Appendix. These seasonal factors were specified judgmentally to satisfy the following characteristics. For six patterns, the seasonal factors were the same for all eight years. These patterns paired three levels of seasonal variation, corresponding to the 50th, 75th, and 95th percentiles observed for the M3 data, with two levels of seasonal skewness, corresponding approximately to the 50th and 96th percentiles observed for the M3 data.

For the other three seasonal patterns, the variation among seasonal factors increased over time, with rates of increase corresponding to the 50^{th}, 75^{th}, and 95^{th} percentiles observed for the M3 data. The seasonal factors within a calendar year sum to 12.0 for every seasonal pattern.

The combination of ten nonseasonal models with nine seasonal patterns produced 90 sets of conditions which are representative of the collection of monthly M3 series with respect to the factors relevant to this study. For each set of conditions, we generated 500 series randomly.

Table 2 Simulation design

Overall design

- Simulation consists of 500 replications for each of 90 monthly time series patterns.
- Each series consists of 96 periods (eight years).
- Each time series pattern is the product of two components: a nonseasonal model (10 models) and a seasonal pattern (9 seasonal patterns), as follows:

Components	Levels[1]				
Non seasonal model: 10 models determined by 2 levels of trend combined with 5 parameter sets					
Trend (y/n)	No trend: ARIMA(0, 1, 1) = simple exponential smoothing		Trend: ARIMA(0, 2, 2) = Holt's linear exponential smoothing, with initial trend = 1.0 (90th)		
Parameter set	1	2	3	4	5
StDev of random component	0.005 (5th)	0.02 (25th)	0.06 (50th)	0.15 (75th)	0.3 (95th)
Smoothing coefficients (α for SES; α,β for Holt)	α = 0.8 β = 0.15	α = 0.6 β = 0.025	α = 0.5 β = 0.015	α = 0.2 β = 0.02	α = 0.125 β = 0.02
Seasonal pattern: 6 patterns with constant seasonals, and 3 patterns with increasing seasonal variation					
Constant seasonals					
SD(seas)	0.033 (50th)	0.085 (75th)	0.25 (95th)		
Skew(seas)	0.35 (42nd)	1.64 (96th)			
Increasing seasonal variation[2]					
Rates of increasing seasonal variation[3]	1.5% (50th)	4.5% (75th)	11.6% (90th)		

[1] Figures in parentheses indicate the corresponding percentiles for the M3 data.

[2] For the three patterns of increasing seasonal variation, the 5th-year seasonal factors were the same as those of the three constant-seasonals patterns with SD(seas) = .033, .085, and .25 and with Skew(seas) = .035.

[3] Rates are percentage increases in the standard deviations of the seasonal factors for successive years.

4.2 Simulation Results

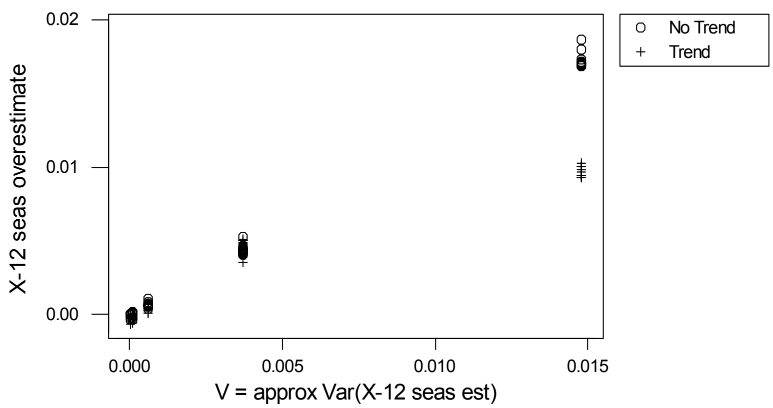
4.2.1 Measurement of estimation error

For each simulated series, we used X-12-ARIMA to estimate all 96 seasonal factors (one for each period of data). Then we applied both shrinkage methods to produce two additional sets of estimated seasonal factors. The accuracy of each set of estimates was measured by comparing them to the known seasonal factors underlying that series; thus, there are 96 estimation errors for each series that is generated. We summarized estimation error over two time frames: (1) overall, for the entire set of 96 seasonals and (2) for the most recent year, i.e., for only the last 12 seasonals. (For some users, accurate seasonal adjustment is most critical for the most recent year.) Unless stated otherwise, the results that follow pertain to the entire eight years. The accuracy of a method over a given time frame was summarized by its mean squared error (MSE) and its mean absolute percentage error (MAPE). We define MSE for a given combination of conditions as the average over 500 replications of the squared estimation errors for all seasonals in a specific time frame. Similarly, we define MAPE as the average of the absolute values of the percentage errors over 500 replications for all estimated seasonals in the time frame. To compare the accuracy of any two methods for a given set of conditions, we determined the ratio of their MSEs and the ratio of their MAPEs. We note that, while MSE is inappropriate for evaluating forecasting accuracy over multiple, disparate series (e.g., Armstrong and Collopy, 1992), MSE is appropriate for the estimation of model parameters for a given series. Indeed, MSE remains the standard metric used in the literature and by mainstream software for parameter estimation and optimization of forecasting methods. In this research, comparisons of MSEs for two methods of estimating seasonals always apply to their performances over 500 series generated from one fixed set of conditions. Additionally, the scale is similar across series, because multiplicative seasonals are always centered around 1.0.

4.2.2 X-12-ARIMA overestimation of seasonal variation

Does X-12-ARIMA exaggerate seasonal variation? If so, how much and under what conditions? Figure 2 shows that this is indeed the case. The vertical axis represents the average difference for each experimental condition between (1) the statistical variance among X-12-ARIMA seasonals and (2) the variance among actual seasonals. Note that all values are either positive, indicating overestimation of seasonal variation by X-12-ARIMA, or close to zero. The horizontal axis represents the approximate variance of the X-12-ARIMA seasonal estimator [V in (2)]. (Figure 2 reflects five levels of V, but the first two levels are indistinguishable graphically – a consequence of squaring the standard deviations that defined the five levels of random variation in the simulation.) The plotting symbol indicates whether the pattern includes trend or not. X-12-ARIMA's overestimation of seasonal variation is directly proportional to sampling error, as suggested by statistical theory (see Appendix). Under no conditions did X-12-ARIMA underestimate seasonal variation. When sampling error is large, the overestimation of seasonal variation is greater for no-trend series than for series with trend. The reason for this is unknown at present. Perhaps it is related to the findings of Ittig (1997), who claims that ratio-to-moving average models contain a systematic error when trend is present. These results suggest that greater damping may be required for no-trend series.

Figure 2 X-12-ARIMA overestimate of seasonal variation



- X-12 seas overestimate = Var(X12-ARIMA seasonals) – Var(actual seasonals)
- V = .1644*Variance of random component (approximately: see Appendix, section 2.1)

4.2.3. Effect of damping on the accuracy of estimated seasonals

Does damping produce more (or less) accurate estimates of seasonal factors than X-12-ARIMA? Table 3 summarizes the ratios of the MSEs for global damping vs. X-12-ARIMA, local damping vs. X-12-ARIMA , and local vs. global. Part (a) characterizes relative performance over the entire length of the series, while part (b) pertains to the last year only. Accuracy over the entire series is almost always greater with either shrinkage method than with X-12-ARIMA alone. In the best case, MSE is reduced by 73% (based on the smallest MSE ratio: .268 for local damping vs. X-12-ARIMA); in the worst case, it is increased by 2.4%. Thus, damping seasonals is virtually a “no lose” decision. Comparing the two shrinkage methods, local damping is more accurate than global damping for most cases. Regarding accuracy for only the most recent year, the results were similar. For most series, damping seasonals improved accuracy. The range of results is wider, probably due to greater sampling error when using only one year. Similar results were found when using MAPE as the accuracy measure.

Table 3 Relative accuracy of seasonal estimators: Distribution of MSE(first method)/MSE(second method) over all 90 simulation conditions

(a) Over all 8 years

Methods	N	Mean	Min	Q1	Median	Q3	Max
Global/X-12	90	0.821	0.292	0.687	0.932	0.996	1.005
Local/X-12	90	0.753	0.268	0.566	0.842	0.975	1.024
Local/Global	90	0.907	0.725	0.844	0.917	0.985	1.047

(b) Over the most recent year only

Methods	N	Mean	Min	Q1	Median	Q3	Max
Global/X-12	90	0.855	0.237	0.767	0.969	1.003	1.223
Local/X-12	90	0.789	0.270	0.640	0.824	0.986	1.443
Local/Global	90	0.914	0.635	0.820	0.910	0.994	1.326

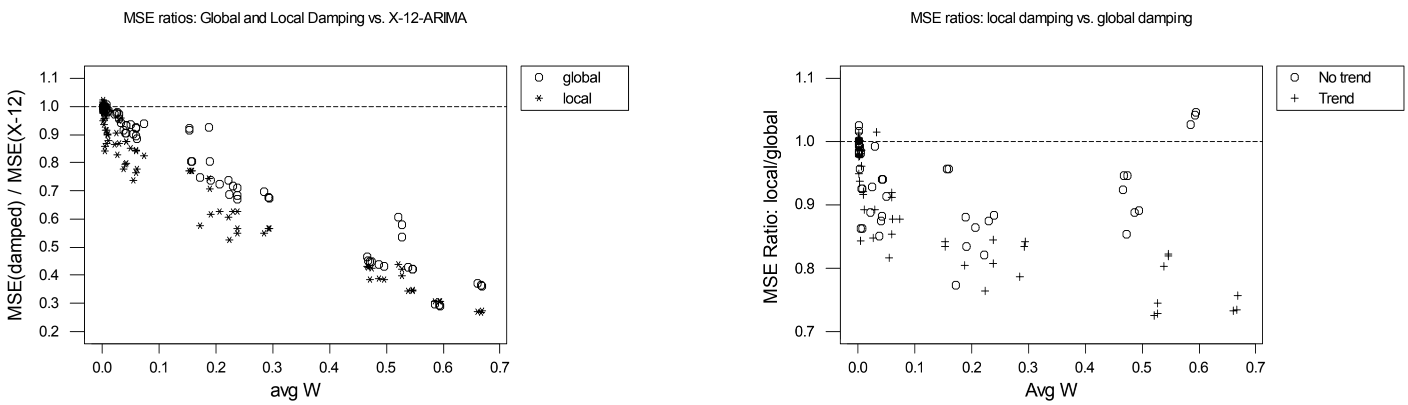
(Values less than 1.0 indicate the first method was more accurate.)

How do the characteristics of a series affect the impact of seasonal damping on the accuracy of estimated seasonals? Figure 3(a) is a plot of the MSE ratios for global and local damping vs. X-12-ARIMA. MSE ratios below 1.0 indicate greater accuracy for damped seasonals. Several observations are immediately apparent. The gain in accuracy obtained by using either damping method is closely associated with the damping parameter W^{JS}. When W^{JS} is near 0.0 with either damping method, little or no seasonal damping is merited and there is little gain in accuracy. As W^{JS} increases, more damping occurs and greater gains in accuracy are achieved. Since the primary determinants of W^{JS} are the amounts of random variation and seasonal variation, these are the predominant factors influencing the need for shrinkage and the associated gain in accuracy. Local damping usually provides a greater gain in accuracy than does global smoothing. For local damping, the relationship between the gain in accuracy and W^{JS} is not quite linear; as W^{JS} increases, the benefit of local damping increases, but at a diminishing rate. The presence or absence of trend [not depicted in Figure 3(a)] has no discernible effect on accuracy gains with local damping. With global damping, gains in accuracy for series with trend are usually less than for no-trend series. For no-trend series, global damping becomes as effective as local damping if W^{JS} is sufficiently large (say, $W^{JS} \geq .5$).

Figure 3(b) provides a direct comparison of global and local damping and illustrates the effect of trend. MSE ratios below 1.0 favor local damping. Local damping is more accurate for almost all conditions and is never less accurate by a substantive amount. For trend series, the relative advantage with local damping increases, but at a diminishing rate, as W^{JS} increases. For no-trend series, the advantage for local damping grows for W^{JS} up to about .25 but diminishes as W^{JS} increases further. The differences between results for trend and no-trend series reflect the diminished effectiveness of global damping in the presence of trend observed in Figure 3(a). Skewness among seasonals, not depicted in Figure 3(b), plays a minor role and is discussed in Section 4.3.

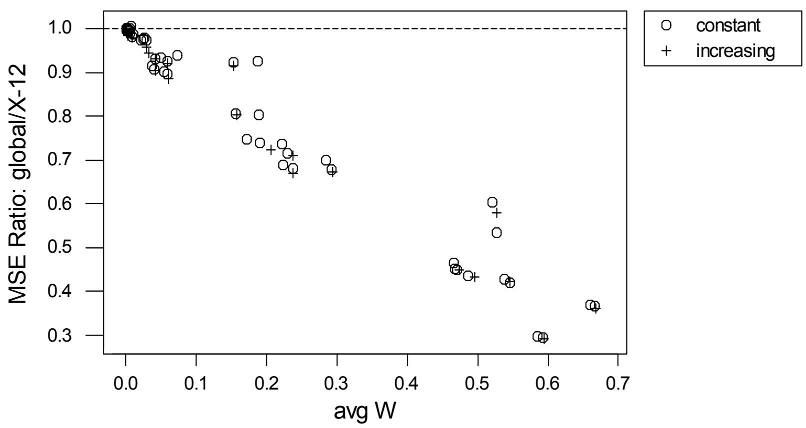
Figure 3 The effect of seasonal damping on the accuracy of estimated seasonal factors

(a) (b)



Is the benefit of seasonal damping affected when seasonal variation increases over time? Figure 4 provides the MSE ratios for global damping vs. X-12-ARIMA for both types of series. Given a damping weight W^{JS}, the benefit of global damping was essentially the same for both types of seasonal pattern. This was true for local damping as well, and we observed similar results for estimating only the last year's seasonals. Additionally, the relative performances of local and global damping were unrelated to whether seasonal variation was constant or increasing.

Figure 4 The effects of using seasonal damping: constant seasonal variation vs. increasing seasonal variation



Patterns similar to Figure 3(a) and Figure 4 emerge when we narrow the focus to the most recent year. Results (not shown) are somewhat more erratic, at least partly due to increased sampling error. For three conditions, the MSE ratio exceeded 1.1 for both shrinkage methods, suggesting that seasonal damping decreased accuracy. Each of these cases involved random variation at its greatest level (encouraging more shrinkage) and seasonal variation at its greatest level (encouraging less shrinkage). For these cases, only moderate damping was applied (with values of W^{JS} between .15 and .20), and overall accuracy over the entire series was improved by damping. The explanation for these instances of poor results in unknown at this time.

In comparing the two damping methods, a pattern similar to Figure 3(b) emerges, but the effect of trend is reversed. For no-trend series, local damping is always more accurate, with reductions in MSE ranging from 10% to 35%, depending on W^{JS}. For series with trend, the results are erratic and local damping is more accurate for the majority of series. Thus, when evaluated over all eight years, local damping's relative advantage was greater for trend series than for no-trend series. But when evaluated over the most recent year, its relative advantage was greater for no-trend series. The reasons for this are not apparent and require additional research in the future.

4.3 Statistical Significance of Simulation Results

The statistical significance of the simulation results is examined through regression modeling. Results for each of three comparisons are shown in Table 4. The dependent variable for each analysis is the MSE ratio (Model 1: for global damping/X12-ARIMA alone; Model 2: for local damping/X12-ARIMA alone; Model 3: for local damping/global damping). The independent variables include both direct effects (terms for the individual predictors) and some indirect effects (one interaction term and one quadratic term). There are direct effects for W^{JS}, trend (= 0 if no trend), skewness, and rate of increasing season variation (= 0 if seasonals were constant). We did not include direct effects for random variation and seasonal variation because, based on our prior research (Miller and Williams, forthcoming) and the results of Section 4.2 in this paper, their effects on the benefit of damping are expressed through the resultant value of W^{JS}. All models are calculated on the ratio of the *average* MSEs over 500 replications. (Variability due to replications within design factors is not considered). Thus the models show how well the design factors explain the corresponding MSE ratios. All reported models accurately portray the relationships observed for Figures 2 – 4 and have appropriately distributed residuals. All models have very high adjusted-R^2, showing that the factors explain a large percentage of the variation in the MSE ratios. When we repeated this modeling exercise using the ratio of the *median* MSEs as dependent variable, the coefficients were similar, the p-values were almost the same, and the fits were slightly better (i.e., greater adjusted R-square values).

Model 1 associates the design factors with the benefit of global damping vs. X-12-ARIMA alone. The only significant factors are W^{JS} – there is some evidence of a quadratic effect – and the interaction between trend and W^{JS}. The net effect of these factors mirrors the effects of global damping as discussed in Section 4.2 and illustrated by Figure 3(a). No other design factors seem to affect the impact of global damping.

Model 2 has to do with local damping vs. X-12-ARIMA alone. The most significant factor is W^{JS}. The significant quadratic term indicates the nonlinear effect described in Section 4.2. Thus, the benefit of damping is mostly determined by the global damping parameter. The significant skewness measure suggests that the benefit of local damping is somewhat greater when the seasonal pattern is skewed (such as peaks in December for retail sales). No other design factors were significant.

Model 3 involves the effect of design factors on the relative performances of local damping vs. global damping. The significant factors are W^{JS}, the square of the W^{JS}, skewness of seasonals, and the interaction of trend and W^{JS}. The net effect of these factors mirrors the relative benefits of local and global damping as discussed in Section 4.2 and illustrated by Figure 3(b).

Table 4 Models of the relative accuracy of seasonal adjustment methods versus series characteristics

Response Variable:	Model 1 MSE(X12-global) / MSE(X12)		Model 2 MSE(X12-local) / MSE(X12)		Model 3 MSE(X12-local) / MSE(X12-global)	
		(p)		(p)		(p)
R^2-adj	.978		.956		.708	
F	651.23	(.000)	321.65	(.000)	36.94	(.000)
Effect	**Coefficient**		**Coefficient**		**Coefficient**	
Intercept	0.994	(.000)	0.969	(.000)	0.973	(.000)
W^{JS}	-1.267	(.000)	-1.727	(.000)	-0.581	(.000)
W^{JS}**2	0.208	(.068)	1.131	(.000)	1.074	(.000)
Trend(y/n)	0.0012	(.894)	-0.0026	(.844)	0.0063	(.601)
SkewSeas	0.0024	(.706)	-0.021	(.030)	-0.027	(.003)
IncrSeas(y/n)	-0.164	(.122)	0.011	(.941)	0.185	(.198)
Interaction: Trend*W^{JS}	0.201	(.000)	0.0059	(.909)	-0.389	(.000)

5. Comparisons of seasonal adjustment methods based on real data

Even if seasonal damping improves seasonal adjustments for data exhibiting a stable pattern, as in the simulation, one may wonder how well they work with real data, for which outliers, level shifts, trend shifts, changes in variation, and other pattern disturbances are common. Thus, we investigated the performance of the proposed methods vis-a-vis X-12-ARIMA using real data. Since the accuracy of a seasonal estimator cannot be measured directly with real data, we took the following approach. In forecasting a time series, it is a common practice to perform seasonal adjustment of the series, to forecast the deseasonalized series, and then to reintroduce seasonality to produce a final forecast. With this approach to forecasting, forecasting accuracy depends to some extent on the accuracy of the estimated seasonal factors. Of course, many other factors affect forecasting accuracy as well, such as choice of forecasting model, estimation of that model's parameters, and the occurrence of pattern breaks during the forecasting horizon.

5.1. Analysis plan

The analysis in this section is based on the 1,428 monthly time series used in the M3-Competition. For each series we (1) withheld 18 months of data (the same data that were withheld in the M3-Competition); (2) seasonally adjusted the remaining data using X-12-ARIMA alone as well as X-12-ARIMA adjusted by global and local damping; (3) forecasted each deseasonalized series using damped-trend linear exponential smoothing (Gardner and McKenzie, 1985); (5) reseasonalized the forecasts; (6) compared the reseasonalized forecasts to the withheld actuals and recorded the MAPE over several forecasting horizons.

Damped-trend exponential smoothing performed the best among conventional, simple methods used in the M3-competition. The model coefficient values were the combination of values that minimized the sum of one-step-ahead squared errors in model fitting, chosen from the following ranges: α between .01 and .9, β between 0 and .15; ϕ between .9 and 1.0. For reseasonalizing the forecast, we used the X-12-ARIMA seasonal values for the most recent 12 months of actual data along with their counterparts from global and local damping. Each forecast was evaluated over five horizons: 1 month, 3 months, 6 months, 12 months, and 18 months. Since the scales of the series differed substantially, MAPE was used to measure forecast error over a given horizon. For example, over a 3-month horizon,

$$\text{MAPE}(3) = 100 \; \frac{\left[\left| e_1 / X_1 \right| + \left| e_2 / X_2 \right| + \left| e_3 / X_3 \right| \right]}{3}$$

where $e_t = X_t - F_t$

X_t = actual for period t

F_t = forecasted value for period t

$= \hat{S}_t$ * F-deseas$_t$

$\hat{S}_t$ = estimated seasonal factor for period t

F-deseas$_t$ = forecasted deseasonalized value for period t

Since each error measure [e.g., MAPE(3) above] reflects forecasting errors month-by-month rather than cumulatively, it is especially sensitive to the accuracy of seasonal factor estimates.

5.2. Results

The results are summarized in Table 5. Part (a) shows the average MAPE values corresponding to the three methods of seasonal adjustment. When seasonal damping is used, the average MAPE is reduced in every case, with reductions between 1.6% and 6.0% depending on the forecasting horizon and the shrinkage method used. Neither shrinkage method dominates the other.

The statistical significance of the results is addressed in part (b). Although average MAPE is a common basis for assessing forecasting accuracy, we found this approach to be unsuitable for determining statistical significance because the distribution of MAPEs was severely skewed over the 1428 series in this set. Therefore, for each series, we determined the difference between the MAPE values of two competing methods as a percentage of the MAPE values. That is,

$100 * \frac{MAPE_1 - MAPE_2}{(MAPE_1 + MAPE_2)/2}$. These values were distributed symmetrically and were not biased toward either method. Assessment of statistical significance is based on this measure of accuracy, labeled 'avg. E'. Forecasting error was reduced in all cases when seasonal damping was used. On average, global damping provided reductions of 1.6% to 2.3%, while local damping provided reductions of 2.3% to 5.0%. Using local damping reduced error by 0.5% to 3.3% when compared to global damping. All comparisons are highly significant statistically.

As another way of comparing the methods, we observed 'percent better', the percentage of series for which a given method produced a smaller MAPE than a competing method. Using either shrinkage method instead of X-12-ARIMA provided a smaller MAPE for 59% to 65% of the series, depending on the forecasting horizon. The MAPE value was lower for local damping than for global damping for 52% to 56% of all series. With one exception, all of these comparisons were highly significant. These forecasting results are consistent with those of Miller and Williams (forthcoming) and Armstrong and Collopy (2000), both of whom damped classical decomposition seasonals for approximately sixty monthly series from the M-competition (Makridakis, et al, 1982).

Table 5 Forecasting accuracy for three methods of seasonal adjustment

(a) Average MAPEs

Seasonal adjustment method	Average MAPE over horizon 1	1 -- 3	1 -- 6	1 -- 12	1 -- 18	No. obs.
X-12 alone	15.62	15.87	16.52	17.45	21.55	1428
X-12 w/global damping	15.07	15.32	15.96	16.90	20.84	1428
X-12 w/local damping	14.68	15.13	15.83	16.93	21.21	1428
Ratio of avg. MAPEs:						
X12-global / X12 alone	0.965	0.966	0.966	0.969	0.967	
X12-local / X12 alone	0.940	0.954	0.958	0.970	0.984	
X12-local / X12-global	0.974	0.988	0.992	1.001	1.018	

(b) Statistical significance

Seasonal adjustment methods compared		1	1 -- 3	1 -- 6	1 -- 12	1 -- 18
X-12-global vs. X-12	avg. E[1]	-1.791	-2.309	-2.28	-1.876	-1.585
	(p)[2]	(0.035)	(0.000)	(0.000)	(0.000)	(0.000)
	percent better[3]	0.587	0.608	0.644	0.645	0.632
	(p)[4]	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X-12-local vs. X-12	avg. E	-5.36	-4.097	-3.837	-2.601	-2.079
	(p)[2]	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	percent better	0.586	0.611	0.632	0.636	0.613
	(p)[4]	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X-12-local vs. X-12-global	avg. E	-3.29	-1.763	-1.564	-0.73	-0.5
	(p)[5]	(0.004)	(0.000)	(0.000)	(0.004)	(0.044)
	percent better	0.546	0.560	0.549	0.524	0.516
	(p)[6]	(0.000)	(0.000)	(0.000)	(0.076)	(0.234)

[1] $E = \frac{MAPE(1st\ method) - MAPE(2nd\ method)}{avg.\ MAPE\ (both\ methods)}$

[2] for paired-t test of H0: mean = 0.0 vs. Ha: mean < 0.0)

[3] Percent of series for which the MAPE of 1st method is smaller than MAPE for 2nd method

[4] for sign test of H0: proportion better = 0.5 vs. Ha: proportion better > 0.5)

[5] for paired-t test of H0: mean = 0.0 vs. Ha: mean ≠ 0.0)

[6] for sign test of H0: proportion better = 0.5 vs. Ha: proportion better ≠ 0.5)

Note:Tests comparing results with and without damping are one-sided since their purpose is to determine the strength of evidence that damping *improves* forecasting accuracy. Tests comparing the two shrinkage methods were two-sided since neither was supposed a priori to be more accurate.

6. Conclusions

In a simulation of a variety of time series patterns that reflect the characteristics of the 1,428 monthly series used in the M3 competition, we found that X-12-ARIMA overestimated seasonal variation. The amount of overestimation was determined by the sampling error associated with the X-12-ARIMA seasonal estimator and was greater for no-trend series than for series with trend. When either global or local damping is used to adjust the X-12-ARIMA seasonals, accuracy improves. The amount of improvement for a given series is determined primarily by the extent to which random variation dominates seasonal variation as characterized by the damping weight W^{JS}. There were no conditions for which seasonal damping degraded accuracy appreciably. For series with little random variation and great seasonal variation, X-12-ARIMA seasonals remain virtually undamped.

Although local damping is more complex than global damping, it produces greater gains in accuracy, and these gains are unaffected by the presence of trend. The benefit of global damping is somewhat greater with no-trend series than for series with trend, perhaps because X-12-ARIMA overestimation of seasonal variation was greater for no-trend series. These basic results were approximately the same regardless of the error measure that was used (MSE or MAPE) or the 'window' for accuracy measurement (the entire series or the last year of the series).

With real data, forecasting accuracy improved when seasonal damping was used in the seasonal adjustment process. For the 1,428 monthly series in the M3-competition, the MAPE of forecasts was consistently lower when seasonal damping was used for seasonal adjustment prior to forecasting than when X-12-ARIMA alone was used. These results support the simulation findings that seasonal damping improves the accuracy of X-12-ARIMA seasonal adjustments of real data.

The results of this research indicate that seasonal damping should be applied routinely to X-12-ARIMA seasonals. The choice between damping methods is a trade off between simplicity of execution (favoring global damping) and expected gains in accuracy (favoring local damping). Additional gains in accuracy may become possible in the future as alternative variations for implementing the damping procedures are explored in future research.

References

Armstrong, J.S., (1978). *Long-Range Forecasting*, John Wiley and Sons.

Armstrong, J.S., (2001). Extrapolation for time-series and cross-sectional data. In J.S. Armstrong (Ed.), *Principles of Forecasting* (p.225). Kluwer Academic Publishers.

Armstrong, J.S. and Collopy, F. Speculations about seasonal factors. (Available online: http://hops.wharton.upenn.edu/forecast/, accessed November 22, 2002)

Bunn, D.W., and Vassilopoulos, A. I. (1999). Comparison of seasonal estimation methods in multi-item short-term forecasting. International Journal of Forecasting, 15, 431-443.

Canavos, G. C. (1984). *Applied probability and statistical methods,* Little, Brown & Company.

Dagum, E. B. (1988). The X-11-ARIMA/88 seasonal adjustment method. Methodology Branch, Statistics Canada, Ottawa ON, Canada.

Dagum, E. B. (1980). The X-11-ARIMA seasonal adjustment method. Statistique Canada, Catalogue 12-564E.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach. Journal of the American Statistical Association, 68, 117-130.

Efron, B. and Morris, C. (1975). Data analysis using stein's estimator and its generalizations. Journal of the American Statistical Association, 70, 311-319.

Findlay, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen, B. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. Journal of Business and Economic Statistics, 16, 127-177.

Gardner, E.S., and McKenzie, E. (1985). Forecasting trends in time series. Management Science, 31, 1237-1246.

Greis, N. P., and Gilstein, C. Z. (1991). Empirical Bayes methods for telecommunications forecasting. International Journal of Forecasting, 7, 183-197.

Ittig, P. (1997). A seasonal index for business. Decision Sciences, 28,335-355.

James, W., and Stein, C. (1961). Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1 (University of California Press, Berkeley, CA), 361-379.

Ladiray, D., and Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method*. Springer.

Lemon, G.H. and Krutchkoff, R.G. (1969). Smooth empirical Bayes estimators: with results for the binomial and normal situations. Proceedings of the Symposium on Empirical Bayes Estimation and Computing in Statistics, Texas Tech University Mathematics Series No. 6, 1969, 110-140.

Macauley, F.R. (1930). The smoothing of time series. National Bureau of Economic Research.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., and Winkler, R. (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. Journal of Forecasting, 1, 111-153.

Makridakis, S., and Hibon, M. (2000). The M3-Competition: results, conclusions and implications. International Journal of Forecasting, 16, 451-476.

Makridakis, S., Wheelwright, S.C., and Hyndman, R. J. (1998). *Forecasting Methods and Applications, Third Edition*, John Wiley and Sons.

Miller, D.M., and Williams, D. Shrinkage Estimators of Time Series Seasonal Factors and Their Effect on Forecasting Accuracy. International Journal of Forecasting (forthcoming).

Morris, C.N. (1983). Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association, 78, 47-55.

Pierce, David A. (1980). A survey of recent developments in seasonal adjustment. The American Statistician, 34, 125 - 134.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proceedings of the Third Berkeley Symposium on Probability and Statistics, 1, 157-163.

Rutherford, J. R., and Krutchkoff, R.G. (1969). An empirical Bayes smoothing technique. Biometrika, 56, 1, 133–137.

Shiskin, J., Young, A.H., and Musgrave, J.C. (1967). The X-11 variant of the Census II method seasonal adjustment program. Bureau of the Census, Technical Paper No. 15.

Stein, C, (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Probability and Statistics, 1, 197-206.

U.S. Census Bureau. *X-12-ARIMA Reference Manual, Version 0.2.10*. (Available online: http://www.census.gov/srd/www/x12a/x12down_pc.html, accessed November 22, 2002)

Appendix

In this appendix, we provide a brief overview of X-12-ARIMA as used in this research, an explanation of the properties of X-12-ARIMA that cause it to overestimate seasonal variation, the proposed shrinkage methods, the simulation study, and the forecasting analysis.

1. X-12-ARIMA

X-12-ARIMA offers a choice among several decomposition models. The multiplicative form is $X_t = C_t*S_t*D_t*E_t*I_t$, where C_t is the trend-cycle series at t, and S_t and I_t are the seasonal and irregular (i.e., random) components, respectively. D_t and E_t are optional components that represent the effects of trading days and the Easter holiday, respectively. X-12-ARIMA provides an option to pre-adjust the original series for irregularities such as level shifts, then proceeds iteratively to estimate the model components, search for disruptive effects (extreme values, trading-days, and holidays), reestimate the components from a corrected series, search again for disruptive effects, etc. These steps are detailed by Ladiray and Quenneville (2001). In the final iteration, a series of "SI ratios" are obtained by dividing the adjusted original data by the corresponding trend-cycle values. Each SI ratio represents the product of the seasonal (S) and irregular (I) components. The final seasonal factor for a given month and year is a centered, weighted moving average of the SI ratios for that month over several (usually seven) consecutive years. The final seasonal factors are normalized so that their average over each 12-month period is approximately 1.0. They appear in Table D10 of X-12-ARIMA output and are estimated for the SI ratios in Table D8 with extreme values replaced by those printed in Table D9.

How does the X-12-ARIMA algorithm lead to overestimation of seasonal variation? Suppose one selects independent random samples from $k \geq 3$ populations in order to estimate the population means. Stein (1955) showed that, provided the sampling distributions of the k sample means are normal distributions with equal variances, the sum of the squared estimation errors associated with the *k* sample means is reduced by adjusting (i.e., shrinking) the sample means toward their global mean. This idea is suggested by basic analysis of variance theory for this situation, where the sample size is *n* for all *k* samples. The expected value of the mean square for treatments is (Canavos, 1984):

$$E\left[\frac{n\sum_{i=1}^{k}\left(\overline{X_i}-\overline{X}\right)^2}{k-1}\right] = \sigma^2 + \frac{n\sum_{i=1}^{k}\left(\mu_i-\overline{\mu}\right)^2}{k-1} \qquad (4)$$

or (dividing by n),

$$E\left[\frac{\sum_{i=1}^{k}\left(\overline{X_i}-\overline{X}\right)^2}{k-1}\right] = \frac{\sigma^2}{n} + \frac{\sum_{i=1}^{k}\left(\mu_i-\overline{\mu}\right)^2}{k-1}. \qquad (5)$$

where $\overline{X}$ is the mean of the sample means, $\overline{\mu}$ is the mean of the population means, and σ^2 is the variance of the random component. Result (5) tells us that the variation among sample means is expected to exceed the variation among the population means by $\frac{\sigma^2}{n}$, the variance of the sampling distribution of each sample mean. Thus, for a given sample size, the excess variation among the sample means is determined by the amount of random variation. This suggests the possibility that (1) accuracy in estimating the population means might be improved by adjusting (i.e., shrinking) the sample means toward their global mean and (2) the amount of shrinkage needed, on a percentage basis, is determined by the amount of sampling error $\left(\frac{\sigma^2}{n}\right)$ as a proportion of the observed variation among the sample means $\left(\frac{\sum_{i=1}^{k}\left(\overline{X_i}-\overline{X}\right)^2}{k-1}\right)$. In other words, for a given sample size, the shrinkage needed depends on the extent to which random variation dominates variation among the means.

The X-12-ARIMA development of final seasonal factors is very similar, but not identical, to the Stein/analysis of variance conditions: In the final iteration, we have 12 samples of SI ratios -- one sample for each month – each with sample size equal to the number of years of data. The final seasonal factors are sample means. Since the seasonal factors are means, the assumption of normal sampling distributions is likely to be an acceptable approximation. However, other assumptions may be violated. The samples of SI ratios may not be independent due to autocorrelation in the trend-cycle and random components. The variances of the sample means (the final seasonal factors) are not equal because (1) they are affected by the magnitudes of the underlying seasonal factors and (2) random variation may vary over time. This paper can be viewed partly as an attempt to determine whether shrinkage methods work in this application despite these violations of the Stein/analysis of variance assumptions.

2. Global damping: Estimation of V and A_j for determining W_j^{JS}.

2.1 Estimating V

V is the variance of the sampling distribution of the X-12-ARIMA seasonal estimator, which we assumed to be the same for all periods. The final X-12-ARIMA seasonal estimator is a centered, weighted moving average of either 5, 7, or 11 SI values. The length of this moving average is determined by the relative amount of random variation versus the amount of seasonal variation. For the sake of simplicity, we assumed for every series that a 7-point centered moving average had been used. The sampling distribution of an X-12-ARIMA seasonal estimator is assumed to be the normal distribution whose mean is the corresponding true seasonal factor and whose variance is $V = .1644*\text{Var(SI)}$, where Var(SI) is assumed to be the same for all periods. That is,

$$V = \text{Var}(S_t) = \sum_{j=1}^{7}\left[W_j^2 * Var(SI)\right] = \sum_{j=1}^{7} W_j^2 * Var(SI) = .1644 * Var(SI). \qquad (6)$$

where W_j (j = 1, 2, …, 7) = 1/15, 2/15, 3/15, 3/15, 3/15, 2/15, 1/15, and $\sum_{j=1}^{7} W_j^2 = .1644$. (This implies an assumption, surely erroneous to some degree, that the SI ratios are independent.)

Var(SI) is estimated as follows:

$$\hat{Var}(SI) = \frac{1}{J(12-1)} \sum_{j=1}^{J} \sum_{k=1}^{12} \left(SI_{j,k} - S_{j,k}\right)^2 \qquad (7)$$

Thus, the estimated variance of an SI ratio is the average of J variance estimates, one for each calendar year. Alternatively, Var(SI) could be estimated directly from the variation among the Irregular factors. For any month k and year j, we have $Var_{jk}(SI) = Sjk^2\, Var(I) \approx Var(I)$ if, as before, we ignore the effect of the seasonal factor on the variance of the seasonal estimator. With this approach, V is determined approximately by the variance of the random component: V = .1644*Var(I), which could be estimated directly from the X-12-ARIMA Irregulars. Although this approach is simpler, it has not yet been sufficiently explored.

2.2 Estimating A_j

A_j represents the variation among the true seasonal factors for year j and is defined as follows:

$$A_j = \frac{\sum_{k=1}^{12}\left(S_{j,k}^{tr} - 1.0\right)^2}{12-1}, \tag{8}$$

where S_{jk}^{tr} is the true seasonal factor at year j, month k. Note that this is essentially the same as the right-hand term of (5), with the true seasonal factor substituted for μ_i.

We assume A_j can differ from year to year, so we developed a unique estimate for each calendar year. Expression (5) indicates that the expected mean square among *k* independent sample means equals the mean square for their corresponding population means plus the variance of the sampling distribution of the sample mean. Therefore, to estimate A_j for year j, we calculate the corrected sum of squares for the twelve X-12-ARIMA seasonals for year j (based on deviations from 1.0) and divide by 12-1 = 11. From this, we subtract the estimated V. If the result is negative, we set the estimated A_j to zero. Thus, for year j:

$$\hat{A}_j = \text{ the greater of } \frac{\sum_{k=1}^{12}\left(S_{j,k} - 1.0\right)^2}{12-1} - \hat{V} \quad \text{or} \quad 0. \tag{9}$$

Expression (9) leads to the development of a distinct damping weight W_j^{JS} for each year. Thus, the damping of a January seasonal is always in regard to "future" seasonals, while the damping of a December seasonal is in regard to "past" seasonals. It might be possible, and better, to use a centered approach in which damping for a month is in relation to both past and future seasonals. We chose our approach for its comparative simplicity.

3. Local damping: Definition and estimation of damping weights

With local damping, the seasonal factor for a given month and year is a weighted average of the twelve X-12-ARIMA seasonals for that year: $S_1, S_2, \ldots, S_{12}$, as shown in expression (3). In developing the damped factor for month k^*, the weight W_{j,k,k^*} given to $S_{j,k}$ is based on the statistical likelihood L_{k,k^*} for observing the value S_{k^*} from a probability distribution whose mean is S_k and whose variance is V. W_{k,k^*} is obtained by dividing this likelihood value by the sum of all 12 likelihood values, as follows:

$$W_{k,k^*} = \frac{L_{k,k^*}}{\sum_{k=1}^{12} L_{k,k^*}} \qquad (10)$$

Since each X-12-ARIMA seasonal estimator is a centered mean of SI ratios for that season, it seems reasonable to assume that its sampling distribution is approximately the normal distribution. Thus,

$$L_{k,k^*} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\frac{S_{k^*} - S_k}{\sigma}\right)^2\right] \qquad (11)$$

where σ is estimated by $\hat{V}^{1/2}$, and $\hat{V}$ is as defined previously for global damping.

Expressions (10) and (11) lead to the development of damped seasonals based only on the 12 months within the same calendar year. As with global damping, a centered approach in which the development of the damped seasonal for a given month is based on both past and future seasonals might be more accurate albeit more complex. This is a worthy subject for future research.

4. Details of the simulation

4.1 Generating data: the nonseasonal component

The data for the nonseasonal component of the simulation (the 10 ARIMA models) were generated using SAS. We generated nonseasonal, no-trend series randomly from ARIMA(0, 1, 1), for which simple exponential smoothing provides optimal forecasts. We generated nonseasonal, trend series randomly from ARIMA(0, 2, 2), for which Holt's linear exponential smoothing is optimal. For each set of SES and Holt coefficients in the experimental design, we determined the corresponding ARIMA parameter values and used them to generate nonseasonal data. Random errors were generated from a normal distribution with mean 0 and standard deviations as specified in the simulation design (Table 2).

4.2 Generating data: Seasonal factors used in the simulation

Table 6 provides the specific factors comprising the nine seasonal patterns used, along with the amount of seasonal variation (standard deviation of seasonals) and the absolute value of the coefficient of skewness for each. Patterns 7, 8, and 9 exhibited increasing variation among seasonal factors. For these, the information in Table 6 represent the fifth year.

Table 6 Seasonal patterns used in the simulation

	Patterns with constant seasonal variation						Patterns with increasing seasonal variation[1]		
	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6	Pattern 7	Pattern 8	Pattern 9
SD(seas):	0.033	0.033	0.085	0.085	0.250	0.250	0.033	0.085	0.250
abs(skew):	0.35	1.67	0.37	1.61	0.34	1.63	0.35	0.37	0.34
Jan	0.956	1.012	0.938	0.941	1.176	1.627	0.956	0.938	1.176
Feb	0.976	0.985	0.907	0.986	1.374	1.048	0.976	0.907	1.374
Mar	0.977	1.084	1.048	0.943	1.221	1.305	0.977	1.048	1.221
Apr	0.984	1.033	0.933	0.931	1.097	0.833	0.984	0.933	1.097
May	0.990	0.975	0.956	0.993	0.935	0.904	0.990	0.956	0.935
Jun	1.029	0.977	1.072	1.222	0.910	1.012	1.029	1.072	0.910
Jul	1.002	0.988	0.842	1.028	0.745	0.873	1.002	0.842	0.745
Aug	1.047	0.969	0.990	1.036	0.867	1.032	1.047	0.990	0.867
Sep	1.051	0.970	1.099	0.983	0.706	0.852	1.051	1.099	0.706
Oct	1.033	1.002	1.116	1.058	0.666	0.998	1.033	1.116	0.666
Nov	0.996	1.015	1.056	0.887	0.906	0.807	0.996	1.056	0.906
Dec	0.958	0.989	1.043	0.991	1.396	0.709	0.958	1.043	1.396

[1] Results for these series represent the 5th year (out of 8 years)

4.3 Specifications used in running X-12-ARIMA

Each series that we generated was input into X-12-ARIMA. We used the multiplicative modeling option in all cases. Options for dealing with outliers, level shifts, and calendar effects were unnecessary because these disruptive effects were never introduced into the data. The ARIMA component of X-12-ARIMA is used to extend the series at the ends and thus avoid end-point problems. In specifying the nonseasonal component of ARIMA, we used the same model forms that we had used to generate data. To specify the seasonal component of ARIMA, we used an integrated moving average of order one. Thus, the ARIMA models were specified as either ARIMA(0, 1, 1)$(0, 1, 1)_{12}$ or ARIMA(0, 2, 2)$(0, 1, 1)_{12}$.

The models we used to generate data do not match the multiplicative model within X-12-ARIMA (or any other of its options). Rather, we generated data in a manner that is consistent with methods that are commonly used in forecasting. Indeed, real data are not generated by either ARIMA or exponential smoothing processes or by any of the options within X-12-ARIMA. Rather, they are the outcomes of unstable processes affected by changing external factors. Indeed, the strength of X-12-ARIMA lies in its being a "robust nonparametric method, applying iterative estimation" rather than being based on an explicit model (Ladiray and Quenneville (2001). We think our approach to simulation reflects seasonal adjustment and forecasting in the real world more accurately than if we had matched the data-generating process perfectly to an X-12-ARIMA option.

5. Specifications used in running X-12-ARIMA in the forecasting analysis

For seasonal adjustment prior to forecasting, we input each M3 series into X-12-ARIMA using its options to adjust for outliers, level shifts, and calendar effects. We always specified a multiplicative decomposition model for practical reasons. Most, if not all, of the monthly M3 series exhibit characteristics suitable for using multiplicative seasonals, as described by Armstrong (2001). If additive seasonals are more appropriate for some of these series, then that is a limitation on the conclusions of this research. In specifying the ARIMA model, we chose the best-fitting model among ARIMA(0, 1, 1), ARIMA(1, 1, 2), ARIMA(0, 2, 2), ARIMA(0,1,1)$(0, 1, 1)_{12}$, ARIMA(1, 1, 2)$(0, 1, 1)_{12}$, and ARIMA(1, 2, 2)$(0, 1, 1)_{12}$. The first three are nonseasonal models for which simple exponential smoothing, damped-trend linear exponential smoothing, and Holt's linear exponential smoothing, respectively, provide optimal forecasts. The latter group of three are their seasonal counterparts.